Design, Manufacture, and Application of DNA Microarrays to Study Gene Expression Phenotypes of Lysine-Producing *Corynebacterium glutamicum*

by

Christopher Roberge

B.S. in Chemical Engineering (1993)
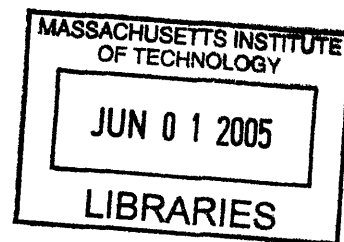Massachusetts Institute of Technology. Cambridge, MA

Submitted to the Department of Chemical Engineering in partial fulfillment of the
requirements for the degree of

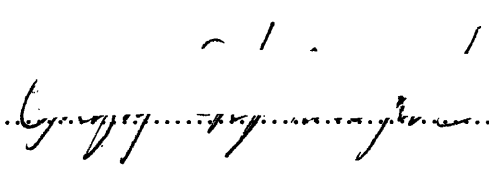Doctor of Philosophy in Chemical Engineering

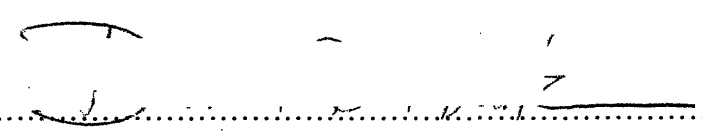at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Signature of the Author..
Department of Chemical Engineering
March 14, 2005

Certified by...
Gregory Stephanopoulos
Bayer Professor of Chemical Engineering
Thesis Supervisor

Accepted by...
Daniel Blankschtein
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Abstract

Corynebacterium glutamicum partial genome DNA microarrays were constructed that were capable of assaying the transcriptional profile of the genes of pathways involved in central carbon metabolism and lysine biosynthesis. It was found that to ensure arrays of high quality, protocols applying the arrays should include DNase treatment of RNA samples, additional RNA filtration purification steps, and the use of gene specific primers in the formation of labeled cDNA through reverse transcription. After implementing these procedures, the accuracy and reproducibility of the array data were validated. The microarrays were used to explore the effects of the over-expression of the key anaplerotic enzyme pyruvate carboxylase and the use of different medium carbon source compositions, both of which have been shown to influence the yields of biomass on carbon and of lysine on biomass. Three different strains of *C. glutamicum* that were grown on six different minimal medium formulations that varied in their balance of glucose and lactate were assayed by isolating total mRNA samples from cultures in three different phases of growth and lysine production. Genes associated with glycolysis and the pentose phosphate pathway showed decreased transcript concentrations as the available carbon source was shifted from glucose to lactate, while those associated with the TCA cycle and the glyoxylate bypass demonstrated increased transcription. As the cultures stopped generating biomass and began generating lysine, mRNA of genes associated with lysine synthesis and export was measured at elevated concentrations. Reduced gene expression trends seen for aspartokinase and aspartate semialdehyde dehydrogenase suggest that the enzymes are bottlenecks to lysine production, particularly when pyruvate carboxylase is over-expressed and lactate is the available carbon source. This over-expressing strain also had higher transcription levels of the genes of biotin synthesis, and lower transcription levels of the acyl-coA carboxylases dtsR1, dtsR2, accC, and accD. Other results implied that malic enzyme is co-expressed with pyruvate carboxylase to better allow cultures grown on lactate to produce NADPH in the absence of significant pentose phosphate pathway flux. Also, the transcriptional and flux profiles of a pair of *C. glutamicum* strains grown on two different medium compositions of

isotopically labeled glucose and lactate were determined simultaneously from the same set of actively growing and lysine-producing cultures. Flux maps for each of the four combinations of strain and medium were constructed using calculations derived from metabolite balances and GC-MS measurements of the isotopic distributions within biomass hydrolysates of the pseudo-steady-state cultures. Comparisons of the two sets of data showed that 19 of 28 pairs of flux and transcription measurements had trends with good agreement with one another. Different pathways of the metabolic network were found to be controlled via transcription in varying degrees. On average, the Embden-Meyerhof-Parnas pathway was shown to be less likely to be regulated though transcription than the pathways of the tricarboxylic acid cycle and central carbon anaplerosis. In the split pathway available to the cells for producing lysine, the succinylation branch showed an increase in flux for only the case of a pyruvate carboxylase over-expressing strain that was grown on lactate, while the alternate dehydrogenation branch showed a complementary decrease in flux. These flux changes were matched by changes in transcription that only occurred for the same culture and growth medium. Through these findings we have demonstrated the application of *C. glutamicum* DNA microarrays to the determination of how the cells regulate their responses at the transcriptional level to changes in both gene over-expression and medium composition.

Acknowledgements

The work of this thesis benefited enormously from the direct and indirect contributions of many people, whose influence helped to shape and support my research. I'll always be grateful for what they allowed me to accomplish.

Firstly, I need to thank my advisor, Greg Stephanopoulos, for treating me with the perfect balance of patience and prodding. He gave me the freedom to find my own solutions to the problems that I encountered, but also pointed me in the best directions to get help and advice when I most needed it. He and the other members of my thesis committee—Charles Cooney, Alan Grossman, Kristala Prather, Anthony Sinskey, and George Stephanopoulos—gave so much of their time and advice. Their recommendations and the discussions that I had with them were incredibly valuable in streamlining my work and giving it focus. I especially want to thank Profs. Cooney and Prather for assisting me by joining my committee later in the process and for being such insightful and encouraging committee members.

In the Stephanopoulos lab group I was most fortunate to be able to work with Mattheos Koffas and Kohei Miyaouko. Mattheos gave me a great introduction into the lab, helped me to learn many of the techniques that I would need again and again, and acted as my resident *C. glutamicum* expert. Kohei was an invaluable coworker who helped me design and carry out all of the flux analyses work, even once I was hundreds of miles away. When I say that I couldn't have finished the thesis the way it is now without him, I couldn't be more literal.

Many thanks go to Merck & Co., Inc., for the extremely generous financial and career development support without which I could not have pursued this degree. I will always be grateful to Michel Chartrain and Barry Buckland, my mentors and bosses at Merck, who championed my return to MIT and made this opportunity a reality for me. I also want to thank Jeffrey Moore, Ali Shafiee and David Robinson of MRL, who steadfastly and patiently supported my completion of my graduate work since my return to Merck in January 2004. Special thanks to my colleagues David Pollard and Brendan Grau, who

helped me on different Merck projects when I needed to take time off to carry out the last of my experiments.

Finally, my growing family has been the steadiest source of inspiration and support that I have. Throughout my time at MIT they have been some of my most valuable advocates and teachers. I want to thank my parents, Kenneth and Barbara who have always been more sure of what I am capable of than I often am. So much of what I achieved was the result of trying things I might not have without having them as my constant cheerleaders. I want to thank my parents-in-law, Dr. Daniel and Sarah Park for keeping me in their prayers, and for even enlisting the patronage of the Roman Catholic Church of St. John Nam to aid in the cause. The depths of their thoughtfulness continue to mean so much to me. I want to thank Lauren, for wishing me "Good Luck, Daddy!" on my way to my defense, and for so often providing a good tickling session to keep everything in perspective. And most of all I want to thank my best friend and wife, Keum, for always being so loving and immensely giving. Knowing her has been an absolute blessing, and being able to share so much with her has brought me more joy than anything else. There is no companion that I would rather have with me on the journeys of my life, and this thesis was certainly a major one. With her beside me, I look forward to starting down the next path.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Motivation

The majority of the worldwide supply of the amino acid L-lysine is produced through the batch fermentation of lysine-producing strains of the bacteria *Corynebacteria glutamicum*. The original techniques used to develop these strains and improve the usefulness of *C. glutamicum* in industrial processes involved repeated cycles of mutagenesis and selection. Although these methodologies did bring about enhancements to the properties of the bacteria, many of the fundamental explanations for these adjustments remained unknown. Later, with the introduction of more advanced genetic engineering tools, specific genes and regulatory elements within the microbe could be added, deleted, or altered [Katsumata *et al.*, 1984; Ozaki *et al.*, 1984; and Santamaria *et al.*, 1984]. Some of the targets of this genetic engineering have included the removal of feedback inhibition from regulatory strategies used by the wild type strain [Morbach *et al.*, 1996], the deletion of enzymes for the purpose of redirecting the flow of metabolites towards specific pathways [Marx *et al.*, 2003], and the over-expression of enzymes in attempts to amplify desired fluxes [Jetten *et al.*, 1995; de Graaf *et al.*, 2001; and Hartmann, 2003]. In this way investigators have attempted to gain an accurate sense of the metabolic networks that operate in the cells, the flux of metabolites through the pathways that make up these networks, and the regulatory and energetic mechanisms that in part determine the quantities of these fluxes.

Additional means of solving these problems were developed in the 1990s in the form of metabolic flux analysis and metabolic control analysis. These were methods used for quantifying the flow of metabolites through pathways, as well as the changes that occur to fluxes in response to genetic and environmental perturbations [Stephanopoulos and Vallino, 1991]. To accurately measure the *in vivo* fluxes that are present under various circumstances, several studies employed increasingly sophisticated experimental procedures involving isotopic tracers and mass- and nuclear-magnetic-resonance-

spectroscopy [Vallino and Stephanopoulos, 1993; Sonntag *et al.*, 1995; Park *et al.*, 1997; Marx *et al.*, 1997; and Petersen *et al.*, 2000].

As the study of *C. glutamicum* has evolved, it has become increasingly apparent that it is necessary to examine strain improvement strategies in ways that recognize the network-wide ramifications of flux alterations. As the flow of material through the metabolic pathways of the organism is altered, the requirements for energy and reducing equivalents may be affected, and metabolites that fuel one set of pathways may be diverted to serve as substrates for another. This is particularly true for the compounds phosphoenolpyruvate, pyruvate, oxaloacetate, and malate, that are positioned at the intersections of pathways involving glycolysis, gluconeogenesis, the tricarboxylic acid cycle, the glyoxylate bypass, and anabolic pathways leading to the formation of, for example, amino acids such as lysine. Although several strains have been produced by targeting the nodes involving these four metabolites for the purposes of up- or down-regulating fluxes, the overall network behavior in these strains is poorly understood and represents a prime candidate for further study with newer tools.

It is essential to the success of metabolic engineering that these tools provide as much detailed information about the inner workings of cellular metabolism as is possible. For this reason, the quantitative assessment of the concentrations of intracellular and extracellular metabolites, the fluxes through the interconnected pathways, and the transcription of genes catalyzing and regulating individual reactions, is of vital importance. This allows for a better determination of both how the cells are currently operating, and where the areas of strain and process improvement with the most promise lie.

Recent advances in the knowledge of the genomics of *C. glutamicum*, including the complete sequencing of the DNA of the species [Kalinowski *et al.*, 2003], have provided an excellent opportunity to probe the workings of the cells on a more comprehensive genetic level. New genome-based approaches to the study of *C. glutamicum* should lead to an increasingly broad, yet more targeted, view of those aspects of the physiology of the organism that are most promising for strain improvement through metabolic engineering. One of the key techniques to be applied for this purpose is that of DNA microarrays, which allow for the simultaneous quantification of the expression levels of all genes

found within an organism. Through the data provided by DNA microarrays, valuable insight can be gained into the gene expression phenotypes associated with desired culture properties, as well as the transcriptional responses to various process demands or environmental stresses of interest.

1.2 Objectives

The general goal of the research work of this thesis was to create and apply partial genome DNA microarrays specific for *Corynebacterium glutamicum* to the study of various strains and fermentation conditions known to affect the production of lysine by the organism. In accomplishing this, the following specific objectives were set:

(1) Design, construct, and validate *C. glutamicum* DNA microarrays, determining which parameters of the protocols used to manufacture and apply the arrays significantly impact the robustness of the technique, and quantifying the reliability of the generated data.

(2) Apply the DNA microarrays to the measurement of gene expression phenotypes of different genetically engineered strains in response to internal and external perturbations.

(3) Evaluate the relationship between transcriptional profiles and flux profiles as descriptors of metabolic network function, describing the correlation between the two types of data sets.

1.3 Approach

To achieve the first of the objectives listed in the previous section, partial genome *C. glutamicum* DNA microarrays were produced that were capable of measuring the mRNA abundances of genes directly involved in the pathways converting glucose and lactate into lysine. These interconnected pathways included the reactions of glycolysis, gluconeogenesis, the pentose phosphate pathway, the tricarboxylic acid cycle, the

glyoxylate bypass, central carbon anaplerosis, and lysine and biotin biosynthesis. Also included in the arrays were spots used to test the effects of printing sequences of different length as well as sequences originating from different species. The protocols that were optimized to reduce nonspecific results and data variability included those of the polymerase chain reaction used to amplify the DNA sequences printed onto the array surface, the printing pattern and use of replicates among spots comprising the array, the isolation and purification of total RNA samples from cultures being studied, the creation of labeled cDNA through reverse transcription with fluorescent nucleotide analogs, the hybridization of the cDNA to the spots of the array surface, and the filtering and normalization associated with microarray data analysis. The microarrays were validated through the use of several types of replicates and the testing of the ability of arrays to accurately describe known biological behaviors.

To achieve the second of the objectives, the DNA microarrays were applied to the study of three different strains grown on six different media compositions and sampled during three different phases of the fermentation process. The tested *C. glutamicum* strains included one (ATCC 21253) that had been designed for lysine over-production through the deletion of genes providing access to competing amino acid synthesis pathways, and another (ATCC 21253[pKD7]) that had been given the trait of over-expressing the key anaplerotic enzyme pyruvate carboxylase. In this way, the arrays could be used to test the effects of over-expression of a single important gene on the behavior of the entire gene population associated with the metabolic network. Also studied were media that differed in their composition of glucose and lactate as available carbon sources. Lactate had been previously shown to influence the activities of some genes within the network and the productivity of lysine achieved by cell cultures.

To achieve the third objective, measurements of metabolite balances and mass isotopomer distribution patterns within cells that had been cultivated on isotopically labeled substrates were used in metabolic network analyses to calculate the flux rates associated with the reactions of the network. These flux values were then compared to transcript concentration values from DNA microarray assays of the same set of culture samples.

1.4 Description of Thesis

The remainder of the thesis is divided into eight additional chapters, covering background information, experimental results related to the approaches described in the previous section, conclusions, references, and supplemental material. Chapter 2 presents background material for both *C. glutamicum* and DNA microarrays. The reactions involved in the metabolic network, and some of the strategies that the cells use to regulate them, are discussed. Also, the previously published effects of various carbon source substrates as well as the over-expression of the genes aspartokinase and pyruvate carboxylase are described. Chapter 3 involves the work done in developing and manufacturing the DNA microarrays used in all experiments of the thesis work. The many procedures involved in the microarray printing and use are detailed, and the results of investigations into improving these procedures are presented. Chapter 4 deals with the experimental validation of the arrays through different types of replicates and by using the arrays to describe the transcriptional profiles of *C. glutamicum* cultivated with either glucose or lactate as the only available carbon source. Glucose and lactate are again studied in the experiment discussed in Chapter 5, along with the effects of pyruvate carboxylase over-expression. Here, several different media are used, with various mixtures of the two carbon source substrates. In the experiment of Chapter 6, RNA samples for DNA microarray analysis were taken at three different time points so that changes in gene transcription related to the phase of the culture could be revealed. Chapter 7 describes the work done in comparing the transcriptional profile and flux profile of a set of four different cultures. This allowed for a measure of the degree to which flux and mRNA levels correlate with one another, and a discussion of the ability of DNA microarray data to accurately reflect the behavior of metabolic flux. Chapter 8 gives a summary of the experimental results from the previous five chapters, lists the primary conclusions of the thesis, and suggests promising areas for future work. Chapter 9 is comprised of a list of references and appendices giving the abbreviations used throughout this text and the list of reactions used in the metabolic network analysis model of Chapter 7.

Chapter 2: Background

2.1 Corynebacterium glutamicum

2.1.1 Overview

Corynebacterium glutamicum is a gram-positive, non-sporulating, non-pathogenic bacterium that was first noted for its ability to produce and excrete high concentrations of the amino acid, glutamate [Kinoshita *et al.*, 1957]. Today the organism is being utilized industrially for the production of several amino acids, such as L-lysine which is used predominantly as a feed additive. Lysine is typically added to agricultural feed stocks such as barley, corn, and wheat that are naturally lacking in the nutritionally required amino acid. A consequence of supplementing these stocks with lysine is that manure generated from domesticated livestock contains significantly less nitrogen, and thus, has a diminished environmental impact. The demand for lysine has increased by a factor of 20 worldwide since the 1980s to a level on the order of 10^6 tons per year, with the current rate of annual market growth near 10% [Hermann, 2003].

Along with *Mycobacterium*, *Streptomyces*, *Nocardia*, *Propionibacterium*, and *Arthrobacter*, *Corynebacteria* belong to the actinomycetes family. Typical characteristics of actinomycetes such as *Corynebacteria* include a relatively high GC content in their genomic DNA, morphology similar to a small and slightly curved rod, facultatively anaerobic to aerobic growth, cell walls composed primarily of arabinose and galactose, and the presence of mycolic acids [Collins and Cummings, 1986]. *Corynebacteria* have been isolated from a wide variety of sources including soil, water, and plant and animal material, and they can be cultivated without significant difficulty provided that the growth medium is supplemented with biotin [Abe *et al.*, 1967]. In addition to its industrial significance, another factor that had made the organism a very attractive target for research was its apparent lesser metabolic complexity relative to that found in other species. Until only recently, it was believed that the cells contained relatively few

isozymes, simplifying the task of understanding and influencing the control strategies at work in regulating the operation of metabolic reactions. With the entire genome sequence of the species now available, however, and the assigning of functions to previously unknown open reading frames underway, this belief has received new scrutiny [Kalinowski *et al.*, 2003].

A large body of research has been compiled dealing with the species, resulting in improvements both to the genetic makeup of microbial strains and the operating protocols of fermentation processes used to grow the cultures and generate products such as lysine, glutamate, and threonine [Ikeda, 2003; Kimura, 2003; Pfefferle *et al.*, 2003; and Debabov, 2003]. It has been found that the production of lysine is very dependent on the intracellular supplies of pyruvate, oxaloacetate, and NADPH reducing equivalents to fuel the reactions and pathways associated with biosynthesis of the amino acid. Classical strain improvement techniques, such as those of random mutagenesis and selection, have targeted these key areas and yielded producers of lysine that have been very successfully integrated into commercial processes. Still, there is a great potential for further advances using the newer techniques for improving strains that have arisen in the past decade with more advanced tools of genetic engineering, metabolic engineering, and bioinformatics.

This chapter will present background material concerning the reactions of several of the major reaction pathways that constitute the primary metabolism of the cells and are necessary for the support of both biomass and lysine production. As a great deal of the focus of the experiments of this thesis lies with the ways in which the cells control enzyme activity, especially at the level of gene transcription, special mention will be made of some other modes known to be utilized by the cells for enzyme regulation. The effects of various carbon sources on the functioning of the metabolic network, as well as the results of over-expressing a pair of key enzymes, will also be discussed, as they factor strongly in the studies and results detailed in later chapters. Finally, some additional background is given in the area of DNA microarrays and their applicability in probing the expression of genes whose protein products are responsible for catalyzing the reactions just described.

Table 2-1: Glycolytic reactions of the Embden-Meyerhof-Parnas pathway and the pyruvate dehydrogenase coupling reaction. Enzyme catalysts are glucokinase (glk), phosphoglucoisomerase (gpi), phosphofructokinase (pfk), fructose bisphosphate aldolase (fba), triosephosphate isomerase (tpi), glyceraldehyde 3-phosphate dehydrogenase (gap), phosphoglycerate kinase (pgk), phosphoglycerate mutase (gpm), enolase (eno), pyruvate kinase (pyk), and pyruvate dehydrogenase (pdh).

Reaction	Catalyst
Glucose + ATP → Glucose 6-phosphate + ADP	glk
Glucose 6-phosphate ↔ Fructose 6-phosphate	gpi
Fructose 6-phosphate + ATP → Fructose-1,6-bisphosphate +ADP	pfk
Fructose-1,6-bisphosphate ↔ Glyceraldehyde 3-phosphate + Dihydroxyacetone phosphate	fba
Dihydroxyacetone phosphate ↔ Glyceraldehyde 3-phosphate	tpi
Glyceraldehyde 3-phosphate + NAD ↔ Bisphosphoglycerate + NADH	gap
Bisphosphoglycerate + ADP ↔ 3-Phosphoglycerate + ATP	pgk
3-Phosphogycerate ↔ 2-Phosphoglycerate	gpm
2-Phosphoglycerate ↔ Phosphoenolpyruvate	eno
Phosphoenolpyruvate + ADP → Pyruvate + ATP	pyk
Pyruvate + CoA +NADP → Acetyl-CoA + CO_2 + NADPH	pdh
Overall: Glucose + 2 ADP + 2 P_i + 2 NAD + CoA + NADP → 2 Acetyl-CoA + 2 ATP + 2 NADH + 2 CO_2 + 2 NADPH + 2 H^+ + 2 H_2O	

2.1.2 Biochemistry

2.1.2.1 Embden-Meyerhof-Parnas Pathway

The Embden-Meyerhof-Parnas pathway is composed of the series of reactions used by the cells to convert glucose to two molecules of pyruvate while producing energy in the form of ATP. The steps of the pathway, as well as the enzymes used in their catalysis, are given in Table 2-1. It is also possible for the cells to operate a slightly modified form of the pathway in the opposite direction, with glucose as the final product rather than the starting material. This gluconeogenic operation, the reactions of which are given in Table 2-2, is used when cells are grown on carbon sources other than glucose. Some metabolites of the glycolytic pathway and the related pentose phosphate pathway described in section 2.1.2.2 are required for cellular growth and maintenance. Because glycolysis is inaccessible for substrates other than glucose, when glucose is unavailable the cells must use this pathway in its energetically unfavorable direction to produce these

Table 2-2: Gluconeogenic reactions of the Embden-Meyerhof-Parnas pathway. Enzyme catalysts are enolase (eno), phosphoglycerate mutase (gpm), phosphoglycerate kinase (pgk), glyceraldehyde 3-phosphate dehydrogenase (gap), triosephosphate isomerase (tpi), fructose bisphosphate aldolase (fba), fructose bisphosphatase (fbpc), phosphoglucoisomerase (gpi), glucose phosphatase (gpc).

Reaction	Catalyst
Phosphoenolpyruvate $\leftrightarrow$ 2-Phosphoglycerate	eno
2-Phosphoglycerate $\leftrightarrow$ 3-Phosphoglycerate	gpm
3-Phosphoglycerate + ATP $\leftrightarrow$ Bisphosphoglycerate + ADP	pgk
Bisphosphoglycerate + NADH $\leftrightarrow$ Glyceraldehyde 3-phosphate +NAD	gap
Glyceraldehyde 3-phosphate $\leftrightarrow$ Dihydroxyacetone phosphate	tpi
Glyceraldehyde 3-phosphate + Dihydroxyacetone phosphate $\leftrightarrow$ Fructose-1,6-bisphosphate	fba
Fructose-1,6-bisphosphate + H_2O $\rightarrow$ Fructose 6-phosphate + P_i	fbpc
Fructose 6-phosphate $\leftrightarrow$ Glucose 6-phosphate	gpi
Glucose 6-phosphate + H_2O $\rightarrow$ Glucose + P_i	gpc
Overall: 2 Phosphoenolpyruvate + 2 ATP + 2 NADH + 2 H_2O $\rightarrow$ Glucose + 2 ADP + 2 NAD + 2 P_i	

intermediates. The three most energetically irreversible reactions of glycolysis—those that involve the actions of glucokinase, phosphofructokinase, and pyruvate kinase—do not function in gluconeogenesis, and the reverse reactions are catalyzed instead by glucose phosphatase, fructose bisphosphatase, and phosphoenolpyruvate carboxykinase, discussed in section 2.1.2.4.4.

As was stated in section 2.1.1, the concentration of pyruvate and its use as a substrate for other reaction pathways are important in determining lysine productivities. For example, a mutant strain engineered to possess reduced activity of the pyruvate dehydrogenase complex was found to have a greatly enhanced productivity of lysine [Shiio *et al.*, 1984]. The activity of pyruvate kinase is positively influenced by high concentrations of AMP and phosphoenolpyruvate and negatively influenced by ATP [Jetten *et al.*, 1994]. This influence of ATP amounts to a type of feedback inhibition on glycolysis as a whole, as ATP is one of the major products of the functioning of the pathway.

Table 2-3: Reactions of the pentose phosphate pathway. Enzyme catalysts are glucose 6-phosphate dehydrogenase (zwf), phosphogluconolactonase (pgl), phosphogluconate dehydrogenase (gnd), ribulose phosphate epimerase (rpe), ribulose phosphate isomerase (rpi), transketolase (tkt), and transaldolase (tal).

Reaction	Catalyst
Glucose 6-phosphate + NADP → 6-Phosphogluconolactone	zwf
6-Phosphogluconolactone + H_2O → 6-Phosphogluconate + H^+	pgl
6-Phosphogluconolactonate + NADP → Ribulose 5-phosphate + NADPH + CO_2	gnd
Ribulose 5-phosphate ↔ Xylulose 5-phosphate	rpe
Ribulose 5-phosphate ↔ Ribose 5-phosphate	rpi
Xylulose 5-phosphate + Ribose 5-phosphate ↔ Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate	tkt
Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate ↔ Erythrose 4-phosphate + Fructose 6-phosphate	tal
Erythrose 4-phosphate + Xylulose 5-phosphate ↔ Glyceraldehyde 4-phoshate + Fructose 6-phosphate	tkt
Overall: 3 Glucose 6-phosphate + 6 NADP + 3 H_2O → 2 Fructose 6-phosphate + Glyceraldehyde 3-phosphate + 6 NADPH + 6 H^+ + 3 CO_2	

2.1.2.2 Pentose Phosphate Pathway

The pentose phosphate pathway is used by the cells for two primary purposes—the generation of reducing equivalents in the form of NADPH necessary for reductive biosynthesis reactions, and the formation of ribose-5-phosphate, an essential component of nucleotides and nucleic acids. The pathway, the reactions of which are given in Table 2-3, has been shown by flux analysis to be one of the dominant pathways of cultures producing lysine from glucose carbon substrates [Marx et al., 1996]. Accordingly, it is the primary source of NADPH in C. glutamicum during growth and lysine production on glucose. Evidence for this includes the low fluxes that have been measured for the reactions of malic enzyme and isocitrate dehydrogenase, the other two enzymes of central carbon metabolism that catalyze the formation of the reducing equivalent [Park et al., 1997].

Although the pentose phosphate pathway is dispensable for lysine formation, another work found that by deleting the phosphoglucoisomerase gene, thereby blocking the entry of glucose into glycolysis and redirecting flow into the pentose phosphate pathway, a subsequent increase in lysine yields was realized [Marx et al., 2003]. In general the

partitioning of central carbon flux between the pentose phosphate pathway and the glycolytic reactions of the Embden-Meyerhof-Parnas pathway has been shown to be very influential on the growth and product formation realized by *Corynebacteria*, with strains of *C. glutamicum* designed for lysine production and *C. ammoniagenes* designed for inosine production that have each been engineered for enhanced pentose phosphate pathway activity exhibiting greater productivities but reduced growth rates [Marx *et al.*, 2003 and Kamada *et al.*, 2001]. The ability of the pathway to generate reducing equivalents can help determine not only whether a culture is better suited for generating product or biomass, but also which product is the one most easily generated. For example, lower pathway activity has been linked with glutamate production, while elevated pathway activity has been shown to promote lysine production [Sahm *et al.*, 2000]. A likely explanation for this is that while glutamate synthesis requires one mol of NADPH per mol product, lysine synthesis requires a much greater four mol of NADPH per mol product.

Glucose-6-phosphate dehydrogenase, which catalyzes the first step of the pentose phosphate pathway, is regulated at least in part through feedback inhibition by the primary product of the pathway, NADPH [Moritz *et al.*, 2000]. The enzyme is also inhibited by elevated oxaloacetate concentrations, and induced by increased glucose [Sugimoto and Shiio, 1987]. Transketolase has been found to be inducible by the carbon sources acetate, fructose, and citrate [Sugimoto and Shiio, 1989].

2.1.2.3 Tricarboxylic Acid Cycle

The tricarboxylic acid cycle, described in Table 2-4, is used by the cells to oxidize acetyl-CoA completely to CO_2, while generating the reducing equivalents NADH and $FADH_2$. Acetyl-CoA is derived from pyruvate of the Embden-Meyerhof-Parnas pathway through the pyruvate dehydrogenase coupling reaction detailed in Table 2-1. Several intermediates of the tricarboxylic acid cycle also have significant roles as precursors for biosynthetic pathways.

The succinate dehydrogenation step is catalyzed by a succinate dehydrogenase complex

Table 2-4: Reactions of the tricarboxylic acid cycle. Enzyme catalysts are citrate synthase (citA), aconitase (aco), isocitrate dehydrogenase (icd), α-ketoglutarate dehydrogenase (sucA), succinyl-coA synthase (sucC), succinate dehydrogenase (sdhABCD), fumarase (fum), and malate dehydrogenase (mdh).

Reaction	Catalyst
Acetyl-CoA + Oxaloacetate → Citrate + CoA	citA
Citrate ↔ Isocitrate	aco
Isocitrate + NAD → α-Ketoglutarate + NADH + CO_2	icd
α-Ketoglutarate + NAD + CoA → Succinyl-CoA + NADH + CO_2	sucA
Succinyl-CoA + GDP → Succinate + GTP + CoA	sucC
Succinate + FAD ↔ Fumarate + $FADH_2$	sdhABCD
Fumarate + H_2O ↔ Malate	fum
Malate + NAD ↔ Oxaloacetate + NADH	mdh
Overall: Acetyl-CoA + 3 NAD + FAD + GDP + H_2O → 3 NADH + $FADH_2$ +GTP + 2 CO_2	

that is encoded by the four genes sdhA, sdhB, sdhC, and sdhD [Kalinowski *et al.*, 2003]. The activity of citrate synthase, located at the entry point of the tricarboxylic acid cycle, is inhibited by high concentrations of ATP, aconitate [Eikmanns *et al.*, 1994], isocitrate [Shiio *et al.*, 1977], and glutamate [Shiio and Ujigawa, 1979]. Isocitrate dehydrogenase, located at the branch point between the tricarboxylic acid cycle and the glyoxylate bypass pathway, has been demonstrated to be synergistically inhibited by glyoxylate and oxaloacetate [Shiio and Ozaki, 1968].

The activity of α-ketoglutarate dehydrogenase, located at the metabolic branch point between the anabolic pathways of glutamate synthesis and the catabolic pathways of the TCA cycle, is lowered when cultures are grown under conditions of biotin limitation [Shingu and Terui, 1971]. A similar repression of the dehydrogenase activity is seen in response to the addition of penicillin or detergents to the culture medium [Kawahara *et al.*, 1997]. This is particularly significant for amino acid synthesis because of the three reactions sharing α-ketoglutarate as a substrate or product. The activity of α-ketoglutarate dehydrogenase was found to have a strong effect on the synthesis of glutamate by *C. glutamicum* cultures [Shimizu *et al.*, 2003]. It has been shown that by attenuating the activity of this enzyme the flux through the downstream tricarboxylic acid cycle reactions is decreased, and the production of glutamate is increased. The dehydrogenase is inducible by high levels of glutamate [Shiio and Ujigawa-Takeda, 1980], suggesting that

if glutamate concentrations become excessive, the metabolism of the cells shifts to divert α-ketoglutarate away from further glutamate synthesis and towards the remaining reactions of the tricarboxylic acid cycle.

2.1.2.4 Anaplerotic Pathways

2.1.1.4.1 Overview

Because all carbon skeletons that complete the tricarboxylic acid cycle are converted to carbon dioxide, there is no associated net production of any of the constituent metabolites. Since several of these metabolites are substrates for other anabolic pathways, such as those of amino acid synthesis, anaplerotic reactions must be used to replenish pools of, in particular, malate and oxaloacetate.

A pair of enzymes—phosphoenolpyruvate carboxylase and pyruvate carboxylase—is available to the cells for use in generating oxaloacetate from intermediates of the glycolytic pathway. Conversely, oxaloacetate may be converted to phosphoenolpyruvate or pyruvate through reactions catalyzed by phosphoenolpyruvate carboxykinase and oxaloacetate decarboxylase, respectively. Another route for synthesizing pyruvate from a tricarboxylic acid cycle intermediate is through the action of the NAD^+-dependent or $NADP^+$-dependent malic enzyme. Previous studies have shown that *C. glutamicum* is unique, in that it is possible for all three of these types of reactions to operate simultaneously [Marx *et al.*, 1997 and 1999].

Another pair of anaplerotic reactions converts isocitrate to malate in a two-step route, rather than the five-step route used by the tricarboxylic acid cycle. Details of these two reactions, as well as of the five reactions linking pyruvate and phosphoenolpyruvate to oxaloacetate and malate, are given in Table 2-5.

Table 2-5: Reactions of the anaplerotic pathways. Enzyme catalysts are pyruvate carboxylase (pyc), phosphoenolpyruvate carboxylase (ppc), oxaloacetate decarboxylase (oad), malic enzyme (malE), pyruvate carboxykinase (pck), isocitrate lyase (aceA), and malate synthase (aceB).

Reaction	Catalyst
Pyruvate + CO_2 + ATP → Oxaloacetate + ADP + P_i	pyc
Phosphoenolpyruvate + CO_2 + H_2O → Oxaloacetate + P_i	ppc
Oxaloacetate → Pyruvate + CO_2	oad
Malate + NADP → Pyruvate + CO_2 + NADPH	malE
Oxaloacetate + ATP → Phosphoenolpyruvate + ADP + CO_2	pck
Isocitrate → Succinate + Glyoxylate	aceA
Glyoxylate + Acetyl-CoA + H_2O → Malate + CoA	aceB

2.1.2.4.2 Pyruvate Carboxylase and PEP Carboxylase

Pyruvate carboxylase activity in *C. glutamicum* was first theorized by Tosaka *et al* in 1979, although 18 years passed before this activity was eventually confirmed [Peters-Wendisch *et al.*, 1997] and the gene encoding the enzyme was identified [Koffas *et al.*, 1998 and Peters-Wendisch *et al.*, 1998]. Pyruvate carboxylase is the primary anaplerotic means for the cells to regenerate intracellular oxaloacetate pools that have been depleted through flow into either the TCA cycle or biosynthetic pathways such as that of lysine production [Park *et al.*, 1997; Peters-Wendisch *et al.*, 1997]. Measurements of the *in vivo* metabolic fluxes of pyruvate and PEP carboxylase present in *C. glutamicum* growing on glucose found the contribution of pyruvate carboxylase to oxaloacetate formation to be at least 3-fold greater than that of PEP carboxylase [Petersen *et al.*, 2000]. This is despite *in vitro* data suggesting that PEP carboxylase is the more active of the two enzymes [Peters-Wendisch, 1993 and 1997]. The same study also concluded that catalysis by pyruvate carboxylase accounted for >80% of all anaplerotic activity within the network. Also, the deletion of PEP carboxylase was shown to inhibit growth or lysine production only under conditions of biotin limitation [Peters-Wendisch *et al.*, 1993 and Gubler *et al.*, 1994].

Relative to pyruvate carboxylase, PEP carboxylase appears to play a less significant role in the metabolic flux within *C. glutamicum*, but the very different responses of the two enzymes to changes in concentrations of other intracellular compounds suggests that the cells uses them in tandem in implementing a regulatory strategy. For example,

pyruvate carboxylase is inhibited by elevated levels of ADP, AMP, and acetyl-CoA [Peters-Wendisch *et al.*, 1997] while PEP carboxylase is conversely activated by the presence of acetyl-CoA as well as fructose-1,6-bisphosphate and inhibited by malate, succinate, α-ketoglutarate, aspartate [Mori and Shiio, 1985; Peters-Wendisch, 1996], and glutamate [Delaunay *et al.*, 2004]. Also, there is evidence that PEP carboxylase is actually the dominant anaplerotic enzyme of the microorganism under conditions of oxygen depravation [Inui *et al.*, 2004]. When both of these carboxylases are deleted from the genome of *C. glutamicum*, the organism is no longer capable of growth on glucose, indicating that together they are the only two required anaplerotic enzymes on this carbon source [Eggeling and Sahm, 1999].

2.1.2.4.3 Malic Enzyme

Malic enzyme catalyzes the decarboxylation reaction linking malate of the tricarboxylic acid cycle with pyruvate of the Embden-Meyerhof-Parnas pathway. Activity of the enzyme, which shows slight inhibition by oxaloacetate and glutamate, is associated with the generation of the reducing equivalent NADPH [Gourdon *et al.*, 2000].

2.1.2.4.4 Phosphoenolpyruvate Carboxykinase

Phosphoenolpyruvate carboxykinase catalyzes the conversion of oxaloacetate to phosphoenolpyruvate in a reaction that is assumed to not operate in the reverse direction due to strong inhibition by ATP [Jetten and Sinskey, 1993]. An isotopomer analysis and metabolite balancing study of *C. glutamicum* grown on glucose showed that phosphoenolpyruvate carboxykinase was the primary reverse anaplerotic component of the metabolic network, accounting for >90% of the flow of metabolites from the tricarboxylic acid cycle to either phosphoenolpyruvate or pyruvate as measured in an *in vivo* analysis of fluxes during growth on glucose [Petersen *et al.*, 2000]. That work described the carboxykinase flux as one component of a futile cycle along with that of pyruvate kinase and pyruvate carboxylase when glucose is the medium carbon source,

consuming one molecule of ATP for each circuit from phosphoenolpyruvate to pyruvate, oxaloacetate, and then again to phosphoenolpyruvate. For growth on substrates that require gluconeogenesis, though, the activity of phosphoenolpyruvate carboxykinase has been found to be essential [Reidel *et al.*, 2001].

2.1.2.4.5 Glyoxylate Bypass

The two enzymes catalyzing the glyoxylate bypass pathway—isocitrate lyase and malate synthase—have been isolated and characterized, with the lyase being shown to be allosterically inhibited by several intermediates of the Embden-Meyerhof-Parnas pathway, such as 3-phosphoglycerate, 6-phosphogluconate, phosphoenolpyruvate, and fructose-1,6-bisphosphate [Reinscheid *et al.*, 1994a and 1994b]. The genes for isocitrate lyase and malate synthesis are located in close proximity to one another in the *C. glutamicum* genome, and they are believed to share a regulatory sequence, from which they are controlled together at the transcriptional level [Wendish *et al.*, 1997].

A novel gene, whose product, glxR, appears to have a regulatory role in controlling the transcriptional promotion of malate synthesis, was recently discovered and characterized [Kim *et al.*, 2004]. The sequence of this putative regulator was found to have a motif with strong resemblance to that of known cAMP-binding domains of other proteins. When lysine-producing cultures grown on acetate were exposed to both cAMP and glxR, the activities of both the isocitrate lyase and malate synthesis enzymes decreased by approximately 90%. Although it is very common for cAMP to be involved in various regulatory strategies functioning in many different species, in most cases the concentration of the molecule is associated with conditions in which a key substrate, such as glucose, has limited availability. In contrast, in *Corynebacteria*, cAMP is found to be at its highest concentration during exponential growth on glucose, when the carbon source is present in abundance [Lynch *et al.*, 1975; Peters *et al.*, 1991; and Kim *et al.*, 2004]. This also coincides with the condition for which *C. glutamicum* has been shown to have its lowest glyoxylate bypass activity [Sonntag *et al.*, 1995].

Table 2-6: Reactions of the lysine biosynthetic pathways. Enzyme catalysts are glutamate dehydrogenase (gdh), aspartate aminotransferase (aat), aspartokinase (ask), aspartate semialdehyde dehydrogenase (asd), dihydropicolinate synthase (dapA), dihydropicolinate reductase (dapB), tetrahydropicolinate succinylase (dapD), succinyldiaminopimelate aminotransferase (dapC), diaminopimelate epimerase (dapF), diaminopimelate dehydrogenase (ddh), and diaminopimelate decarboxylase (lysA).

Reaction	Catalyst
α-Ketoglutarate + Ammonia + NADPH ↔ Glutamate + NADP + H_2O	gdh
Glutamate+ Oxaloacetate ↔ Aspartate + α-Ketoglutarate	aat
Aspartate + ATP ↔ Aspartyl phosphate	ask
Aspartyl phosphate + NADPH ↔ Aspartate semialdehyde + P_i + NADP	asd
Aspartate semialdehyde + Pyruvate → Dihydropicolinate + 2 H_2O	dapA
Dihydropicolinate + NADPH + H^+ → Tetrahydropicolinate + NADP	dapB
Tetrahydropicolinate + Succinyl-CoA + H_2O → Succinyl-amino-oxopimelate +CoA	dapD
Succinyl-amino-oxopimelate + Glutamate → Succinyl-diaminopimelate + α-Ketoglutarate	dapC
Diaminopimelate → Meso-diaminopimelate	dapF
Tetrahydropicolinate + Ammonia + NADPH + H^+ → Meso-diaminopimelate + NADP + H_2O	ddh
Meso-diaminopimelate → Lysine + CO_2	lysA

Overall for succinylation route: Oxaloacetate + Pyruvate + Glutamate +Ammonia + ATP + 3 NADPH + NAD → Lysine + Succinate + ADP + 3 NADP + NADH + P_i + 2 CO_2 + H_2O

Overall for dehydrogenation route: Oxaloacetate + Pyruvate + 2 Ammonia + ATP + 4 NADPH + H^+ → Lysine + ADP + 4 NADP + + P_i + CO_2 + 4 H_2O

2.1.2.5 Lysine Synthesis Pathway

C. glutamicum utilizes a split pathway in the synthesis of lysine, by allowing the tetrahydropicolinate lysine precursor to be converted to diaminopimelate by either a one-step dehydrogenation reaction catalyzed by meso-diaminopimelate dehydrogenase (ddh), or by a four-step succinylation branch catalyzed by tetrahydrodipicolinate succinylase (dapD), succinyldiaminopimelate aminotransferase (dapC), succinyldiaminopimelate desuccinylase (dapD), and diaminopimelate epimerase (dapF) [Schrumph *et al.*, 1991]. The reactions involved in each of these routes, as well as those for steps upstream and downstream of the split portion of the pathway, are described in detail in Table 2-6. Flux analyses of the branched pathway found that both options are utilized by the cells in parallel with individual contributions typically ranging from 30-70% of the combined flux [Sonntag *et al.*, 1993]. Cultivation time and the availability of ammonium ions required by the ammonium-incorporating dehydrogenase were shown to be two of the

greatest factors in influencing which of the two branches is the preferred route [Wehrmann *et al.*, 1998]. Of the two pathways, the dehydrogenase route is the less costly from an energetics viewpoint and is the preferred route when ammonium is present in excess. The existence of the alternative succinylation route greatly increases the metabolic flexibility that the organism can use in adapting to different growth environments, and helps ensure that not only lysine but also the lysine precursor diaminopimelate, itself an important component of cell wall synthesis, are generated in sufficient quantities under a wide range of conditions.

Many of the genes for enzymes catalyzing the reactions of the lysine biosynthetic pathway are present in clusters within the *C. glutamicum* genome, increasing the probability that their expression levels are regulated simultaneously by the cells. Examples of these clusters include one containing the genes for aspartokinase and aspartate semialdehyde dehydrogenase [Cremer *et al.*, 1991 and Follettie *et al.*, 1993] and another containing the genes for dihydropicolinate synthase and reductase [Patek *et al.*, 1997].

The results of several experiments have led to the conclusion that three of the enzymes in the lysine biosynthetic pathway—aspartokinase, dihydropicolinate synthase, and the lysine exporter—are responsible for catalyzing steps likely to be rate limiting [Cremer *et al.*, 1991; Jetten *et al.*, 1995; and Malumbres and Martin, 1996]. In particular, aspartokinase is one of the most heavily regulated enzymes catalyzing the anabolic pathway, displaying susceptibility to feedback inhibition by the downstream products lysine and threonine [Shiio and Miyajima, 1969 and Kalinowski *et al.*, 1991]. A key strategy implemented in the course of improving strains for lysine over-production was the replacement of the native aspartokinase enzyme with a mutated form selected for its resistance to this feedback inhibition [Shiio *et al.*, 1970; Tosaka *et al.*, 1978; and Follettie *et al.*, 1993]. This selection was based on resistance to S-(2-aminoethyl) L-cysteine, a lysine antimetabolite.

Table 2-7: Reaction of lactate consumption. Enzyme catalyst is lactate dehydrogenase (ldh).

Reaction	Catalyst
Lactate + NAD $\leftrightarrow$ Pyruvate + NADH + H^+	ldh

Table 2-8: Reactions of biotin biosynthesis. Enzyme catalysts are 7,8-diaminopelargonic acid aminotransferase (bioA), dethiobiotin synthase (bioD), and biotin synthase (bioB).

Reaction	Catalyst
Adenosylmethionine + 8-Amino-7-oxononanoate $\leftrightarrow$ Adenosyl-4-methylthio-2-oxobutanoate + 7,8-Diaminononanoate	bioA
7,8-Diaminononanoate + ATP + CO_2 $\leftrightarrow$ Dethiobiotin + ADP + P_i	bioD
Dethiobiotin + Sulfur $\leftrightarrow$ Biotin	bioB

2.1.2.6 Other Pathways

The previous several sections describe the reactions directly involved in the pathways used by the cells to convert glucose substrate to biomass and lysine products. There are other reactions and pathways that are also significant to the work done in this thesis, and they are described in this section. The first of these is the dehydrogenation of lactate in a NAD-dependent reaction that allows lactate substrate to enter central carbon metabolism at the point of pyruvate. The reaction is shown in Table 2-7.

Biotin is a cofactor required by carboxylation enzymes, such as the pyruvate carboxylase and phosphoenolpyruvate carboxylase described in section 2.1.2.4.2. A portion of the pathway for synthesizing biotin is given in Table 2-8. The pathway is incomplete because *C. glutamicum* is incapable of generating the 8-amino-7-oxononanoate that is a substrate for the first reaction listed. This is why the organism requires biotin supplementation in its growth medium, as was described in section 2.1.1.

Table 2-9 lists the reactions used by the microorganism to produce trehalose, a disaccharide that is typically used by cells as a glucose storage compound. In fermentations operated to produce glutamate and lysine, however, trehalose has been found to accumulate extracellularly [Walker *et al.*, 1982 and Inbar *et al.*, 1985].

Table 2-9: Reactions of the trehalose biosynthesis pathway. Enzyme catalysts are phosphoglycerate mutase (gpm), UDP glucose pyrophosphorylase (galU), trehalose phosphate synthase (otsA), and trehalose phosphatase (otsB).

Reaction	Catalyst
Glucose 6-phosphate ↔ Glucose 1-phosphate	gpm
Glucose 1-phosphate + UTP ↔ UDP-glucose + PP_i	galU
Glucose 6-phosphate + UDP-glucose ↔ Trehalose 6-phosphate + UDP	otsA
Trehalose 6-phosphate + H_2O ↔ Trehalose + P_i	otsB
Overall: 2 Glucose 6-phosphate + UTP + H_2O ↔ Trehalose + UDP + P_i + PP_i	

Table 2-10: Reactions of the acetate biosynthesis pathway. Enzyme catalysts are phosphate acetyltransferase (pta) and acetate kinase (ack).

Reaction	Catalyst
Acetyl-CoA + P_i ↔ Acetyl phosphate + CoA	pta
Acetyl phosphate + ADP ↔ Acetate + ATP	ack
Overall: Acetyl-CoA + ADP + P_i ↔ Acetate + CoA + ATP	

Measurements of trehalose, as well as of acetate, alanine, glutamine, valine, and phenylalanine which are all described below, were important for this work in the metabolite balancing of the flux analysis experiment of Chapter 7.

Acetate production in *C. glutamicum* was determined to take place through the actions of a phosphate acetyltransferase and an acetate kinase, as is shown in Table 2-10 [Shiio *et al.*, 1969].

The route of biosynthesis of alanine in the cells is given in Table 2-11. Of the potential routes of alanine dehydrogenation, aspartate decarboxylation, and pyruvate transamination, the latter was found to be in effect in *C. glutamicum* for generating the amino acid [Tosaka *et al.*, 1978].

Glutamine is synthesized in the energy-dependent reaction of Table 2-12 through the addition of ammonia to glutamate.

A four-step process, shown in Table 2-13, is used to generate valine from pyruvate and glutamate precursors [Tsuchida and Momose, 1975].

Table 2-11: Reaction of alanine biosynthesis. Enzyme catalyst is alanine aminotransferase (gpt).

Reaction	Catalyst
Pyruvate + Glutamate ↔ Alanine + α-Ketoglutarate	gpt

Table 2-12: Reaction of glutamine biosynthesis. Enzyme catalyst is glutamine synthetase (glnA).

Reaction	Catalyst
Glutamate + Ammonia + ATP ↔ Glutamine + ADP + P_i	glnA

Table 2-13: Reactions of the valine biosynthesis pathway. Enzyme catalysts are acetohydroxyacid synthase (ilv) acetohydroxyacid isomeroreductase (ilvC), acetohydroxyacid dehydratase (ilvD), and transaminase B (ilvB).

Reaction	Catalyst
2 Pyruvate ↔ Acetolactate + CO_2	ilv
Acetolactate + NADPH + H^+ ↔ Dihydroxyisovalerate +NADP	ilvC
Dihydroxyisovalerate ↔ α-Ketovaline + H_2O	ilvD
α-Ketovaline + Glutamine ↔ Valine + α-Ketoglutarate	ilvB
Overall: 2 Pyruvate + Glutamine + NADPH + H^+ ↔ Valine + α-Ketoglutarate + NADP+ H_2O + CO_2	

Finally, phenylalanine and α-ketoglutarate are produced from phosphoenolpyruvate of the Embden-Meyerhof-Parnas pathway and erythrose 4-phosphate of the pentose phosphate pathway through the series of reactions described in Table 2-14.

2.1.3 Effects of Carbon Source

Results from several studies have revealed that the activities of many of the enzymes described in the previous sections, as well as the resulting growth and product formations of *C. glutamicum* cultures, are very strongly influenced by the carbon source applied to the cultivation of the bacteria. *C. glutamicum* is capable of growing on many different

Table 2-14: Reactions of the phenylalanine biosynthesis pathway.
Enzyme catalysts are DAHP synthase (aroG), dehydroquinate synthase (aroB), dehydroquinate dehydratase (aroQ), shikimate dehydrogenase (aroE), shikimate kinase (aroK), EPSP synthase (aroA), chorismate synthase (aroC), chorismate mutase (aroH), prephenate dehydratase (pheA), and aspartate aminotransferase (aat).

Reaction	Catalyst
Phosphoenolpyruvate + Erythrose 4-phosphate + H_2O $\leftrightarrow$ DAHP + P_i	aroG
DAHP $\leftrightarrow$ dehydroquinate + P_i	aroB
Dehydroquinate $\leftrightarrow$ Dehydroshikimate + H_2O	aroQ
Dehydroshikimate + NADPH + H^+ $\leftrightarrow$ Shikimate + NADP	aroE
Shikimate + ATP $\leftrightarrow$ Shikimate 3-phosphate + ADP	aroK
Shikimate 3-phosphate + Phosphoenolpyruvate $\leftrightarrow$ EPSP + P_i	aroA
EPSP $\leftrightarrow$ Chorismate + P_i	aroC
Chorismate $\leftrightarrow$ Prephenate	aroH
Prephenate $\leftrightarrow$ Phenylpyruvate + H_2O + CO_2	pheA
Phenylpyruvate + Glutamate $\leftrightarrow$ Phenylalanine + α-Ketoglutarate	aat
Overall: 2 Phosphoenolpyruvate + Erythrose 4-phosphate + NADPH + H^+ $\leftrightarrow$ Phenylalanine + α-Ketoglutarate + ADP+ NADP + H_2O + CO_2 + 4 P_i	

carbon sources, and of co-utilizing several different pairs of these substrates, including glucose and fructose [Dominguez *et al.*, 1993], glucose and lactate, glucose and pyruvate [Cocaign *et al.*, 1993], and glucose and acetate [Wendisch *et al.*, 2000]. This is in contrast to the diauxic behavior more common to many other bacteria, in which the carbon source that is capable of supporting higher growth rates and yields is preferentially consumed until its supply has been exhausted and the other carbon source present in the mixture begins to be metabolized [Monod, 1949].

In a flux analysis experiment comparing the cases of growth on glucose with growth on acetate, several pathways were found to be influenced by the change in carbon source [Wendisch *et al.*, 2000]. Cultures grown on acetate were found to have higher fluxes in the tricarboxylic acid cycle, the glyoxylate bypass pathway, and gluconeogenesis. The higher tricarboxylic acid cycle flux can be attributed in part to the activities of the four genes of the succinate dehydrogenase complex, which are over-expressed during growth on acetate as the sole carbon source [Muffler *et al.*, 2002], but repressed for growth on mixtures of acetate and propionate [Hüser *et al.*, 2003]. When cultures are grown on carbon sources such as acetate or ethanol that enter the metabolic network in the form of acetyl-CoA, the glyoxylate cycle is required to meet the anaplerotic demands of the cells,

and gluconeogenesis is required to allow for the formation of sugar phosphates using TCA cycle intermediates as substrates [Gerstmeir et al., 2003]. Growth on mixtures of glucose and acetate demonstrated that growth rates of cultures decreased as the amount of acetate present in the medium was increased [Wendisch et al., 2000]. Two possible explanations offered for this observation are the interference with methionine synthesis by intracellular acetate [Roe et al., 2002] and the acetate-influenced uncoupling of the transmembrane pH gradient of the cells [Baronofsky et al., 1984].

Other studies investigated metabolism for growth on fructose in place of glucose, finding that the yields of both biomass and lysine were lower when fructose rather than glucose was used as the carbon source, while the formation of lactate, glycerol, and dihydroxyacetone were increased [Dominguez et al., 1998 and Kiefer et al., 2003]. Also, the flux through the pentose phosphate pathway was lower, and through pyruvate dehydrogenase and the tricarboxylic acid cycle was higher upon using fructose as a substrate.

Some of the enzymes most affected by changes in carbon source are those, such as pyruvate carboxylase, malic enzyme, and the enzymes of the glyoxylate bypass pathway, that serve anaplerotic roles within the cells. Results published previously indicate that by culturing C. glutamicum on lactate or pyruvate rather that glucose as the sole carbon source, the activity of pyruvate carboxylase is increased [Peters-Wendisch et al., 1998 and Koffas et al., 2002]. This is in agreement with similar findings relating growth on lactate to enhanced pyruvate carboxylase activity in other species such as Rhodopseudomonas spaeroides, Saccharomyces cerevisiae, Azotobacter vinelandii, and Rhodobacter capsulatas [Payne and Morris, 1969; Young et al., 1969; Scrutton and Taylor, 1974; and Willison, 1998]. Another link between the activity of pyruvate carboxylase and lactate as a carbon source is the finding that pyruvate carboxylase activity is essential for growth on lactate in all cases, expendable for growth on glucose when the anaplerotic role of the enzyme can be instead fulfilled by phosphoenolpyruvate carboxylase, and expendable for growth on acetate in all cases [Peters-Wendisch et al., 1998]. Other studies have shown malic enzyme to also have a higher specific activity on lactate as well as on fructose and a lower activity on acetate than on glucose [Dominguez et al., 1998 and Gourdon et al., 2000], leading to speculation that the ability of the

enzyme to generate NADPH may increase the value of its reaction to cells grown on carbon sources that cause reduced pentose phosphate pathway flux.

The isocitrate lyase and malate synthase enzymes that comprise the glyoxylate bypass pathway have been shown to be regulated in part by the supply of carbon source available to the cells [Reinscheid et al., 1994a; Reinscheid et al., 1994b; and Wendisch et al., 1997]. The activities of these enzymes were found to be unnecessary for growth on glucose, but essential for growth on acetate [Reinscheid et al., 1994a and 1994b]. Subsequent flux analyses confirmed that there was in fact negligible flux through the glyoxylate bypass pathway for the case of growth on glucose [Sonntag et al., 1995 and Marx et al., 1996]. Although of all of the carbon sources tested in that study, acetate was shown to bring about the largest degree of up-regulation relative to glucose, lactate was found to cause a 2-fold to 4-fold increase in activity of these glyoxylate bypass enzymes [Wendisch et al., 1997].

2.1.4 Effects of Over-expression of Aspartokinase and Pyruvate Carboxylase

The body of knowledge discussed above relating to the reactions of the metabolic network of central carbon metabolism and lysine biosynthesis, the enzymes responsible for catalyzing these reactions, and the control strategies employed by the cells in regulating the enzyme activities, has been extremely valuable in allowing for the rational selection of targets for strain improvement. Nevertheless, the degree to which the productivity of lysine has been improved through metabolic engineering in the past two decades has been underwhelming given the amount of detailed information that is now available about the characterization of individual reactions and enzymes. A frequent focus on individual steps within a pathway, rather than concerted approaches involving multiple reactions, may be a source of the lesser gains achieved from genetic manipulation. As examples, the individual over-expression of the key enzymes

aspartokinase and pyruvate carboxylase yielded decidedly mixed results for cell growth and lysine production.

The major control point of the lysine biosynthesis pathway lies in the activity of aspartokinase, the enzyme responsible for catalyzing the first step towards the formation of all amino acids derived from the starting material aspartate. In part as a result of the feedback inhibition of the enzyme, discussed in section 2.1.2.5, the flux from aspartate to aspartyl semialdehyde can be a limiting bottleneck to the overall flow of material towards lysine biosynthesis. Nevertheless, attempts to alleviate this limitation through the over-expression of the aspartokinase gene have led to lower specific productivities of lysine in resulting recombinant strains [Jetten et al., 1995 and Hua et al., 2000]. This is believed to be caused by a depletion of the oxaloacetate intermediate of the tricarboxylic acid cycle, resulting in a decreased availability of the compound for use in other pathways vital to the growth of cells, such as the tricarboxylic acid cycle itself and, in the case of growth on carbon substrates other than glucose, gluconeogenesis. Very similar results were seen for a strain engineered to over-express dihydropicolinate synthase, another enzyme that catalyzes a potentially rate limiting step of lysine biosynthesis [Eggeling et al., 1998]. This strain also had a higher lysine productivity than was realized with its parent strain, but this increase came at the expense of a decrease in culture growth.

Soon after the activity of the pyruvate carboxylase in the species was ascertained, its gene was cloned and sequenced [Peters-Wendisch et al., 1998; Koffas et al., 1998]. Because of its importance in supplying precursors to lysine production within the bacteria, it was assumed that up-regulation of the enzyme would lead to an enhancement in lysine yields. In contrast to expectations, when the lysine-producing strain ATTC 21253 (American Type Culture Collection, Manassas, VA) was transformed with a plasmid, p(KD7), that conferred pyruvate carboxylase over-expression, a decrease in lysine yield along with an increase in growth yield resulted [Koffas et al., 2002]. In addition, the changes in yield were much more pronounced when the cells were grown in a medium that contained lactate, rather than glucose, as its carbon source. Results from this study are given in Table 2-15. The underlying metabolic explanation for these results is most likely related to discordance between the higher pyruvate carboxylase activity in the transformed strain and the lower, native activities of the enzymes catalyzing the

Table 2-15: Effect of over-expression of pyruvate carboxylase on growth and lysine production by *C. glutamicum* strains. Cultures of 21253 and 21253(pKD7) were grown using a minimal medium with lactate or glucose as the carbon source. Data are shown for the maximum specific lysine production rate (q_m) and the final cell concentration (X_f) [Koffas *et al.*, 2002].

Strain	Glucose		Lysine	
	q_m (mg lysine/ (g cell · hr))	X_f (g/l)	q_m (mg lysine/ (g cell · hr))	X_f (g/l)
21253	46	4.2	63	2.5
21253 (pKD7)	46	4.2	23	4.0

subsequent reactions of the lysine biosynthetic pathway. These lower activities, and in particular that of the aspartokinase bottleneck discussed above, prevent the majority of the excess oxaloacetate created from an increased pyruvate carboxylase flux from serving as a substrate for increased lysine production. Instead, the oxaloacetate is presumably fueling additional tricarboxylic acid cycle flux, leading to the production of more ATP, reducing equivalents, and biosynthetic precursors that will support increased cell growth.

These findings are not surprising given the theoretical results of metabolic control analysis [Kascer and Burns, 1973 and 1995] showing that often there is not a single rate limiting step present among the reactions that constitute a particular pathway. Instead, metabolic engineering success is much more likely to occur when the activities of multiple genes have been taken into consideration [Cremer *et al.*, 1991; Colon *et al.*, 1995; Albrecht *et al.*, 1999; and Koffas *et al.*, 2003]. This type of strategy allows for the beneficial adjustment of cellular properties without unnecessarily depleting intracellular pools of metabolites required by the cells for multiple, and often competing, purposes [Koffas *et al.*, 2003]. To better gauge the coordinated behaviors of multiple enzymes functioning within a metabolic network, a more global approach is required. The powerful new technique of DNA microarrays, just now beginning to be applied to metabolic engineering, provides such a broader view.

2.2 DNA Microarrays

The relatively recent development and refinement of the powerful biological assay that is the DNA microarray has greatly expanded the accessibility of global quantifications of the state of a cell culture. By providing a means for the simultaneous assay of the entire population of mRNA species, or the transcriptome, present within a cell, DNA microarrays can give insight into the gene expression phenotypes that are associated with the responses of the cell to, for example, different environmental stresses or process demands. Microarray analyses are somewhat analogous to the more familiar metabolic engineering techniques of metabolic flux analysis and metabolic control analysis, in that all three attempt to describe the functioning of the metabolic network as a whole, rather than in only describing the behaviors of individual components. However, in allowing for the simultaneous monitoring of transcription of all genes of interest, DNA microarray technology affords one clear advantage over other metabolic engineering tools, such as *in vivo* flux measurements. In many of the reaction network configurations that are more interesting to study, quantitative data on flux rates can only be obtained through the imposition of significant simplifications or assumptions. These include, for example, cases in which parallel reactions are combined together into a single representative flux [Marx *et al.*, 1996; Jucker *et al.*, 1998] and in which some reactions are removed entirely through the study of deletion mutants [Park *et al.*, 1997]. No such simplifications or assumptions are required for the microarray analysis of the behavior of an entire reaction network that can, theoretically, span the entire genome of an organism. Instead, the limitations of the technique manifest themselves in the precision of the data provided by the microarrays, the ability of bioinformatics techniques to discern significant patterns from within the potentially very large data sets, and the degree to which information obtained about the transcriptional state of the organism accurately depicts the behavior of the metabolic fluxes present in a reaction network.

A good review of DNA microarrays, their history, and their application to different biological studies can be found in Lucchini *et al.* (2001). The major aspects of the technology are covered in this section, to provide a background for the discussion of protocol improvements found in the following chapter. DNA microarrays are themselves

enhancements of the long-established procedures of Northern and Southern blots used for the detection of desired DNA and mRNA sequences using fluorescent labeling [Southern, 1975 and Alwine *et al.*, 1977]. The two advancements that were most responsible for the technological evolution from the macroarray origin of these blots to the microarrays of the type designed and applied in the work of this thesis were improvements to robotics and the development of dual fluorescent labeling. The innovations made in the area of robotics allowed for the printing at high density of very small and numerous spots of DNA onto a single glass slide, so that probes for potentially thousands of gene sequences or their replicates could be contained within a small area [DeRisi *et al.*, 1996]. The advent of dual fluorescent labeling enabled the simultaneous hybridization of two differently labeled populations of cDNA onto an array to provide information about relative mRNA abundances [Schena *et al.*, 1995].

There are two primary array-based technologies currently used for the detection and quantification of gene expression levels—one using oligonucleotides and the other using complementary DNA (cDNA). Oligonucleotide DNA arrays were not used in any of the work related to this thesis, and as such their manufacture and application will not be discussed. For a good review of oligonucleotide DNA arrays, see for example the review by Lipshutz *et al.* (1999). The following paragraphs present a brief overview of the technical bases for creating and using cDNA microarrays of the type involved in this thesis.

In creating a cDNA microarray, individual DNA sequences derived from the open reading frames of an organism are spotted onto an array surface that is typically a glass microscope slide. These sequences may represent all or part of the genome of the organism being studied, and are typically obtained in concentrations sufficient for printing onto an array by amplification with the polymerase chain reaction (PCR).

An RNA sample is then isolated from a sample of cells being cultivated under the conditions of interest to a particular experiment. In this way, the content of the aggregate mRNA population within the culture sample can give information about the response of the cells on a transcriptional level to stimuli such as culture phase and growth environments. Because the mRNA of bacterial species such as *C. glutamicum* lacks the polyadenylated region inherent to the mRNA of mammalian systems, the mRNA must be

isolated from the cell cultures as a component of total RNA samples. Within these samples, mRNA constitutes approximately 3% of the total RNA [De Saizieu *et al.*, 1998]. This increases the amount of total RNA that needs to be isolated to obtain sufficient mRNA quantities, and decreases the specificity of subsequent reverse transcription reactions in creating fluorescently labeled DNA complementary to mRNA sequences.

The sequences contained within the RNA sample are fluorescently labeled by creating cDNA, via reverse transcription, that is composed in part of fluorescent nucleotide analogs. This fluorescent cDNA is then applied in the form of a solution to the DNA probes spotted onto the surface of the microarray and allowed to hybridize under stringent conditions chosen to enforce the specificity of binding by relying on the mutual selectivity of complementary strands of nucleic acids. After hybridization has occurred, the microarrays are assayed by laser excitation of the fluors contained in target DNA that is now bound to the array surface, causing emission with characteristic spectra, the intensity of which is quantified by a scanning confocal laser microscope. The amount of fluorescence observed at each spot within the array is then linearly related to the abundance of the mRNA transcript within the cell sample that is of the same gene sequence printed at that particular spot.

By using a dual labeling strategy, the makeup of two RNA samples can be probed simultaneously with the same microarray. Two cDNA solutions can be prepared, each labeled with distinct fluorescent nucleotide analogs with non-overlapping excitation and emission properties, and co-hybridized onto the array. In this way, the relative expression levels of genes can be determined for a pair of cultures. Signal ratios are computed by calculating the ratio of fluorescence intensity measured at the fluor emission wavelength of one of the cDNA populations to the intensity measured at the wavelength of the second population. The signal ratios are then transformed into a log-space so that positive values indicate increased transcript concentrations in one of the two RNA samples relative to the concentration of the same transcript sequence in the other sample. Negative values indicate decreased transcript concentration in this first sample relative to that in the second sample, and a signal ratio value close to zero indicates that the transcript of that particular gene is equally represented in each of the two sampled cultures.

Other studies performed in applying DNA microarrays to the species of *C. glutamicum* have been published during the time in which the work of this thesis was performed. These publications include discussions of the techniques used in the development and validation of *C. glutamicum* microarrays [Loos *et al.*, 2001]; the application to the investigation of growth on alternative carbon sources such as acetate [Muffler *et al.*, 2002 and Hayashi *et al.*, 2002], propionate [Hüser *et al.*, 2003], ribose [Wendisch, 2003], and valine [Lange *et* al., 2003]; and the examination of transcriptional responses to stresses such as heat shock [Muffler *et al.*, 2002], and phosphate starvation [Ishige *et al.*, 2003]. The results presented in Chapters 3-7 represent independent discoveries in improving the protocols of *C. glutamicum* DNA microarray design, manufacture, and application, as well as the first description of the gene expression profiles of cells grown on lactate as a carbon source. Also described are the results of microarray analyses of cells engineered to over-express a key enzyme associated with growth and lysine production, and one of the first comparisons between the transcriptional data obtained from microarrays and the flux data obtained from metabolic network analysis.

Chapter 3: Development of *C. glutamicum* DNA Microarrays and Related Protocols

3.1 Selection and Amplification of DNA Sequences

To investigate the gene expression phenotypes associated with the production of biomass and lysine for various *Corynebacterium glutamicum* cultures and growth conditions, a partial genome DNA microarray was developed which monitored the transcriptional behavior of genes related to multiple pathways involved in the conversion of glucose and lactate substrates to lysine product. From the entire metabolic network of the organism these pathways together form a subset that comprises the majority of the key reactions transforming carbon skeletons to those of the aspartate-family amino acids, while generating much of the energy, reducing equivalents, and synthesis intermediates required by the cells for maintenance and growth. This network subset includes the Embden-Meyerhof-Parnas pathway used for glycolysis and gluconeogenesis, the pentose phosphate pathway used for producing reducing equivalents and converting hexoses to pentoses, the tricarboxylic acid cycle used for aerobic respiration, the glyoxylate bypass pathway and other anaplerotic reactions in the oxaloacetate/pyruvate node used for replenishing metabolic intermediates, and the lysine biosynthetic pathway used for generating amino acid product. Other genes included on the microarray include those of the biosynthetic pathway of biotin—a cofactor of pyruvate carboxylase, one of the main focuses of these experiments—as well as genes encoding several other carboxylases, the

activities of which may influence amino acid production through mechanisms not yet understood. The complete list of gene sequences sourced from *C. glutamicum* that were chosen for inclusion on the microarrays is given in Table 3-1.

After determining which genes would be represented by the spots of the partial genome microarray, suitable templates needed to be selected for use in the repeated copying rounds of PCR. In this work, the restriction-deficient *Corynebacterium glutamicum* strain E12 [Follettie and Sinskey, 1996] was the source of genomic DNA used as the template in these PCR amplifications. To obtain sufficient amounts of the genomic DNA, the strain was grown in Luria-Bertani medium [Sambrook and Russel, 2001] prior to genomic DNA isolation according to the protocol of Treadway et al (1999). E12 was chosen as the source strain in part because of the extensive experience of our laboratory in its cultivation, and because of the large and biologically pure culture stock of the strain in our inventory. The primary reason, though, was that although E12 had been altered on a genetic level from the phenotype of the wild strain, these modifications were not made in the pathways and specific genes involved directly in amino acid biosynthesis. This is not the case with the genomes of, for example, ATCC 21253 or ATCC 21799. These strains either lack individual genes for enzymes that catalyze reactions present in amino acid anabolic pathways, or contain genes for which the sequences have been changed in order to adjust protein properties and functions. As a result, the use of these genomes may not yield DNA capable of the hybridization events necessary for generating data relating to the transcription of certain genes of interest. By amplifying the wild type gene sequences from E12 instead, the DNA microarray will have the broadest applicability possible in probing the behavior of the transcriptome of the metabolic network represented by the array.

3.1.1 Use of *Mycobacterium tuberculosis* Gene Sequences

Until very recently, the available information about the genetic makeup of *C. glutamicum* was very limited. Only with the advent of modern genetic engineering

Table 3-1: Microarray spots composed of DNA sequences derived from *Corynebacterium glutamicum*.

Gene	Enzyme	NCBI ID	Primer 1	Primer 2	Size (bp)
aat	aspartate aminotransferase	E16763	agtggatttgtcctgttccg	gattgaccttgttagggcca	500
accA	acetyl-coA carboxylase beta subunit	NC_003450	acgtcgtcaagctttctggt	acgccttgtcgctctatgat	500
accC	acyl-coA carboxylase	NC_003450	cgcgaaggctatctgtaagg	gtggaatgggataacggttg	500
accD	acetyl-coA carboxylase beta subunit	NC_003450	ctcgacttcctgcctttgac	gaggttcttggaacccatca	500
aceA	isocitrate lyase	X75504	ccacacctaccctgaccagt	ggctcgagaccattcttgac	500
aceB	citrate synthase	X78491	atggtaccgagctggttctg	ttgttctcgtaggcctgctt	500
aco	aconitase	NC_003450	ctcccgtgagctttacgaag	atgccagcagccttgtagtt	500
asd	aspartate semialdehyde dehydrogenase	L16848	gcacggaaatcgaggtagaa	ttgtaagcgattggggaaac	500
ask	aspartokinase	L16848	gttgctggttttcagggtgt	gttgatttctgcatcagcca	500
bioA	7,8-diaminopelargonic acid aminotransferase	D14083	ggctaccacggagacacatt	attagccataaaggtggggc	500
bioB	biotin synthase	U31281	tgccatttctgctcacagtc	tcaaggaagttcatgggggac	500
bioD	dethiobiotin synthase	AP005282	tgggaaaaccttctccacag	caggattttgagggatcgaa	500
birA	biotin-(acetyl-coA carboxylase) ligase	NC_003450	caatcgtgtctgtgctcgtt	ccgatcgcttcacctaagag	500
citA	citrate synthase 1	NC_003450	atcactgatcctctacgccg	tgttttcgtactgctccacg	500
citA2	citrate synthase 2	NC_003450	tttgaaagggatatcgtggc	gccttatcaagctgtgcctc	500
citEa	citrate lyase subunit	NC_003450	actttttggctgacgtggag	tccaactggttagcctctgg	500
dapA	dihydrodipicolinate synthase	Z21502	ggagtagagcacttcggcac	gaccgccaaaatcgtaggta	500
dapB	dihydrodipicolinate reductase	Z21502	tcaccactcctaacgctgtg	tgcttgatggtcaaggtctg	500
dapD	tetrahydrodipicolinate succinylase	AJ004934	ccatcaacctagacggcatt	tgcctgcagtgatgtagagg	500
dapE	succinyl-diaminopimelate desuccinylase	X81379	ggtcatgcttgctggtcata	tcgcagaaaacgatgttgag	500
dapF	diaminopimelate epimerase	AX137620	gatgcgcgcctagatttc	gtgggggaagaattcctgat	500
ddh	meso-diaminopimelate dehydrogenase	Y00151	tggatccaggaatgttctc	ctgtgaggaagcggtgaaat	500
dtsR1	propionyl-coA carboxylase	AB018531	agatcatggagctggcaatc	ccagtggtgtgtaagagcga	500
dtsR2	propionyl-coA carboxylase	AB018531	gatggccgtaaggtctgtgt	aacccaatccaaagcatctg	500
eno	enolase	AX136862	tcgacgacgaagcaatgatc	gccgtccttgaagaactcag	500
fba	fructose bisphosphate aldolase	X17313	aatcatccagttctccaccg	acgttgcctggcttgtaaac	500
fum	fumarase	NC_003450	acgtcaaggagctcaaggaa	agccgattgcagagttcagt	500
gap	glyceraldehyde 3-phosphate dehydrogenase	X59403	ttccacccttctcaagttcg	aacgatgttgactgctgctg	500
gnd	6-phosphogluconate dehydrogenase	E13660	cgacctggattcctacctca	tcaagttcagcgtttgcatc	500

gpi	phosphoglucoisomerase	AX136015	atggagaccaatggcaagtc	cctcttcaccagagacagcc	500
gpm	phosphoglyceromutase	AX108748	atacacctccttgctgcgtc	tactacggaaccgtcttcgg	500
icd	isocitrate dehydrogenase	X71489	tctccgacccaatcatcttc	gatgcggaaggtcttgtcat	500
lysA	diaminopimelate decarboxylase	E16355	aaccccactgttcgtagtcg	tcaggttttctgcgttgttg	500
lysE	lysine export	X96471	gccgatcgtgctcgatatta	catcagtttgatggccaatg	500
lysG	lysine export regulator	X96471	aacgcagattcgctatccac	catcggtgtgtcaatgggta	500
malE	malic enzyme	AF234535	cttacaccccaggtgttgct	gttggtcagcatgtccacag	500
pck	phosphoenolpyruvate carboxykinase	AJ269506	actacgcatccaacccaatc	ccacccttgttaccatgtc	500
pdh	pyruvate dehydrogenase subunit	NC_003450	ctgagccagaaggactggac	gagacgtcgatcttgccttc	500
pfk	phosphofructokinase	AX109130	aagtcacctgggttgtcctg	taagacacggcattcagcag	500
pgk	phosphoglycerate kinase	X59403	gaatcattgcctccctacca	ggatccaccgagaaccacta	500
ppc	phosphoenolpyruvate carboxylase	X14234	caggagattggatccctcaa	gcgtcgcatcatctctacct	500
pyc	pyruvate carboxylase	Y09548	gtggacatcttccgcatctt	agtacggctcgaggtcagaa	500
pyk	pyruvate kinase	L27126	ccatgggctttggaaatcta	ggtcaggcctatggttgaaa	500
rpi	ribose 5-phosphate isomerase	NC_003450	ataccttggagcagaccacg	attccagtgcgctcgtaatc	500
sdhA	succinate dehydrogenase subunit A	AX113263	cgaaatggttgacgtcattg	tatgcagcgttgttcagagg	500
sdhB	succinate dehydrogenase subunit B	AX113263	ccactgttcttcgaatccgt	gtggttgacgtggagggtat	500
sdhC	succinate dehydrogenase subunit C	AX113263	tcacatgatcggaaacctga	ggtgattcccaggtcagaga	500
sucA	α-ketoglutarate dehydrogenase	D84102	ggatctggaccgtaccttca	aaccgaggtggtacttcacg	500
sucC	succinyl-CoA synthase alpha subunit	NC_003450	ccgtaattttcgttcctcca	ccacaactggttttgtcacg	500
sucD	succinyl-CoA synthase beta subunit	NC_003450	attgtggcatcaacaccaga	gggttcacctcaacgagtgt	500
tal	transaldolase	AX076274	tctgtggcttccttcttcgt	tacttcaggcgagcttccat	500
thrA	homoserine dehydrogenase	Y00546	ttaggattcggaacagtcgg	gttgcctcagccaaagaatc	500
thrB	homoserine kinase	Y00546	tggcaattgaactgaacgtc	agcatactgtggctggctct	500
thrC	homoserine synthase	X56037	atgcccagctgagtaaatgg	tgactacgtcttggcaatcg	500
tkt	transketolase		ggagaccgacgcttactacg	gatgccagcttcaacagaca	500
tpi	triosephosphate isomerase	X59403	cgcccttccaaaggaatact	ggatagccttgcacacttcc	500
zwf	glucose-6-phosphate dehydrogenase	E13655	aaccagctgtttgagccact	agggatacagtcatgtcgcc	500

techniques did it become possible to identify, isolate, and manipulate individual open reading frames for various enzymes within the organism for the first time. Although the

number of genes whose sequences had been determined with these tools grew consistently throughout the 1980s and 1990s, a detailed description of the entire genome of the species was not published until September of 2002 [Nakagawa, 2002]. As a result, when the microarrays described in this thesis were first designed, it was not possible to represent the entire lysine production network and all of its relevant enzymes with *C. glutamicum* DNA. For this reason, the gene sequences of a second species, the related *Mycobacterium tuberculosis*, were considered as substitutes for those undiscovered in *C. glutamicum*. In this way, an attempt was made to fill in the gaps that would otherwise be present in formulating a quantitative description of gene expression profiles.

Other work has been carried out with varying degrees of success in the use of DNA microarrays that are specific for one species in investigating the transcriptional profiles of another distinct species. The majority of these experiments have involved higher organisms. Examples include mammalian studies using human microarrays to study pig [Moody *et al.*, 2002] and monkey [Zou *et al.*, 2002] mRNA levels, fish studies using trout microarrays to probe gene expression in Artic char [Wiseman *et al.*, 2003], and plant studies comparing sequences from pine, spruce, and tobacco [Van Zyl *et al.*, 2002]. In creating a heterologous bacterial microarray, *M. tuberculosis* was chosen for two reasons—its close phylogenic proximity to *C. glutamicum*, and the large body of knowledge that exists about its genomic makeup. The two species are closely related, with both belonging to the Actinomyces family. They each are gram-positive, non-motile, catalase-positive microbes that have rod-like morphologies and produce characteristic long chain fatty acids termed mycolic acids. In addition, there is precedent for the use of *M. tuberculosis* DNA sequences to examine *C. glutamicum*, with the pyruvate carboxylase gene of *C. glutamicum* first being isolated by using primers designed with sequence information from *M. tuberculosis* [Koffas *et al.*, 1998]. The complete genome of *M. tuberculosis* H37Rv was published in 1998 [Cole *et al.*, 1998], making the organism even more useful for our purposes, and numerous experiments since then have demonstrated the use of both partial and full genome microarrays for the species [Behr *et al.*, 1999 and Schoolnik, 2002].

Table 3-2 lists the genes that are represented on the microarray using *M. tuberculosis* DNA. Included in the table are the primer pairs that were used to amplify the sequences

Table 3-2: Microarray spots composed of DNA sequences derived from *Mycobacterium tuberculosis*.

Gene	Enzyme	NCBI ID	Primer 1	Primer 2	Size (bp)
aat	aspartate aminotransferase	BX842573	tgttggcggacgaaatctac	tgggtgaccaggatcttctc	500
aceA	isocitrate lyase	BX842573	gaccagttggcctctgagaa	gcgtgatgaactggaacttg	500
aceAa	isocitrate lyase subunit	BX842578	ggaccgcaaccaacactatc	gccaagcgcatagagaagat	500
aceAb	isocitrate lyase subunit	BX842578	ttggtcgaatctccctatcg	gcgacttcatcttcgaggtc	500
aceB	malate synthase	BX842578	caaggacgtgatcctggaat	aaccccagcacatcttcaac	500
aceE	pyruvate dehydrogenase E1 component	BX842579	cgtatttgcccgacattgat	acgtcgtcgatccggtact	2653
asd	aspartyl semialdehyde dehydrogenase	BX842583	gctggatatcgcgttgttct	ctcgaagcgcaacttttgat	500
ask	aspartokinase	BX842583	accaaggatgtcacgacgtt	ggtgaaggtgatgtcggtct	500
bioA	7,8-diaminopelargonic acid aminotransferase	BX842577	gatgaaccacgtcatgttcg	atcgaagatcagcagcacct	500
bioB	biotin synthase	BX842577	gaggattgccatttctgctc	gttgaggaagttcagcggga	500
bioD	dethiobiotin synthase (probable)	BX842577	acgatcctggtcgtcacc	atcaccagccctgcacat	500
bioF	8-amino-7-oxononanoate synthase (probable)	BX842577	tggcctccaacgactatctc	gaccggtgtcgaagatgaac	700
dapA	dihydropicolinate synthase	BX842580	gggagaaaatcgagctgctg	gaaggcggacaacaactctc	500
dapB	dihydropicolinate reductase	BX842580	aggtcgtcatcgacttcacc	atggtcagagtctccccctc	500
dapE	succinyl-diaminopimelate desuccinylase	BX842575	cttcatctggccgctacact	acacgtcatggacatgttgc	500
dapF	diaminopimelate epimerase	BX842580	cacggtacccagaacgactt	caccaatcgtcagcgaga	836
devb	glucose-6-phosphate 1-dehydrogenase	BX842576	catcgagatcttccccgata	agcagccagagcgtgttc	699
eno	enolase	BX842575	gccgattatcgagcaggtta	accgaggaaatgccaggt	1268
fba	fructose bisphosphate aldolase	BX842573	gtgccaatcgatgagaacct	cactccgtcgtagttggtga	500
fum	fumarase	BX842575	caattaccgcatcgagcac	atagccgatcgctgtcgag	1400
gap	glyceraldehyde 3-phosphate dehydrogenase	BX842576	tcctgggagttaacgacgac	ggtcaacccagagtcgaaaa	500
gltA	citrate synthase 2 (probable)	BX842575	cgacccaatccgacatctac	tagaaggtgcccggtaacac	500
gltA2	citrate synthase 3	BX842574	ctacggttgacgttcggatt	gaagtcgacgttggggtaaa	500
gnd	6-phosphogluconate dehydrogenase	BX842578	cgtgatgggttccaacatc	gtacttcggtgcggtctga	1394
gnd2	6-phosphogluconate dehydrogenase 2	BX842575	gcaactaggaatgatcggtct	ccttgttggcgaagtcgt	965
gpm	phosphoglycerate mutase 1	BX842573	aatgccctcaacctgttcac	cgggtccagatacgtaccac	648
gpi	phosphoglucoisomerase	BX842575	gatcggaaacacccatcttc	cggtcgagctgtccgact	1544

49

icd1	isocitrate dehydrogenase (probable)	BX842582	gcgtatgcgatcaagaaaca	gtctttgaacatcccgtcgt	500
icd2	isocitrate dehydrogenase	BX842572	ccgacaccaacatcatcaag	atcctgcatctgctcttcgt	500
lpdA	dihydrolipoamide dehydrogenase	BX842582	gacccgcatcgtgatcct	gtgtggcaccaaggctagtt	1456
lpdB	dihydrolipoamide dehydrogenase	BX842574	ccgtgagggtgaaacgtatg	ccagcaggtaaaacgaatctc	1448
lysA	diaminopimelate decarboxylase	AF126720	ggacgagatcatgctgctg	cattgacccttcacctccag	1347
mdh	malate dehydrogenase	BX842576	gagcgctagtcctctcaagg	aactcggcggttgacttgt	945
opcA	unknown function, may aid G6PDH	BX842576	cggtcaacaagaagctcgac	tcaccggtactgcaccttct	878
pyc	pyruvate carboxylase	BX842581	ggtcaccgataccacattcc	acgatctgctcagccagttt	500
pdhA	pyruvate dehydrogenase E1 component subunit	BX842580	gccatctgggatgttgatgt	agtcagtccgcgccagtt	1082
pdhB	pyruvate dehydrogenase E1 component subunit	BX842580	gatgagacgcttgcggtag	gttcgacgcagtccaacag	994
pdhC	dihydrolipoamide acetyltransferase	BX842580	ggtgaggacagcatcaggtc	agatcccgcagctcacac	1139
pfkA	phosphofructokinase I	BX842581	tgcggattggagttcttacc	cgtcgtcgtaacggctct	1011
pfkB	phosphofructokinase II	BX842578	gcgaatcatcactttgacca	tattgatcctgcccgacttc	945
pgk	phosphoglycerate kinase	BX842576	caaggatctactcgccgaag	gggttgcgatgtcatagacc	500
pyk	pyruvate kinase	BX842577	gcgactacgacgatcacaag	atcacctcgtggaccagttc	500
rbsK	ribokinase	BX842579	gcaaacgccagtgagactaa	gacaccggacaccaaagtt	843
rpe	ribulose-phosphate 3-epimerase	BX842576	gtcgatcctagccgctgat	taggctcaggtggagtgagg	652
rpi	ribose-5-phosphate isomerase	BX842579	gaccacgccggatatgag	tactcggcgagaatgtcgat	413
Rv0462	probable dihydrolipoamide dehydrogenase	BX842573	cactatgacgtcgtcgttctc	aattgatcatgtggccaacc	1384
Rv2419c	putative phosphoglycerate mutase	BX842579	ggcatggacaaacggactac	aggacatcgctggagacct	640
Rv3837c	putative phosphoglycerate mutase	BX842584	agtcctatggcaacgtcgag	ccactgcacacaactccatc	569
sdhA	succinate dehydrogenase A subunit	BX842582	ttgatctgccaacaccgata	tttgaaatcaaggcggacat	1728
sdhB	succinate dehydrogenase B subunit	BX842582	ccggacgtcgaaactttg	tcagcgggtgaacatcag	780
sdhC	succinate dehydrogenase C subunit	BX842582	ggtcgtgggtatgccatc	atcggaagtgctcccacat	333
sdhD	succinate dehydrogenase D subunit	BX842582	ctggacaacccacgatcac	gagatgttcgggtcgaatgt	386
sucA	α-ketoglutarate	BX842576	gaacgggtgttcaacgaagt	gcccacagtgtagccgtact	500

50

	dehydrogenase				
sucB	dihydrolipoamide succinyltransferase	BX842579	ctcaaacaggaaggcgacac	gtcctaaatcggcctcgaac	1591
sucC	succinyl-CoA synthase	BX842575	caaggagttattcgccaagc	gtcgccaccagtgtcacc	1095
sucD	succinyl-CoA synthase	BX842575	gcagggacaacaaggtcatt	gtcttgccgaccttcacac	838
tal	transaldolase	BX842576	gctcagaaccccaacctg	ttcctggagtagctcgttcc	1083
thrA	homoserine dehydrogenase	BX842576	gtcgatctgtgagcgcataa	atcaacggtttccgagagtg	500
thrB	homoserine kinase	BX842576	tcgtacaaacggattcgtca	ggtcggcaactctgaatctg	500
thrC	homoserine synthase	BX842576	gcggcaactaatctctccaa	gtgtagcccttccagtacgc	500
tkt	transketolase	BX842576	ctcgatttccaccaaggagt	atgagaatgacgtcgggttc	500
tpi	triosephosphate isomerase	BX842576	gtattacgaccgggttgacg	caacgaggccaactcttttc	500
zwf	glucose-6-phosphate 1-dehydrogenase	BX842575	gcgtcagatctgttggtgat	cgatcgtaccgatggatttc	1289
zwf2	glucose-6-phosphate 1-dehydrogenase	BX842576	ggacaagcgattacccagaa	cgcagcatctccaaggat	1461

from genomic DNA, as well as the sizes of the sequences defined by these primer pairs. The sizes belong to one of two categories—500-base pairs and full open reading frames—that will be discussed in more detail in the following section. Those genes that were represented by full open reading frame sequences were unidentified in *C. glutamicum* when the microarrays were first developed. Genes that were represented by 500-base pair, or partial open reading frame, *M. tuberculosis* sequences were also present on the array in the form of a *C. glutamicum*-derived spot. In this way, a comparison could be made between the results obtained from two sets of spots, each corresponding to the same gene but with sequences originating from two different species. As further *C. glutamicum* gene sequences were made available through publications, patents, and submissions to the National Center for Biotechnology Information (NCBI), they were utilized in creating additional spots on the microarrays, increasing the degree to which the metabolic network was described using native DNA.

Although *M. tuberculosis* and *C. glutamicum* are closely related, the degree of similarity, or homology, between the two species at the level of individual genes must be taken into consideration. For the genes that were printed onto the microarrays in the form of both *M. tuberculosis* and *C. glutamicum* DNA sequences, there was in fact a wide range of homologies found, as is shown in Table 3-3. These sequence homologies were calculated by comparing the full open reading frames of the same gene in each of the two

Table 3-3: Homologies between gene sequences from *M. tuberculosis* and *C. glutamicum*.

Gene	% Homology
aceA	75.7
fba	71.7
ask	67.2
pyk	66.9
gltA	66.3
gap	65.9
asd	65.7
tpi	65.1
dapB	64.7
sucA	64.5
pgk	63.2
pca	62.8
thrA	60.8
bioA	59.6
dapA	58.8
dapE	58.8
thrB	53.6
thrC	48.2
bioD	47.6
icd	47.4
aat	45.5

species with the "Blast 2 Sequences" software [Tausova and Madden, 1999]. The algorithm of the software, which can be accessed interactively at the web site of the NCBI, generates a score of % homology by performing a pairwise DNA-DNA sequence comparison. The results show that while there are genes, such as that of isocitrate lyase, that have greater than 75% sequence homology between the two species, there are others, such as that of aspartate aminotransferase, that have close to 45% sequence homology. One goal of this research was to determine if a relationship exists between the similarities of gene sequences from different species and the relative behavior of spots derived from those sequences that are present on the DNA microarray.

3.1.2 Use of Sequences of Different Size

The genes encoding the enzymes at work in the metabolic network being studied vary widely in the lengths of their open reading frames. Some genes, such as those of ribose-5-phosphate isomerase and some of the succinate dehydrogenase subunits, are composed of as few as 400 base pairs. Others, such as that of pyruvate carboxylase and α-ketoglutarate dehydrogenase, span greater than 3500 base pairs. In the design of the microarrays described in this work, all *C. glutamicum* genes printed on the arrays are present in the form of 500-bp fragments. The reasons for this were twofold—to eliminate any possibility that genes of different sizes would be any more or less prone to microarray results that were skewed by problematic events such as DNA folding or nonspecific or incomplete hybridizations, and to allow for the standardization of the procedure used in creating DNA printing stocks through PCR as will be discussed in the following section.

The possibility of nonspecific binding may increase if the sizes of the DNA printed onto the microarrays are too small to effectively discriminate against cDNA that is not perfectly complementary. Alternatively, if the printed DNA sequences are too large, there may be regions within these sequences that complement enough of a cDNA to allow binding to occur even if the sequence as a whole does not compliment that particular cDNA strand in its entirety. Another problem that may occur if the printed DNA sequences are too large is that regions of a DNA strand within a spot may bind to other regions within its own strand in a self-folding mechanism. This would have the effect of reducing the available regions to which complimentary cDNA could bind. To allow for the study of the effects that printed sequence size has on the quality of microarray results obtained, several *C. glutamicum* genes were represented on the arrays both by spots containing 500-bp sequences and a distinct set of spots containing full-gene sequences, enabling a direct comparison of the results obtained by the two methodologies. Table 3-4 lists the genes for which *C. glutamicum* DNA sequences comprising the entire open reading frames were included on the microarrays.

As has been described earlier, the genes whose sequences were unidentified in *C. glutamicum* at the time of microarray manufacture were represented by full open reading frame sequences of *M. tuberculosis* DNA. The choice was made to include the entire

Table 3-4: Microarray spots composed of full open reading frame DNA sequences derived from *Corynebacterium glutamicum*.

Gene	Enzyme	NCBI ID	Primer 1	Primer 2	Size (bp)
ask	aspartokinase	L16848	cagaaatatggcggttcctc	tcatcttcacggatcagcac	1300
bioD	dethiobiotin synthase	AP005282	tgggaaaaccttctccacag	ttcttaaaggcatcaagggc	700
dapE	succinyl-diaminopimelate desuccinylase	X81379	aggattagatctcctcggcg	tcgctcaggtactgcttcaa	1200
fba	fructose bisphosphate aldolase	X17313	ctcgatcgtgctaaggaagg	taagagcgtgggtcgtaagc	1000
fum	fumarase	NC_003450	gcaggaattccgtattgagc	tcttcctcggtgagcttctc	1500
gap	glyceraldehyde 3-phosphate dehydrogenase	X59403	cgaggtagttgcagtcaacg	cgttgtcgtaccaggaaaca	900
gnd	6-phosphogluconate dehydrogenase	E13660	gtagacaccaccgttgacca	ggagcctttaagcttccacc	1900
icd	isocitrate dehydrogenase	X71489	gctgctcgcgacctactc	cttcttcagtgcgtcaacga	2176
lysA	diaminopimelate decarboxylase	E16355	acctcgctgaagaatacgga	gtgagaggatgtcgtcgagc	1400
malE	malic enzyme	AF234535	catgaggaaatcttcgaggc	gagcttcatttcaggggtga	1100
pfk	phosphofructokinase	AX109130	ccgagtattgattccacgct	ggtcaggacgaggaatggta	1000
pgk	phosphoglycerate kinase	X59403	gacttcaatgtccccctcaa	ccttgccttcaaggtactcg	1200
ppc	phosphoenolpyruvate carboxylase	X14234	ctgcgagcgttgcttaaagt	cgaaagattggactgcacac	3194
pyc	pyruvate carboxylase	Y09548	cactttggcttgaagtcgtg	ggtttaggaaacgacgacga	3588
pyk	pyruvate kinase	L27126	gatggaattctgcgtttggt	catcatgtcatcggtatcgc	1300
sdhA	succinate dehydrogenase subunit A	AX113263	actcggatacgacgtcaagg	acggattcgaagaacagtgg	1900
sucA	2-oxoglutarate dehydrogenase	D84102	aggcataatcagccaacgac	tgcaccttagcaacaccagt	3710
sucC	succinyl-CoA synthase alpha subunit	NC_003450	ctcaattcagattcccgcat	cataagcttcgcggtttcac	900
thrA	homoserine dehydrogenase	Y00546	catctgccccaagctttaac	tgattgccttaacaacaggc	1400
tpi	triosephosphate isomerase	X59403	cgtaagccacttatcgctgg	gacgatctcagcgacggttt	800
zwf	glucose-6-phosphate dehydrogenase	E13655	tacaaggagcttttcgacgc	ttcattggtggactcggtaa	2300

coding regions in the printing stocks of these particular spots to ensure that partial gene sequences were not randomly chosen that had significantly lower homology to the unknown sequence of the analogous *C. glutamicum* gene. The goal was to avoid a situation in which the *M. tuberculosis*-derived spots gave results that were any less useful

as predictors of *C. glutamicum* transcriptional levels than was necessary, in the event of a close correlation between sequence homology and microarray signal ratio results.

3.1.3 Polymerase Chain Reaction Protocol

In amplifying the selected DNA sequences from *C. glutamicum* and *M. tuberculosis* genomic DNA, that have lengths between three and five megabase pairs, it was found that the success of these amplifications was very strongly influenced by adjustments made to the PCR protocol. This was in contrast to results typically seen for amplification from DNA templates of much smaller sizes, such as plasmids that typically consist of between 1 and 20 kilobase pairs. In these cases, the reactions are much more robust, and more significant changes to the PCR parameters are needed to move from a regime in which the desired sequence is created in large quantities to one in which the specificity or productivity of the reaction decreases. This lack of robustness when genomic DNA is used as the template can be attributed to the strong DNA-DNA interactions that intensify with the greatly increased template length. Because of these interactions, PCR parameters must be selected that favor the destabilization of the double-stranded genomic DNA enough to facilitate the dissociation of the two strands, but conversely do not prevent the annealing of the PCR primers to the single-stranded genomic DNA. These primers were 20-mers (Sigma-Genosys, The Woodlands, TX) that were chosen by the Primer3 software [Rozen and Skaletsky, 2000] as having a melting temperature of 60°C and product sizes of 500 base pairs or that of the entire open reading frame, with no preference given for the position of the sequence within the open reading frame.

To assay the success of the PCR reactions used in the construction of the DNA microarrays, agarose gel electrophoresis of 2-μl aliquots of each PCR reaction was used in visualizing the quality and quantity of DNA to be printed. In these gels, positive results are confirmed by the presence of significantly bright, singular bands indicative of sufficient DNA concentration of only the gene sequence of interest. Examples of PCR results are shown in Figure 3-1 in the form of an electrophoresis gel that has been used to analyze PCR products of multiple reactions used to amplify four different DNA

Figure 3-1: Gel electrophoresis analysis of PCR reactions used to amplify DNA sequences for creating microarray printing stocks. Each quadrant of the gel contains one lane of a 1-kb DNA ladder standard and six lanes of samples from PCR reactions used to amplify one particular sequence. Each of these six reactions was carried out at a different annealing temperature—54.0, 55.5, 57.0, 58.5, 60.0, and 62.5 °C from left to right. The lowest bands in each lane consist of unincorporated material.

sequences. The gel includes cases in which the target gene was amplified sufficiently well, in which no product was detected in any significant concentration, and in which the specificity of the reaction was poor, with multiple different sequences created.

These latter negative results were corrected by making various adjustments to the PCR method. In optimizing the protocol used for amplification of regions of genomic DNA, three parameters were found to be crucial in determining the success of the reactions— the concentration of genomic DNA, the temperature used in annealing the 20-mer primers to the genomic DNA template, and the selection and concentrations of adjuvant co-solvents used to destabilize the DNA-DNA interactions within the reaction mixture. Regarding the first of these, it was found that at template concentrations above 100 µg/l,

no amplification occurred. As destabilization agents, a combination of 5 % (v/v) glycerol and 3 % (v/v) dimethyl sulfoxide (DMSO) was shown to markedly increase the PCR yields. And an annealing temperature of 58.4 °C gave the most consistent success rate in creating material suitable for printing, requiring no change for nearly all of the 500-bp sequences to be amplified.

Of these three parameters, the optimum annealing temperature was the most dependent on the size of the sequence to be amplified. Frequently, when sequences larger than 500-bp were desired, tests would need to be conducted involving a gradient of annealing temperatures to determine at what temperature sufficient product could be created. As an example of the specificity of the relationship between the best annealing temperature and the particular sequence size being amplified, note that in Figure 3-1 two of the genes give the best PCR productivity or specificity at the higher of the temperatures tested, while another gene only generates PCR product at the lower of the tested temperatures.

Listed below, and in multiple other sections throughout this chapter, is a detailed step-by-step description of one of the protocols used in creating the *C. glutamicum* partial genome DNA microarrays, and in applying them to a typical experiment. These protocols have been reproduced in a form distinct from the standard text for easy future reference.

Protocol for PCR Amplification of DNA to be Printed Onto the Arrrays
- PCR reaction mixtures are prepared with the following composition:
  - 10 μl 10x PCR buffer
  - 2 μl primer solution
  - 2 μl genomic DNA
  - 2 μl 10 mM dNTP mix
  - 3 μl 50 mM magnesium chloride
  - 0.5 μl 1 U/ μl Taq DNA polymerase
  - 5 μl glycerol
  - 3 μl DMSO
  - 72.5 μl de-ionized water
- The PCR reactions are incubated in a thermal cycler with the following program:
  - 95 °C for 10:00 minutes
  - 58.4 °C for 0:45 minutes
  - 72 °C for 2:30 minutes
  - 95 °C for 0:30 minutes
  - go to second step 39 times
  - 72 °C for 10:00 minutes

o   < 4 °C until purification

3.2 Amplified DNA Purification and Printing

3.2.1 PCR Product Isolation

To isolate the DNA sequences produced by the PCR amplifications described above, two different isolation techniques were tested. In the first, a QIAquick PCR Purification Kit from Qiagen (Valencia, CA) was used to separate the partial and full gene sequences of interest from the buffer, reverse transcriptase enzyme, cofactors, PCR primers, and unincorporated nucleotides present in the PCR reaction mixtures. In this procedure, the reaction mixture is applied to a silica gel membrane contained in a microfuge tube. The PCR product is then washed and eluted through the use of centrifugation and buffers developed to promote the selective adsorption and desorption of DNA molecules within particular size ranges. In the second technique, a phenol/chloroform/isoamylalcohol mixture is used in an extraction in which the extraneous reaction components partition into the organic layer. In both techniques, the resulting DNA present in an aqueous solution is later concentrated by ethanol precipitation.

It was found that although the PCR Purification Kit may have resulted in DNA of a higher purity due to the more complete removal of undesired materials, the yield of the DNA was significantly greater when the extraction route was instead used for product isolation. Enhancement of the purity of the DNA sequences after PCR amplification is desired; however the likelihood that small concentrations of the other components of the reaction would adversely affect the quality of the resulting DNA microarrays was small. In the printing process to follow, DNA is selectively cross-linked to the array surface, which is subsequently and repeatedly washed, reducing the chance that buffer salts and the enzyme would remain. Also, any incidentally printed unincorporated nucleotides would be too small, and PCR primers would be too identical in sequence to the desired

DNA, to cause significant nonspecific hybridization in the use of the microarrays. In contrast, the quality of the microarrays is very dependent on the concentrations of DNA present in printing stock solutions. It was found that if these stocks had titers outside of the range of 100-400 mg/l, then the fluorescent signal resulting from labeled cDNA hybridization at the resulting array spots would be either too dim or too bright to yield statistically significant data. Because of this, the phenol/chloroform/isoamylalcohol extraction protocol was adopted for use in purifying each of the sequences generated by PCR amplification.

Two solutions were tested for use in the resuspension of the DNA sequences after they had been concentrated by ethanol precipitation. Those solutions were 5% (w/v) saline/sodium citrate (SSC) buffer, and 50% (v/v) DMSO. The use of the solution containing DMSO resulted in printed spots with slightly better morphologies that more often approached symmetrical circles without gaps present within their borders. Also, because of its high boiling point, DMSO gave the benefit of reducing the evaporation rate of the printing stock solutions. Since these solutions were required to have their surfaces exposed to ambient air for several hours each time a new supply of microarrays was produced, this lower evaporation rate was a welcome property.

Protocol for the Isolation of Sequences Produced by PCR Amplification
- To each PCR reaction is added 100 μl phenol/chloroform/isoamyl alcohol solution.
- The mixtures are vortexed for 5 seconds and centrifuged for 5 minutes at 13,000 RPM.
- To each supernatant is added 50 μl chloroform.
- The mixtures are vortexed for 5 seconds and centrifuged for 2 minute at 13,000 RPM.
- To each supernatant is added 10 μl 3 M pH 5.2 sodium acetate and 250 μl ethanol.
- The mixtures are incubated at -20 °C for > 1 hour and centrifuged for 15 minutes at 13,000 RPM.
- To each pellet is added 400 μl -20 °C 70% ethanol and the tubes are centrifuged for 7 minutes at 13,000 RPM.
- The pellets are air-dried for 20 minutes and resuspended in 20 μl 50% DMSO solution.
- The resulting solutions are distributed in 96-well plates.

3.2.2 Printing Pattern and Intraslide Replicates

The DNA sequences resulting from PCR amplification were printed onto amino-silane-coated glass slides (GAPS II; Corning Inc. Life Sciences, Acton, MA) with the use of an arraying robot (GMS417; Affymetrix, Santa Clara, CA). The robot used a pin-and-ring mechanism in which a head containing four small ring loops was lowered toward a 96-well plate containing printing stock solutions. The rings were submerged into the solutions and removed in such a way as to pick up films of liquid containing the DNA to be printed. Four pins then pierced the films and tapped the surface of glass slide three times to deposit the DNA, printing four spots.

Due to the large number of different sequences that must be printed onto a slide with limited available surface area, most full genome microarrays have only a small minority of their component DNA sequences present in more than one spot per array. One advantage afforded by the smaller number of genes represented by partial genome microarrays such as the one developed in this work is that it is much easier to print multiple spots containing identical sequences. In this particular case it was determined that each sequence would be printed onto each array in six different locations using two of the four printing pins of the robot. A representative microarray printing pattern is depicted in Figure 3-2, showing the six locations of the spots for one of the printed sequences.

By including six different spots on the array in this manner, experimental error is decreased in several important ways. The widely dispersed positions of the spots within the array create six different x-axis locations and two different y-axis locations where the spots may be found. In this way in the event that a location-specific phenomena renders a certain area of the microarray unusable, the probability that all or several of the spots associated with a particular gene are affected is minimized. Examples of such location-specific phenomena include the depositing of salts or other debris onto the array in the process of slide handling or hybridization, slippage of the coverslip used to contain the hybridization mixture in contact with the printed array area, and the premature evaporation of this hybridization mixture. Also, by using one pin to create three of the six spots and a second pin to create the remainder, the risk is reduced that a mechanical

identical sequence
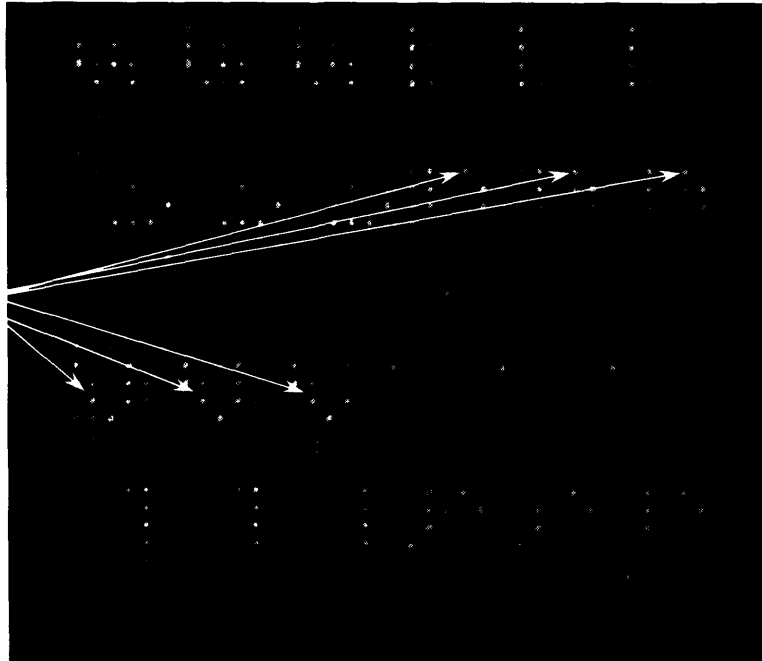
Figure 3-2: Example of DNA microarray spot printing pattern with replicates. The locations of six spots of the array in which the same DNA sequence has been printed have been marked.

error with the head assembly causes improper printing for all six of the spots. Finally, because the first three of the six spots are printed for all of the sequences before the 96-well plates containing the DNA stock solutions are repositioned and the robot is reset to print the second half of the array, there is a delay of as much as several hours between the printing times of the first and last spots for each sequence. Again, this provides insurance that a time-specific error in the printing process would not invalidate all spots providing transcriptional information about an individual gene.

After printing, the DNA was UV cross-linked to the slides at 650 µJ (Stratalinker 2400; Stratagene, La Jolla, CA) and blocked by incubating for 20 min in a mixture of 3 g/l succinic anhydride, 50 % (v/v) 1-methyl-2-pyrrolidinone, and 0.1 M boric acid at pH 8. The arrays were next washed with water and then ethanol, dried, and stored in the dark at room temperature. To validate that printing had taken place properly the fluorescent dye SYBR Green (Molecular Probes, Eugene, OR) was used to nonspecifically label the arrays. SYBR Green has a very high affinity for double-stranded DNA [Schneeberger *et al.*, 1994], and by using it to bind to the DNA on the microarrays and then scanning the

arrays at a wavelength of 532 nm with a GenePix scanner (Axon Instruments, Union City, CA), a visualization of the arrays could be created. In this way the locations and morphologies of the spots could be ascertained, and the quantity of DNA printed at each spot could be estimated through its qualitative correlation with fluorescent intensity.

Protocol for Printing DNA Microarrays
- The PCR products are spotted onto Corning CMT-GAPS II amino-silane-coated glass slides with 6 replicates per sample.
- The DNA sequences are cross-linked to the slides by exposing to 90 mJ UV radiation.
- A solution of 6 g/l succinic anhydride in 1-methyl-2-pyrrolidinone is prepared.
- Immediately after the succinic anhydride is dissolved, the solution is added to an equal volume of 0.2 M pH 8 boric acid to create a blocking solution.
- The microarray slides are submerged in the blocking solution and incubated in the dark for 20 minutes.
- The slides are rinsed first with water and then with ethanol.
- The slides are dried with compressed air and stored in the dark.
- Presence of DNA on a printed microarray is confirmed by applying 1 ml SYBR to the array surface for 10 minutes in the dark, washing first with TAE buffer and then with ethanol, drying the slide by centrifuging for 3 minutes at 500 RPM, and scanning the slide at 550 nm.

3.3 RNA Isolation

3.3.1 Initial RNA Isolation and Purification Protocol

To obtain from the cell cultures the messenger RNA that would be used to evaluate the gene expression phenotypes, total RNA samples were isolated in the course of the experiments. To do this, culture samples were stored overnight at -80°C to prevent further mRNA synthesis or degradation. No sooner than one day later, the samples were thawed and centrifuged for 10 min at 4000 rpm and 4°C. This freezing and thawing served to weaken the cell walls of the gram-positive organism and facilitate cell lysis and the release of the intracellular RNA. The pellets were resuspended in 3 ml RLT buffer

(Qiagen, Valencia, CA) and 30 μl 2-mercaptoethanol. The resulting slurries were added to sterile 6-ml glass vials that contained 3 ml of 100-micron-diameter glass beads, and the vials were agitated for 4 cycles of 2 min on and 2 min off in an orbital glass bead mill (5100 Mixer Mill; SPEX Certiprep, Metuchen, NJ) at 4°C to lyse the cells. To settle the beads and excess cell debris, the vials were then centrifuged for 1 min at 4000 rpm at 4°C. Total RNA in the supernatants from the vials was then isolated according to the Qiagen RNeasy Midi Bacterial protocol. Lithium chloride was used to precipitate the RNA in the eluants, and the pellets were resuspended in 20 μl sterile water. RNA samples were stored at -80°C until used to create labeled cDNA.

3.3.2 Effects of Growth Phase on RNA Yield

An experiment was run to determine the effect that the time of sampling during the course of the growth of a culture had on the quantity and quality of RNA that could be isolated from those samples. *C. glutamicum* cultures were grown for three days, during which time the biomass concentration was monitored by measurement of the optical density (OD) at 600 nm. The biomass was estimated with the equation

$$\text{biomass [grams dry cell weight per liter culture]} = \text{OD}(600) \times 0.25$$

RNA was isolated from samples taken during the early-exponential, late-exponential, and stationary growth phases of the culture. The results, shown in Figure 3-3, indicate that by far the greatest yield of RNA was achieved when samples were taken during early-exponential growth. The suggestion is that actively growing and multiplying cells, of the type found in exponentially growing batch cultures or steady-state continuous cultures, possess the highest abundance of RNA and provide the best opportunity for quantifying transcriptional levels. The quality of the RNA isolated, as measured according to the procedures outlined in section 3.3.4 below, was consistent among all samples regardless of culture phase. For this reason, in most of the experiments detailed in the following chapters, microarray analyses were performed on cultures during exponential growth, as the cells are still producing more biomass, and are just beginning to produce lysine.
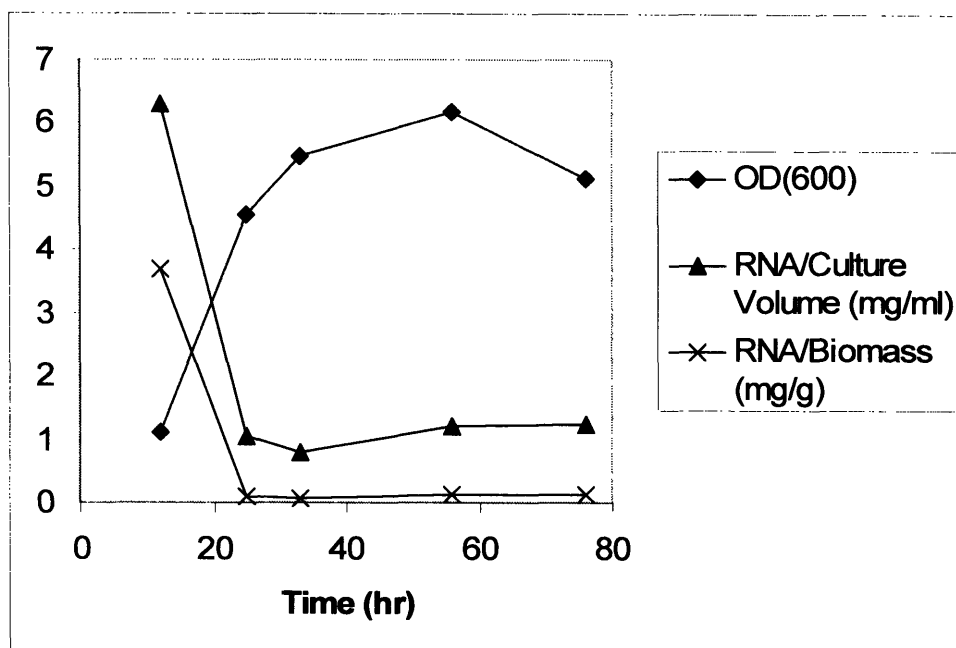
Figure 3-3: Effect of culture growth phase on yield of RNA isolated from culture samples. The trends depict biomass formation as measured by the optical density, and the yields of isolated RNA on biomass and culture volume from samples taken at the indicated time points.

3.3.3 Identification and Reduction of Genomic DNA Contamination

In carrying out early tests of the internally developed DNA microarrays using specific hybridization of labeled cDNA rather than nonspecific binding with SYBR staining, it was found that in many cases the quality of the results as measured by signal variance and the signal-to-noise ratio was not acceptable. Before implementing several improvements to the protocols described below, the majority of the arrays exhibited fluorescent signals that were less than 20% greater than the local backgrounds of the spots, and that had signal-to-noise ratios of less than 2. This measure of the confidence that signal values close to the background level are statistically significant is commonly accepted to have a minimum value of 3 for results to be considered accurately quantified. Accordingly, the variance, a measure of the deviation of results obtained from multiple

spots printed with the identical DNA sequence, rose to unacceptable values approaching 50% with these arrays. The two areas where improvements in purity yielded dramatic improvements in the quality of microarray data were those of RNA isolation, detailed in this and the following section, and labeled cDNA preparation, described in section 3.4.

As a means of testing the quality of the mRNA isolated from culture samples, cDNA was created in reverse transcription reactions in which the mRNA served as templates. PCR reactions with some of the same primer pairs used in generating the DNA printed onto the arrays were then used to attempt to amplify selected sequences from this cDNA. If the mRNA was of good quality and transcription of the genes specific to the primer pairs had occurred, then the PCR reactions would amplify these gene sequences to concentrations high enough to be visualized on agarose gels similar to those described in section 3.1.3. As negative controls, cases were carried out in which no reverse transcriptase enzyme was added to the initial reactions. Here, there will be no formation of cDNA, and there should be no template suitable for amplification in the subsequent PCR reactions.

Results from the experiment are summarized in Table 3-5. In total, three pools of RNA and five PCR primer pairs were tested, and in all 15 of these combinations the desired sequence was amplified and its concentration and correct size and were verified by agarose gel electrophoresis. However, in 12 of the 15 negative control combinations, amplification also occurred, even though no new template was created. The explanation for this is that some genomic DNA had likely been isolated along with the RNA, and the open reading frames of this genomic DNA, rather than cDNA formed in reverse transcription, is serving as the PCR template. The presence of genomic DNA has a very detrimental effect on the quality of microarray data, because the arrays will no longer be measuring the presence of gene sequences on only mRNA. As the level of contaminating genomic DNA increases, the relevance of the fluorescent intensities measured at the microarray spots decreases, because the concentration of fluorescently labeled cDNA of a particular gene is less dependent on the abundance of its related mRNA within the total isolated RNA pool.

The genomic DNA contamination was removed by incubating each RNA sample for 30 min at 30°C in the presence of 5 U DNase. The DNase was then inactivated by incubating

Table 3-5: PCR amplification of sequences found on cDNA formed by reverse transcription. RNA samples isolated from three different cultures were used as templates in six reverse transcription reactions. One reaction for each RNA sample was a negative control that did not contain reverse transcriptase. PCR reactions were then run for each of the reverse transcriptions to amplify five randomly chosen sequences from any DNA that was present. Amplification results were visualized with agarose gel electrophoresis, and assigned qualitative descriptors of "-" for no amplification, "+" for some amplification, and "++" for strong amplification.

Gene	RNA 1	RNA 1 Control	RNA 2	RNA 2 Control	RNA 3	RNA 3 Control
gpm	++	++	++	++	++	++
pfk	+	-	+	-	+	+
gpi	+	+	+	+	+	++
malE	+	-	+	+	+	++
sdhA	++	++	++	++	++	++

the samples for 10 min at 65°C. The samples were diluted with 2 ml RLT buffer and 20 μl 2-mercaptoethanol and the Qiagen RNeasy Midi Cleanup protocol was followed to additionally purify the RNA. The eluants were again precipitated with lithium chloride, and the precipitates were again resuspended in 20 μl sterile water. In this way a small drop in the overall yield of RNA was observed, but the purity of the RNA was greatly improved.

3.3.4 Identification and Reduction of RNA Degradation

Two methods were used for the determination of the purity of isolated RNA, one involving electrophoresis and the other involving spectroscopy. In the first, 2-μl aliquots from each sample were studied with agarose gel electrophoresis. High quality RNA was identified by the sole presence of two distinct bands in such a gel, each representing a type of ribosomal RNA that together make up most of the total RNA isolated from the cell cultures. Figure 3-4 shows two examples of gels used in this way to assay RNA purity. In the left gel, in addition to the two bands of ribosomal RNA, large regions of
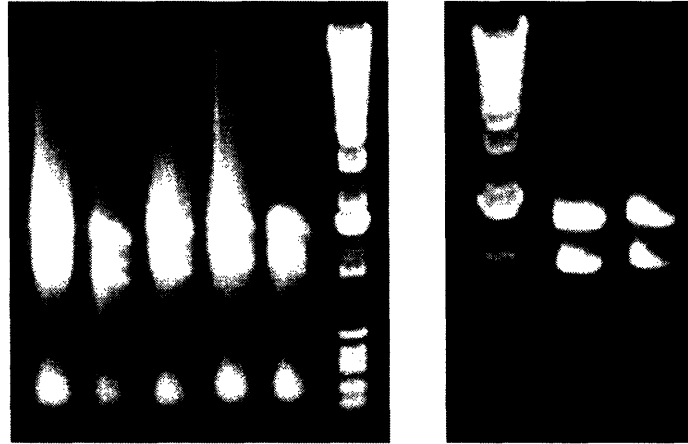
Figure 3-4: Electrophoretic assay of isolated RNA quality. Each of the two gels shown contains one lane of a 1-kb DNA ladder standard and multiple lanes of RNA samples isolated from cell cultures. RNA of acceptable quality produces two distinct bands corresponding to two sizes of ribosomal RNA.

additional fluorescence can be seen both below and above these bands. Fluorescence above the ribosomal RNA bands is caused by genetic material with larger masses—most likely the genomic DNA shown in the above section to be a potential contaminant in the RNA samples. The fluorescence below the ribosomal RNA bands is caused by material with less mass and is most likely the result of degradation of the RNA that has been isolated. In the second RNA assay, a BioPhotometer (Brinkmann, Westbury, NY) was used to measure the optical density of samples at both 260 nm and 280 nm. A ratio of OD(260)/OD(280) in the range of 1.7-2.1 was associated with RNA of high quality, with ratios above 2.1 suggesting contamination with genomic DNA and ratios below 1.7 suggesting RNA degradation. These optical density readings were also used to quantify the amount of purified total RNA that had been isolated from each cell culture sample.

RNA degradation is an extremely common occurrence, both *in vivo* and *in vitro*. DNA must have good stability within cells to be able to serve as a constant genetic template for both DNA replication involved in cell division and DNA translation involved in the formation of RNA transcripts. Conversely, the much lower stability of RNA serves a different but equally important role for the cells. By increasing RNA instability, the makeup of the total intracellular pool is able to change in a relatively short period of time,

allowing the instantaneous concentrations of available transcripts to be controlled by the cells in such a way as to meet the particular demands for gene expression at that time. One method that the cells use to achieve a high RNA turnover rate is a system of RNases that have high activity both intracellularly and extracellularly. One of the goals of the RNeasy RNA isolation kit used is to inactivate and/or remove these RNases as a means of increasing the stability of total RNA samples. That the RNA showed significant degradation nonetheless suggests that *C. glutamicum* has a particularly active and robust RNase population that required further modifications to the RNA isolation protocol to eliminate.

The addition of a second filtration pass through the Qiagen RNeasy Midi Cleanup column as described in section 3.3.3 was one strategy used to help increase RNA stability. Another was the use of RNAwiz (Ambion, Austin, TX), a proprietary stabilization solution that operates by passing through the walls of the cells and inhibiting both the degradation of existing RNA and the synthesis of new mRNA. After implementing both of these changes, agarose gel assays of all RNA samples produced results similar to those shown in the gel on the right in Figure 3-4.

3.3.5 Final RNA Isolation and Purification Protocol

Protocol for RNA Isolation and Purification

- Stop solutions of 1.6 ml ethanol and 84 μl RNAwiz from Ambion are prepared and stored at -80 °C.
- Each 15-ml culture sample is added to one stop solution aliquot and stored at -80 °C for > 16 hours.
- The samples are thawed at 25 °C for 10 minutes and centrifuged for 10 minutes at 4000 RPM and 4 °C.
- The pellets are resuspended in 3 ml RLT buffer from Qiagen containing 10 μl 2-mercaptoethanol per ml RLT buffer.
- The cells are mechanically lysed with a glass bead mill in which the 3 ml cell slurry is added to 3 ml 0.1 mm glass beads and the mixture is subjected to 4 cycles of milling for 2 minutes and incubating in ice for 2 minutes.
- The mixture is centrifuged for 1 minute at 4000 RPM and 4 °C to remove excess beads and cell debris.

- The supernatant from the vial (approximately 1.5 ml) is mixed with 0.7 volumes of ethanol (approximately 1.05 ml) and applied to an RNeasy Midi Column from Qiagen.
- The column is centrifuged for 3 minutes at 4000 RPM and 25 °C.
- The column is washed with 4 ml Qiagen buffer RW1 by centrifuging for 3 minutes at 4000 RPM and 25 °C.
- The column is washed twice with 2.5 ml Qiagen buffer RPE by centrifuging for 1 minute for the first wash and 5 minutes for the second wash, both at 4000 RPM and 25 °C.
- The column is eluted twice with 250 µl sterile water by applying to the column for 1 minute and centrifuging for 3 minutes at 4000 RPM and 25 °C.
- The RNA in the eluant is precipitated by adding 0.5 volumes of 4 M LiCl, incubating for 30 minutes at -20 °C, and centrifuging for 12 minutes at 13000 RPM.
- The precipitate is washed with 750 µl 70% ethanol, dried for 20 minutes at room temperature, and resuspended in 20 µl sterile water.
- DNA in the solution is digested by adding 5 U RNase-free DNase and incubating for 30 minutes at 30 °C.
- The DNase is inactivated by incubating the sample for 10 minutes at 65 °C.
- The sample is added to 2 ml RLT/2-mercaptoethanol and 1.4 ml and applied to an RNeasy Midi Column.
- The column is centrifuged for 3 minutes at 4000 RPM and 25 °C.
- The column is washed twice with 2.5 ml RPE by centrifuging for 1 minute for the first wash and 5 minutes for the second wash, both at 4000 RPM and 25 °C.
- The column is eluted twice with 250 µl sterile water by applying to the column for 1 minute and centrifuging for 3 minutes at 4000 RPM and 25 °C.
- The RNA in the eluant is precipitated by adding 0.5 volumes of 4 M LiCl, incubating for 30 minutes at -20 °C, and centrifuging for 12 minutes at 13000 RPM.
- The precipitate is washed with 750 µl 70% ethanol, dried for 20 minutes at room temperature, and resuspended in 20 µl sterile water.
- RNA quantity and quality are assayed by gel electrophoresis and spectrophotometry.
- Purified RNA samples are stored at -20 °C.

3.4 Creation of Labeled cDNA and Hybridization

3.4.1 Effects of Fluorescent Label and cDNA Purification Protocol on Signal Properties

After producing a pure sample of RNA that has been isolated from a culture, the next step in the microarray process is the formation of labeled cDNA through a reverse transcription reaction using the RNA as a template. Two different fluorescent labeling techniques, one using CyDyes (Amersham Biosciences, Piscataway, NJ) and the other using Alexa fluors (Molecular Probes, Eugene, OR), were studied in the course of optimizing the protocol for creating this cDNA. The CyDyes tested as cDNA labels were Cy-dUTP nucleotide analogs that are incorporated into the DNA during the reverse transcription. The Alexa fluors are succinimidyl esters that are chemically bonded to aminoallyl-dUTP nucleotide analogs after the aminoallyl-dUTP has been incorporated into the cDNA in an earlier process step.

The two-step labeling process of the Alexa fluors seems to offer multiple advantages over that of the CyDyes. The DNA is purified at the end of each of the two main steps in the Alexa procedure, compared to only one purification step in the CyDye procedure. This should produce a cleaner solution of labeled cDNA that will be applied to the microarrays and should in turn reduce concentrations of extraneous material that could bind nonspecifically to the arrays. In addition, because the same aminoallyl-dUTP nucleotide analog will be inserted into both of the cDNA pools created for hybridizing to a particular array, the subsequent fluor content of each pool should be identical on a fluor per base basis. In contrast, with the CyDyes two distinct compounds—Cy3-dUTP and Cy5-dUTP—are used directly in labeling the two cDNA populations for a single array. Because Cy5-dUTP is a larger molecule than Cy3-dUTP, there are more steric hindrances to its incorporation into cDNA. As a result, the concentration of Cy5-dUTP is lower than that of Cy3-dUTP on a fluor per base incorporation basis and the cDNA labeled with Cy5-dUTP will have a lower signal intensity when visualized on the array. Also, because

Alexa fluors have a stronger fluorescent emission intensity per molecule than CyDyes, and they incorporate into DNA at a greater frequency, less cDNA needs to be produced in the reverse transcription reaction and hybridized to the arrays in order to create strong enough signals for accurate quantification. Since less cDNA needs to be produced, a smaller amount of RNA is needed to serve as a template. Because high yields of RNA on sampled biomass are not always realized, as was discussed in section 3.3.2, this is an advantage in the overall microarray protocol.

Initially, the use of Alexa fluors appeared to provide a very good alternative to the use of CyDyes as fluorescent DNA labels, not only for the reasons discussed above, but also because the signal-to-background and signal-to-noise ratios seen for arrays hybridized with Alexa-labeled cDNA were consistently and significantly higher than those seen with arrays using CyDye. However, further testing revealed that although the signals associated with these Alexa arrays were very intense, they were also much less specific relative to the sequences found on the individual spots. One measure of this lack of specificity involved the use of SpotReport Alien DNA sequences (Stratagene, La Jolla, CA). These artificial sequences have been designed to have no significant homologies with any published sequences of prokaryotic or eukaryotic open reading frames or genomes. Four of these artificial DNA sequences were printed onto the microarrays, and mRNA spikes complementary to two of the four sequences were added to the total RNA isolated from culture samples. In this way, the spots associated with these two sequences were used as positive controls, since it was known that the reverse transcription reaction should produce labeled cDNA that would hybridize to them. The other two sequences were used as negative controls, since there should be no sequences present in the RNA isolated from *C. glutamicum* that can be used for generating labeled cDNA that would significantly hybridize to them. When the Alexa fluors were used in the generation of cDNA from RNA templates that contained the Alien mRNA spikes, the positive control spots fluoresced with intensities that were, on average, in the 91st percentile of all spot intensities in the array. Contrary to expectations, though, the negative control spots also showed strong fluorescence, with average intensities that were in the 74th percentile of all spot intensities on the array.

One potential source for nonspecific fluorescence could be the presence of unincorporated label remaining from the reverse transcription reaction. In an attempt to remove this material from the final cDNA solutions applied to the microarrays, three different cDNA purification procedures were tested. In the first and most basic, ethanol precipitation was used to selectively precipitate the cDNA sequences of interest prior to their resuspension in a concentrated form in a hybridization solution. In the second and third, the QIAquick PCR Purification Kit and phenol/chloroform/isoamylalcohol extraction techniques described previously in more detail in section 3.2.1 were tested. Figure 3-5 presents the results of this experiment, from which two important trends can be seen. Firstly, the addition of a solvent extraction prior to ethanol precipitation brings about increases in signal-to-background and signal-to-noise ratios, while switching from the extraction method to the column filtration method of the Qiagen kit yields still further improvements. Secondly, as the cDNA purification method was refined, the degree to which the Alexa system provided stronger signals than the CyDye system decreased, eventually to the point at which the CyDyes were as good as or better than the Alexa fluors when column filtration was used.

Further tests using both the CyDyes and the filtration purification of labeled cDNA showed that the benefits of this combination extended beyond those of signal enhancement. The difference between the positive and negative control signal intensity percentiles tripled when compared to typical behavior with the Alexa fluors and minimal cDNA purification as described above. Also, the dynamic range of the signal ratios was increased from ± 1.6 to ± 5.3. This dynamic range is a measure of the difference between the signal ratio for the printed sequence whose spots fluoresced most preferentially in one of the two measured wavelengths, and the signal ratio for the sequence whose spots fluoresced most preferentially in the other wavelength. In this way it is a comparison of the values generated for transcripts that are much more prominent in one or the other of the two mRNA samples studied with a particular microarray. The signal ratios are, like all reported in this work, measured in a $\log_2$-scale, so the observed increase in dramatic range represents a 13-fold improvement in the specificity of the patterns of fluorescence within the microarrays. These results showed that the reliability that the data are giving an accurate depiction of the transcriptional state of the cell cultures was increased
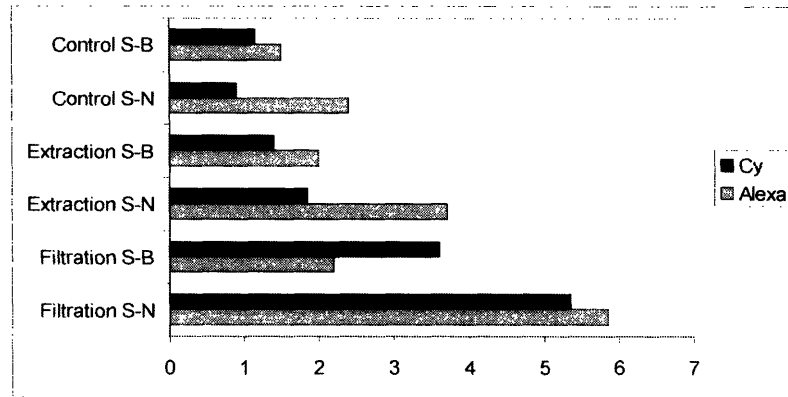
Figure 3-5: Effects of cDNA fluorescent label and purification on microarray signal qualities. A batch of RNA isolated from a culture sample was subjected to six reverse transcription reactions, three using CyDyes and three using Alexa fluors, to created fluorescently labeled cDNA. Each reaction in the two sets of three were purified according to a different protocol—ethanol precipitation (Control), solvent extraction (Extraction), and silica gel column filtration (Filtration). The six purified solutions of labeled cDNA were then used to hybridize six different arrays, and the signal-to-background (S-B) and signal-to-noise (S-N) ratios of each array were calculated and presented in the bar graph.

dramatically.

3.4.2 Application of Gene-Specific Primers

One strategy to decrease the likelihood of nonspecific binding of fluorescent label to the DNA microarray spots involves the removal of unincorporated fluors as described in the previous section. Improvement to another aspect of the purity of the cDNA solutions applied to the arrays proved to be even more significant. This involved the use of gene-specific primers to eliminate the production of cDNA sequences that are not being assayed by the microarrays as being representative of genes of interest [Talaat *et al.*, 2000]. In typical microarray protocols, the reverse transcription of RNA to form labeled cDNA is by design a nonspecific event, with random DNA hexamers used as primers in the amplification of sequences found in the RNA. While this theoretically allows for the

capturing of all open reading frame sequences that are represented by the mRNA in the samples, it also allows for sequences of other RNA components, such as ribosomal RNA and transfer RNA, to be transcribed into labeled cDNA that will have the opportunity to hybridize to the array. Any lack of homology between these undesired sequences and those of the spots printed onto the microarray should reduce the degree to which nonspecific hybridization will take place, but this hybridization is believed to occur at some at least minor background level across the entire population of spots. Because the undesired sequences will exist in each of the two cDNA solutions applied to each array, the degree of nonspecific fluorescence will be more or less equal in both wavelengths used to scan the arrays, and the $\log_2$ signal ratios derived for each spot will trend towards zero as nonspecific hybridization increases, severely depressing the resulting dynamic range.

In the system described above, the discrimination between desired and undesired sequences exists solely at the level of hybridization. A better alternative is to discriminate at the earlier step of transcription. By replacing the primer set of random hexamers with a set containing primers specific for the sequences actually being investigated by the DNA microarrays, the creation of undesired labeled sequences should be stopped. For extremely large microarrays, such as those testing all open reading frames within the genome of a particular organism, the number of primers needed to provide two for each gene on the array adds an undesired level of complexity to the creation of a specific primer set. Also, because the total concentration of primer in the reverse transcription reaction should be fixed at 2 μM, as the number of included primers grows the concentration of each individual primer within the mixture will decrease, eventually reaching a point at which their low levels impact the degree to which individual gene sequences are transcribed. Another advantage, then, in the use of a partial genome array is that the number of primers required to prepare a gene-specific primer mix is much more manageable. Also, because the arrays were developed internally, the same primers that were used to initially generate the printing stocks through PCR amplification were readily available for use in a gene-specific primer mix. Tests confirmed that instituting a protocol in which labeled cDNA was formed using an equimolar mixture of each of the 262 PCR primers used in the manufacture of the DNA arrays improved the quality of the

74

microarray results in several ways. Spot signal intensities were increased, array background noise was decreased, and the dynamic range of the signal ratios was greatly enhanced.

3.4.3 Final Reverse Transcription and Hybridization Protocols

Protocol for Creation of Labeled cDNA by Reverse Transcription and Hybridization of cDNA to Microarrays
- Two reverse transcription reaction mixtures are prepared with the following composition:
  - o 10-30 μg RNA
  - o 1 μl gene-specific primer solution
  - o 1 μl RT-dNTP mix
    - 5 mM dATP
    - 5 mM dCTP
    - 5 mM dGTP
    - 3 mM dTTP
  - o 2 μl of 2 mM Cy-dUTP (Cy3-dUTP for one reaction and Cy5-dUTP for the second reaction)
  - o 4 μl of 5X First Strand Buffer
  - o 2 μl of 10X dithiothreitol solution
  - o 1 μl SuperScript II reverse transcriptase
  - o H_2O to bring total reaction volume to 20 μl
- The reverse transcription reactions are incubated for 2 hr at 42 °C.
- To inactivate the enzyme the reactions are boiled for 5 min.
- NaOH is added to the solutions to a concentration of 0.3 M and they are incubated for 15 min at 65 °C to dissociate the new cDNA from the RNA and degrade the RNA template.
- HCl is added to the solutions to a concentration of 0.3 M to neutralize the base.
- The samples are mixed with 500 μl binding buffer PB and applied to QIAspin microfuge columns from Qiagen.
- The columns are centrifuged for 1 min at 13,000 RPM.
- The columns are washed with 750 μl washing buffer PE by centrifuging for 1 min at 13,000 RPM and dried by centrifuging for another 1 min at 13,000 RPM.
- The columns are eluted twice with 100 μl sterile water by applying to the columns for 1 min and centrifuging for 5 min at 13,000 RPM.

- The eluants are combined, and 40 µl 3 M pH 5.2 sodium acetate and 1 ml ethanol are added to the mixture.
- The mixture is incubated at -20 °C for > 1 hour and centrifuged for 15 minutes at 13,000 RPM.
- To the pellet is added 750 µl of -20 °C 70% ethanol and the tube is centrifuged for 7 minutes at 13,000 RPM.
- The pellet is air-dried for 20 minutes and resuspended in 30 µl GlassHyb hybridization buffer from Clontech (Palo Alto, CA).
- The resulting solution is pipetted between LifterSlip coverslips (Erie Scientific, Portsmouth, NH) and the printed surface of a DNA microarray.
- The microarray slide is placed into a GAPS hybridization chamber (Corning Inc. Life Sciences, Acton, MA) and incubated for 16 hr at 60 °C.
- After this hybridization, the slide is washed in 2X SSC buffer for 5 min.
- The slide is then washed in 0.1X SSC + 0.1% SDS buffer for 5 min.
- The slide is then washed in 0.1X SSC buffer for 5 min.
- The slide is dried by centrifugation for 5 min at 500 RPM and room temperature.
- A GenePix scanner and its related GenePix 3.0 software package is used to scan the slide at 523 nm and 635 nm.

3.5 Data Filtering and Normalization

3.5.1 Use of Alien DNA Sequences as Controls

The signals observed in analyzing DNA microarrays are subject to variability from a wide range of sources, such as the differences between incorporation frequencies for the two labels used, the amount of RNA isolated from cell culture samples, the efficiency of hybridization, and the settings used in scanning the arrays. To account for this variability, the data for each array is normalized so that signals within the same array and across a series of arrays can be compared to one another properly.

One common normalization technique involves scaling all signal ratios from an array by multiplying all values by a normalization value chosen so that the modified data set has a mean value of zero. The biological significance of this type of normalization is that it balances the degree of up-regulation and down-regulation within the transcriptome

measured by the arrays so that there is no net change. For microarrays that describe an entire genome, this is often valid because it is a reasonable estimation to consider the size of the entire population of mRNA within the cells to remain relatively constant at different time points. It is assumed that the composition of the transcriptome, rather than its total concentration, is what varies over time as the cells adapt to changing conditions and gene expression demands. For the smaller transcriptome subset that is studied by a partial genome microarray such as the one described in this work, the assumptions underlying this normalization are not as readily justifiable. Because only a fraction of the entire transcriptome is considered, even if the concentration of the entire transcriptome remains unchanged, it may be that the arrays involve predominantly those genes for which transcript concentration rises or falls.

Another normalization strategy does not consider the signal ratios for all spots present on an array, but instead on a few spots chosen specifically to be used as normalization controls. The spots are, for example, composed of DNA sequences corresponding to genes whose transcription should not change from one culture sample to another. Cell maintenance genes are a common classification used for this purpose, as they are involved in cellular processes marked by their rigidity in response to various intracellular and extracellular stimuli. Alternatively, sequences foreign to the organism can be used for these spots, freeing the normalization from reliance on any assumptions whatsoever about the biological behavior of transcription. One such set of foreign sequences are the Alien SpotReport sequences described in section 3.4.1.

To test the different normalization protocols as they apply to the microarrays that were created for this work, four Alien sequences were printed onto several arrays. The available mRNA spikes for three of the four sequences were added to one of the two RNA samples to be measured with the array, and spikes for a different set of three of the four sequences were added in equal amounts to the other sample. As a result, one of the four sequences would be expected to fluoresce at only one of the two measured wavelengths, a second sequence would fluoresce only at the other wavelength, and the third and fourth sequence would fluoresce equally at both wavelengths. The results from this experiment are shown in the histograms of Figure 3-6. In the first graph, the signal ratios have been normalized by setting the mean of all ratios at zero, and in the second,
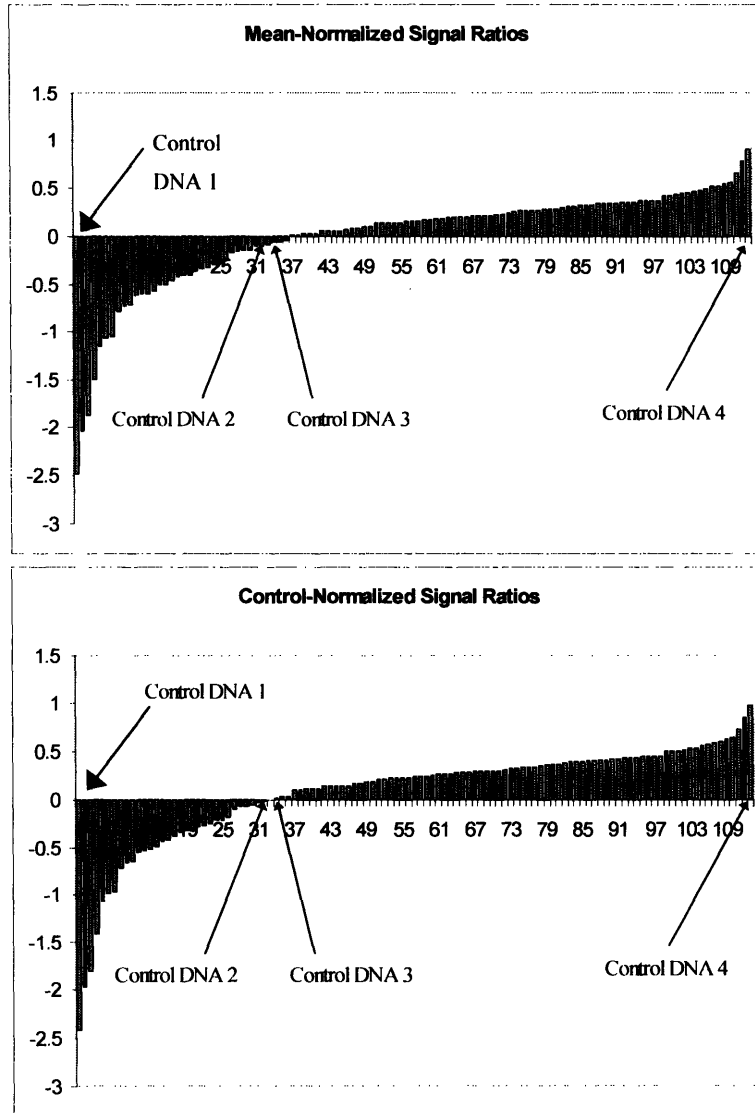
Figure 3-6: Comparison of microarray signal normalization techniques. Two distinct RNA populations were used to generate two samples of labeled cDNA that were hybridized to a DNA microarray. The signal ratios were calculated for each printed sequence as means of their associated spots. The upper histogram depicts all signal ratios after adjusting by normalizing to give a zero average of all signal ratios. The lower histogram depicts all signal ratios after adjusting by normalizing to give a zero average of signal ratios for control 2 and control 3. Note also that spots for control 1 properly gave the largest negative signal ratio and those for control 4 properly gave the largest positive signal ratio.

the normalization factor was chosen to set the average of just the two signal ratios for the Alien controls fluorescing in both channels at zero. That the histograms are almost

identical to one another provides a justification for the interchangeability of the normalization techniques in this case. This became important as the labeled cDNA formation protocol was altered to use gene-specific primers. Because the sequences of the Alien DNA are unpublished trade secrets, no non-random primers could be designed to allow for the transcription of the Alien mRNA spikes. The Alien sequences were then removed from the array printing pattern, and the arrays in all experiments described below were normalized based on balancing the signal ratios derived for all remaining sequences.

3.5.2 Data Filtering and Normalization Protocols

Prior to normalizing the data calculated for each sequence on the array, a two-step filtering process was used to discard those spots on the array for which the fluorescence values were not deemed to be sufficiently accurate. In the first step, data from all spots that showed a signal-to-background ratio of less than 1.25 in at least one of the two scanned fluorescent channels were discarded. For spots with signal intensities lower than this threshold, fluorescence at the spot was deemed not significantly distinguishable from either the background fluorescence local to the spot or the noise present throughout the array. In the second data filtering step, the variance of the signal ratio of each spot with respect to the signal ratios of the other spots representing the same printed DNA sequence was calculated. Data for all spots for which this variance was greater than one standard deviation larger than the mean of all such calculated variances were then also discarded. This step provided for the objective removal of the spots for which fluorescence values were most likely affected by factors other than transcriptional abundance, such as location-specific errors in hybridization or washing, mechanical errors in printing, and others detailed in section 3.2.2. Then, as described in the previous section, for each of the distinct DNA sequences that were printed on the microarray, the mean of the ratio of medians for each remaining spot was determined and transformed into $\log_2$-space. These values were then normalized such that their mean was zero, and they were reported as the transcriptional signals associated with the genes encoded by the DNA sequences.

Chapter 4: Validation of DNA Microarrays and Effects of Carbon Source on Gene Transcription

4.1 Experimental Design

After several *C. glutamicum* partial genome DNA microarrays were manufactured according to the optimized protocols described in Chapter 3, a series of experiments were carried out to test the statistical significance, experimental reproducibility, and biological consistency of the data that they were capable of generating. To do this it was decided to use the microarrays to compare the gene expression profiles of cells in *C. glutamicum* cultures that were grown on two different carbon sources. As was detailed in section 2.1.3, it is known that when *C. glutamicum* is grown on carbons other than the standard glucose, there are several impacts on both the intracellular enzyme activities and the extracellular culture properties that are observed. For example, cultures grown on fructose rather than glucose had lower lysine and biomass yields resulting from, among other factors, increased tricarboxylic acid cycle flux and decreased pentose phosphate pathway flux [Kiefer *et al.*, 2004]. Recently, the underlying bases for some of these observations have been explored by using microarrays to explore the transcriptomes of cultures grown on various substrates. Examples can be found in the microarray investigations of cultures grown with the use of acetate [Muffler *et al.*, 2002 and Hayashi *et al.*, 2002] or propionate [Hüser *et al.*, 2003] as carbon sources.

For this study, two minimal media compositions that contained either glucose or lactate as the sole carbon source were used. These two substrates were chosen because there have been several reports that outline significant differences seen between cultures that

have utilized them. Of particular interest to this work is the increase in pyruvate carboxylase activity seen in both native strains and strains engineered for the over-expression of the enzyme, when lactate is consumed in place of glucose [Koffas *et al.*, 2002]. To date there has been no study of the behavior of the cells on the transcriptional level, though, seeking to explain this or other lactate-associated behaviors.

For this first set of experiments to test the abilities of the arrays to accurately describe differences between two distinct culture conditions, the strain E12, the genome of which was the source of the DNA sequences printed on the arrays, was selected as the model strain. Cultures were cultivated in flasks containing a modified form of the minimal medium supplemented with threonine, methionine, and biotin that was developed by Park (1996) specifically for its ability to support *C. glutamicum* growth. This medium recipe, given in Table 4-1, was used for this and all other experiments described in the following chapters. A major modification from the medium of Park is the removal of sodium citrate, leaving as the only carbon source either 20 g/l glucose or 42 g/l 60% (w/w) sodium DL-lactic acid syrup (Sigma-Aldrich Corp., St. Louis, MO). These amounts set the initial carbon concentration at 680 mM.

Four cultures were grown in 250-ml round-bottom flasks containing 50 ml medium that were incubated at 30°C and 250 rpm on a rotary shaker (New Brunswick Scientific, Inc., Edison, NJ). Two of these cultures were fed glucose as the carbon source and two were fed lactate. Two samples of total RNA were isolated from each culture during mid-exponential growth that occurred at approximately 16 hr post-inoculation. One RNA sample from each culture was used in a reverse transcription reaction to generate cDNA labeled with Cy3-dUTP, and the other sample was used in generating cDNA labeled with Cy5-dUTP. Four microarrays were then hybridized with four pairs of cDNA solutions:

(1) Cy3-cDNA representing the first glucose culture and Cy5-cDNA representing the first lactate culture

(2) The same cDNA used with microarray (1)

(3) Cy3-cDNA representing the first lactate culture and Cy5-cDNA representing the first glucose culture

(4) Cy3-cDNA representing the second glucose culture and Cy5-cDNA representing the second lactate culture

Table 4-1: Composition and preparation of *C. glutamicum* minimal medium. Parts A, B, and C listed below are autoclave sterilized separately from one another and then diluted to 1X with sterile water. The components of part D are then filter sterilized and added individually.

6X Part A	
Glucose	120 g/l
or 60% (w/w) Sodium DL-lactate syrup	252 g/l
Sodium chloride	6 g/l
Magnesium sulfate heptahydrate	1.2 g/l
EDTA disodium salt dihydrate	0.45 g/l
Ferrous sulfate heptahydrate	0.15 g/l
Calcium chloride dihydrate	0.3 g/l
100X Mineral salt solution	60 ml/l

100X Mineral Salt Solution	
Magnesium sulfate monohydrate	225 mg/l
Sodium tetraborate decahydrate	30 mg/l
Ferric chloride hexahydrate	200 mg/l
Zinc sulfate heptahydrate	50 mg/l
Cupric chloride dihydrate	20 mg/l
Ammonium molybdate tetrahydrate	10 mg/l
Adjust pH to 2 with HCl	

6X Part B	
Potassium phosphate dibasic	48 g/l
Potassium phosphate monobasic	6 g/l

6X Part C	
Ammonium sulfate	30 g/l

Part D	
D-Biotin	1 mg/l
Thiamine-HCl	1 mg/l
L-Threonine	150 mg/l
L-Methionine	80 mg/l
L-Leucine	100 mg/l

4.2 Reproducibility Study

The four DNA microarrays produced by the hybridizations described above allowed for the comparison of data reproducibility in four important ways. In the first, data from each array was examined to confirm that identical DNA sequences printed in multiple spots

within the array gave similar results. Section 3.2.2 described the methods used for generating the six distinct spots within each array that were composed of the same DNA sequence, as well as the benefits of this type of array printing scheme. This test sought to prove that barring some form of experimental error of the type mentioned in section 3.2.2 that would affect a subset of the spots of an array, these six spots do indeed behave identically. The second comparison involved microarrays (1) and (2), described in the above section. Because these arrays were hybridized with the same two cDNA batches, any discrepancies between their results could not be attributed to factors associated with either the RNA samples isolated from the cultures or the cDNA produced by reverse transcription. Instead, any differences would be caused by factors related to the production of the two arrays, the hybridization of the cDNA to the arrays, or the washing and scanning of the arrays after hybridization is stopped. The third comparison involved microarrays (1) and (3), and was of a type commonly referred to as a dye swap. Here, the same two RNA samples are being assayed by each microarray, but the labeling strategies used in each array have been reversed. This test allows for an estimation of the degree to which there is any dye bias for the gene sequences present on the array. Dye bias relates to a preference that a particular sequence has for incorporating one of the two fluorescent nucleotide analogs into its cDNA. If this preference is significantly different among the population of genes studied by the array, then the array-wide normalization strategies discussed in section 3.5.1 will be insufficient in correcting the raw fluorescent signal data for all individual genes. The fourth and final comparison involved microarrays (1) and (4) and takes into account any biological variability that may occur from one culture to another. As with the dye bias quantified by the dye swap test, it is assumed that there is some level of biological variability inherent to any system that will be studied by microarrays. What is to be determined is the significance of this variability and its impact on the accuracy of obtained results.

In quantifying the results of the four reproducibility tests that were performed, the coefficient of variation was considered the key measure of the accuracy of the data. This was calculated by first using the GenePix scanner and software to determine the dual fluorescence intensities and the median of ratios for each spot on the array. The spot data was then filtered and normalized according to the protocols of section 3.5. Prior to the

values being transformed into $\log_2$-space, the mean and standard deviation for each printed sequence was computed using the signal ratio values for each of the up-to-twelve spots that represented the sequence on one or two arrays. The coefficients of variation for the sequences were then derived as the ratios of the standard deviations to the mean values obtained for the sequences. In the case of each of the four validation tests, the median of all coefficients of variation across the array or arrays considered was less than 20%, indicating successful validation of the internally created DNA microarrays. Also, as more arrays were included in a particular validation test, the median coefficient of variation decreased. For example, the spot-to-spot test of variability between identical spots within the same array yielded a median coefficient of variation of 18%. When data from the spots of a second, identical array were also considered to test array-to-array variability, the median coefficient of variation decreased to 16%, indicating that any source of additional variability brought about by differences from one array to the other were smaller than the additional confidence in the measurements brought about by a doubling in the sample size. These results demonstrated that measurement error, technical variability, dye bias, and biological variability that will be present in some amount in all experiments carried out with the arrays, will not influence the results to the degree that the statistical significance of the findings will be in question.

A graphical depiction of the results of the dye swap and biological variability validation tests is given in Figure 4-1. The plots show the average signal ratios for each DNA sequence as derived from one microarray on one axis, and the signal ratios derived from a second microarray on the second axis. The first graph presents data from the two arrays that were hybridized with cDNA derived from two different pairs of cultures. In this graph, L5G3 refers to the data set obtained with Cy5-cDNA based on a lactate culture and Cy3-cDNA based on a glucose culture. Similarly, L5G3-2 refers to the data set obtained from two other cultures that were labeled in the same way. From the plot, which closely approximates the ideal case of a scatter along the line y = x, it can seen that the agreement between the results from the two arrays was very high, and that the standard deviations of the data at each individual point is quite low relative to the overall spread of values. This gives confirmation that when RNA samples are isolated from distinct
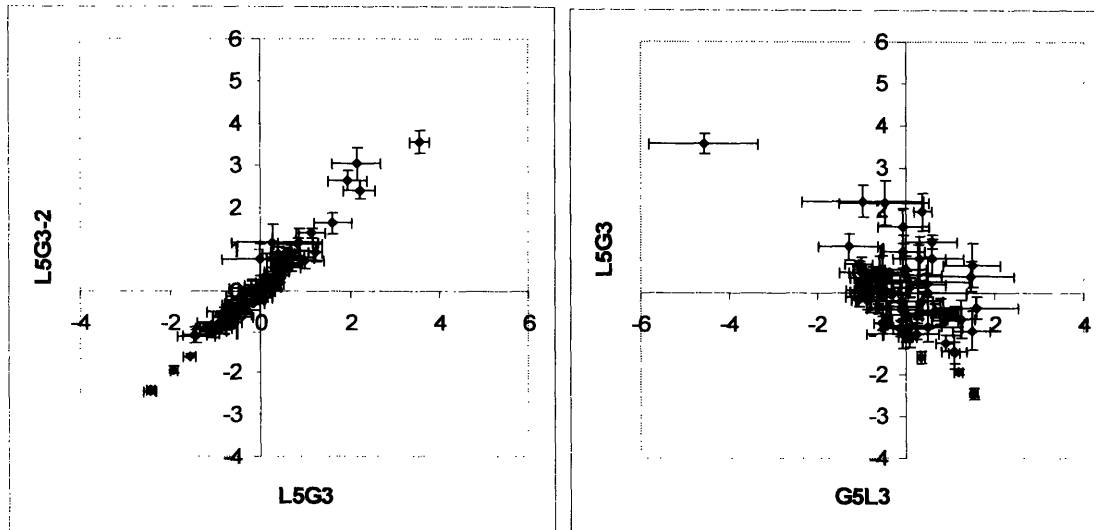
Figure 4-1: Reproducibility tests of DNA microarrays. Four cultures of E12 were grown on minimal media with two receiving glucose and two receiving lactate as the carbon source. Axes depict the signal ratios from the microarrays on a $\log_2$-scale. L5G3 signifies an array hybridized with cDNA derived from a lactate culture labeled with Cy5-dUTP and from a glucose culture labeled with Cy3-dUTP. L5G3-2 signifies an array hybridized with cDNA labeled similarly to L5G3 but derived from different cultures. G5L3 signifies an array with the opposite labeling scheme.

cultures that represent the same cultivation conditions and sampling times, the transcriptional behavior of those two cultures as described by the DNA microarray signal ratios will be significantly similar.

The second graph presents data from the dye swap experiment in which two arrays were hybridized with cDNA derived from the same pair of RNA samples, but labeled in opposite ways. Here, L5G3 again refers to the data set obtained with Cy5-cDNA based on a lactate culture and Cy3-cDNA based on a glucose culture. The axis labeled G5L3 refers to the data set obtained with Cy5-cDNA based on the same glucose culture and Cy3-cDNA based on the same lactate culture. With these complementary schemes, the numerators and denominators of the resulting signal ratios from the two arrays will be inverted, and when the data are transformed into a $\log_2$-scale, one data set should be equal in magnitude but negative in sign relative to the other. In fact, the points within the

plot do offer a reasonable approximation of the ideal case of the line y = -x., although the fit is not as fine as is the case with the first graph.

At least part of the explanation for this lies in the different behaviors of the two RNA populations in the reverse transcription reaction. Although the quality of the RNA samples, as verified by gel electrophoresis and spectroscopy as described in section 3.3.4, was good for both the cultures grown on a glucose carbon source and those grown on a lactate carbon source, the samples gave qualitatively different results when being used as templates for the formation of labeled cDNA. Using the reverse transcription and cDNA isolation protocol of section 3.4.3, the cDNA that has been created is purified away from other components of the reverse transcription reaction, such as unincorporated label, through the use of microcentrifugation columns that discriminate based upon molecule size. After the smaller materials are allowed to pass through the column in the binding and washing steps, the larger cDNA molecules remain attached to the column matrix, and because of their incorporated label they give a blue or red cast to the appearance of the column surface, with the color depending on which CyDye has been used to label that particular cDNA population. When RNA isolated from glucose cultures served as the template in cDNA production, the intensity of color observed on the column prior to product elution was less than that seen when lactate culture RNA was used. This lesser coloration could be due to either a smaller amount of cDNA formed in the reverse transcription, or less incorporation of the fluorescently labeled nucleotide analog into the cDNA that was formed. When the Cy5-cDNA was the cDNA solution derived from the glucose culture, the arrays to which the cDNA was hybridized exhibited signal-to-background and signal-to-noise ratios that, although certainly acceptable, were lower than was seen when the glucose culture was the source of the template for the Cy3-cDNA. Another measure of this somewhat lower quality of data can be seen in the generally larger x-axis standard deviation error bars associated with the points of the second graph in Figure 4-1. Again, although the two axes present information about the same two cultures, the data has less variability when the glucose culture RNA is used in generating the cDNA labeled with Cy3-dUTP. This finding proved to be true in general, and in all future experiments, every pair of RNA samples that were being studied with the microarrays was subjected to a similar dye swap. In this way, it could be ensured that if

one of the two RNA samples of the pair proved to be more problematic in the reverse transcription reaction, there would be one array in which the apparently more robust Cy3-cDNA would be based on this RNA, and the results would demonstrate less variability.

4.3 Biological Consistency Study

The experiments described in the previous section were carried out to prove that the data obtained through the use of the new *C. glutamicum* DNA microarrays were reliably reproducible and statistically significant. At least equally important to the validation of the arrays was a confirmation of the biological significance of their findings. To do this, the results from the set of arrays used in the tests just described were examined for consistency with known and expected biological behavior of cells grown on glucose and on alternative carbon sources.

Of all of the pathways represented on the partial genome microarrays the one most directly affected by the change from glucose to a lactate carbon source is the Embden-Meyerhof-Parnas pathway. As is described in section 2.1.2.1, the EMP pathway is used by the cells primarily for the glycolytic conversion of glucose to a pair of molecules of pyruvate, while generating a small amount of energy in the form of two molecules of ATP, and reducing equivalents in the form of NADPH. A modified form of the pathway, bypassing the most highly energetically irreversible steps of glycolysis, can also operate in a reverse. gluconeogenic direction, producing glucose from pyruvate and lactate precursors. Gluconeogenesis is an energetically unfavorable pathway that is of little importance for cells that have glucose readily available to them in the form of a medium component. Cells that do not have access to glucose, though, require the pathway because of its role in producing glycolytic intermediates that are essential building blocks to some cellular materials. Nonetheless, the importance of gluconeogenesis to cells grown on substrates other than glucose is smaller than the importance of glycolysis to cells grown on glucose because the latter uses its pathway not only to generate these EMP pathway intermediates but also to produce energy and, most importantly, to generate pyruvate. In

contrast, cells grown on lactate substrate are capable of generating pyruvate directly from lactate through the activity of lactate dehydrogenase. As a result, it was believed that in comparing the transcript concentrations of genes associated with the EMP pathway as measured with the DNA microarrays, higher levels would be seen in the culture that was grown on glucose than in the culture grown on lactate.

Table 4-2 presents data for these EMP pathway genes by listing the signal ratios derived from relevant array spots containing both partial and full open reading frame sequences. Those spots for which the signal ratio was greater than two standard deviations removed from zero have been highlighted, as these were considered statistically significant results indicating genes whose mRNA levels were higher in one of the two samples. Several interesting findings can be deduced from the data of the table. Foremost among these is a general pattern of induction of EMP pathway gene transcription in the case of growth on glucose, and/or repression of transcription of the genes in the case of growth on lactate. For example, note that in the L5G3-2 microarray all spots associated with these genes but one showed significantly more fluorescence in the Cy3 wavelength, with the Cy3-labeled cDNA being derived from RNA isolated from samples of the glucose culture. Conversely, in the G5L3 microarray, all signal ratios of significant magnitude were for sequences yielding more fluorescence in the Cy5 wavelength, which in this array was again the wavelength associated with the glucose culture. In all, for the 60 sets of spots derived from EMP pathway genes that were present on the four microarrays, 45 of the spot sets, or ¾ of the total, measured significantly higher transcription for growth on glucose. Among the remaining ¼ of the sequences, 14 did not measure a significant difference between transcriptional levels for growth on the two carbon sources, and only one measured significantly higher transcription for growth on lactate. By thus accurately depicting the relationship between the gene expression profiles of the two cultures, these results were considered a clear validation of the consistency of the transcriptional data quantified by the C. glutamicum DNA microarrays with respect to actual biological behavior.

A comparison of the quality of the data for each of the four arrays shows that there is much more similarity between the array pairs that used identical labeling schemes than there is between array pairs that are dye swaps of one another. For the two arrays that

Table 4-2: Relative transcriptional levels of EMP pathway genes for E12 grown on glucose and lactate. Four cultures of E12 were grown on minimal media with two receiving glucose and two receiving lactate as the carbon source. EMP pathway signal ratios from microarrays investigating the four cultures are given in the table on a $\log_2$-scale. Negative signal ratios indicate more fluorescence in the Cy5 wavelength, while positive signal ratios indicate more fluorescence in the Cy3 wavelength. Gene sequences are 500-bp unless otherwise noted as being the full gene sequence. Values given in black type on light grey background fluoresced significantly more intensely in the Cy3 wavelength. Values given in white type on dark grey background fluoresced significantly more intensely in the Cy5 wavelength. Values given in black type over white background did not fluoresce significantly more intensely in either wavelength. Significant signal ratio values were defined as being greater than two standard deviations removed from zero. L5G3 signifies an array hybridized with cDNA derived from a lactate culture labeled with Cy5-dUTP and from a glucose culture labeled with Cy3-dUTP. L5G3-2 signifies an array hybridized with cDNA labeled similarly to that of L5G3 but derived from different cultures. G5L3 signifies an array with the opposite labeling scheme. G5L3-2 signifies an array hybridized with cDNA labeled similarly to that of G5L3 but derived from different cultures.

Gene	L5G3	L5G3-2	G5L3	G5L3-2
eno	-1.5827	-1.60764	0.341567	0.517188
fba	-1.05754	-0.96981	0.091934	0.116375
fba (full)	-1.00681	-0.9083	0.249736	0.012971
gap	-2.4424	-2.40659	1.546058	1.359307
gap (full)	-2.45508	-2.46381	1.539526	1.3435
gpi	-0.52108	-0.59121	0.025475	-0.04531
gpm	-1.06974	-1.06014	-0.00774	0.161965
pfk	-0.62438	-0.64254	1.255223	0.636102
pfk (full)	-0.91885	-0.79383	1.495258	1.261263
pgk	-0.65385	-0.34644	0.887281	0.832779
pgk (full)	-1.23935	-0.91337	0.895475	1.226801
pyk	0.8405	1.127788	0.312621	1.064184
pyk (full)	-1.93227	-1.94356	1.182092	1.11088
tpi	-1.46836	-1.07221	1.084423	1.642629
tpi (full)	-1.44374	-1.11193	1.086573	1.356285

were hybridized with Cy5-cDNA derived from glucose culture RNA samples and Cy3-cDNA derived from lactate culture RNA samples, 20 of the sequences, or approximately 65% of the total number, gave signal ratios that were significantly removed from zero on a $\log_2$-scale. For the two arrays that were hybridized with Cy3-cDNA derived from the

glucose culture and Cy5-cDNA derived from the lactate culture, 26 of the sequences, or approximately 85% of the total number, gave signal ratios indicating significant gene expression differences. Also, the signal ratio with the largest absolute magnitude for the EMP pathway genes on the G5L3 and G5L3-2 arrays was 1.6, which corresponds to a 3-fold difference between the mRNA abundance of that particular gene in the two cultures studied. For the L5G3 and L5G3-2 arrays, the signal ratio with the largest absolute magnitude was -2.5, corresponding to a 5.7-fold transcriptional difference between the two cultures. This higher maximum signal ratio magnitude, as well as the larger number of printed sequences for which the signal ratios were significantly removed from zero, are features resulting from the greater dynamic range realized by hybridizing the L5G3 and L5G3-2 arrays with Cy3-cDNA rather than Cy5-cDNA that was based on RNA sampled from the glucose culture. As was discussed in the previous section, the RNA isolated from the glucose culture did not produce cDNA solutions containing as much label in the reverse transcription reactions as RNA isolated from the lactate culture. Any inadequacies brought about by this were exacerbated when Cy5-dUTP rather than Cy3-dUTP was used as the fluorescently labeled nucleotide analog in creating cDNA from the glucose culture RNA template. The effects of this can clearly be seen in the higher quality results of the L5G3 and L5G3-2 DNA microarrays.

One other trend seen in the data from the four microarrays of this experiment is the close agreement between the signal ratios calculated from spots printed with partial gene sequences and from those printed with full gene sequences. Six of the nine EMP pathway genes represented on the arrays were present in both sizes of sequence, and as can be seen from the values of Table 4-2, with five of these six pairs, the results generated from the spots were extremely similar. The lone exception is that of pyruvate kinase. For this gene the full sequence spots produced significantly negative signal ratio values in the L5G3 and L5G3-2 arrays and significantly positive values in the G5L3 and G5L3-2 arrays, indicating gene over-expression on glucose in all cases. In contrast, the partial 500-bp sequence gave results that neither agreed well with those of the full sequence, nor possessed much internal agreement, with spots on two of the four arrays showing no significant relative over-expression, spots on a third array showing over-expression for growth on lactate, and spots on the fourth array showing over-expression for growth on

glucose. Because of the eight arrayed sequences representing pyruvate kinase, five show up-regulation for growth on the glucose carbon source and two show no significant carbon source effect, it can be assumed that there is some error associated with those spots containing the partial pyruvate kinase sequence that does not in general impact the relationship between results derived from the printing of either partial or full gene sequences.

4.4 Effects of Spotted DNA Sequence Size

As was detailed in section 3.1.2, several genes, in addition to the six EMP pathway genes described above, have been represented on the C. glutamicum partial genome DNA microarrays with two sets of six spots each. One set of these spots is composed of a 500-bp sequence that has been chosen from within the open reading frame of the gene, and the second set of spots is composed of the entire open reading frame sequence. The validation experiments that were carried out with the microarrays and described in this chapter provided the first opportunity to compare the data obtained from these different sequence sizes. From the previous section, the data from Table 4-2 shows very good agreement between the two sets of array spots for five of the six genes of the EMP pathway. The data for these genes and all others on the array that were printed onto the array in multiple sequence sizes are given in Figure 4-2. Each point on the graph gives data for a single gene, with the x-axis value given by the average signal ratio as derived from the partial gene sequence spots, and the y-axis value given by the average signal ratio as derived from the full gene sequence spots.

It can be seen from the very close proximity of almost all points on the graph to the ideal case of a line with equation $y = x$, that the results given by the two pairs of spots are virtually identical. Because a selection of genes were chosen that have a wide range of sequence lengths, this shows that there is no effect of printed sequence length in the range from 500 base pairs to 3700 base pairs on the proportional hybridization of two populations of labeled cDNA. As with the results of Table 4-2, again the only gene for

Figure 4-2: Effects of spotted sequence size on DNA microarray results. Similarity of DNA microarray results obtained from spots containing DNA sequences of different sizes representing the same gene. Axes depict signal ratios from lactate-grown and glucose-grown E12 on a $\log_2$-scale as measured with the L5G3-2 DNA microarray described in the text.

which a poor fit between the results from the two printed sequence sizes exists, is pyruvate kinase. Because the size of the pyruvate kinase gene is 1300 base pairs, it is not at an extreme within the range of sizes covered by this experiment, making it very unlikely that a size effect is the cause for the data discrepancy. For all other genes, the validation that all spots will give statistically identical results regardless of sequence size effectively doubles the number of spots available to gather data from. This extra level of redundancy adds to the usefulness of the array by decreasing the chance that many or all of the spots associated with a particular gene will be unreadable in scanning the array or will be removed by the data filtration procedures of section 3.5.2.

4.5 Effects of Spotted DNA Origin Species

The other comparison that was now capable of being performed with the data collected through the validation experiments was that of results obtained from spots which were printed with DNA sequences originating from the two different species, *C. glutamicum* and *M. tuberculosis*. The inclusion of *M. tuberculosis* DNA sequences on the *C. glutamicum* partial genome microarray served to attempt the complete transcriptional description of a metabolic network for which initially not all associated *C. glutamicum* genes had been characterized. The design of an interspecies array also allowed for a study of the effects that differences in sequence homology would have on differences in fluorescent signal intensities and ratios. The use of *M. tuberculosis* DNA in the microarray is described in more detail in section 3.1.1.

Figure 4-3 presents the data for all genes that were printed onto the array in the form of DNA sequences obtained from the genomes of both species. As in Figure 4-2, each point on the graph gives data for a single gene, here with the x-axis value given by the average signal ratio as derived from the *C. glutamicum* sequence spots, and the y-axis value given by the average signal ratio as derived from the *M. tuberculosis* sequence spots. Also as in Figure 4-2, the ideal case is a line with equation y = x, but here the deviation between the actual measured values and this ideal is more pronounced than it was for the comparison of sequence size effects. The trend of the data is certainly suggestive of a positive correlation between the sets of signal ratios derived from the two groups of array spots, with the majority of the points located in either the upper right or lower left quadrants where the x-values and y-values share the same sign. The intercept of the line is shifted away from the origin, though, and if a curve were fit to the data, its slope would approach zero as the signal ratios derived from the *C. glutamicum* spots continued to increase. This plateau in the signal ratios obtained from the *M. tuberculosis* spots indicates that there is a limit to the degree of differentiation that exists between the fluorescence of the two cDNA populations hybridized to the spots. Also, the graph shows that the standard deviations of the signal ratios are higher on average for the *M. tuberculosis* spots than for the *C. glutamicum* spots. Together, these aspects of the data suggest that nonspecific fluorescence is a significant factor for the spots printed with *M. tuberculosis* DNA.
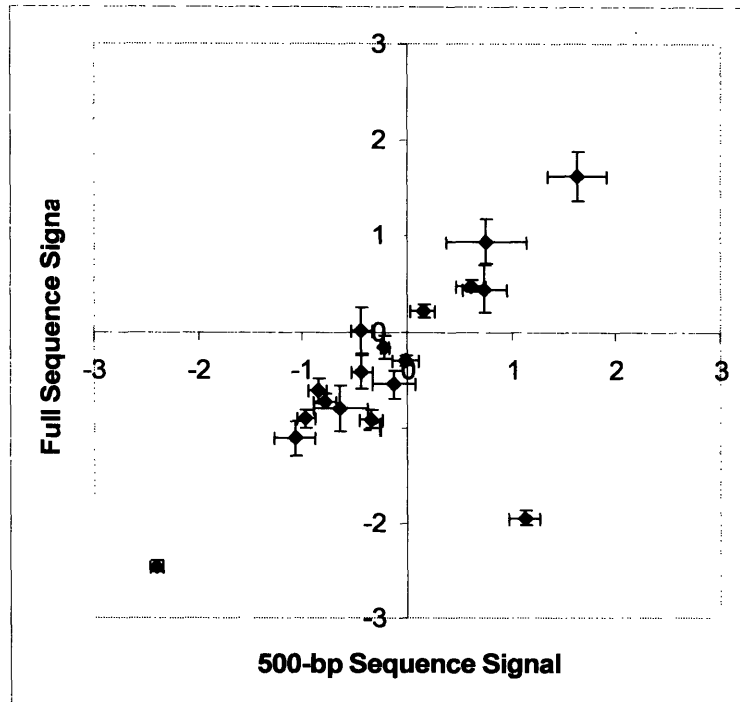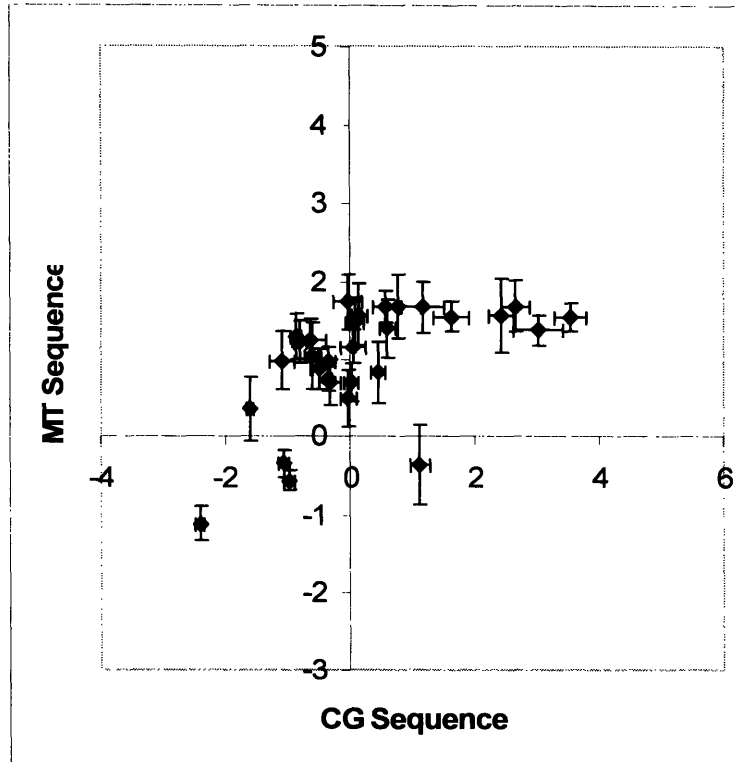
Figure 4-3: Effects of spotted sequence origin species on DNA microarray results. Similarity of DNA microarray results obtained from spots containing DNA sequences amplified from the species *C. glutamicum* and *M. tuberculosis* and representing the same gene. Axes depict signal ratios from lactate-grown and glucose-grown E12 on a $\log_2$-scale as measured by the L5G3-2 DNA microarray described in the text.

In quantifying the fluorescence at both measured wavelengths for the spots of the array, it was seen that those spots whose sequences were based on *M. tuberculosis* DNA had on average lower intensities than was seen for spots printed with native DNA. The lower signal-to-background and signal-to-noise ratios for these spots rendered the fluorescence caused by the hybridization of the desired labeled cDNA more susceptible to being partially confounded with undesired background fluorescence. This would in turn result in a decrease in the dynamic range of the data set, as was described in section 3.4.2, and an increase in the variability and standard deviations of the data. The lower hybridization specificity is systemic to all spots with *M. tuberculosis* DNA, and there is no discernible correlation for the degree of homology that exists between a pair of sequences from the

two species and the similarity between the signal ratios of spots printed using those two sequences.

The conclusion that must be drawn from this test is that in the case of the two species whose genomes were applied in creating this array, the use of DNA sequences from one species to probe the transcriptional behavior of the second species through a microarray provides fair, but not very accurate results. In the absence of other means to test the concentrations of mRNA present in a cell sample, this technique could yield important information about, for example, the subpopulation of the transcriptome that is up-regulated or down-regulated in certain strains or under certain conditions. The technique would be much less useful in determining the degree to which this up- or down-regulation occurs, especially in relation to that of the other genes being studied.

4.6 Other Transcriptional Effects of Carbon Source

4.6.1 Overview

In measuring the changes to the transcriptome of the *C. glutamicum* E12 strain in response to the availability of glucose or lactate as the sole carbon source, the pathway that was most useful in determining the biological consistency of the array data was that of the Embden-Meyerhof-Parnas pathway, as was discussed in detail in section 4.3. Once the partial genome DNA microarrays had been validated by this consistency study and other tests described earlier in this chapter, the results obtained from the arrays for the other pathways that comprise the metabolic network of central carbon metabolism and lysine production could be examined with greater confidence. Accordingly, the signal ratios from all spots within the array were measured to determine which genes showed significantly different expression levels between growth on glucose and growth on lactate for *C. glutamicum* E12. Table 4-3 gives the results in the form of fold-changes of

Table 4-3: Significant changes in gene transcript concentrations for E12 grown on glucose and lactate. Four cultures of E12 were grown on minimal media with two receiving glucose and two receiving lactate as the carbon source. Positive changes indicate higher mRNA levels for growth on lactate and negative changes indicate higher mRNA levels for growth on glucose.

Pathway	Gene	Enzyme	Fold Change
glycolysis			
	gpi	phosphoglucoisomerase	-1.2
	fba	fructose bisphosphate aldolase	-1.5
	gpm	phosphoglyceromutase	-1.5
	pgk	phosphoglycerate kinase	-1.6
	pfk	phosphofructokinase	-1.7
	eno	enolase	-2.0
	tpi	triosephosphate isomerase	-2.5
	pyk	pyruvate kinase	-2.9
	gap	glyceraldehydes 3-phosphate dehydrogenase	-3.8
biotin synthesis			
	bioB	biotin synthase	14.0
	bioA	7,8-diaminopelargonic acid aminotransferase	3.7
	bioD	dethiobiotin synthase	1.5
glyoxylate bypass			
	aceB	malate synthase	4.9
	aceA	isocitrate lyase	4.9
TCA cycle			
	aco	aconitase	2.1
	citA	citrate synthase 1	1.6
	citA2	citrate synthase 2	1.7
	sdhC	succinate dehydrogenase subunit C	1.5
	sdhB	succinate dehydrogenase subunit B	1.4
	sdhA	succinate dehydrogenase subunit A	1.4
lysine biosynthesis			
	dapD	tetradihydrodipicolinate succinylase	1.6
	ddh	meso-diaminopimelate dehydrogenase	1.6
	lysA	diaminopimelate decarboxylase	1.4
	dapB	dihydropicolinate reductase	-1.4
	dapF	diaminopimelate epimerase	-1.4
oxaloacetate node			
	malE	malic enzyme	1.9
	pck	phosphoenolpyruvate carboxykinase	1.4
pentose phosphate pathway			
	zwf	glucose-6-phosphate dehydrogenase	-1.4

transcript level for those genes for which the signal ratio was greater than two standard deviations removed from zero. The data has been converted from a set of signal ratios in log_2-space to measures of the fold changes that exist between the concentration of a

particular mRNA in the cultures grown on glucose and the concentration of the same mRNA in the cultures grown on lactate. For example, from the table it can be seen that for the EMP pathway genes, the expression level changes ranged from a small 20% decrease for glucose-6-phosphate isomerase to an almost 4-fold reduction for glyceraldehyde-3-phosphate dehydrogenase. Results for the other pathways are discussed in more depth in the following sections.

4.6.2 Pentose Phosphate Pathway

As the entry point into the pentose phosphate pathway is phosphorylated glucose, the accessibility of the pathway for cells grown on carbon sources other than glucose is severely limited. Because of this the expectation for the transcriptional profile of the pathway was that of a down-regulation of gene expression when cultures were fed lactate instead of glucose. In fact, glucose-6-phosphate dehydrogenase, catalyzing the initial reaction of the pentose phosphate pathway, did show less transcription for growth on lactate, although this was the only gene in the pathway that had a significant expression level change. It may be that what control of metabolic flux through the pathway exists at a transcriptional level is concentrated at this entry point. With less catalyst present for this first of a series of reactions, the degree to which the transcription of the genes catalyzing the subsequent reactions can influence the flux through the pathway as a whole could be greatly reduced. A similar control strategy for other pathways of the metabolic network of *C. glutamicum* was suggested by the microarray experiments of Krömer *et al.* (2004) that found that with the strains and conditions tested, the expression of genes associated with the initial steps of the glycolytic, lysine biosynthesis, and TCA cycle pathways were among the most affected by changes in cultivation time.

4.6.3 Tricarboxylic Acid Cycle

In contrast to the decreases that occurred in the mRNA amounts of several genes of the Embden-Meyerhof-Parnas and pentose phosphate pathways, there were significant increases detected in the mRNA amounts of three of the enzymes of the tricarboxylic acid cycle—aconitase, citrate synthase, and succinate dehydrogenase—for growth on lactate rather than glucose carbon source. That the cells would up-regulate the transcriptional concentrations of several TCA cycle enzymes for growth on lactate is not surprising given the shifts that occur in the ability of the cells to derive energy and reducing equivalents from elsewhere in the metabolic network. In the earlier sections of this chapter, the lesser significance of glycolysis and the pentose phosphate pathway for *C. glutamicum* cultures grown on lactate has been discussed extensively. Because for cells grown on glucose these two pathways are major sources for both ATP and NADPH, the two important compounds must be generated through other means when lactate is the available carbon source. The TCA cycle provides an alternative method for both of these molecules to be produced from lactate, and the higher mRNA titers seen are likely to coincide with an increased flux of carbon through the pathway when lactate is fed to the cell cultures.

As was discussed in the previous section, often the most significant point of control that a cell has over the flow of material through a reaction pathway lies at the entryway into that pathway. The results shown in Table 4-3 for the TCA cycle genes may provide another example of this phenomenon, as the greatest up-regulation of transcript titers was seen for citrate synthase and aconitase, the first two enzymes of the pathway. Together these catalysts support the conversion of acetyl-CoA, produced by the pyruvate dehydrogenase coupling reaction, along with oxaloacetate to isocitrate. Since lactate is converted to acetyl-CoA in a relatively simple two-step process involving lactate dehydrogenase and pyruvate dehydrogenase, there may be more intracellular acetyl-CoA present for growth on lactate than on glucose, for which there can be loss of carbon to carbon dioxide and biosynthesis intermediates produced in the pentose phosphate pathway in generating acetyl-CoA. One method that the cells use to up-regulate flow through this pathway involves acetyl-CoA as an inducer [Eikmanns *et al.*, 1994]. The

higher concentration of acetyl-coA that may be present for growth on lactate relative to growth on glucose may then have a positive influence on the expression level of citrate synthase. Another interesting finding related to the control of transcription by the cells is that the sequences of the two isoenzymes of citrate synthase, citA1 and citA2, generated nearly identical signal ratios to one another on the microarray, as did the three subunits of succinate dehydrogenase—sdhA, sdhB, and sdhC. These very similar transcriptional measurements, along with the related functions that the isoenzymes or subunits serve within the cellular metabolism, suggest that the within each of the two gene groupings a regulatory strategy is shared, so that the cells can influence the expression of the related genes simultaneously.

Finally, the three enzymes that were found in this set of experiments to have higher mRNA levels for growth on lactate can together catalyze a sequential series of reactions capable of converting acetyl-CoA to citrate, isocitrate, succinate, and then fumarate. To complete this set of four reactions and bypass those catalyzed by isocitrate dehydrogenase, α-ketoglutarate dehydrogenase, and succinyl-CoA synthase, one other enzyme—isocitrate lyase—must also be active for growth on lactate. This enzyme is one of two involved in the glyoxylate bypass pathway, and as a result the pattern of the three over-expressed genes in the TCA cycle provides evidence supporting an enhanced flux through glyoxylate when lactate is the available carbon source.

4.6.4 Glyoxylate Bypass

The implication presented in the previous section that the importance of the glyoxylate bypass pathway increases in *C. glutamicum* cells if the carbon source available to them in the growth medium is lactate rather than glucose, is further supported by both the transcriptional measurements of the DNA microarrays and knowledge about the metabolic need of the cells. If in fact there is a higher flux through the tricarboxylic acid cycle for the case of growth on lactate, as was suggested by the data discussed in the previous section, then the utility of the glyoxylate bypass pathway to the cells increases dramatically. This is because the complete operation of the TCA cycle converts all

carbon entering the pathway in the form of acetyl-CoA into carbon dioxide. Anaplerotic pathways, such as the glyoxylate bypass, will then be necessary to replenish the intracellular pools of key intermediates such as malate and oxaloacetate, which serve as starting materials for other important pathways including that of lysine biosynthesis. Agreeing with this expectation, the mRNA of both genes of the glyoxylate bypass pathway—isocitrate lyase and malate synthase—had almost 5-fold higher concentrations in cultures grown on lactate.

The identical degree to which the expression levels of the two genes were up-regulated is a sign of co-regulation of the pair of enzymes. This agrees with observations by Wendisch *et al.* (1997) that the specific activities of the lyase and synthase were influenced by the growth media through shared transcriptional regulation, and more specifically, the control of transcriptional initiation. Another means for controlling the activity of a glyoxylate bypass enzyme is through the inhibition of isocitrate lyase by the glycolytic intermediates fructose-1,6,-bisphosphate, 3-phosphoglycerate, and phosphoenolpyruvate [Reinscheid *et al.*, 1994a]. As was discussed in section 4.3, Embden-Meyerhof-Parnas pathway activity will be reduced for growth in media not containing glucose. This will in turn cause the concentrations of these metabolites to decrease, and remove a source of down-regulation of the glyoxylate bypass pathway.

4.6.5 Malic Enzyme and Phosphoenolpyruvate Carboxykinase

The DNA microarray results also indicated that two other enzymes responsible for anaplerotic reactions in the oxaloacetate/malate node of the metabolic pathway—malic enzyme and phosphoenolpyruvate carboxykinase—had higher transcript levels for the case of growth on lactate. These enzymes, described in sections 2.1.2.4.2 and 2.1.2.4.3, convert the TCA cycle intermediates malate and oxaloacetate, respectively, to the EMP pathway intermediates pyruvate and phosphoenolpyruvate, respectively. Flux of metabolites from the TCA cycle to the EMP pathway is most obviously required to

provide substrates for the reactions of gluconeogenesis, which for the reasons mentioned in section 4.3 will be of limited value for cell cultures that have glucose available as a component of the growth medium. This agrees with the results of a study that quantified *in vivo* anaplerotic fluxes at work in *C. glutamicum* grown on glucose, and found that the amount of pyruvate formed by malate and oxaloacetate under these conditions is only 10% of that formed by the action of pyruvate kinase on phosphoenolpyruvate [Petersen *et al.*, 2000]. Another investigation focused on the role of malic enzyme within the metabolic network of *C. glutamicum* by exploring the relationship between growth on different carbon sources and the activity of the enzyme, in part through the use of a strain in which the malic enzyme gene was deleted [Gourdon *et al.*, 2000]. It was found that of glucose, lactate, and acetate, lactate was the only carbon source for which diminished growth characteristics were seen in the absence of malic enzyme activity. Along with the 90% increase in malic enzyme transcript concentration for growth on lactate detected in the work presented here, this suggests that a very important role for the enzyme exists when lactate is the available carbon source. The nature of this role was examined in greater detail in later experiments described in the following chapters.

4.6.6 Lysine Synthesis Pathway

Of all pathways of the metabolic network that were represented on the *C. glutamicum* partial genome DNA microarray, the lysine biosynthesis pathway gave results that were the least conclusive from a transcriptional point of view. Three enzymes— tetradihydrodipicolinate succinylase, meso-diaminopimelate dehydrogenase, and diaminopimelate decarboxylase—showed more mRNA present in samples of cell cultures that were grown on lactate. However, two others—dihydropicolinate reductase and diaminopimelate epimerase—showed less mRNA under the same conditions. This mixture of up- and down-regulation in the pathway implies that the control of its flux does not rest primarily with the transcription of the related enzyme genes. Several other control strategies have been identified as operating within the lysine synthesis pathway, such as the feedback inhibition of aspartokinase by the lysine and threonine end products

of amino acid biosynthesis. These types of controls may instead prove to be the dominant ones at work in allowing the cells to manipulate lysine synthesis flux.

4.6.7 Biotin Synthesis Pathway

Somewhat surprisingly, the largest increase in mRNA concentration seen as a result of switching *C. glutamicum* E12 growth from a glucose medium to a lactate medium was a 14-fold transcriptional change for the gene biotin synthase. Additionally, the other two genes of the biotin biosynthesis pathway printed onto the DNA microarrays—7,8-diaminopelargonic acid aminotransferase and dethiobiotin synthase—also exhibited up-regulation with a lactate carbon source. In one sense the over-expression of these biotin biosynthesis genes fits well with other observations made of the behavior of the *C. glutamicum* during growth on lactate. In particular, the transcriptional increases may be related to the higher activity of pyruvate carboxylase noted in section 2.1.4. This is because for the enzyme to be active, pyruvate carboxylase requires biotin as a cofactor. However, *C. glutamicum* requires supplementation of its growth medium with biotin because it lacks the enzyme 7,8-diaminopelargonic acid synthetase, preventing the native formation of biotin from components of central carbon metabolism [Hatakeyama *et al.*, 1993]. It therefore seems likely that there is a some type of vestigial regulatory scheme at work, providing a link between the higher activities of pyruvate carboxylase that have been recorded for growth on lactate, and the transcriptional up-regulation of those genes that in an ancestral strain might have allowed for increased biotin generation.

102

Chapter 5: Effects of Gene Over-Expression and Mixed Carbon Sources on Transcriptional Profiles

5.1 Experimental Design

The results of the experiments described in the previous chapter show very clearly that numerous significant changes can occur in the aggregate transcriptional behavior of a cell culture in response to an external change in the type of carbon source used in the culture growth medium. Also, from studies described in section 2.1.4, it is known that the activity of the key anaplerotic enzyme pyruvate carboxylase is influenced by the concentration of lactate present in the medium. In this chapter, an experiment is described in which the extracellular macroscopic culture properties and the intracellular gene expression profiles of cultures grown on varying mixtures of glucose and lactate were determined. In addition, the effects of gene over-expression were also considered by using two strains—21253 and 21253(pKD7), described in section 2.1.2—that differed in their expression of the pyruvate carboxylase gene. The goals of this study were to determine the trends of transcriptional up-regulation and down-regulation employed by the cells as a glucose carbon source is incrementally replaced with lactate, to explore the influence that the forced over-expression of one gene important to the operation of the metabolic network would have on these trends, and to draw conclusions regarding the metabolic implications of the transcriptional responses of the cells as well as the control strategies used by the cells in initiating these responses.

To gauge the effects that the carbon source has in different ratios of glucose and lactate, a series of twelve cultures were grown in minimal media; six for each of the two strains

tested. The culture flasks had 0%, 20%, 40%, 60%, 80%, and 100% of their 20 g/l carbon source in the form of lactate, with the remainder consisting of glucose. In this way, the total concentration of carbon fed to the bacteria was kept constant at 680 mM in each case. As is described in section 4-1, all cultures were grown on a slightly modified version of the minimal medium developed by Park. Cultures of 21253(pKD7) were further supplemented with 50 mg/l kanamycin to select for cells with at least one functional copy of the plasmid conferring pyruvate carboxylase over-expression. The cultures were grown in 250-ml round-bottom flasks containing 50 ml medium each that were incubated at 30°C and 250 rpm on a rotary shaker.

Samples with a volume of 1 ml were taken periodically for assaying the concentrations of lysine, biomass, glucose, and lactate. The lysine measurements were made by centrifuging the samples for 1 min at 13,000 rpm and assaying the supernatant with the use of a reverse phase HPLC column (Hypersil AA-ODS; Agilent, Palo Alto, CA) capable of resolving all amino acids excreted by the cells into the culture broth. In this assay the amino acids were derivatized with ortho-phthaldialdehyde (OPA), passed through the column along with a mobile phase of acetonitrile and 0.1 M sodium acetate pH 5 in a gradient mixture at a rate of 0.45 ml/min, and detected by ultraviolet absorbance at a wavelength of 338 nm. The biomass measurements were made by measuring the absorbance of the culture samples at a wavelength of 660 nm in a spectrophotometer (Ultrospec 2100 pro; Amersham Biosciences, Piscataway, NJ), diluting the samples sufficiently to maintain the optical density measurements below 0.9 absorbance units. This ensured that the optical density was linearly related to the biomass concentration according to the formula given in section 3.3.2. The glucose and lactate measurements were made by assaying the sample supernatants with the enzymatic oxidation probes of a YSI 2300 STAT Plus analyzer (YSI Inc., Yellow Springs, OH) Because the lactate measurements derived from this assay only reported the concentrations of L-lactate present in the medium, the values were doubled to estimate the total concentration of lactate. The initial charge of lactate to the medium is in the form of a DL-lactic acid syrup, and it was assumed that the culture consumes both enantiomers of the carbon source substrate at an equal rate.

Samples with a volume of 15 ml were taken for use in RNA isolation and microarray analyses of the transcriptional profiles of the twelve cultures. The time point for these samples was chosen to be at mid-exponential growth phase—approximately 22 hr post-inoculation—so that the transcriptome would reflect the demands of the cells related to both culture growth, which would be nearing its end, and lysine production, which would have recently begun. The fluorescently labeled cDNA created from each RNA sample was hybridized along with labeled cDNA created from a stock of control RNA to two microarrays. This control RNA was obtained by pooling together all twelve cultures of the experiment after growth and lysine production had completed, and isolating enough RNA from this pool to be used in all required microarrays.

5.2 Biomass and Lysine Yields

5.2.1 Yield of Biomass on Carbon

Previous work done with the recombinant strains 21253 and 21253(pKD7) had demonstrated that when the two strains were grown on glucose as the carbon source, there was no significant difference between their growth rates [Koffas *et al.*, 2002]. The study found that when the strains were grown on lactate as the carbon source, though, the transformant 21253(pKD7) grew at a faster rate than the parental strain 21253. The growth characteristics of the twelve cultures studied in this experiment are presented in Figure 5-1, confirming these earlier findings as well as showing for the first time the behavior of the strains when grown on mixed carbon sources. Agreeing with expectations, in the case of 100% lactate, 21253(pKD7) showed a 30% higher yield of biomass on carbon than that of the control strain. Additionally, it was found that the two strains showed very similar growth behavior in the case of 100% glucose, with the disparity between the biomass yields increasing along with the amount of lactate fed to the cultures. Both strains showed a decrease in growth as the media contained more

Figure 5-1: Yields of biomass on carbon source. 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Yields were based on final biomass and initial carbon source concentrations.

lactate, but the decline in yield for the pyruvate carboxylase over-expressing strain was approximately 20% while that for the parental strain was 40%. The higher yields of biomass seen for the transformant here and in earlier experiments may support the theory that decreased growth rates of *C. glutamicum* on lactate are at least in part attributable to intracellular pyruvate accumulation [Cocaign-Bousquet *et al.*, 1993]. The over-expression of the pyruvate carboxylase gene in 21253(pKD7) may serve to relieve any growth limitations caused by this excess pyruvate.

In order to support the production of biomass, as well as lysine, the cells must generate sufficient energy in the form of ATP to serve as cofactors for the anabolic reactions. For growth on glucose, a key supply of this energy can be found in the reactions comprising the pathway of glycolysis. As was discussed in section 4.3, many of the enzymes responsible for catalyzing the glycolytic reactions can also operate in the reverse direction, allowing cells grown on alternative carbon sources to generate those intermediates of the Embden-Meyerhof-Parnas pathway that are required as substrates in many important anabolic processes of cell growth and maintenance. For this to occur,
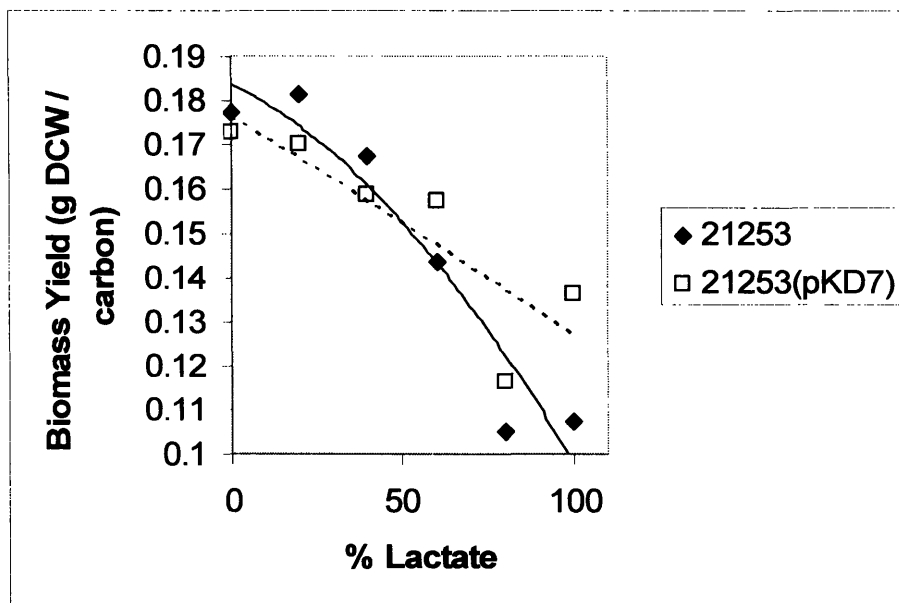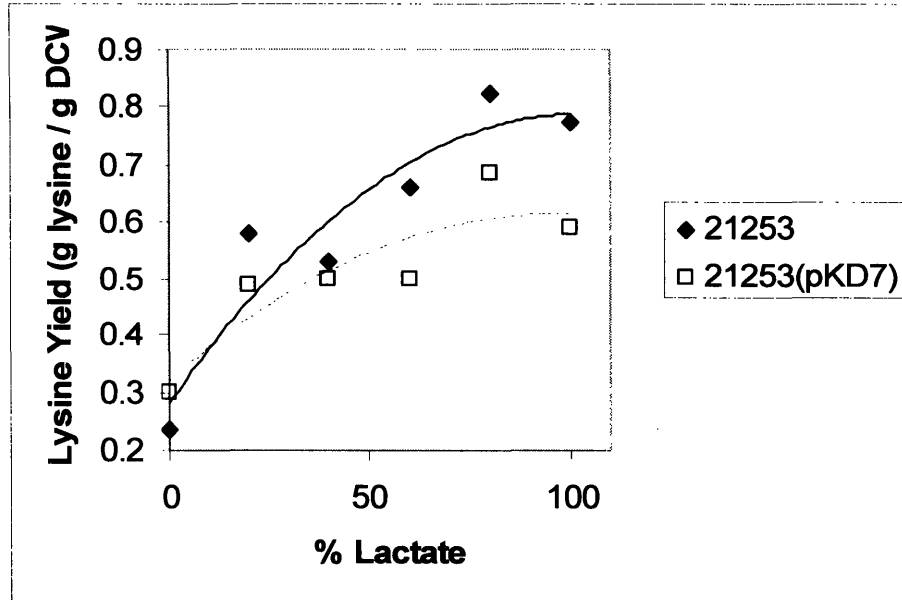
Figure 5-2: Yields of lysine on biomass. 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Yields were based on final lysine and biomass concentrations.

ATP must be consumed, rather than produced, in order to drive the reactions in the energetically unfavorable gluconeogenic direction. This causes energy inefficiencies that may play a role in decreasing the growth yield as more lactate and less glucose is available to the growth medium, and as a result the cells must rely increasingly more on gluconeogenesis and less on glycolysis.

5.2.2 Yield of Lysine on Biomass

The earlier experimentation carried out with 21253 and 21253(pKD7) discovered that the use of glucose in cultivating the strains caused no difference in their lysine production rate, while the use of lactate caused 21253(pKD7) to produce less lysine than was produced by 21253 under identical conditions [Koffas et al., 2003]. The results shown in Figure 5-2, from the twelve cultures grown in this study, again corroborate this previous work while providing new information about the trends of the lysine yields over a range

of carbon source compositions. In trends that mirrored those described above for growth, the lysine yields on biomass of each strain were nearly identical when the medium carbon source was solely glucose. As the percentage of lactate increased, both 21253 and 21253(pKD7) exhibited elevations in specific lysine yields, with once again the more dramatic changes occurring in the control strain. As a result, for growth on lactate there was a significant 30% difference between the yields of lysine on biomass for the two strains. Over the course of changing from a medium containing 100% glucose carbon source to one containing 100% lactate carbon source, the improvement in yield for 21253 was greater than 3-fold, while that of 21253(pKD7) was nearly 2-fold.

Other publications have also described increases in product yields brought about by changes in the carbon source composition of the growth medium. For example, another experiment found a similar increase in lysine production yields when small amounts of acetate are added to medium in which the carbon source is primarily in the form of glucose [Paegle and Ruklisha, 2003]. One recent study reported that the production of the cell wall glycoprotein PS2 also increased if the minimal medium on which *C. glutamicum* was cultivated was supplemented with lactate [Soual-Hoebeke *et al.*, 1999]. The yield increases of these products seen for growth on larger percentages of carbon source substrates other than glucose may provide another explanation for the decreases in biomass yields seen under similar conditions and described in the previous section. The larger specific yields of lysine indicate that the metabolism of the cells is altered in such a way as to divert some material away from the production of cellular building blocks and towards the production of overflow metabolites. This diversion may reduce the ability of the culture to support a high cell density while enhancing lysine specific productivity. It has also been suggested that the decreases in growth yields may bring about an increase in available NADPH that is no longer being utilized as extensively in the generation of biomass [Ruklisha *et al.*, 2001]. This NADPH would then be available to support the increases in product yields that have been seen. As an example, a study of nystatin production in *Streptomyces noursei* found that a decrease in the pentose phosphate pathway flux, and the amount of available NADPH derived from the pathway, signaled a transition to a culture phase with a lower growth rate and a higher productivity [Jonsbu *et al.*, 2001]. For these reasons, some work has been done in the development of strategies

for lowering growth rates as a means for obtaining higher lysine productivities [Kiss and Stephanopoulos, 1991].

5.3 Carbon Consumption Kinetics

Before moving to a description of the intracellular measurements of the transcriptional profiles of the cells in each of the twelve cultures of the experiment that describe their responses to the different mixtures of glucose and lactate carbon source, this section describes the extracellular measurements of how each of those cultures consumes the carbon sources available to them in their growth medium. Of particular interest are the rates of consumption, how they differ from one carbon source to another and from one *C. glutamicum* strain to another, and how these differences relate to the yields of biomass and lysine that were presented in the above two sections.

Figure 5-3 shows the glucose and lactate concentrations present in the culture broths for two of the twelve cases—those of 21253 grown on 0% lactate and 100% lactate—at different time points throughout the growth of the cultures. A number of interesting observations can be made from the four trends of this plot. It can be seen that in the culture that was not fed any lactate, there was lactate detected nonetheless in the culture samples taken between 20 hr and 30 hr post-inoculation. Typical growth and lysine production kinetics for the cultures of this experiment are shown in Figure 5-4. From this graph it can be seen that the time period between 20 hr and 30 hr corresponds to the phase of mid-exponential growth for the cells, when the growth rate of culture is at its maximum. Because lactate generation is an occurrence most commonly associated with growth that operates under oxygen depravation, its appearance in the culture during mid-exponential growth may be an indication that the culture is becoming oxygen limited at these times. The cultures are being grown in batch mode in Erlenmeyer flasks stirred via orbital mixing, a technique known to have poor oxygen mass transfer characteristics. Given this, and that all twelve of the cultures exhibited similar increases in lactate synthesis during these times, it is reasonable to believe that the dissolved oxygen levels

Figure 5-3: Glucose and lactate consumption by 21253 in the case of 100% glucose and 100% lactate carbon source. Strain 21253 was grown on minimal medium containing 20 g/l carbon source that was 0% or 100% in the form of lactate, with the remainder being glucose. The trends in the graph show residual lactate and glucose in the growth medium for these 0% and 100% cases.

of the cultures dropped to low levels during growth and just prior to the beginning of significant lysine production.

Comparing the trends in Figure 5-3 of glucose concentration for the case of the 0% lactate culture and of lactate concentration for the case of the 100% lactate culture shows some key differences between the rates in which these predominant carbon sources are consumed by their respective cultures. The initial lag period, during which carbon consumption rates are very low, is longer for the culture fed entirely lactate than it is for the culture fed entirely glucose. These lag periods take place in the first approximately 20-25 hr of the culture, which, from the trends of Figure 5-4, is the time during which the cells are producing only additional biomass as lysine generation has not yet begun. From section 5.2.1, the biomass yields of cultures grown on 100% lactate are lower than for
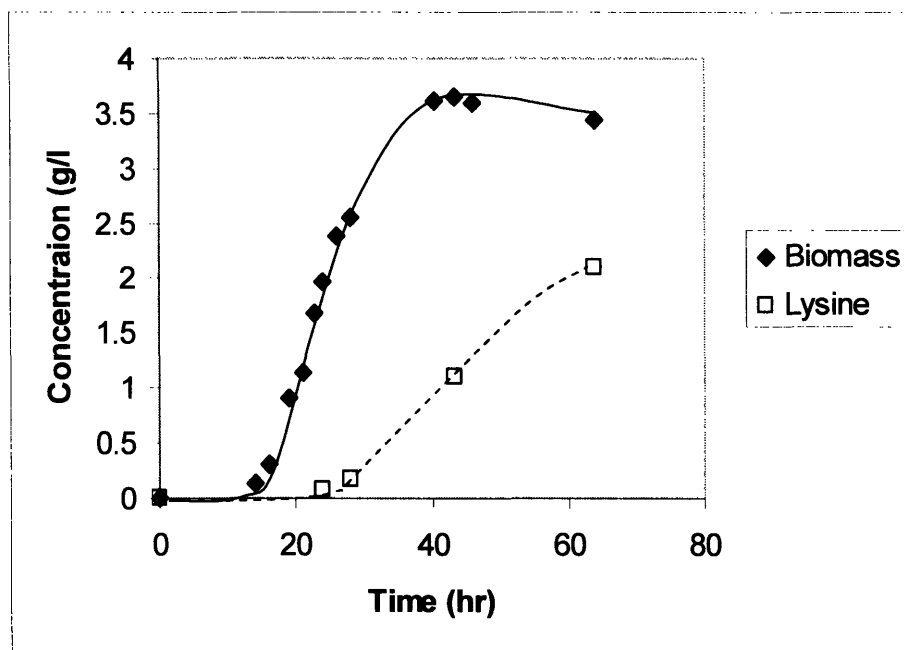
Figure 5-4: Typical growth and lysine production trends. Strain 21253 was grown on minimal medium containing 20 g/l carbon source that was 20% in the form of lactate, with the remainder being glucose.

those grown on 100% glucose. The slower uptake of lactate by the cells utilizing it as their carbon substrate may then be another trait associated with the lower ability of the lactate medium to support culture growth.

After this initial lag period has ended, the consumption rate of lactate in the case of the 100% lactate culture is actually higher than the rate of glucose consumption by the 100% glucose culture. These increased rates occur from approximately 20 hr until approximately 50-60 hr post-inoculation. This then coincides with the time period during which lysine synthesis is taking place, as is shown in Figure 5-4. In contrast to the yield of biomass, the yield of lysine as was described in section 5.2.2 was higher for cultures grown on 100% lactate than for those grown on 100% glucose. This higher lysine yield may very well then be related to the higher carbon consumption rate seen for growth on lactate relative to growth on glucose.

Figure 5-5 presents more data for the glucose and lactate concentrations also present in cultures grown on either 0% lactate or 100% lactate, but here the results are for the pyruvate-carboxylase over-expressing strain 21253(pKD7). A comparison between the
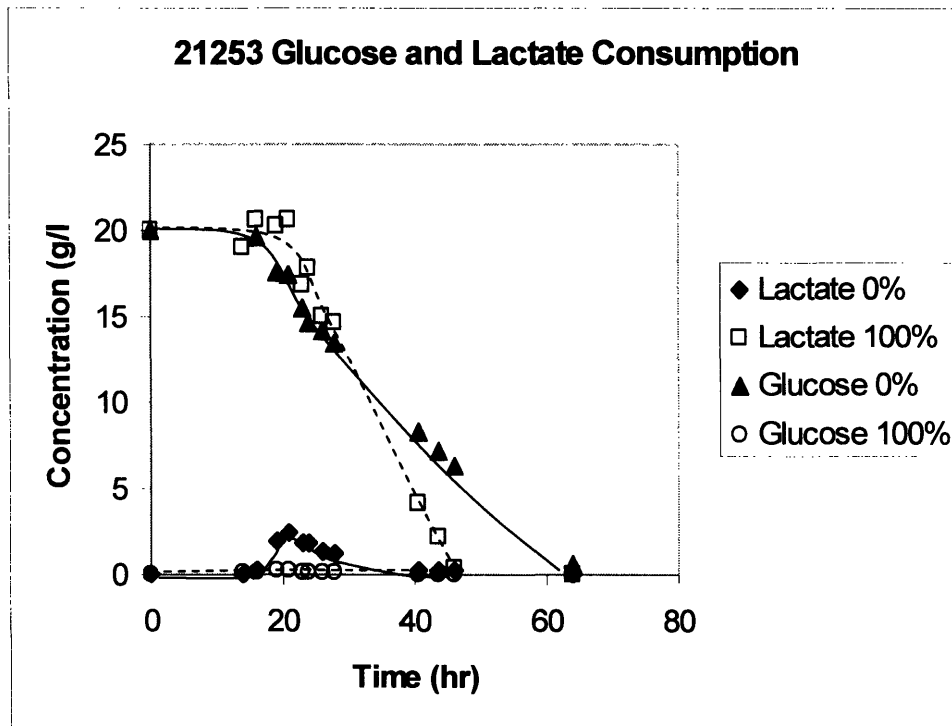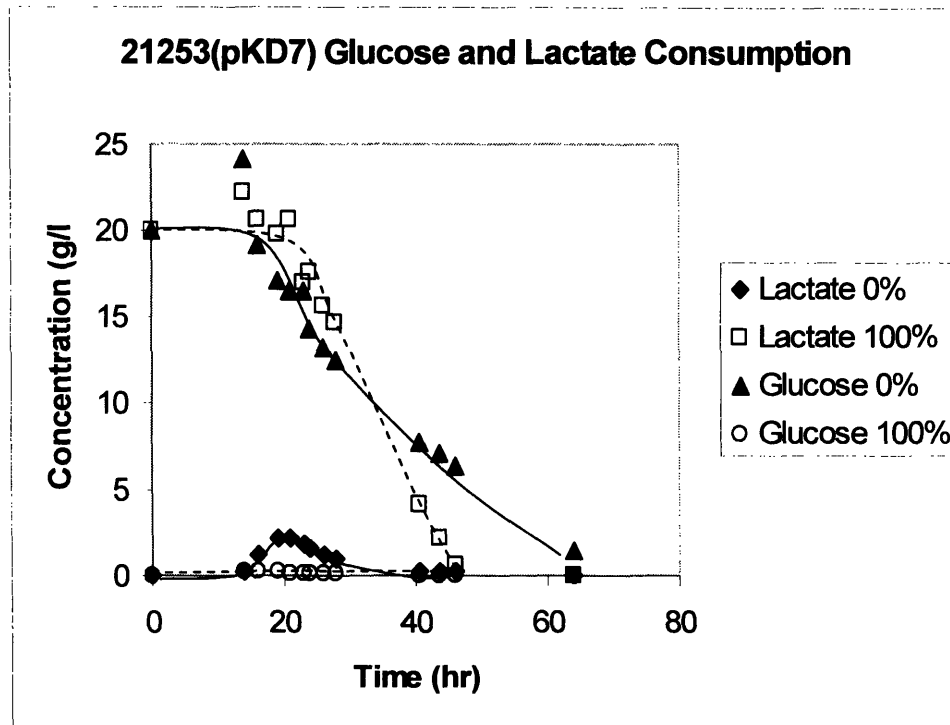
Figure 5-5: Glucose and lactate consumption by 21253(pKD7) in the case of 100% glucose and 100% lactate carbon source. Strain 21253(pKD7) was grown on minimal medium containing 20 g/l carbon source that was 0% or 100% in the form of lactate, with the remainder being glucose. The trends in the graph show residual lactate and glucose in the growth medium for these 0% and 100% cases.

trends of this figure and those for the parental 21253 strain shown in Figure 5-3 shows that the same features discussed in relation to the data for the parental strain are also very noticeable in the behavior of the data for the recombinant strain.

Results for the case of an intermediate mixture of carbon sources are shown in the glucose and lactate trends of Figure 5-6. Here, data is presented for both the 21253 and 21253(pKD7) strains in the case of growth on a medium containing 40% of its 20 g/l carbon source in the form of lactate and 60% in the form of glucose. As with the lactate trends seen in Figures 5-3 and 5-5, there is an increase in lactate concentration seen as the cells produce the carbon source as a side product of metabolism at approximately 20 hr post-inoculation. After this synthesis of lactate, as well as several hours of simultaneous glucose consumption, has taken place, the concentrations of the two carbon

Figure 5-6: Glucose and lactate consumption by 21253 and 21253(pKD7) in the case of 40% lactate carbon source. Strains 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l carbon source that was 40% in the form of lactate, with the remainder being glucose. The trends in the graph show residual lactate and glucose in the growth medium for the cases of the two strains.

sources are essentially equivalent. Within the following period, as the cultures age from 20 hr to 30 hr, the consumption rates for all four combinations of the two carbon sources and two strains remain virtually identical. This indicates that in cultures that contain an equal amount of carbon available in each of glucose and lactate, the cells show no kinetic preference for the consumption of either substrate during this time. After the cells have finished their growth and the cultures have entered into their stationary phases, there is more spread seen among the trends for the four cases. Examination of the data for time points later than 40 hr shows that for a given strain, there is more residual glucose than lactate remaining in the growth medium. This agrees with the results presented earlier for the cases of media with 100% glucose and 100% lactate. There, as here, the lactate is consumed at a faster rate than the glucose during the lysine synthesis phase of the
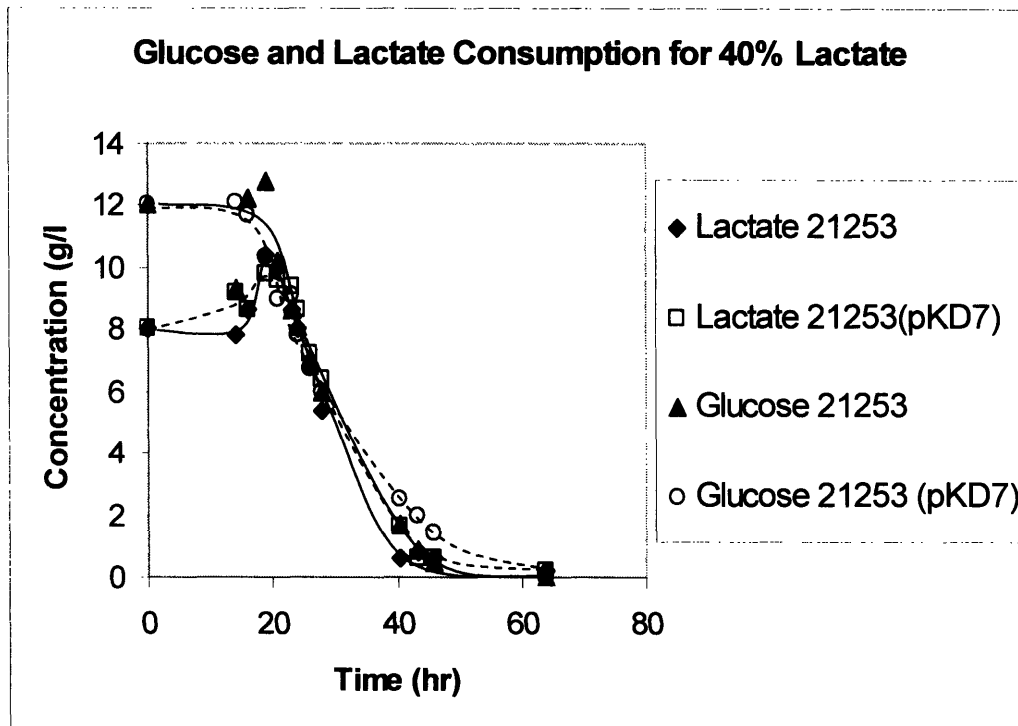
113
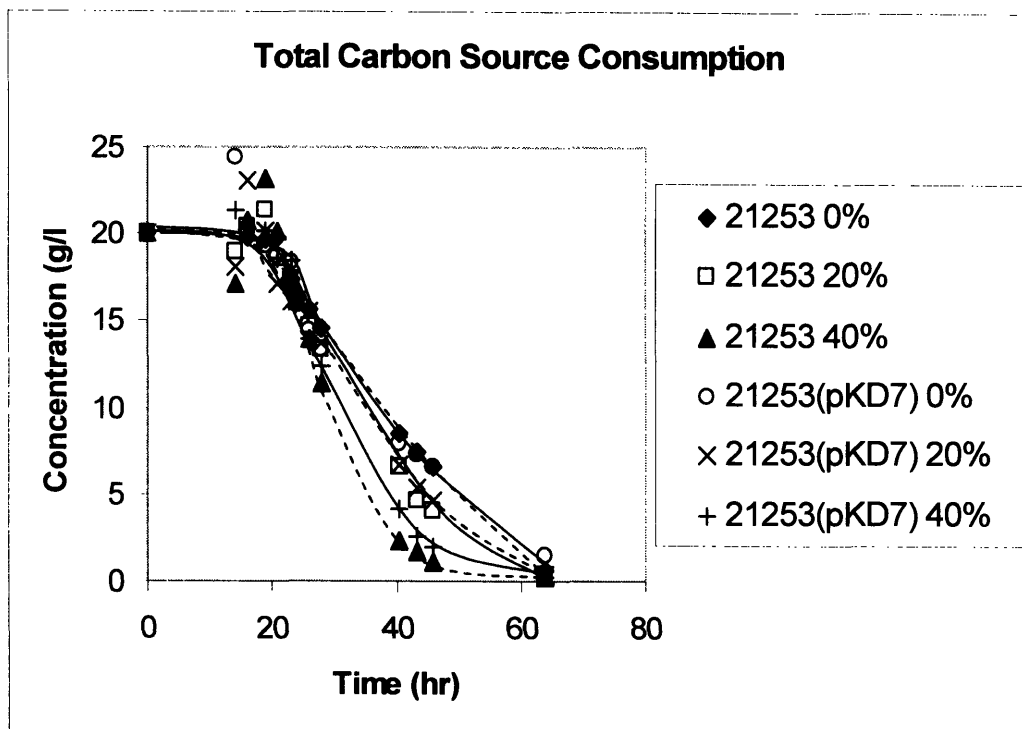
Figure 5-7: Total carbon source consumption by 21253 and 21253(pKD7) in the cases of 0%, 20%, and 40% lactate carbon source. Strains 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l carbon source that was 0%, 20%, or 40% in the form of lactate, with the remainder being glucose. The trends in the graph show residual lactate and glucose in the growth medium for these 0%, 20%, and 40% cases.

cultures, leaving more glucose undigested in the medium. Also, for a given carbon source, there is more residual carbon in the culture broth for the recombinant 21253(pKD7) strain than for the parental 21253 strain. As 21253(pKD7) is the strain with the lower yield of lysine on biomass, this lower overall consumption of carbon during the lysine synthesis phase may be a function of a decreased demand for the carbon substrate required to generate the decreased product concentration.

These differing late-phase carbon consumption kinetics between the two carbon sources and the two *C. glutamicum* strains were phenomena that arose gradually as the percentage of lactate used in the culture media was increased. Figure 5-7 presents overall carbon concentration profiles for both tested strains and the cases of media with 0%, 20%, and 40% lactate. The data of the graph shows that the trends for each strain when grown on a

medium with 100% glucose nearly overlap one another. The same can be said for the pair of trends associated with the 20% lactate medium. As the medium lactate was increased to 40%, though, the trends for the two strains began to diverge. Once again the recombinant 21253(pKD7) strain showed an earlier decrease in consumption kinetics, leaving more residual carbon as the culture entered its stationary phase. This onset of divergence as the ratio of glucose to lactate within the medium is decreased is to be expected, given the lysine yield results of section 5.2. In quantifying all three variables of yield of biomass on carbon, yield of lysine on biomass, and kinetics of carbon consumption, the behavior of the two strains was very similar when grown on 100% glucose. The values of all three of these variables only began to distinguish between the two strains as the amount of lactate used was steadily increased. It is also interesting to note from Figure 5-7 that once again as more lactate and less glucose are fed to the cells, the rate of overall carbon source consumption decreases for both strains.

5.4 Transcriptional Trends

5.4.1 Overview

The previous section detailed several ways in which the carbon consumption rates within the set of twelve cultures changed depending on the strain being cultivated and the composition of the medium used to support growth. These kinetic effects were reproducible from culture to culture, and could be explained in relation to the effects seen in the cell growth and synthesis of lysine by each of the cultures. Still, as the composition of the medium was altered to contain a smaller amount of glucose and a larger amount of lactate, the shifts in overall carbon source consumption kinetics were significant but relatively subtle. The small sizes of the changes in total carbon consumption rates and amounts imply that further changes are occurring in the ways in which the cells utilize the carbon in the metabolic networks after the glucose and lactate have been consumed.

These changes in the flow of material through the metabolic network will likely be brought about in large part through manipulations made to the gene expression profiles of the cells. The *C. glutamicum* partial genome DNA microarrays were used to probe the gene expression phenotypes of the cultures to obtain transcriptional data that could potentially be linked to the macroscopic culture behaviors observed. As is described in greater detail in section 5.1, the mRNA abundances present in the twelve studied cultures were assayed by sampling total RNA pools after 22 hr growth and hybridizing labeled cDNA derived from the RNA to the microarrays. The trends of relative transcript concentrations over the range of carbon substrate compositions tested were then determined for each of the genes represented on the arrays.

5.4.2 Embden-Meyerhof-Parnas Pathway

The transcriptional data for the genes of the Embden-Meyerhof-Parnas pathway, and all pathways studied in this experiment, have been presented in the form of trends of signal ratios determined relative to a constant RNA control sample and transformed into a $\log_2$-scale. Examples of these trends can be seen in the graphs of Figure 5-8. Because the signal ratios each measure the mRNA abundance of a particular sequence in one of the twelve cultures relative to its abundance in a control RNA that has no particular biological significance to the experiment, the absolute values of the signal ratios by themselves are not important. What are important are the relative relationships among the signal ratio values within a given trend line, and between one trend line and another determined for the same gene sequence expressed in the other of the two strains. In this way, the degrees to which the concentration of a transcript is up-regulated or down-regulated as the medium composition is altered, and to which the concentration differs between cultures of the two strains at various medium compositions, can be ascertained.

As a first example, the transcriptional trends for six of the genes associated with glycolysis and gluconeogenesis are presented in Figure 5-8. For these six genes— phosphofructokinase (pfk), triosephosphate isomerase (tpi), glyceraldehyde 3-phosphate dehydrogenase (gap), phosphoglycerate kinase (pgk), phosphoglycerate mutase (gpm),

Figure 5-8: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for glycolytic genes I. Signal ratios in $\log_2$-scale are shown for the genes phosphofructokinase (pfk), triosephosphate isomerase (tpi), glyceraldehyde 3-phosphate dehydrogenase (gap), phosphoglycerate kinase (pgk), phosphoglycer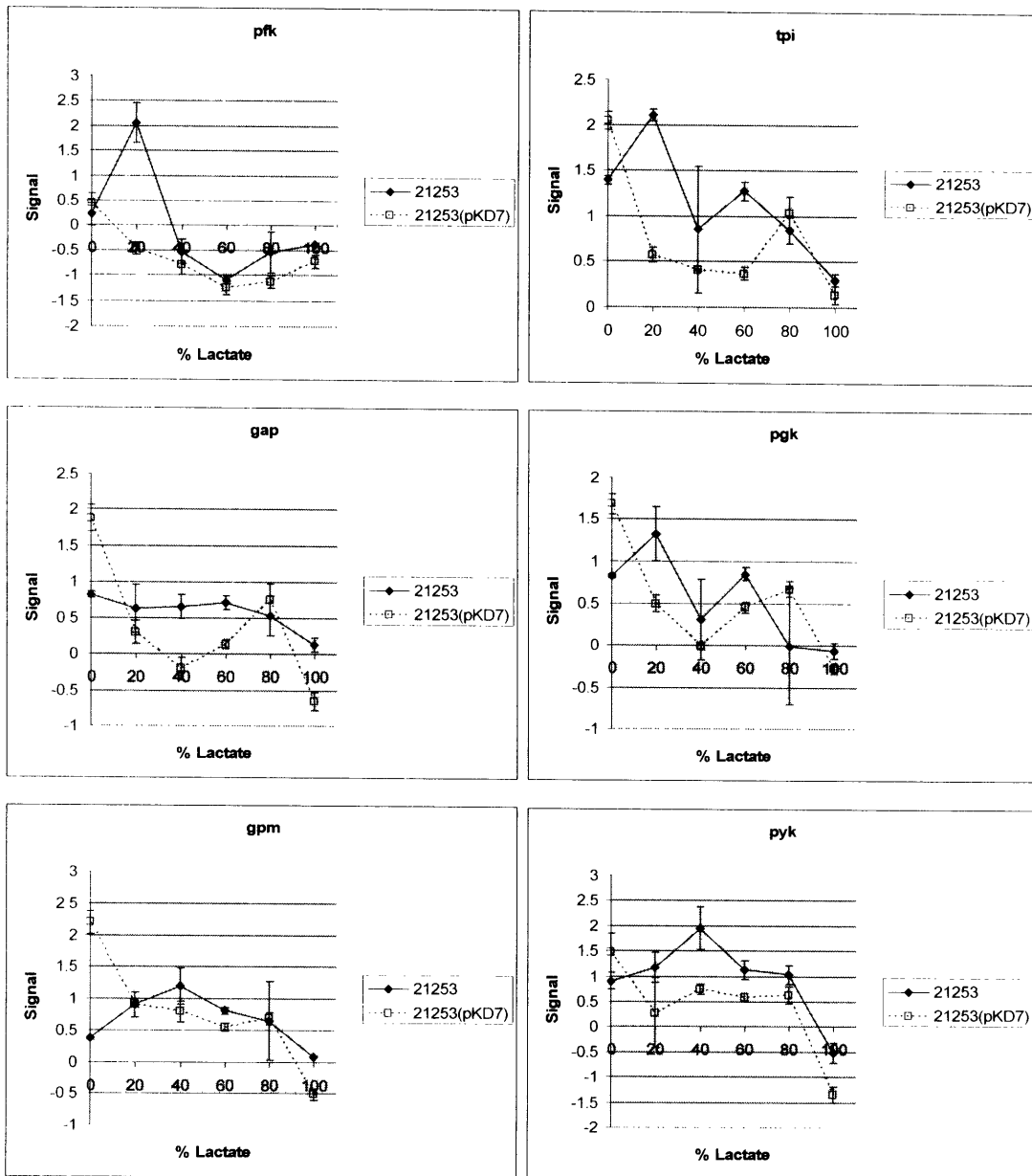ate mutase (gpm), and pyruvate kinase (pyk). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

and pyruvate kinase (pyk)—the mRNA concentrations generally declined as the medium lactate percentage increased. This suggests that the demands by the regulatory system of the cells for transcription of these genes drops as the availability of substrates forces the network to adapt from a glycolytic mode to a gluconeogenic one. Similar results were seen in the results of section 4.3, in which the mRNA levels of all EMP pathway genes declined in the *C. glutamicum* strain E12 when the sole medium carbon source was changed from glucose to lactate. The agreement of the results regarding the expression behavior of these genes in all three strains tested not only adds to the reliability of the data, but also may indicate that the mechanism that the cells use to trigger this down-regulation may be conserved among many of these genes. One such shared regulatory strategy is known to exist for tpi, gap, and pgk, the genes of which are located on the same operon in the *C. glutamicum* genome [Eikmanns, 1992]. A comparison of the transcriptional profiles shown in Figure 5-8 of these three genes in 21253 and 21253(pKD7) over the substrate range tested showed a close similarity as a result of this common regulation.

For the first five of the six EMP pathway genes that showed decreases in transcript abundance, the declines were fairly steady over the entire range of increasing lactate amounts. The mRNA levels of the sixth gene, pyk, also declined overall, but the trend of this gene remained very steady over lactate levels from 0% to 80% in both 21253 and 21253(pKD7) before dropping sharply in the case of total lactate carbon source. This significant drop, also seen in the E12 strain, indicates a regulatory shift that occurs only in the absence of any glucose, with the presence of any amount of glucose possibly sufficient to up-regulate the expression of the pyk gene. Because pyk is one of the three enzymes of glycolysis that the energetics of which render essentially irreversible, the cells have no practical use for the activity of the enzyme when there is no longer any glucose present in the growth medium. This may explain the sudden drop in the demand by the cells of both strains for the transcription of the gene. The steady region of the transcriptional concentration profile could indicate either a lack of regulation or a balancing of repression and induction. One regulatory mechanism in place with regards to pyruvate kinase involves a positive effect by AMP and a negative effect by ATP [Ozaki and Shiio, 1969]. As the amount of glucose available to the cells decreases, the

Figure 5-9: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for glycolytic genes II. Signal ratios in log$_2$-scale are shown for the gene glucose-6-phosphate isomerase (pgi). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

amount of energy in the form of ATP that can be derived from glycolysis will also decrease. This may then influence pyruvate kinase by up-regulating its transcription. At the same time, the flow of carbon to form the phosphoenolpyruvate substrate of pyk from glycolysis will of course decrease as glucose becomes the minority carbon source. Phosphoenolpyruvate concentrations have been shown to be a factor capable of pyk up-regulation [Jetten et al., 1994]. As a result, a decreased intracellular pool of phosphoenolpyruvate may also be a factor in lowering the demand of the cells for the expression of the gene.

In contrast to the trends described above, the trend derived from data for glucose-6-phosphate isomerase (pgi) demonstrated an increase, not decrease, in transcript titer from 0% to 60% lactate before steadying or dropping at higher lactate levels. The graph of Figure 5-9 shows the results associated with this gene for both of the *C. glutamicum* strains. Additionally, from Table 4-3, of all genes associated with glycolysis and gluconeogenesis, pgi showed the smallest degree of down-regulation in E12 grown on lactate instead of glucose. Another investigation found that in a mutant in which both phosphoenolpyruvate carboxylase and pyruvate kinase were deleted, pgi also operated primarily in the reverse, gluconeogenic, direction. As in this work, none of the other
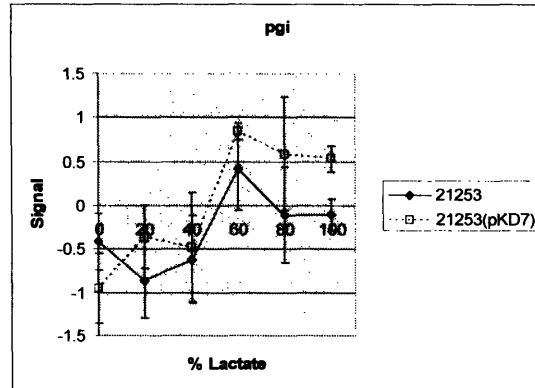
119

Figure 5-10: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for glycolytic genes III. Signal ratios in log_2-scale are shown for the genes fructose bisphosphate aldolase (fba) and enolase (eno). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

glycolytic enzymes were shown to be catalyzing the reverse reaction to a significant degree. Taken together, these may be indicators that pgi activity is a key regulatory factor in driving gluconeogenesis.

The results for the other two EMP pathway genes that have not yet been discussed are given in Figure 5-10. The data for these two genes—fructose bisphosphate aldolase (fba) and enolase (eno)—did not show any noticeable trends indicating a transcriptional response to either the changing availability of the two carbon sources or to the pyruvate carboxylase over-expression of the 21253(pKD7) strain. From Table 4-3, both fba and eno did show lower mRNA abundances in the strain E12 as it was cultivated on lactate in place of glucose. These results would seem to be in closer agreement with the general pattern of EMP pathway down-regulation on lactate that was demonstrated in both the experiment of chapter 4 and the work described in this section, raising the possibility that the data for the two genes may be faulty. It may be, however, that in the strains 21253 and 21253(pKD7) the control of fba and eno occurs at a level other that gene transcription. Even if the lack of clear up-regulation or down-regulation in Figure 5-10 is not caused by error or data variability, the overall picture of EMP pathway flux control is
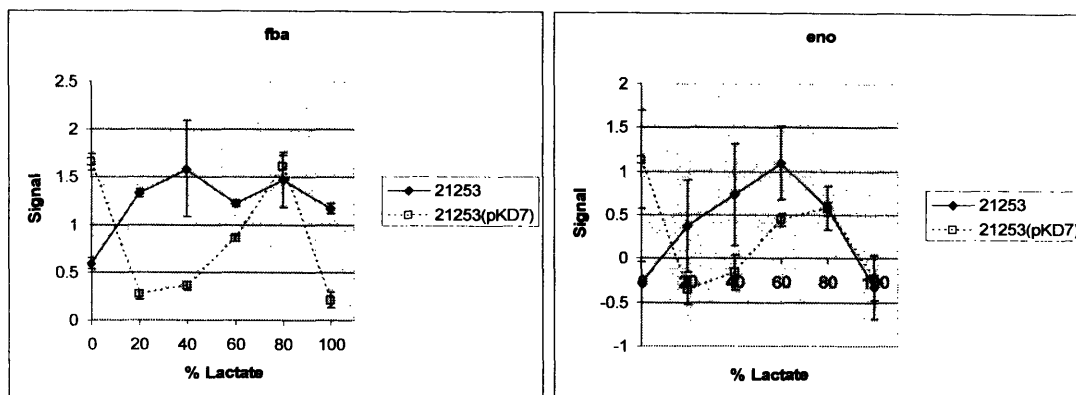
Figure 5-11: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for PPP genes. Signal ratios in $\log_2$-scale are shown for the genes phosphogluconate dehydrogenase (gnd), ribose-5-phosphate isomerase (rpi), and transketolase (tkt). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

one in which the cells actively adjust the concentrations of transcripts of its genes to respond to changes in carbon source availability.

5.4.3 Pentose Phosphate Pathway

From the transcriptional behavior, depicted in Figure 5-11, of the three pentose phosphate pathway genes represented on the *C. glutamicum* DNA microarrays, it can be seen that there was very little change in the amount of mRNA species corresponding to the genes under any of the medium or strain conditions tested. The expression levels of

neither phosphogluconate dehydrogenase (gnd), ribose-5-phosphate isomerase (rpi), nor transketolase (tkt) appeared to be influenced by the modifications made to the ratio of glucose and lactate available as substrates or by the pyruvate carboxylase over-expression of the 21253(pKD7) strain. In the previous experiment carried out with the E12 strain, only one of the three genes—that of gnd—responded to the shift from growth on glucose to growth on lactate. That gene, as is shown in Table 4-3, had 40% less transcript present when the medium carbon source was lactate. As was discussed in section 4.6.2, there is most likely a greatly reduced role for this pathway during growth on lactate. The small or negligible differences in mRNA concentrations assayed for the genes of the pathway supports the conclusion that this reduced role for the most part does not cause the cells to regulate the degree to which pentose phosphate pathway genes are transcribed.

5.4.4 Tricarboxylic Acid Cycle

Figures 5-12 and 5-13 present the data for the nine tricarboxylic acid cycle genes measured with the DNA microarrays. The first set of graphs shows the results for four genes—citrate synthase (citA), aconitase (aco), isocitrate dehydrogenase (icd), and succinyl-coA synthetase (sucC)—that displayed generally increasing transcript levels that coincided with increasing lactate concentrations in the culture broth. In the second set of graphs, trends for five genes—α-ketoglutarate dehydrogenase (sucA), succinate dehydrogenase (sdhA, sdhB, and sdhC), and fumarase (fum)—are shown for which fluctuations in mRNA titers were either insignificant or did not display a discernible pattern with respect to changes in medium composition and pyruvate carboxylase over-expression.

The higher mRNA concentrations that are seen for citA, aco, icd, and sucC in both 21253 and 21253(pKD7), as well as for citA, aco, sdhA, sdhB, and sdhC in E12 as was presented in Table 4-3, are all signs of the increased importance of the TCA cycle as the percentage of glucose relative to lactate decreases. Of the four genes that exhibited increased transcription in 21253 and 21253(pKD7), the one that had the smallest enhancement in mRNA levels for growth on lactate—icd—did not show a significant
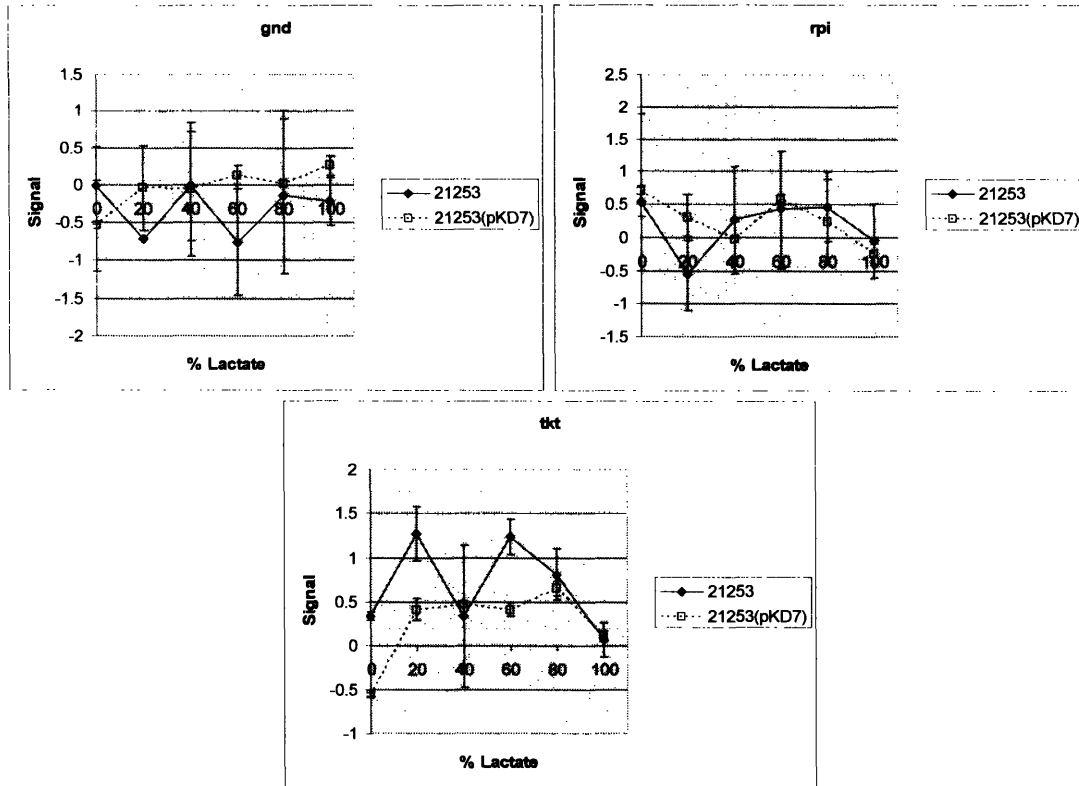
Figure 5-12: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for TCA cycle genes I. Signal ratios in $\log_2$-scale are shown for the genes citrate synthase (citA), aconitase (aco), isocitrate dehydrogenase (icd), and succinyl-coA synthetase (sucC). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

change in the E12 strain under identical growth conditions. Similarly, the succinate dehydrogenase enzymes sdhA, sdhB, and sdhC, had the smallest transcriptional increase in E12 of the five genes with significant up-regulation in that strain, and the same genes did not show noticeable trends with respect to carbon source alterations in 21253 and 21253(pKD7). The discussion in section 4.6.3 of the important role of the TCA cycle for growth of *C. glutamicum* E12 on lactate as a carbon source is still valid for the additional strains 21253 and 21253(pKD7). Once again, as the accessibility of the glycolytic reactions of the EMP pathway is reduced, the cells have a larger incentive to rely on the ability of the TCA cycle to produce the energy-storing molecules of ATP through

Figure 5-13: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for TCA cycle genes II. Signal ratios in $\log_2$-scale are shown for the genes α-ketoglutarate dehydrogenase (sucA), succinate dehydrogenase (sdhA, sdhB, and sdhC), and fumarase (fum). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

oxidative respiration. The lower yields of 21253 and 21253(pKD7) biomass on carbon for growth on lactate instead of glucose that were shown in Figure 5-1 may also be

partially explained by an increased flux through the TCA cycle. This is because as more of the substrate fed to the cells is passed through the TCA cycle, more carbon will be lost as the carbon dioxide byproducts of the isocitrate and α-ketoglutarate dehydrogenation reactions and will not be available to be incorporated into the building blocks of additional biomass.

For each of the four TCA cycle genes that showed up-regulation as the growth media of 21253 and 21253(pKD7) were altered to contain proportionally more lactate, the degree of up-regulation was greater in the case of the strain that had been engineered to over-express pyruvate carboxylase than in the case of the parental strain. Specifically, the mRNA concentrations of citA, aco, icd, and sucC that were assayed from the culture of 21253(pKD7) grown on a medium with 100% of its carbon source in the form of lactate were 300%, 300%, 200% and 400% higher, respectively, than the mRNA concentrations of the genes assayed when the strain was grown on 100% glucose. In comparison, the transcriptional levels of the same four genes were 200%, 60%, 10%, and 60% higher, respectively, in 21253 grown on 100% lactate than in the same strain grown on 100% glucose. The more up-regulated transcription of these genes in the transformant may provide increased flux throughout the TCA cycle, resulting in increased generation of energy and reducing equivalents. As this strain has higher yields of biomass on carbon for the case of growth on 100% lactate as is shown in Figure 5-1, 21253(pKD7) may then benefit from greater TCA cycle gene transcription in ways allowing it to support the higher culture growth that was observed.

One other difference among the trends of the four genes whose transcription was positively influenced by lactate can be seen in comparing the plot for icd with those of citA, aco, and sucC in Figure 5-12. Unlike the consistently upward trends seen for these three latter genes, the trend of icd exhibited a decline as the carbon source composition increased from 80% to 100% lactate. This decrease at high lactate levels implies that in addition to the pressure on the cells to up-regulate gene expression and metabolic flux associated with the TCA cycle as less glucose and more lactate are available, there is another factor that figures into the overall regulatory strategy in operation within the cells to control icd mRNA concentrations. This other factor potentially responsible for down-
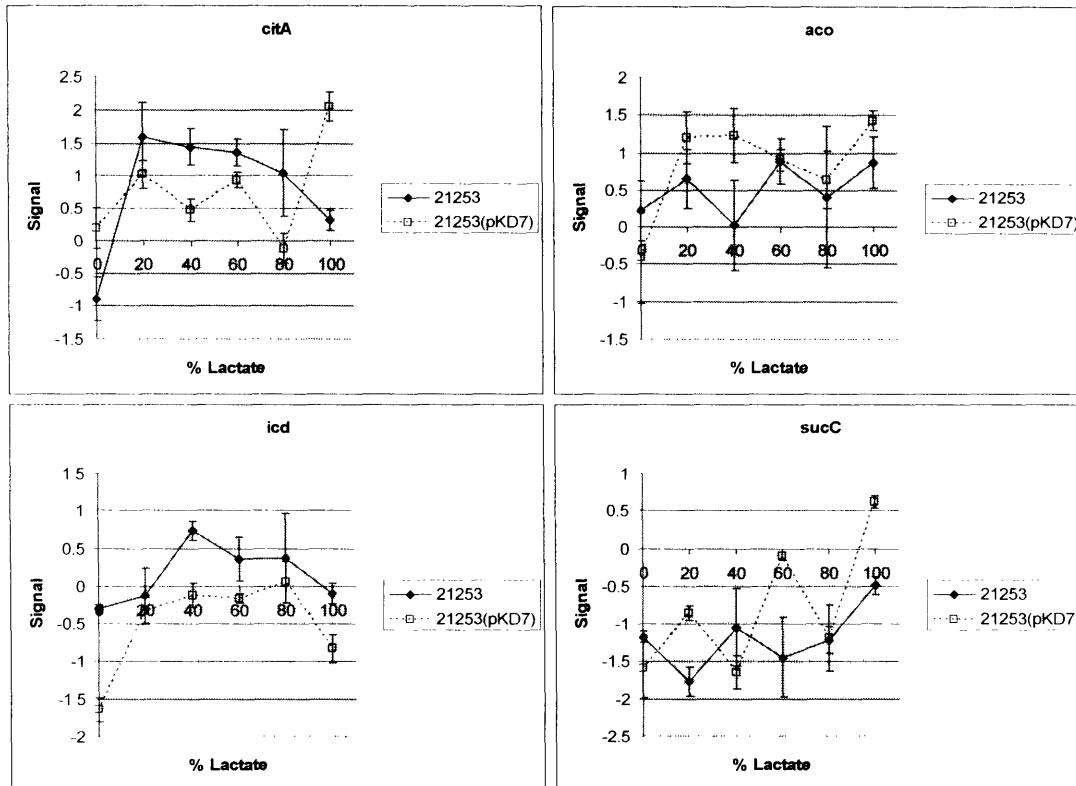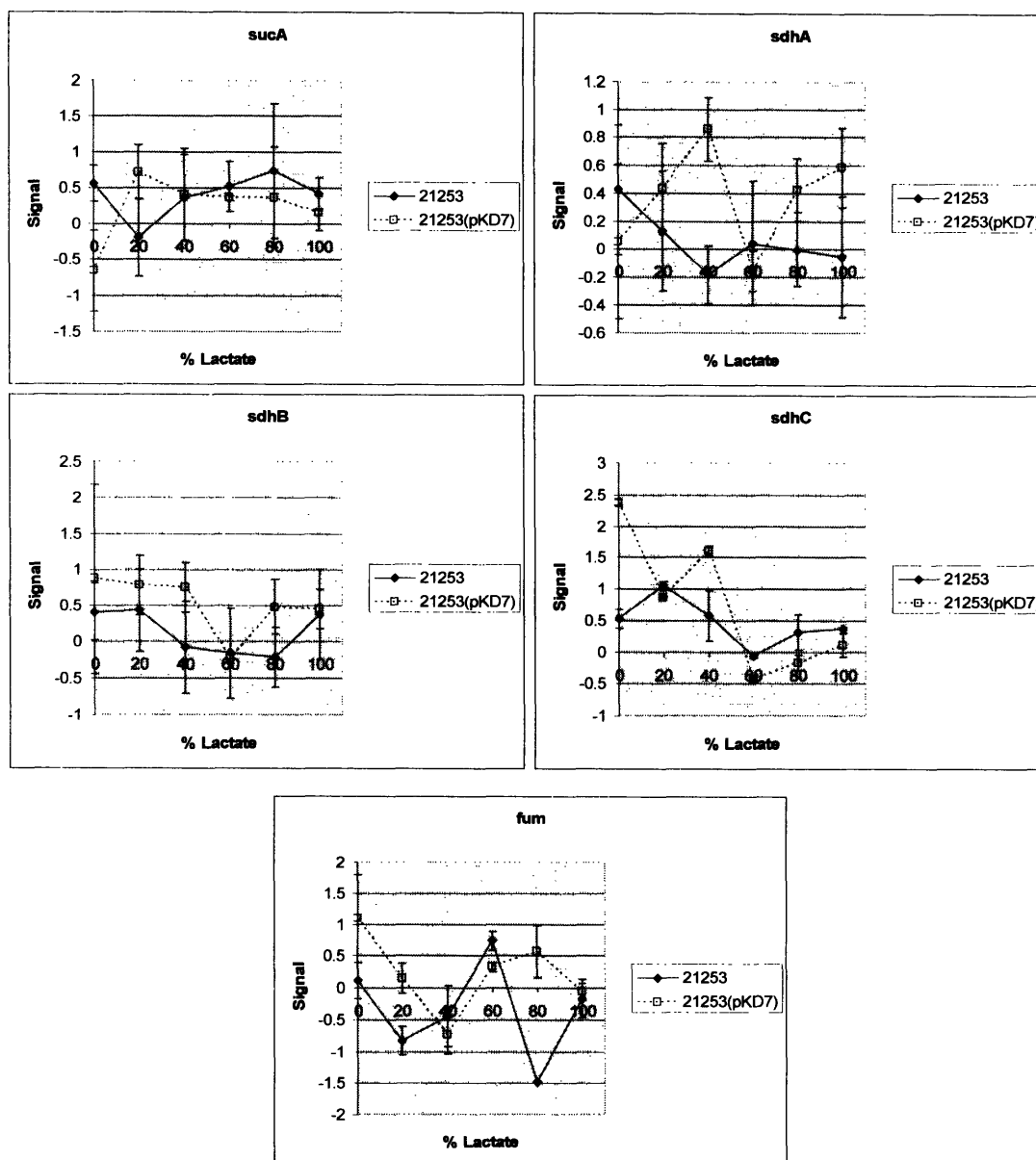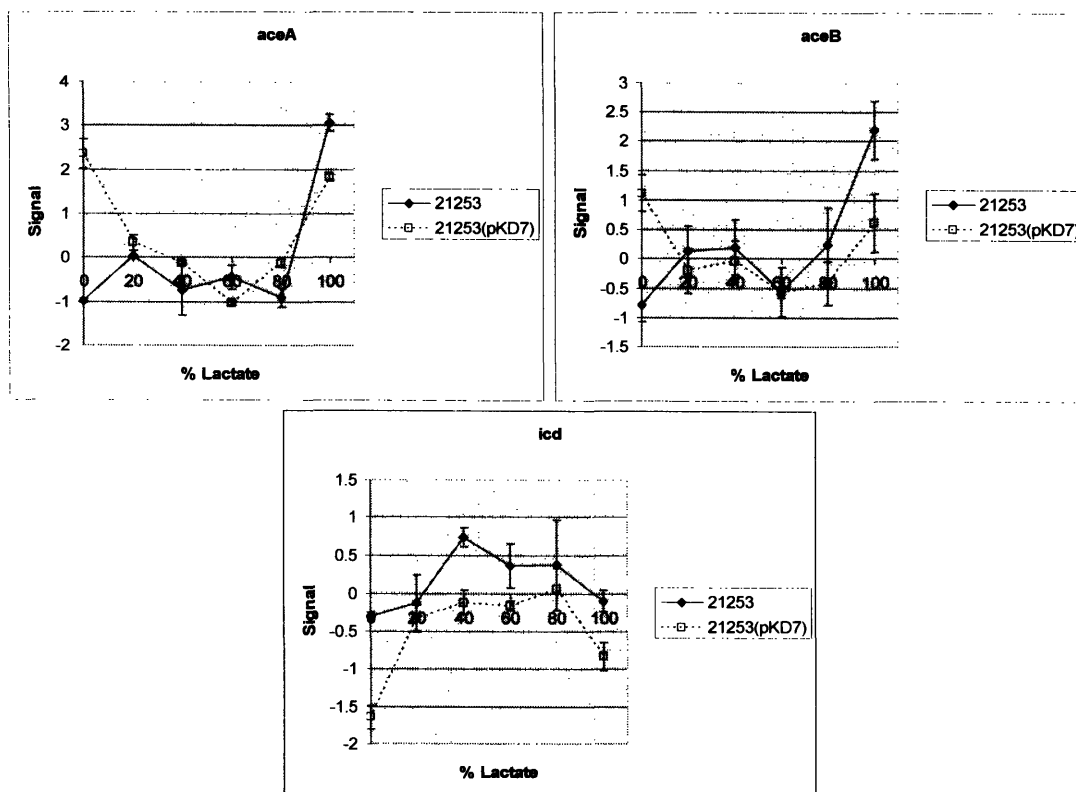
Figure 5-14: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for glyoxylate bypass genes. Signal ratios in $\log_2$-scale are shown for the genes isocitrate lyase (aceA), malate synthase (aceB), and isocitrate dehydrogenase (icd). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

regulating icd transcription at high lactate concentrations is discussed in the following section.

5.4.5 Glyoxylate Bypass

The data for the two genes of the glyoxylate bypass pathway—isocitrate lyase (aceA), and malate synthase (aceB)—are given in two of the graphs of Figure 5-14. The transcriptional trends for this pair of genes within each of the two tested strains were found to also be ones in which the abundance of mRNA increased as the percentage of

lactate constituting the total carbon source was increased, in particular for higher percentages of lactate. This result agrees well with those of the earlier experiment using the *C. glutamicum* E12 strain, discussed in section 4.6.4. In both experiments, an increase in transcription of the glyoxylate bypass pathway genes that is associated with lower availability of glucose is thought to provide enhanced opportunity for anaplerosis. This would then allow the cells to reconstitute intracellular caches of malate and oxaloacetate that would be otherwise be severely depleted by the increased demand of the cells for the TCA cycle operation discussed in the previous section. In the data given in Table 4-3, the transcriptional up-regulation of aceA and aceB in E12 associated with the medium alteration was identical. Comparing the first two graphs of Figure 5-14 shows that within each strain, the transcriptional behavior of the two genes was again virtually identical. This is surely another result of the shared means, described in section 4.6.4, by which the cells regulate the activity of aceA and aceB. Also, by demonstrating that genes known to be co-regulated by the cells do indeed display similar transcriptional trends in this experiment, validity is given to claims made for other instances of co-regulation that are either novel or merely theorized, but have not yet been clearly demonstrated.

Another instance of apparent co-regulation involves the two genes of the glyoxylate bypass pathway and the isocitrate dehydrogenase gene from the TCA cycle that was discussed in the previous section, the trend of which is reproduced in Figure 5-14 for easy comparison. The transcriptional trends for aceA and aceB have nearly identical properties, in which the assayed mRNA amounts remain fairly steady as the lactate increases from 20% to 80% of the carbon source, and then rise significantly between 80% and 100% lactate. This behavior is the opposite of what takes place with the transcriptional abundance of icd in the same medium compositions. The icd mRNA amounts also remain fairly steady from 20% to 80% medium lactate, but then fall sharply between 80% and 100% lactate. Because aceA and icd each catalyze reactions that share isocitrate as a substrate, the simultaneous increase in the concentration of the transcript of one and decrease in the concentration of another may signal a resulting shift in flux from one reaction pathway to another. In other words, it may be that when there is a mixture of glucose and lactate available to the cells for use as carbon supplies, the metabolic network is used in a way that balances the use of the glycolytic and gluconeogenic

reactions of the EMP pathway as well as those of the TCA cycle according to the ratio of glucose and lactate that are present. As the amount of available glucose relative to lactate decreases, the need for TCA cycle flux increases, up to the case in which glucose levels are less than 20% of the carbon source. At this point, the cells continue to rely heavily on the TCA cycle, but now there is need for the anaplerosis of the glyoxylate cycle because of the increased TCA cycle flux. To accommodate this, the cells may simultaneously down-regulate the transcript levels of icd, which had been elevated as a result of the higher need for the TCA cycle, and up-regulate the transcript levels of aceA.

Curiously, in the case of the transformed strain, higher mRNA levels were seen for both aceA and aceB than in the case of the parental strain for growth on a medium with 100% glucose carbon source. As there were no macroscopic extracellular properties, such as markedly different yields of biomass or lysine or rates of glucose consumption, specific to the transformed strain growing in a 100% glucose medium, there is no obvious explanation for this behavior. Nevertheless, because similar results were seen for both of the two glyoxylate bypass genes, and because the standard deviations of the measurements for both genes were relatively low, the validity of the finding is high. Note also that the deviations between the transcriptional levels of these two genes found in the two strains in the case of 100% glucose were mirrored by an analogous, but opposite, deviation between the transcriptional levels of icd in the strains at 100% glucose.

5.4.6 Oxaloacetate/Pyruvate Node

5.4.6.1 Pyruvate Carboxylase and Malic Enzyme

Figure 5-15 presents the results of the mRNA titers found for the first two of multiple genes operating at the metabolic node of oxaloacetate and pyruvate—pyruvate carboxylase (pyc) and malic enzyme (malE). From the first of these graphs it can be seen that for all mixtures of glucose and lactate that were tested, the mRNA abundance of pyc was higher in the strain that had been genetically engineered for its over-expression. This expected result offers another means of validating the results provided by the C.

Figure 5-15: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for pyc and malE. Signal ratios in $\log_2$-scale are shown for the genes pyruvate carboxylase (pyc), and malic enzyme (malE). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

glutamicum partial genome DNA microarrays as well as the data collection and analysis procedures of this experiment.

Also in the recombinant 21253(pKD7) strain, the transcript concentration for malE was almost 3-fold higher for growth on 100% lactate than on 100% glucose. From Table 4-3, the E12 strain similarly possessed higher transcript levels for malE when grown on a medium with lactate as the only carbon source. Section 4.6.5 lists evidence that malE is important for the growth of C. glutamicum on lactate, which the findings from 21253(pKD7) and E12 appear to support. The malE transcriptional trend for the parental 21253 strain, in contrast, showed a much flatter profile, suggesting that there may be a

link between the expression levels of malE and that of the pyruvate carboxylase gene that is not being over-expressed in this strain.

Further exploring the possible link between the transcription of pyc and malE, the third plot of Figure 5-15 superimposes the data of mRNA levels for the two genes in the 21253(pKD7) strain when grown on different mixtures of glucose and lactate. The patterns of increasing and decreasing transcript concentrations in the two trends as the medium contains more lactate and less glucose do in fact match very well. This high degree of similarity between the two profiles provides evidence that these two enzymes serve a coordinated role that is of major importance for cells that are metabolizing lactate, and may also be indicative of co-regulation by the cells to serve a joint purpose. Such a co-regulatory strategy could exist in *C. glutamicum* as an additional means of generating reducing equivalents. As was discussed in section 4.6.2, with a culture medium that presents the bacteria with less glucose and more lactate, the accessibility of the pentose phosphate pathway, the primary source of the reducing equivalent NADPH [Vallino and Stephanopoulos, 1993], is greatly reduced. For growth on glucose, it has been shown that the availability of NADPH is not a limiting factor in the growth of *C. glutamicum* [Moritz *et al.*, 2000], and an increase in its production by gluconate catabolism did not increase the yield of lysine [Vallino and Stephanopoulos, 1994]. Because this is not likely to be the case for growth on lactate, an alternative method is required by the cells to generate the NADPH necessary to support growth and product formation. One proposed pathway involves a cycling of metabolites through reactions catalyzed by both malE and pyc in such a way as to take advantage of the inherent NADPH generation of the conversion of malate to pyruvate [Cocaign-Bousquet and Lindley, 1995]. A linking of these two reactions could happen in the form of a modified TCA cycle, in which they form a two-step process used to transform malate into oxaloacetate, bypassing the one-step malate dehydrogenation. Alternatively, a smaller transhydrogenation cycle could be operating in which malate dehydrogenase catalyzes the reverse reaction, converting oxaloacetate into malate and consuming NADH [Dominguez *et al.*, 1998]. Malic enzyme could then convert the malate to pyruvate while producing NADPH, and pyruvate carboxylase could then replenish the oxaloacetate pool.
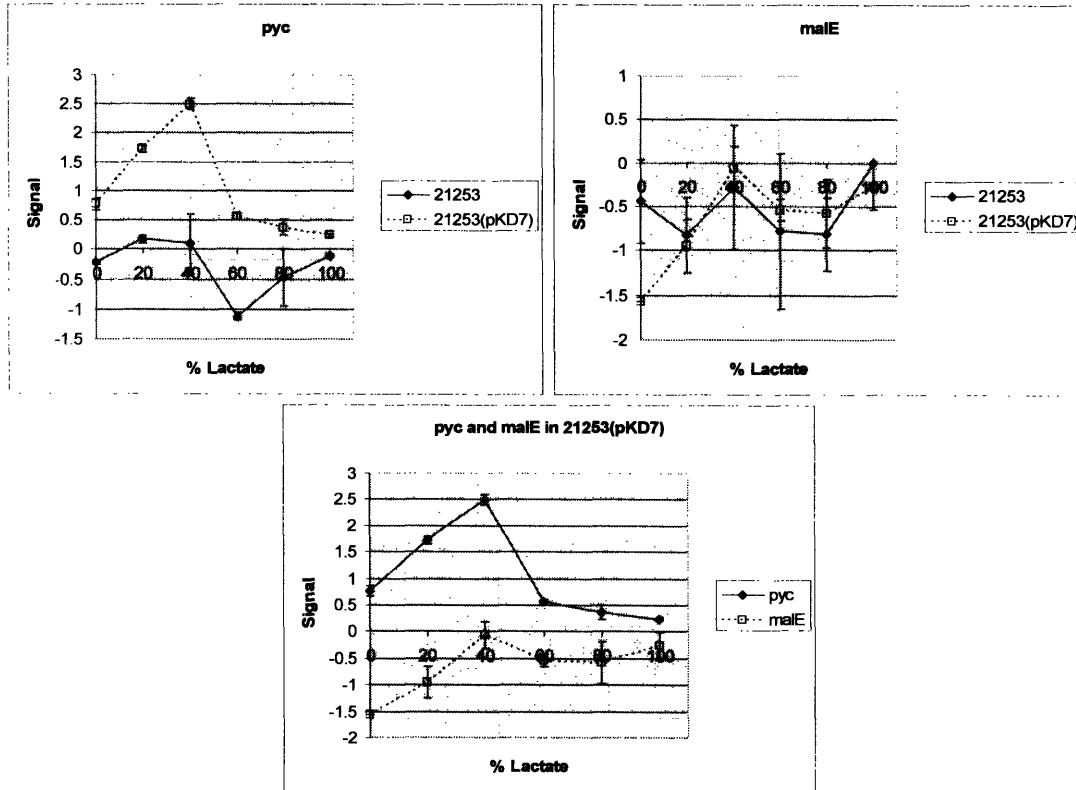
Figure 5-16: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for pck. Signal ratios in $\log_2$-scale are shown for the gene phosphoenolpyruvate carboxykinase (pck). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

5.4.6.2 Phosphoenolpyruvate Carboxykinase

The mRNA titers for phosphoenolpyruvate carboxykinase (pck) increase in each of the two strains, as is shown in Figure 5-16, as the medium is changed from predominantly glucose to predominantly lactate. The role of this enzyme in gluconeogenesis, discussed in section 2.1.1.4.3, is in accordance with its apparent up-regulation under these conditions for all three *C. glutamicum* strains of 21253, 21253(pKD7), and E12, the results from which are given in Table 4-3. As the growth medium carbon source is shifted from glucose to lactate, gluconeogenesis is required to allow the cells to form sugar phosphates from precursors available in the TCA cycle. In replenishing the intracellular pool of phosphoenolpyruvate, which is the starting material for the gluconeogenic reactions of the EMP pathway, pck activity is necessary for growth on carbon sources other than glucose or other glycolytic intermediates. The importance of the enzyme, as well as apparently its transcriptional abundance, then increases as the percentage of lactate in the medium is raised.
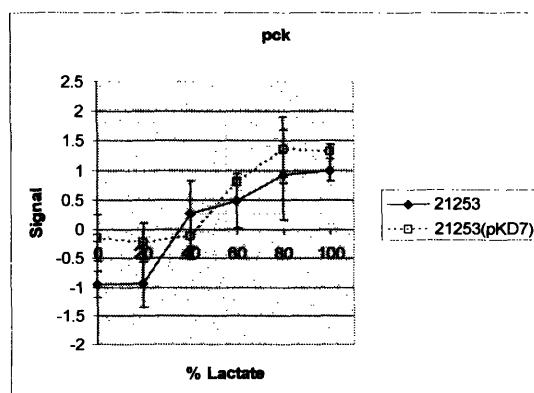
Figure 5-17: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for pdh. Signal ratios in $\log_2$-scale are shown for the gene pyruvate dehydrogenase (pdh). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

5.4.6.3 Pyruvate Dehydrogenase

The graph of Figure 5-17 shows that the mRNA titers of pyruvate dehydrogenase (pdh) also increased as lactate and glucose approached equal concentrations before steadying or dropping as the lactate percentage increased from 60% to 100%. The trends for the two strains here differed somewhat, though, with that of the transformed strain exhibiting a lag before both the increase in transcript concentration and the eventual slight decline. This may be because the transformant has an enhanced alternate means, in the form of the activity of the pyruvate carboxylase that is over-expressed by the strain, of depleting intracellular pyruvate pools without using pdh. It may be that a threshold of pyruvate concentration must exist before the cells trigger responses by controlling the transcription of related genes, such as that of pdh. With increased pyruvate carboxylase activity it may be that more of the available pyruvate is converted away from the coupling reaction with the TCA cycle and towards the oxaloacetate node from which material may flow into, for example, the lysine biosynthetic pathway. As lactate concentration is directly linked to pyruvate concentration through the activity of lactate dehydrogenase, a theoretical increase to the critical level of pyruvate production required to elicit a transcriptional

response could easily translate to the 20-40% medium lactate percentage lags seen in the graph.

5.4.7 Lysine Synthesis Pathway

Given the significant positive effect that the medium lactate carbon source composition was found to have on lysine productivities in 21253 and 21253(pKD7), it was reasonable to expect a pattern of higher mRNA levels for genes in the lysine synthesis pathway when growth occurred on lactate. Instead, as the results of Figures 5-18 and 5-19 show, several of the genes catalyzing reactions in the pathway had no noticeable transcriptional response to changes in carbon source composition or pyruvate carboxylase over-expression, while the genes that did in fact have higher mRNA levels did so only significantly for the transformed strain. The genes that had negligible transcriptional responses included succinyl dihydrodipicolinate reductase (dapB), tetrahydropicolinate succinylase (dapD), and diaminopimelate epimerase (dapF), shown in Figure 5-18. The TCA cycle gene set that did exhibit some up-regulation includes succinyl diaminopimelate desuccinylase (dapE), diaminopimelate decarboxylase (lysA), lysine exporter (lysE), and lysine export controller (lysG), and is shown in Figure 5-19. From the trends of these graphs, the transcript concentrations increased by 2- to 4-fold in 21253(pKD7) as the glucose carbon source of the medium was increasingly replaced with lactate, while the transcript concentrations changed by an insignificant amount in 21253. Also, of these four genes, only lysA also showed significant up-regulation in Table 4-3 for growth on lactate in *C. glutamicum* E12. It is counterintuitive that of the two strains studied in this experiment, the one for which lysine biosynthesis genes were up-regulated for growth on lactate relative to growth on glucose was the strain which, from Figure 5-2, had the lower yield of lysine on biomass. Still, even with the trends for these genes showing upward movement with increasing lactate in 21253(pKD7), the resulting mRNA levels are rarely higher than those for the same genes in 21253.

Although the transcript concentrations of some of the genes of the lysine biosynthetic pathway did increase on lactate for the strain 21253(pKD7), the most significant trends of
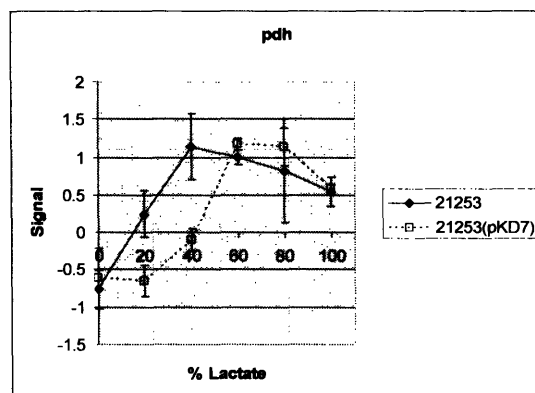
Figure 5-18: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for lysine biosynthesis genes I. Signal ratios in $\log_2$-scale are shown for the genes succinyl dihydrodipicolinate reductase (dapB), tetrahydropicolinate succinylase (dapD), and diaminopimelate epimerase (dapF). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

the pathway were the downward ones seen for aspartokinase (ask) and aspartate semialdehyde dehydrogenase (asd) shown in Figure 5-20. Comparing these trends for the two enzymes within each of the parental and transformed strain independently, shows strong similarity between transcriptional behavior and evidence of co-regulation. Such co-regulation is likely to be a result of the proximity of the genes within the genome of *C. glutamicum*, where their constituent open reading frames and promoters actually overlap [Kalinowski *et al.*, 1990]. Not only did both genes show similar patterns of declining mRNA titers across the range of increasing lactate content, but the magnitudes of the declines were very similar. Comparing the trends for the two enzymes across the two
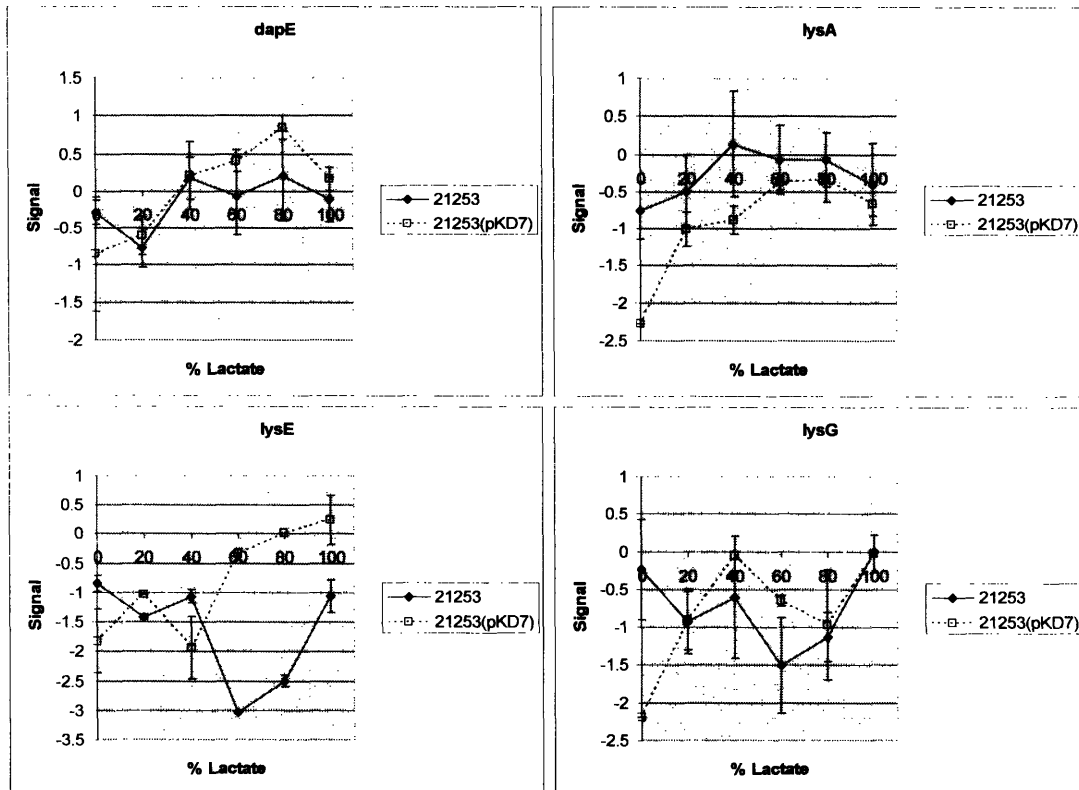
Figure 5-19: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for lysine biosynthesis genes II. Signal ratios in $\log_2$-scale are shown for the genes succinyl diaminopimelate desuccinylase (dapE), diaminopimelate decarboxylase (lysA), lysine exporter (lysE), and lysine export controller (lysG). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

strains, however, showed significant differences. While ask and asd transcription dropped 2- to 3-fold in 21253, the transcription of the pair of genes dropped 6- to 8-fold in 21253(pKD7). Also, the mRNA levels of the two enzymes were nearly identical for the case of growth on 100% glucose. These facts together implicate these two genes as being intimately involved in the effects seen on lysine yields for the two strains, given in Figure 5-2. As with the ask and asd transcriptional levels, the lysine yields for 21253 and 21253(pKD7) are nearly identical for growth on 100% glucose, and then deviate from one another as available glucose is reduced, with the transformed strain exhibiting the more dramatic change. That the ask and asd mRNA concentrations drop as the cultures
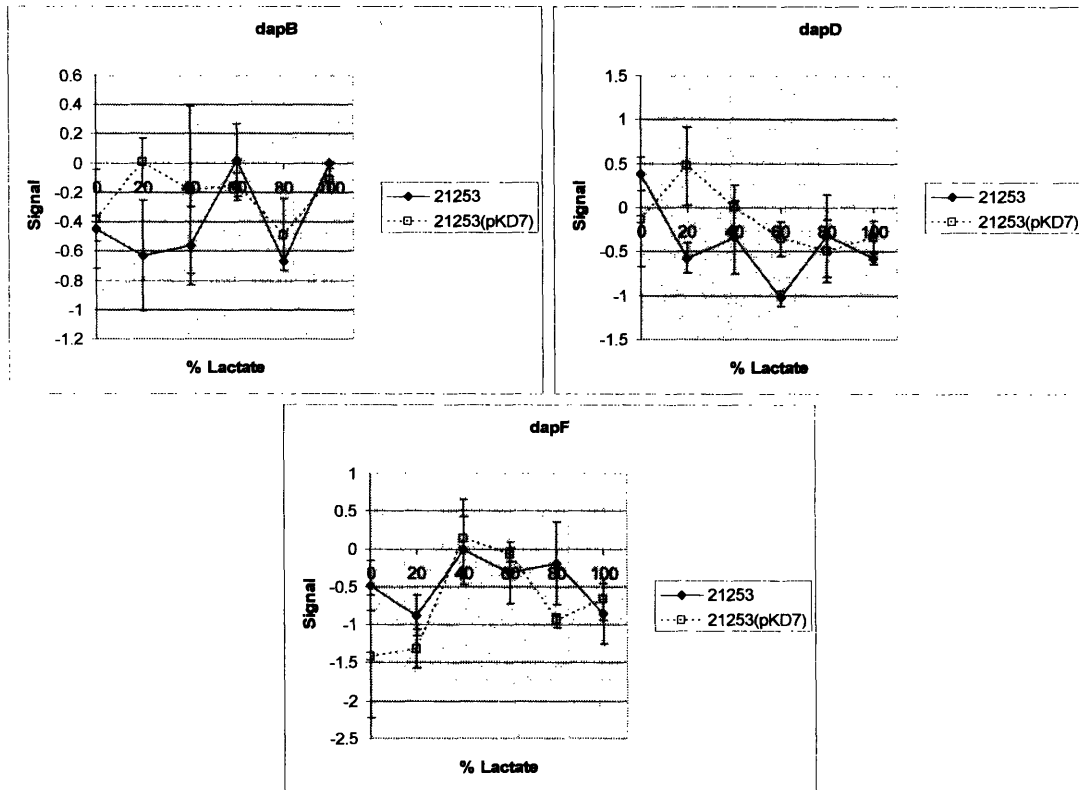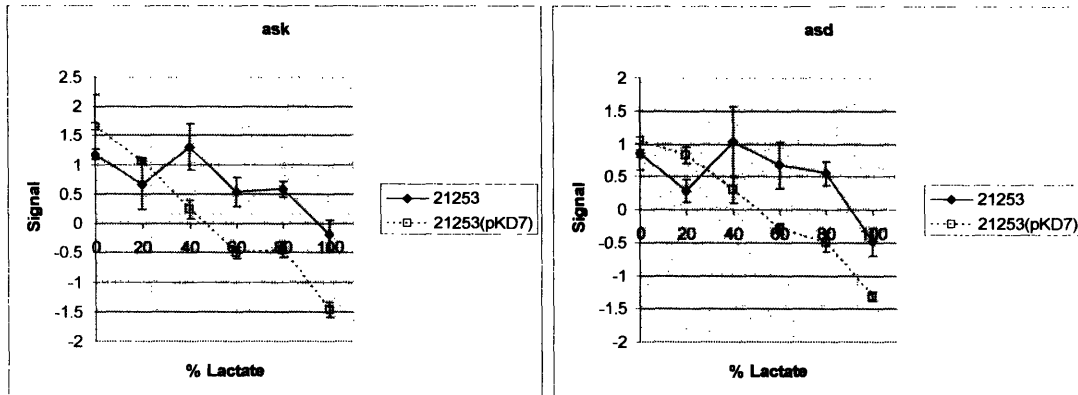
Figure 5-20: Transcriptional trends relative to compositions of mixed carbon sources and pyc over-expression for lysine biosynthesis genes IIi. Signal ratios in log$_2$-scale are shown for the genes aspartokinase (ask) and aspartate semialdehyde dehydrogenase (asd). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l glucose and/or lactate in the percentages indicated. Larger signal ratios indicate more transcript relative to the control RNA sample.

are fed higher percentages of lactate suggests that a bottleneck exists at this point in the flow of material from central carbon metabolism to lysine biosynthesis. That the drop is larger for the transformed strain suggests that this bottleneck is even more pronounced when pyruvate carboxylase is over-expressed. Another strain was recently developed in our laboratory that has been designed to over-express both pyruvate carboxylase and aspartokinase. When this strain is grown on a minimal medium using lactate as its carbon source, an increase in lysine yields is seen relative to that of the parental strain [Koffas, 2003]. The microarray results presented here help to clearly explain why this technique of coordinated gene over-expression proved to be successful.

Chapter 6: Effects of Gene Over-Expression and Culture Phase on Transcriptional Profile

6.1 Experimental Design

To further explore the transcriptional differences that exist between the two *C. glutamicum* strains 21253 and 21253(pKD7), another experiment was carried out in which their cultures were sampled at three different time points representing different phases of growth and lysine production. In this way, we could also see how concentrations of specific mRNA sequences within the cells change over time to adapt to shifting demands on the metabolic networks of the cells. Other experiments that have also investigated changes in *C. glutamicum* gene transcriptional levels present at different time points include studies of the effects of shifts from the growth phase to the lysine production phase of a glucose-fed fermentation [Loos *et al.*, 2001], of the time-dependent cellular response to phosphate starvation [Ishige *et al.*, 2003], and of the dynamic induction of multiple over-expressed genes related to amino acid biosynthesis [Glanemann *et al.*, 2003].

From the results of the previous chapter, of all mixtures of glucose and lactate used as carbon sources in the culture growth medium, the case of 100% lactate caused the greatest differentiation between the parental strain and the recombinant strain engineered for pyruvate carboxylase over-expression. This was true both for macroscopic extracellular properties, such as the yields of biomass on carbon and of lysine on biomass, and for intracellular transcriptional properties, such as the expression profiles of genes associated with the key metabolic pathways of central carbon metabolism and

lysine biosynthesis. For this reason, the cell cultures cultivated for this experiment were fed only lactate as their carbon source.

The cells were grown on the minimal medium described in Table 4-1, with the 21253(pKD7) culture supplemented with 50 mg/l kanamycin to select for cells with functional copies of the pyruvate carboxylase over-expression plasmid. The cultures were grown in 250-ml round-bottom flasks containing 50 ml medium each that were incubated at 30°C and 250 rpm on a rotary shaker. Samples with a volume of 1 ml were taken from the cultures periodically and assayed for cell density and lysine and threonine concentrations according to the procedures of section 5.1. Also, samples with a volume of 15 ml were taken at three different time points for use in RNA isolation and microarray analyses of the transcriptional profiles of the two cultures. The fluorescently labeled cDNA created from each RNA sample was hybridized along with labeled cDNA created from a stock of control RNA to two microarrays. This control RNA was obtained by pooling together both cultures of the experiment after growth and lysine production had completed, and isolating enough RNA from this pool to be used in all required microarrays.

6.2 Biomass and Lysine Yields

The growth and lysine generation trends of the two strains are shown in Figure 6-1. As was the case in the experiment of Chapter 5, the strain 21253(pKD7) was capable of supporting a larger cell concentration than 21253 when grown on an equal amount of lactate as the carbon source. Also again, the specific productivity of lysine in these transformant cells was lower than that seen in those of the parental strain.

In both cultures, lysine production began during late-exponential growth and coincides with the depletion of threonine. The 21253 strain is a threonine auxotroph due to the deletion of the enzyme homoserine dehydrogenase, which catalyzes the conversion of aspartate semialdehyde of the lysine biosynthetic pathway to homoserine, a starting material required for the synthesis of the amino acids threonine, glycine, and serine.
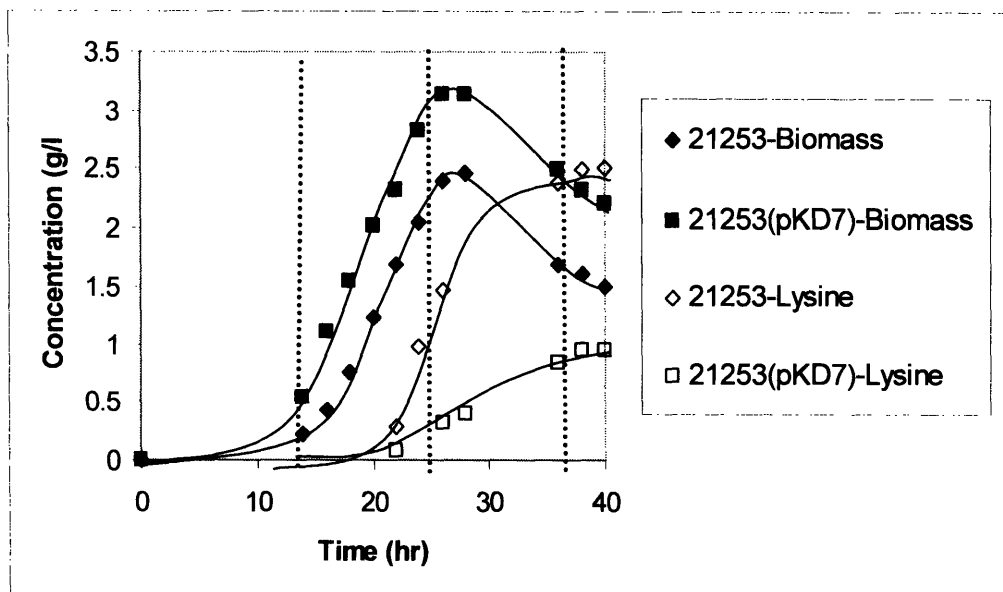
Figure 6-1: Growth and lysine production of 21253 and 21253(pKD7) grown on minimal medium containing 20 g/l lactate. Strains 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l carbon source in the form of lactate. Vertical dashed lines indicate the three time points at which total RNA samples were isolated for use in DNA microarray analyses.

Through the deletion of this enzyme, the diversion of material away from the lysine anabolic pathway is reduced, and the formation of the aspartokinase inhibitor threonine is prevented. As threonine is an essential nutrient for growth, it is provided to the culture in the medium, preventing the production of lysine until its concentration falls to a negligible level. Total RNA samples were taken at 14-, 28-, and 36-hours post-inoculation, and therefore represent the cultures in early-exponential growth, in late-exponential growth as lysine is beginning to be generated, and when the growth and lysine biosynthesis are both essentially stationary. The DNA microarrays were used to probe mRNA that was isolated from each sample, and comparisons were made between data from the different strains as well as from different time points.

139

Figure 6-2: Transcriptional trends relative to culture phase and pyc over-expression for pyc and malE. Signal ratios in $\log_2$-scale are shown for the genes pyruvate carboxylase (pyc) and malic enzyme (malE). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l lactate. Larger signal ratios indicate more transcript relative to the control RNA sample.

6.3 Transcriptional Trends

6.3.1 Genes Up-Regulated With Pyruvate Carboxylase

6.3.1.1 Pyruvate Carboxylase and Malic Enzyme

The transcriptional data for the genes of pyruvate carboxylase (pyc) and malic enzyme (malE) are given in Figure 6-2. As with the plots of section 5.4, the trends display the signal ratios determined relative to a constant RNA control sample and transformed into a

log$_2$-scale. Once again, because the signal ratios each measure the mRNA concentration of a particular sequence in one of the two cultures relative to its concentration in a control RNA pool, the absolute values of the signal ratios are not as important as the relative relationships among the values. These relative relationships can be used to determine the degrees to which the abundance of a transcript is up-regulated or down-regulated as the cultures enter a different phase, and to which the abundance differs between cultures of the two strains within the same phase.

Agreeing with expectations, during both the growth and lysine production phases of the cultures the transcript concentration of pyc was higher in the strain that had been genetically engineered to over-express its gene. As with the similar results of section 5.4.6.1, this provides another validation of the results provided by the *C. glutamicum* partial genome DNA microarrays as well as the data collection and analysis procedures of this particular experiment. Also, transcripts for malE were significantly higher in this strain as well. For the first two time points sampled, these two genes had on average 16-fold more copies of their representative mRNA present within the transcriptome of the cells. In addition, within each strain, the trends for the two genes showed markedly similarly profiles, as can be most easily seen in the third graph of Figure 6-2, which overlays the four trends of the two genes within the two strains. This adds further transcriptional evidence to support the hypothesis suggested by the previously described experiment that pyc and malE are regulated by the cells in a coordinated manner to serve closely related roles within the metabolic network, especially for the case of growth on a lactate carbon source. As was described in section 5.4.6.1, these roles are most likely to be involved in the generation of reducing equivalents through the activity of either a modified tricarboxylic acid cycle or a transhydrogenation cycle, thereby providing an additional means for the cells to compensate for a decreased ability to form NADPH through the activity of the pentose phosphate pathway.
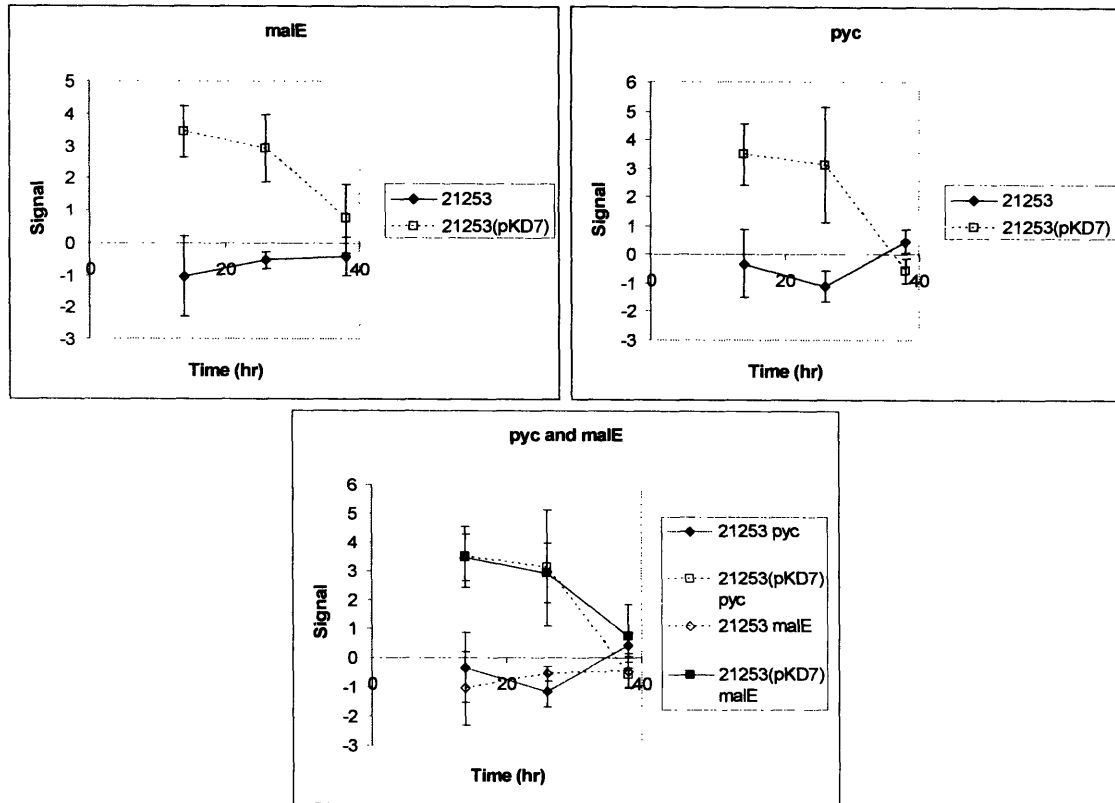
Figure 6-3: Transcriptional trends relative to culture phase and pyc over-expression for citA and sucA. Signal ratios in $\log_2$-scale are shown for the genes citrate synthase (citA) and α-ketoglutarate dehydrogenase (sucA). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l lactate. Larger signal ratios indicate more transcript relative to the control RNA sample.

6.3.1.2 Tricarboxylic Acid Cycle

Figure 6-3 presents data for the two genes of the tricarboxylic acid cycle—citrate synthase (citA) and α-ketoglutarate dehydrogenase (sucA)—that demonstrated significant differences among transcript levels assayed for the two strains. For both of these genes, the higher concentration of mRNA was seen in the pyruvate carboxylase over-expressing 21253(pKD7) strain. As was shown in Figure 6-1, this is the strain that is capable of supporting a greater density of biomass within the culture when grown on lactate. From the description in section 2.1.2.3 of the functioning of the tricarboxylic acid cycle, the pathway serves to generate NADPH, energy, and metabolic intermediates that are the starting materials for amino acids, nucleic acids, and cell wall components, all of which are components necessary for cellular maintenance and growth. The increased transcript levels for citA and sucA then appear to fit well with a model in which the flux rate through the tricarboxylic acid cycle increases to metabolize the larger intracellular pool of oxaloacetate resulting from higher pyruvate carboxylase activity in this strain. The increased tricarboxylic acid cycle flux is then responsible for the greater biomass yield seen with 21253(pKD7).

It is interesting to note that both citA and sucA catalyze reactions that are located at branch points within the metabolic network of the organism. The first of these is positioned at the oxaloacetate node, with increasing activity of the enzyme causing more conversion of oxaloacetate to citrate in the tricarboxylic acid cycle, instead of to aspartate in the aspartate-family amino acid biosynthesis pathway. The second is positioned at the α-ketoglutarate node, with additional activity resulting in further conversion of α-ketoglutarate to succinyl-CoA in the tricarboxylic acid cycle, instead of to glutamate in the glutamate-family amino acid biosynthesis pathway. For the recombinant pyruvate carboxylase over-expressing strain relative to the parental strain, the emphasis is clearly then on greater material flow through the tricarboxylic acid cycle at the expense of flow leading to amino acid anabolism. Again, the ramifications of this can be seen in Figure 6-1, in which 21253(pKD7) has not only a higher biomass yield than 21253, but a lower lysine yield as well.

One other feature of the trends of Figure 6-3 with an apparent link to the macroscopic properties of the cell cultures can be seen in the data plot for citA. As was mentioned earlier, this enzyme catalyzes the tricarboxylic acid reaction positioned at a branch point with the lysine synthesis pathway. This pathway is subject to strong feedback inhibition targeting aspartokinase, the enzyme catalyzing the alternate reaction at the same branch point, as was described in section 6.2. This inhibition is removed with the depletion of threonine from the culture growth medium, which occurs at a time between those of the first and second RNA samples. With the removal of this inhibition, lysine production begins, and presumably more oxaloacetate is directed into the lysine biosynthetic pathway at the expense of the tricarboxylic acid cycle. This time-dependent decrease in the tricarboxylic acid cycle flux specifically at this point in the pathway is reflected in the decrease in transcript levels of citA from the growth phase assayed with the first sample to the lysine formation phase assayed with the second sample. The decrease in citA transcript levels over time, which was also observed by Loos et al. (2001), can be seen in the trends for both strains shown in Figure 6-3.
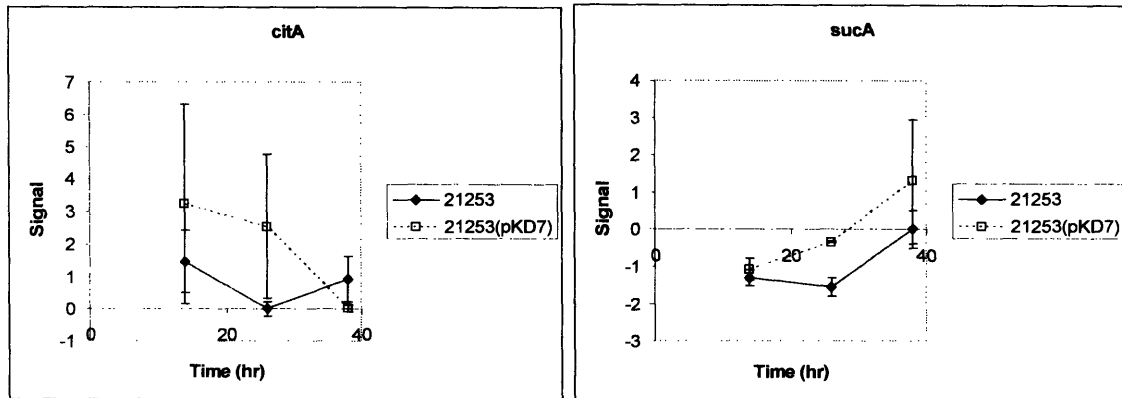
Figure 6-4: Transcriptional trends relative to culture phase and pyc over-expression for bioA, bioD, and bioB. Signal ratios in $\log_2$-scale are shown for the genes 7,8-diaminopelargonic acid aminotransferase (bioA), dethiobiotin synthase (bioD), and biotin synthase (bioB). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l lactate. Larger signal ratios indicate more transcript relative to the control RNA sample.

6.3.1.3 Biotin Synthesis Pathway

The relative abundances of mRNA for the three genes of the biotin biosynthesis pathway extant in *C. glutamicum* are given in the graphs of Figure 6-4. From the plots it can be seen that the transcriptional concentrations are higher in the pyruvate carboxylase over-expressing strain for all three genes during both biomass production and lysine production culture phases. Biotin is an essential cofactor for pyruvate carboxylase activity, with the dependence of the enzyme on biotin stronger than it is for many other carboxylases, such as phosphoenolpyruvate carboxylase. Evidence for this comes from an experiment in which a phosphoenolpyruvate carboxylase-deficient strain of *C.*

glutamicum grown on glucose was shown to have growth and amino acid production rates that suffered significantly more than in the wild-type strain from biotin limitation [Peters-Wendisch *et al.*, 1997]. However, due to the lack of the enzyme 7,8-diaminopelargonic acid synthetase, as was described in section 4.6.7, all known *C. glutamicum* strains are biotin auxotrophs unable to generate an internal supply of the cofactor. Nevertheless, along with the experiment of Chapter 4, this is the second study of the thesis work that found a positive correlation between the engineered over-expression of the pyruvate-carboxylase gene, and the over-expression of biotin synthesis genes due to some type of promotion or induction native to the cells. This adds more evidence to the theory that there is a link between the activity of carboxylases, and in particular pyruvate carboxylase, and the regulation of biotin formation.

Also, it is known that culture environments in which biotin is limited are more favorable for the excretion of glutamate [Takinami *et al.*, 1966]. Under these conditions, the activity of α-ketoglutarate dehydrogenase is reduced relative to that of glutamate dehydrogenase [Kawahara *et al.*, 1997], allowing an increase of flux to glutamate at the expense of flux along the tricarboxylic acid cycle to succinyl-coA and the lysine precursor, oxaloacetate. In this way the increase in the availability of biotin is indirectly and negatively correlated to the production of glutamate, and positively correlated to the production of lysine [Tosaka *et al.*, 1979]. Once again, although the transcription of the biotin synthesis genes can not actually lead to an increase in biotin levels within these cells, there is certainly evidence to suggest that a metabolic shift that increases the yield of lysine would be very compatible with one that increases the intracellular concentrations of biotin. Such a metabolic shift occurs, of course, as the cell cultures transition from the growth phase measured by the first microarray sample to the lysine production phase measured by the second microarray sample. Examining the trends of Figure 6-4 for the three genes in each of the two strains shows that in fact five of the six trends do show increasing mRNA levels from the first to the second data points, possibly relating a control strategy of the cells designed to increase biotin availability with the onset of increased lysine synthesis.
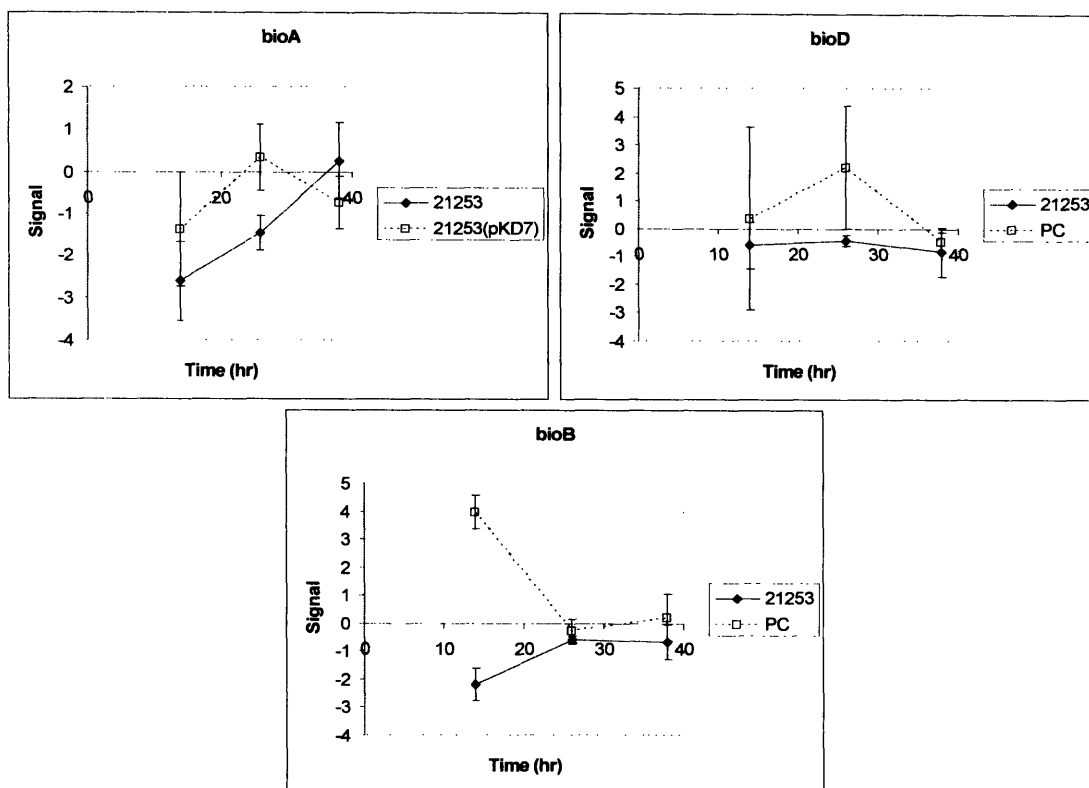
Figure 6-5: Transcriptional trends relative to culture phase and pyc over-expression for dapA, dapD, and lysE. Signal ratios in $\log_2$-scale are shown for the genes dihydrodipicolinate synthase (dapA), tetrahydrodipicolinate succinylase (dapD), and the lysine exporter (lysE). 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l lactate. Larger signal ratios indicate more transcript relative to the control RNA sample.

6.3.2 Genes Up-Regulated in Parental Strain

6.3.2.1 Lysine Synthesis Pathway

Figure 6-5 shows the transcriptional trends over time for the three genes of the lysine biosynthetic pathway—dihydrodipicolinate synthase (dapA), tetrahydrodipicolinate succinylase (dapD), and the lysine exporter (lysE)—that had significantly different transcriptional amounts between either the two strains being studied or the different time points assayed. For dapA and dapD, the parental 21253 strain exhibited higher mRNA abundances than the recombinant 211253(pKD7) strain during both the growth phase and

146

the lysine production phase, and for lysE, the parental strain had higher mRNA than the recombinant strain during the latter lysine production phase. From Figure 6-1, cultures of 21253 possessed significantly higher lysine yields than those of the 21253(pKD7), raising the possibility that the decrease in yields seen in the recombinant strain may be related to the lower transcript levels observed for these three genes encoding proteins involved in lysine synthesis and export.

For each of dapA, dapD, and lysE, higher transcript levels were observed in samples taken during lysine synthesis than in those taken during early growth. This pattern, seen in Figure 6-5, is equally true for samples taken from both strains, and is indicative of a regulatory response made once the inhibitory effects of threonine, which occur upstream to these three genes in the reaction network, are removed. In particular, mRNA for the lysine exporter gene shows a dramatic 14-fold increase in 21253, coincidentally the strain with the higher yield of lysine per cells, and this is the largest time-based increase of any of the genes represented on the microarray. Loos *et al.* (2001) also found that lysine exporter gene transcription increased, in their case by a factor of approximately 2-fold, upon the culture shifting from a growth phase to one involving both growth and lysine production. These findings agree well with those of the earlier exporter gene characterization, which demonstrated that the gene was regulated at the transcriptional level, with mRNA levels positively correlating with those of intracellular lysine [Vrljic *et al.*, 1996].

6.3.2.2 Other Carboxylases

Genes for four enzymes—dtsR1, dtsR2, accC, and accD—with an unconfirmed relationship to amino acid synthesis also exhibited interesting transcriptional behavior, depicted in the plots of Figure 6-6. The enzymes dtsR1 and dtsR2 are not known to be directly involved in the synthesis of amino acids, but their activity levels have been demonstrated to be influential in determining the productivities of glutamate [Kimura *et al.*, 1996, 1997, and 1999] and lysine [Loos *et al.*, 2001] achieved by a culture.
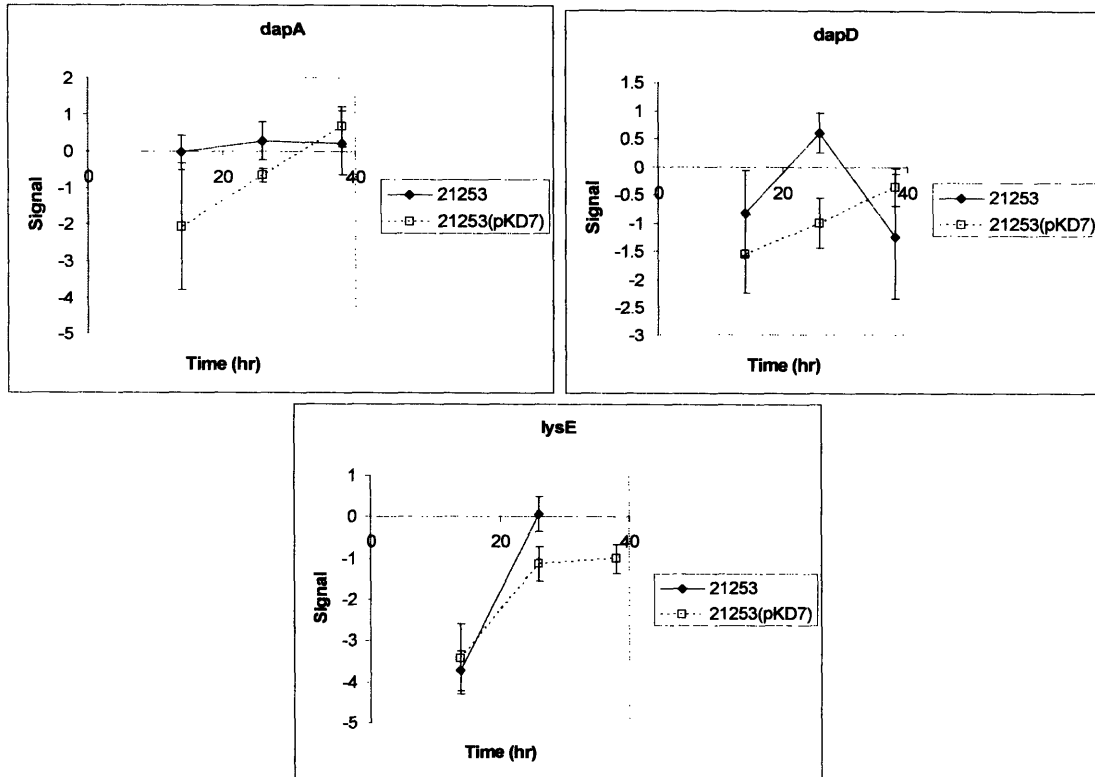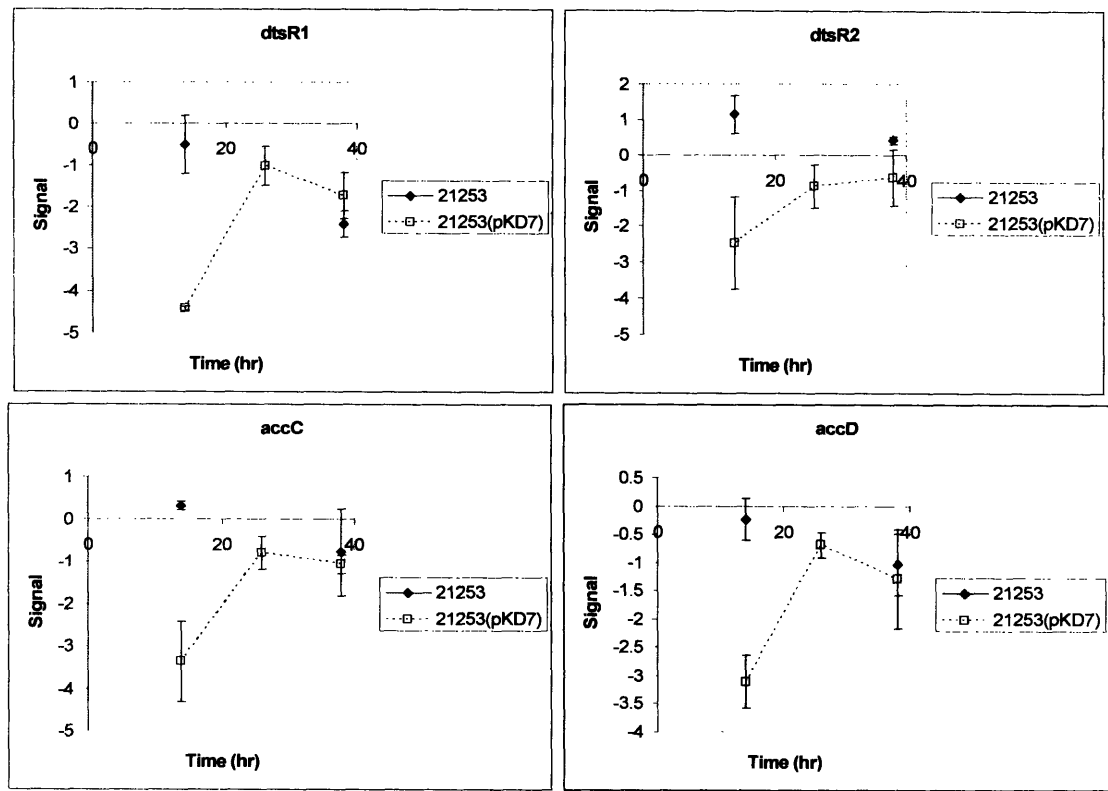
Figure 6-5: Transcriptional trends relative to culture phase and pyc over-expression for dtsR1, dtsR2, accC, and accD. Signal ratios in $\log_2$-scale are shown for the genes of the propionyl-CoA carboxylases dtsR1 and dtsR2 and the acyl-CoA carboxylase subunits accC and accD. 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l lactate. Larger signal ratios indicate more transcript relative to the control RNA sample.

Decreases in the activity of the dtsR1 protein have been shown to correlate well with the induction of glutamate overproduction in *C. glutamicum*, even under conditions of excess biotin which typically depresses glutamate synthesis [Kimura *et al.*, 1997]. Other ways to obviate the repression of glutamate production by biotin include the addition of detergents such as Tween 40 (polyoxyethylene sorbitan monopalmitate) [Duperray *et al.*, 1992] or β-lactam antibiotics such as penicillin [Nunheimeer *et al.*, 1970]. Initially it was believed that these effects of dtsR1 and dtsR2, as well as of the surfactants and antibiotics, on the excretion of glutamate were the result of alterations made to the cell wall composition and its properties such as permeability [Nakao *et al.*, 1970 and Demain

and Birnbaum, 1968] that are already known to occur as a result of biotin limitation [Shiio et al., 1962]. However, more recent evidence implicates the activity of the key metabolic branch point enzyme, α-ketoglutarate dehydrogenase, as the factor affected by these inducers and repressors [Kimura, 2002].

Proteins deduced from the sequence data of dtsR1 and dtsR2 suggest that the genes encode carboxyltransferase domains of acyl-CoA carboxylase, with significant similarity seen between their sequences and those of the β-chain of propionyl-CoA carboxylase [Kimura, 2002]. Further evidence linking dtsR1 and dtsR2 and propionyl-CoA carboxylases can be found in the significant up-regulation seen in the transcript abundances of the two genes in cultures grown on propionate as a carbon source [Hüser et al., 2003]. Two other enzymes, accC and accD, are significantly homologous with the biotin-binding subunit of known propionyl-CoA carboxylases [Wolfgang et al., 1996] and may themselves be involved in forming a complex with one or more of the dtsR enzymes and biotin [Kimura, 2002]. The accC gene product in particular is believed to be the α-subunit of the propionyl-CoA carboxylases [Jäger et al., 1996].

The over-expression of these putative acyl-coA carboxylases has been shown to coincide with diminishing production of glutamate. With less material being diverted from the tricarboxylic acid cycle into the pathways leading to excess glutamate synthesis, more is available to enter those of lysine biosynthesis. It is noteworthy, then that all four of these genes are present in higher transcript levels in samples taken during lysine production than in those taken during growth for the culture of 21253(pKD7). Note that microarrays hybridized with cDNA created from the RNA samples of 21253 during lysine production did not generate data of sufficient quality to be included in Figure 6-6. Also, the transcript concentrations are higher in the strain 21253, with its higher lysine productivity, than in strain 21253(pKD7) for RNA samples isolated from cultures during their growth phase. These findings are corroborated by the results of another microarray study that also found a positive relationship between increased lysine production and dtsR1 up-regulation [Loos et al., 2001].

Chapter 7: Simultaneous Measurement of Transcriptional and Flux Profiles

7.1 Overview

The results of the two experiments described above fit well with prior knowledge and proposed hypotheses describing the functioning of the metabolic network of *C. glutamicum*. The biological significance of these results, though, is not an absolute, because the data do not make an attempt to describe to what degree the enzymes being investigated are regulated through mRNA concentrations. As a result, although information of the type presented in this publication is certainly of great value, it must be remembered that in the absence of corroborating experiments, any correlations made between transcriptional measurements and the activities of enzymes and flux of metabolites is based on assumptions that need to be individually proven.

The use of induction and repression of gene transcription is the most important type of control that cells have over enzyme activities and metabolite flows present within their metabolic networks. This type of control is by far the most economical for the cells to employ, as no excess energy or materials are wasted in the translation of RNA into proteins that will not be utilized at their full potential activity. Because DNA microarrays directly measure the concentrations of transcripts that exist within the cells for given times and conditions, they are then an extremely powerful tool in deciphering the underlying explanations for macroscopic cell physiologies. Nonetheless, gene expression and the resulting mRNA abundances that are eventually assayed by DNA microarrays are only one method available to cells for influencing metabolic behavior.

The vast majority of cell functions occur through the activities of proteins, and not mRNA. Therefore the steps, which are not measured in any way by DNA microarrays, involved in creating a functioning protein from the starting material of an mRNA transcript have the potential to be very significant in determining the phenotype of a metabolic network. The mRNA sequence for a particular gene may be required to undergo some posttranscriptional modification to arrive at its useful form. Also, different mRNA species may have significantly different stabilities and degradation rates. Different factors may be involved in enhancing or inhibiting the initiation of DNA translation into a protein sequence through reverse transcription. As with the mRNA strand, this protein sequence may also be subject to posttranslational modification, such as the addition or removal of various functional groups or cofactors. The protein must then properly fold and may also be required to join with other subunits to form active enzyme complexes. Given this large opportunity for control to be exerted by the cells in each of these operations, the transcriptional data of microarrays should be considered neither a definitive indication of *in vivo* enzyme activities nor of actual reaction and pathway fluxes.

This potential for a significant quantitative disconnect between transcriptional data and the magnitudes of protein activities and material fluxes was demonstrated very well in recent investigations by Glanemann *et al.* (2003) of DNA microarray and enzymatic activity data for selected genes in *C. glutamicum*. In the study, the changes in mRNA levels for three enzymes catalyzing amino acid biosynthesis—homoserine dehydrogenase, threonine dehydratase, and homoserine kinase—were measured with DNA microarrays and competitive reverse transcriptase-polymerase chain reaction over a three-hour period. The two analytic methods gave results in good agreement with one another, but that showed significant discrepancies with activity measurements for two of the three enzymes. Lange *et al.* (2003) identified 13 proteins with reduced intracellular concentrations when *C. glutamicum* growth on a glucose-based medium was supplemented with valine. However, none of the mRNA species associated with these 13 genes showed any significant changes in concentrations when monitored with DNA microarrays. Work by Tian *et al.* (2004) found in comparing gene and protein expression in the model system of mice that measurements of the differential expression of mRNA

only accounted for approximately 40% of the differential expression of proteins both in steady-state cell lines and during dynamic processes. Similar experiments also compared the transcriptome and proteome in yeast [Ideker *et al.*, 2001 and Griffin *et al.*, 2002], *Halobacterium* [Baliga *et al.*, 2002], and human carcinomas [Chen *et al.*, 2002].

All of the above investigations described the correlation or lack thereof between mRNA abundances and protein concentrations or activities. The ultimate determination of the function of a metabolic network, though, is the flow of metabolites brought about by reactions catalyzed by these proteins. To better link the information gathered about the transcriptome of *C. glutamicum* in the work of this thesis with the actual metabolite flux through the reaction network of central carbon metabolism and lysine biosynthesis, an additional experiment was performed in which cultures of 21253 and 21253(pKD7) were grown on both glucose and lactate, and samples of these cultures were used to both collect transcriptional data and construct a metabolic flux map.

7.2 Flux Analyses Background

Metabolic flux is defined as a measure of the rate at which material is converted from one species to another as the result of metabolic network processes [Stephanopoulos, 1998]. Metabolic flux analysis is one of the most powerful tools of metabolic engineering, allowing researchers to arrive at a quantification of the amount of this material flowing through each of the reactions and pathways that comprise a given metabolic network. With this information, conclusions can be drawn concerning the location of key rate limiting steps and flux control points within the network, as well as the most promising targets for engineering to bring about strain improvement [Stephanopoulos, 1998; Bailey, 1998; and Nielsen, 1998].

Some of the earliest approaches to measuring the flux values associated with the reactions and pathways that make up a given metabolic network involved the use of stoichiometric metabolite models along with experimentally observable rates of product formation and substrate consumption and balances of reducing equivalents and ATP

[Vallino and Stephanopoulos, 1993]. These methods provided the first accurate flux maps for several culture and condition types, however the maps were incomplete as the quantifiable fluxes were limited to net values only.

The availability of quantitative flux data for use in metabolic flux analysis was greatly increased with the advent of technologies and protocols using isotopic labeling. By growing cultures on a medium that is continuously or discretely fed one or more ^{13}C-labeled substrates, cell samples that are harvested from the culture will have a labeling pattern in their intracellular metabolites that is indicative of the fluxes present in the metabolic network. Because these isotopic distributions found in intracellular compounds are independent of the concentrations of the compounds themselves, an additional source of information is realized, increasing the number of fluxes that are observable beyond what could otherwise be derived by using only metabolite balancing [Klapa, 2001]. This further enhancement to metabolic flux analysis, in which data from intracellular labeling patterns are added to those measuring metabolite balances, is often referred to as metabolic network analysis.

Because of the difficulties associated with accurately assaying intracellular metabolites, that are often present within the cells in very small quantities, a further improvement was made to this technique involving the measurement of labeling patterns of amino acids, rather than of other metabolites that are intermediates in the pathways of the network. These amino acids can most easily be obtained in sufficient quantities by recovering the proteinogenic constituents found in hydrolysates of biomass [Szyperski, 1995]. The isotopic distributions are measured by using either nuclear magnetic resonance spectroscopy [Zupke and Stephanopoulos, 1995] or gas chromatography and mass spectrometry (GC-MS) [Donato et al., 1993]. Of the two measurement methods, GC-MS is the more sensitive assay, allowing for higher precision in isotopic quantification. It was first used for analyzing a flux network by Christensen and Nielsen (1999 and 2000), has been applied extensively since then [Christiansen et al., 2002; Dauner et al., 2002; Fischer and Sauer, 2003; and Klapa et al., 2003] and is the technique employed in the experiment described in detail in this chapter.

Unlike samples that are used with DNA microarrays, samples to be used in current techniques of metabolic network analyses must be taken from cultures of cells that are

operating under steady-state conditions. This is because the analyses seek to construct a flux map through a combination of isotopic labeling and metabolite balancing, both of which only give a depiction of current *in vivo* conditions when physiological shifts are not occurring. Isotopic labeling experiments are typically carried out using continuous cultures, for which the patterns of labeling and rates of metabolite conversion exist in a steady-state by definition. Some recent studies have instead examined the isotopic distribution patterns and metabolite balances within batch cultures that are experiencing exponential growth [de Graaf *et al.*, 2000 and Pedersen *et al.*, 2000]. This is possible because it has been well-established that cells that are growing exponentially exist in a metabolic pseudo-steady state [Sauer *et al.*, 1999 and Hochuli *et al.*, 1999]. These batch cultures are much easier to operate experimentally, and are more representative of the bioreaction types typically found in industry.

The investigation presented in this chapter was carried out using cultures that were operated in batch mode as is described in the following section. In this experiment we also sought to find sampling times for which exponential growth and lysine production were simultaneously taking place, so that we could describe the behavior of the metabolic network as it functions for both of these purposes. Once the period of time overlapping both growth and product formation was identified, the existence of a metabolic pseudo-steady state was confirmed by examining the labeling patterns of amino acids found in hydrolysates of sampled biomass.

7.3 Experimental Design (in collaboration with Dr. Kohei Miyaoku)

To compare transcriptional and flux data collected simultaneously from the same set of samples, four cultures were grown. Two of these cultures were of the *C. glutamicum* strain 21253, and two were of the recombinant pyruvate carboxylase over-expressing strain 21253(pKD7). One culture for each of the two strains was grown on a minimal medium in which the carbon source was predominantly glucose and another was grown

154

on a minimal medium in which the carbon source was predominantly lactate. The media were composed of a slightly modified version of the recipe developed by Park (1996), as is described in section 4-1. Cultures of 21253(pKD7) were further supplemented with 50 mg/l kanamycin to select for cells with at least one functional copy of the plasmid conferring pyruvate carboxylase over-expression. The glucose-rich medium contained 20 g/l carbon source with the composition 67.5 % unlabeled glucose, 22.5% 1-^{13}C glucose, and 10% 3-^{13}C lactate. The lactate-rich medium also contained 20 g/l carbon source, but with the composition 90% 3-^{13}C lactate, 7.5% unlabeled glucose, and 2.5% 1-^{13}C glucose. All isotopically labeled substrates were obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). The cultures were grown in 250-ml round-bottom flasks containing 50 ml medium each that were incubated at 30°C and 250 rpm on a rotary shaker.

Samples with a volume of 1 ml were taken periodically for assaying the concentrations of biomass, glucose, lactate, and excreted amino acids and other metabolites. These samples were initially taken every 2-3 hr during the lag growth phase of the cultures. As the cultures approached mid-exponential growth, the sampling frequency was increased to once per hour. Biomass, glucose, and lactate concentrations were determined using the spectrophotometric and enzymatic assays described in section 5.1. Concentrations of the amino acids aspartate, glutamate, glutamine, glycine, threonine, alanine, valine, methionine, phenylalanine, leucine, and lysine were quantified with the use of the HPLC assay also described in section 5.1. Trehalose present in the sample supernatants was converted to glucose with the use of trehalose enzyme (Sigma-Aldrich Corp., St. Louis, MO) and then the extracellular glucose, lactate, ammonia, and acetate concentrations present in the culture samples were measured with the use of enzymatic assay kits (Roche-Biopharm, Inc., South Marshall, MI).

The pellets obtained from centrifuging the 1-ml samples taken during mid- to late-exponential growth were also used to isolate derivatized proteinogenic amino acids from biomass hydrolysates. These cell pellets were resuspended in 300 μl 6 M hydrochloric acid each and the resulting suspensions were placed inside vacuum tubes. A vacuum was applied to each tube, which was then incubated at 110°C for 20 hr. The solutions were next cooled, transferred to microfuge tubes, and dried with the use of a Vacufuge

concentrator (Brinkmann, Westbury, NY). The dry pellets were each resuspended in 100 µl of sterile de-ionized water and passed through a 0.2 µm polyvinylidene fluoride (PVDF) filter. The resulting solutions were dried again with the Vacufuge, and then each resuspended in 50 µl silylation-grade dimethyl formamide (DMF). To each sample was added 70 µl (N-methyl-N-[tert-butyldimethyl-silyl]trifluoroacetamide [MTBSTFA] + 1% tert-butyldimethyl-chlorosilane [TBDMCS]), and the mixtures were incubated at 70°C for 30 minutes. The DMF and (MTBSFA + 1% TBDMCS) were purchased from Pierce Biotechnology, Inc. (Rockford, IL). After this derivatization step, the isotopic distributions within the amino acids of the samples were assayed by GC-MS.

To complete the flux map construction, data from these labeling patterns were combined with measurements of the consumption or production of nutrients, extracellular products, and biomass. The software package, MNA-Flux 1.0 (Maciek Antoniewicz), was then used with the combined data to calculate the fluxes operating in the reactions of a metabolic network that had been predefined.

Also, samples with a volume of 15 ml were taken at two different time points with the goal of having one of these sample times occur within the metabolic and isotopic pseudo-steady-state so that its samples could be used in RNA isolation and microarray analyses of the transcriptional profiles of the two cultures. The fluorescently labeled cDNA created from each RNA sample was hybridized along with labeled cDNA created from a stock of control RNA to two microarrays. As in the previous experiments, this control RNA was obtained by pooling together all cultures of the experiment after the fermentation monitoring had completed, and isolating enough RNA from this pool to be used in all required microarrays.
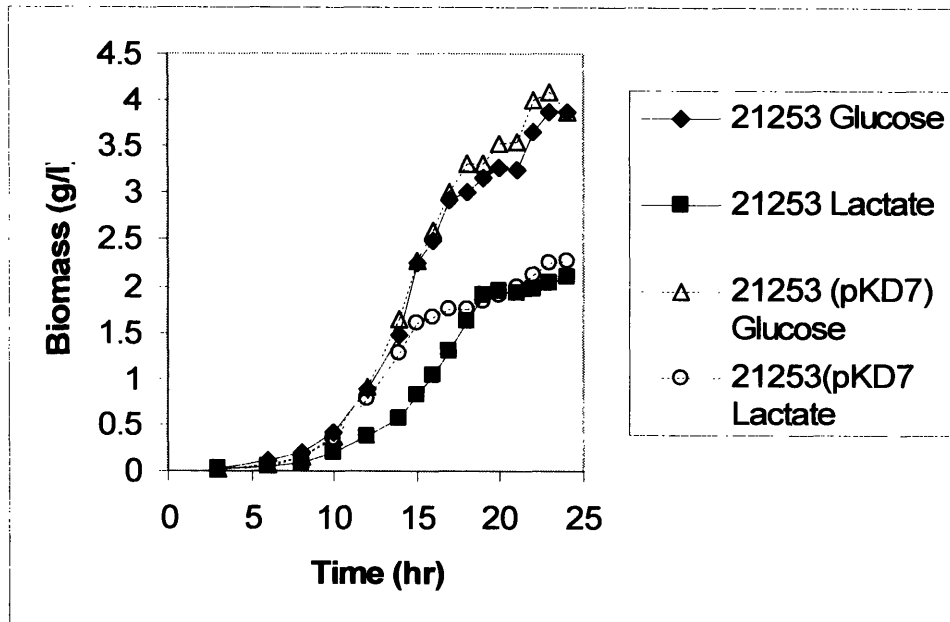
Figure 7-1: Growth trends. Strains 21253 and 21253(pKD7) were grown on minimal medium containing 20 g/l carbon source that was either 90% or 10% in the form of glucose, with the remainder being lactate. The former medium is referred to as glucose and the latter as lactate in the legend of this figure and all other figures and tables of this chapter.

7.4 Extracellular Measurements

7.4.1 Biomass Measurements

The growth curves for each of the four cultures of the experiment are shown in Figure 7-1. From the trends in the graph it can be seen that the exponential growth phase of the cultures occurred in the period from approximately 5 hours to 20 hours post-inoculation. The final biomass concentrations of the cultures are given in Table 7-1. From the results it can be seen that the 21253 and 21253(pKD7) cultures grown on the glucose-dominant medium possessed identical overall cell growth characteristics, with 3.2 g/l dry cell weight produced in 24 hr of culture growth. In contrast, when the strains are cultivated on the lactate-dominant medium, 21253 produces 50% less biomass, with a final concentration of 2.1 g/l, and 21253(pKD7) produces 40% less biomass, with a final

Table 7-1: Effect of carbon source on growth and lysine production by *C. glutamicum* **strains.** Cultures of 21253 and 21253(pKD7) were grown using minimal media with predominantly lactate or glucose as the 20 g/l carbon source. Data are shown for the overall specific lysine productivity (q_m) and the final cell concentration (X_f).

Strain	Glucose		Lactate	
	q_m (mmol lysine/ g cell)	X_f (g/l)	q_m (mmol lysine/ g cell)	X_f (g/l)
21253	2.2	3.2	2.8	2.1
21253 (pKD7)	2.4	3.2	3.6	2.3

concentration of 2.3 g/l. This agrees well with the data of Figure 5-1, in which the parental and recombinant strains show similar yields of biomass on carbon when cultivated on a medium containing glucose as the only available carbon source. In that experiment, when the cultures were instead grown on a medium with lactate as the only available carbon source, the biomass yield of the recombinant strain dropped more significantly than that of the parental strain.

7.4.2 Carbon Consumption Kinetics

Figure 7-2 displays the concentrations of glucose and lactate present in the broths of the four cultures grown in this study. As with the experiment described in Chapter 5, there are no dramatic differences between the rates with which the two different cultures utilize the two different substrates. There are, however, small distinctions among the consumption kinetics that fit well with the growth profiles of Figure 7-1. On average, the two strains have a faster major carbon source uptake rate when cultured on glucose than on lactate. This may very well be related to the higher yields of biomass on this substrate that were described in the previous section. Also, in the case of growth on the predominantly lactate medium, 21253(pKD7) had the higher lactate consumption rate of the two strains. Again, from the data on culture cell densities presented in Table 7-1, this was the same strain that produced a greater amount of biomass when lactate was the more available carbon source.
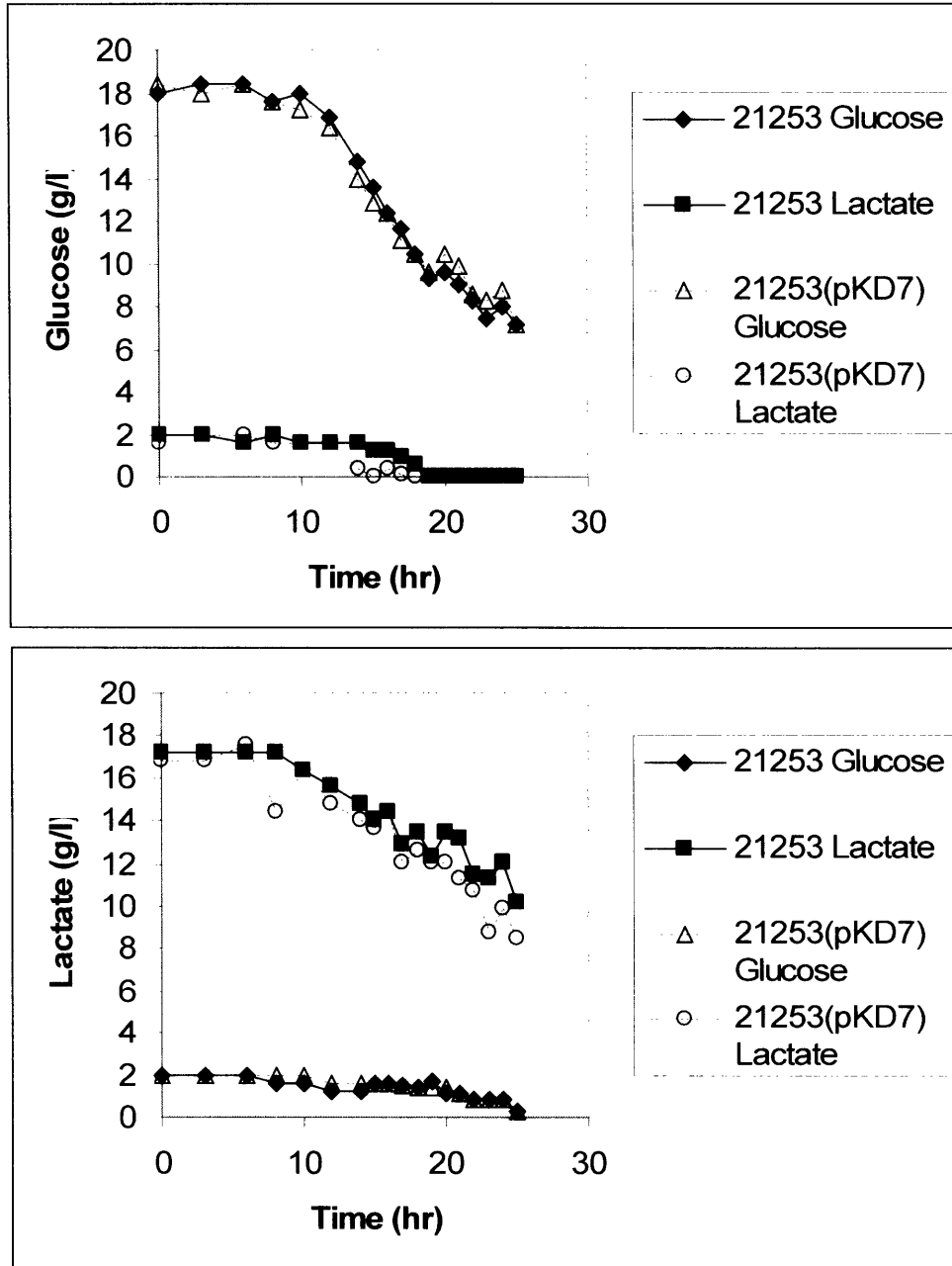
Figure 7-2: Glucose and lactate consumption. Cultures of 21253 and 21253(pKD7) were grown using minimal media with predominantly lactate or glucose as the 20 g/l carbon source. The trends in the graph show residual lactate and glucose during the fermentation

The profiles derived for the consumption of the glucose and lactate substrates also show similarity to the growth profiles in that they each share noticeable lag phases that each span the first 5-10 hours of the fermentation. After this point, the cultures enter their

exponential growth phase, and the carbon depletion rates increase. In each of the eight trends of Figure 7-2, the data for the period from 10 to 20 hours post-inoculation can easily be approximated by fitting with an exponentially decreasing equation to model the consumption kinetics with little error involved as a result of the estimation. This indicates that the glucose and lactate are being used by the metabolic network at more or less constant rates on a per cell basis as the concentration of cells exponentially increases, and adds to the validity of the assumption of a pseudo-steady-state during this time.

7.4.3 Extracellular Amino Acid Measurements

The amino acid concentrations measured in the culture broths using the HPLC assay are shown in the four plots in Figures 7-3 and 7-4. Although data was collected throughout the experiment, the graphs show only those points for the period in which the cultures had aged between 10 and 20 hours. This is because, from the growth, glucose consumption, and lactate consumption kinetics presented in the two previous sections, this is the time period in which the pseudo-steady-state assumptions may be applicable. Verification for this came from both the exponential growth rates of Figure 7-1 and the constant carbon source consumption kinetics of Figure 7-2 in this period. From the amino acid concentration data, lysine production began in each of the four cultures at approximately 16 hours post-inoculation. As a result, the fermentation times between 16 hours and 20 hours were the most significant, with cell culture growth, lysine formation, and potential pseudo-steady-states all occurring. Note that the trends for glycine, threonine, leucine, and phenylalanine are not shown in the graphs as no amount of these four amino acids were detected during this period of simultaneous exponential growth and lysine production.

As with the carbon source data of the previous section, exponential decreasing equations can be fit to the data for each of the studied amino acids, providing an estimate for the concentration kinetics of the compounds. The ideal case would result in concentration trends with exponentially increasing or decreasing slopes. These would be indicative of constant metabolite excretion or consumption rates on a per cell basis,

Figure 7-3: Extracellular amino acid concentrations for 21253 cultures. Cultures of 21253 were grown using minimal media with predominantly lactate or glucose as the 20 g/l carbon source. The trends in the graph show concentrations of amino acids present in the culture broths.
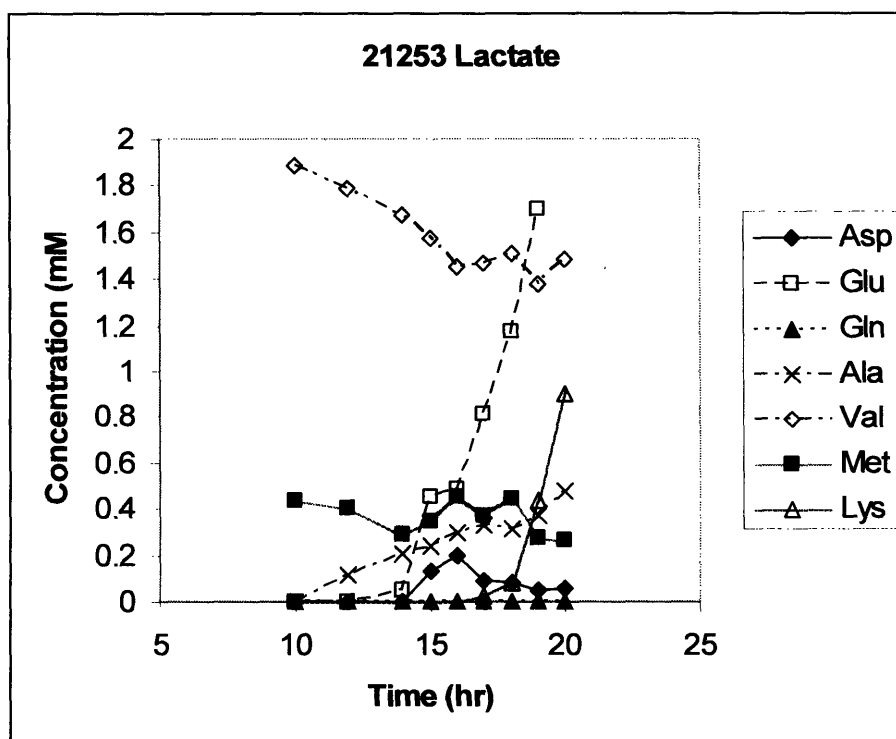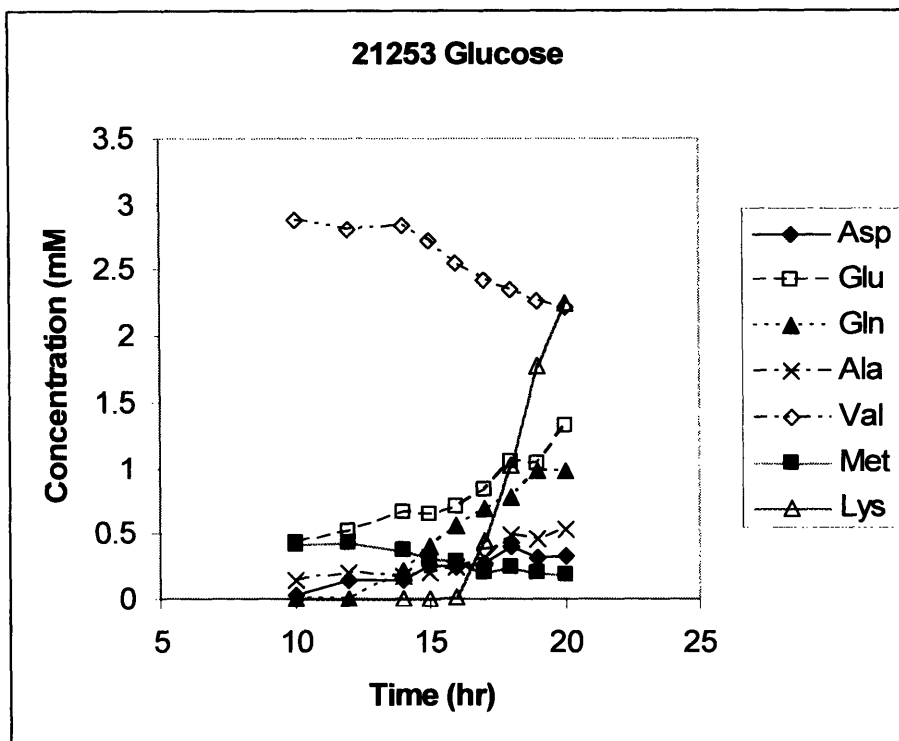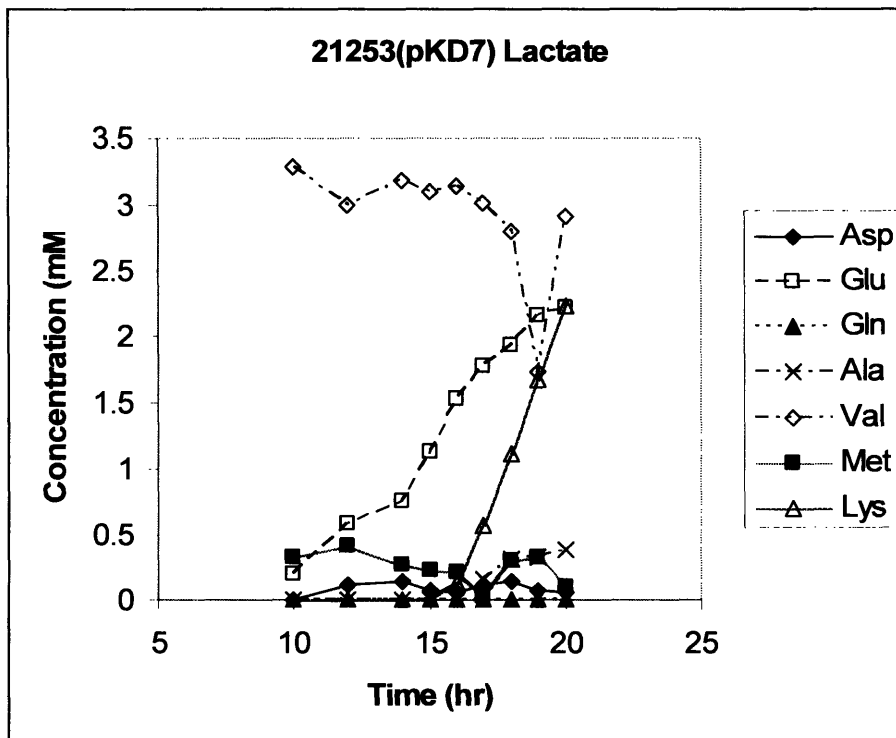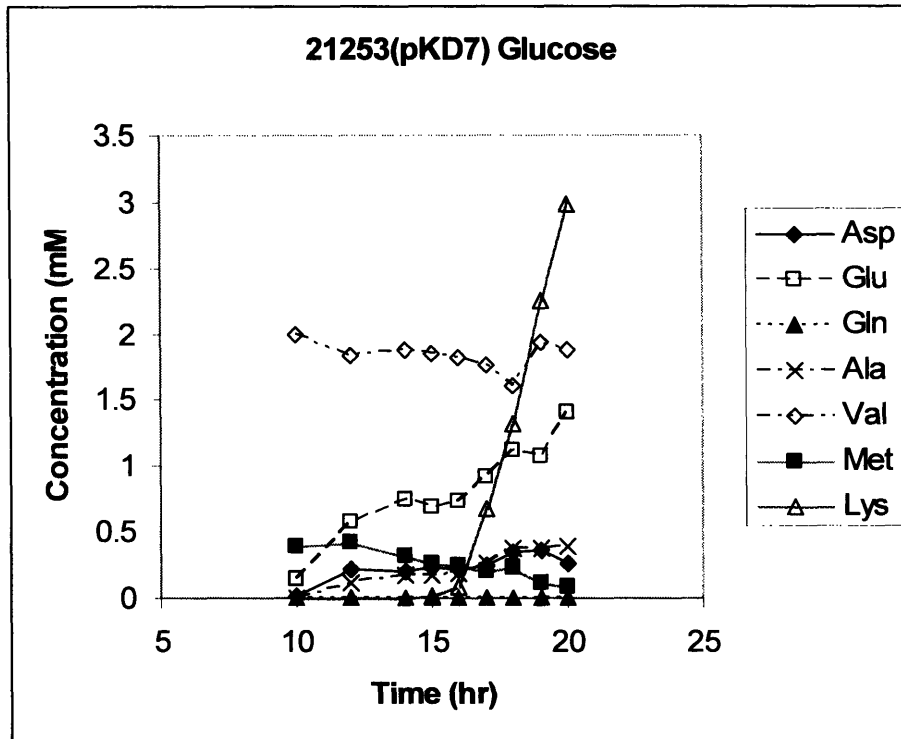
Figure 7-4: Extracellular amino acid concentrations for 21253(pKD7) cultures. Cultures of 21253(pKD7) were grown using minimal media with predominantly lactate or glucose as the 20 g/l carbon source. The trends in the graph show concentrations of amino acids present in the culture broths.

162

providing further validatory evidence for the existence of a pseudo-steady-state. Examination of the HPLC results shows that exponential fits can indeed be made for each of the data sets of Figures 7-3 and 7-4 without incurring significant error associated with the approximations.

Before discussing the measurements of intracellular isotopic labeling distribution patterns, mention should be made of the behavior of lysine production in the four different cultures. The lysine productivities observed during the 24 hours of fermentation that were monitored agreed in most cases with the findings of section 5.2.2. In that earlier experiment, both the 21253 and 21253(pKD7) strains possessed similar lysine yields on biomass when grown on a medium with glucose as the sole carbon source. In the current study, the yields for the two strains were within 10% of one another in the cultures using the predominantly glucose medium, as is shown in the data of Table 7-1. In both of the two investigations, the lysine yields were higher for both strains in media in which lactate was the more available source of carbon. However, while this increase in lysine yield was more prominent in the parental strain than in the transformed strain in the earlier experiment, as well as in the work by Koffas *et al.* (2002) described in section 2.1.4, with the study described here the pyruvate-carboxylase over-expressing strain showed a large 50% increase in lysine yield compared to a smaller 30% yield increase for the parental strain. One possible explanation for this is described in section 7.7.7, potentially accounting for the discrepancy in results that otherwise show very good agreement with earlier findings.

7.5 Identification of Isotopic Steady State

As was discussed previously in section 7.2, samples utilized in metabolic network analyses must be taken from cultures operating under steady-state or pseudo-steady-state conditions. This is because if there are any perturbations made to the system of a growing culture, then the culture may enter a transient phase during which time the isotopic labeling of compounds may gradually adjust to a new steady-state. Because new protein

synthesis is occurring during this transient phase, the labeling pattern of the proteinogenic amino acids of the biomass that are being analyzed will temporarily and partially reflect this transitional state. After the new steady-state has been reached, the transitional labeling patterns will eventually be replaced or diluted by turnover caused by new synthesis.

Representative trends of these labeling patterns, shown in Figure 7-5, demonstrate that the amino acids remained essentially isotopically unchanged during this time interval. Shown in the graphs are the mass fractions of the first two isotopic variants all of the analyzed amino acid fragments from the culture of 21253 grown on mostly glucose carbon source. In the plots, M refers to the molecular weight of the given amino acid fragment being analyzed by the GC-MS. The trends labeled M+0 and M+1 then represent the fraction of all forms of the amino acid fragment that are present in an unlabeled or mono-labeled form, respectively. It can be seen from the nearly constant values for all trends in the graphs that an isotopic steady-state was in fact in existence during the time period of interest. Also, Table 7-2 contains the values for the mass isotopomer fractions of each amino acid fragment after averaging the data from each sample of the pseudo-steady-state. From the table, the standard deviations and coefficients of variation of the fraction values are very low, verifying the relative consistency of the label distribution pattern over time. This confirmed that measurements of the mass isotopomer distribution patterns of this section and the extracellular measurements of section 7.4 would allow for the calculations of valid flux values for the reactions of the metabolic network. Also from section 7.4, DNA microarray results from the analysis of the RNA isolated from samples taken during this isotopic pseudo-steady-state would represent the simultaneous transcriptional responses of the cell cultures to both biomass and lysine production.
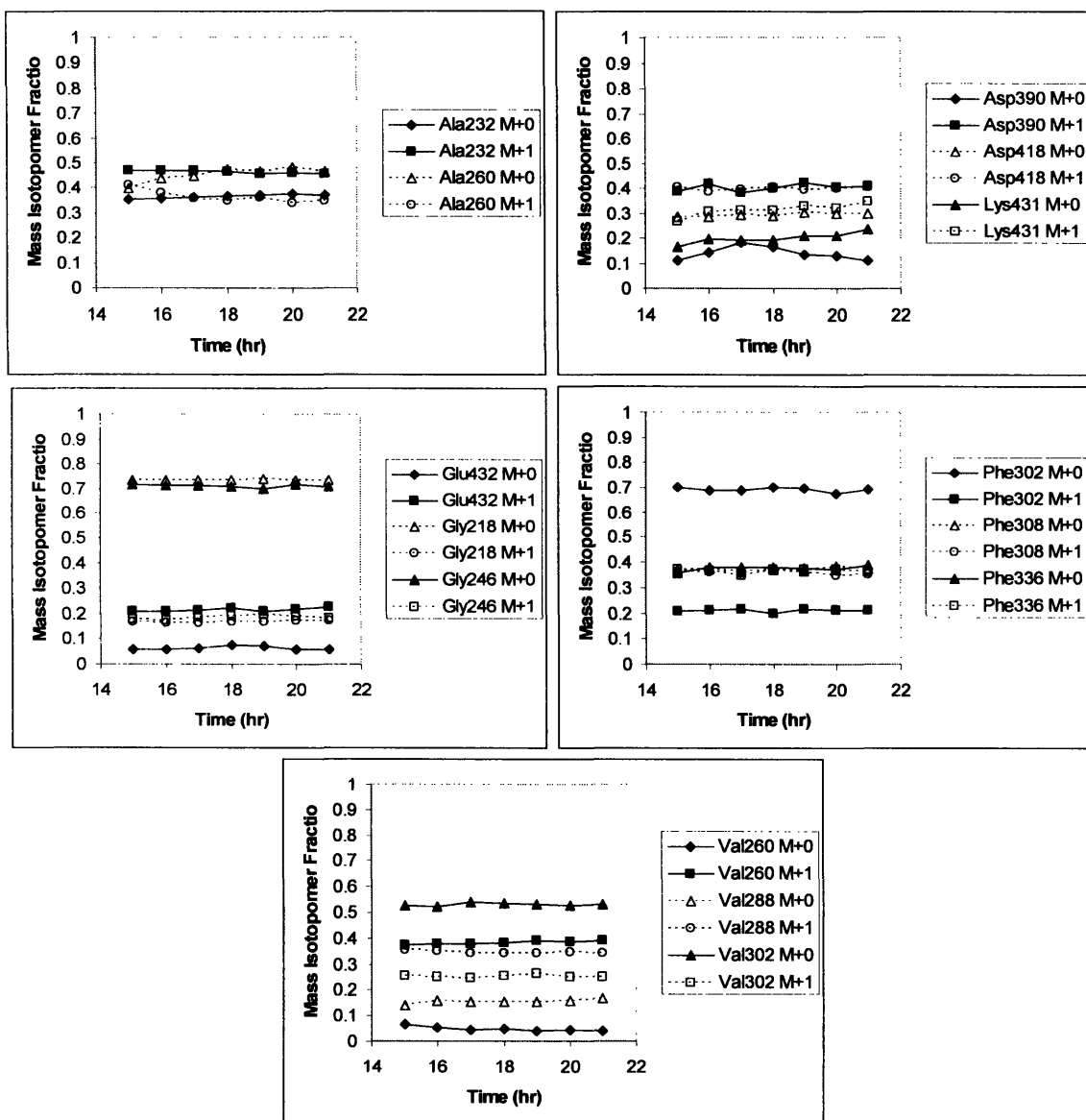
Figure 7-5: Time-dependence of isotopic labeling distribution of proteinogenic amino acids fragments in a pseudo-steady-state culture. A culture of 21253 was grown using a minimal medium with predominantly glucose as the 20 g/l carbon source. Biomass samples taken at 1-hr intervals were hydrolyzed into constituent amino acids, which were then derivatized and analyzed by GC-MS. Mass isotopomer fractions of the unlabeled (M+0) and mono-labeled (M+1) fragments of alanine (ala), aspartate (asp), lysine (lys), glutamate (glu), glycine (gly), phenylalanine (phe), and valine (val) of the indicated molecular weight are shown.

Table 7-2: Pseudo-steady-state mass isotopomer fractions of proteinogenic amino acid fragments. A culture of 21253 was grown using a minimal medium with predominantly glucose as the 20 g/l carbon source. Biomass samples were taken hourly from 15 to 21 hours post-inoculation and hydrolyzed into constituent amino acids, which were then derivatized and analyzed by GC-MS. Mass isotopomer fractions of the unlabeled (M+0), mono-labeled (M+1), etc. fragments of alanine (ala), aspartate (asp), lysine (lys), glutamate (glu), glycine (gly), phenylalanine (phe), and valine (val) of the indicated molecular weight are given after averaging across all time points.

Amino Acid Fragment	Mass Isotopomer	Fraction (%)	Standard Deviation (%)	Coefficient of Variation (%)
Ala232	M+0	36.47	0.90	2.48
	M+1	46.27	0.64	1.39
	M+2	11.93	0.34	2.82
	M+3	3.95	0.13	3.24
Ala260	M+0	45.22	2.79	6.18
	M+1	36.45	2.47	6.78
	M+2	9.88	0.62	6.31
	M+3	3.14	0.28	8.82
Asp390	M+0	13.88	2.60	18.74
	M+1	40.24	1.49	3.71
	M+2	29.84	1.36	4.54
	M+3	11.36	1.18	10.39
	M+4	3.31	0.58	17.51
	M+5	1.08	0.27	24.68
Asp418	M+0	29.25	0.63	2.14
	M+1	39.90	0.60	1.50
	M+2	14.18	0.77	5.42
	M+3	11.31	0.24	2.13
	M+4	3.83	0.20	5.15
Lys431	M+0	20.48	1.93	9.43
	M+1	31.58	2.02	6.38
	M+2	20.89	2.79	13.35
	M+3	16.30	1.49	9.15
	M+4	6.66	0.92	13.81
	M+5	1.58	0.65	41.12
Glu432	M+0	6.21	0.73	11.76
	M+1	21.44	0.72	3.35
	M+2	39.72	0.54	1.36
	M+3	21.28	0.22	1.05
	M+4	8.36	0.44	5.29
	M+5	2.29	0.23	9.99
Gly218	M+0	73.52	0.13	0.18
	M+1	16.78	0.31	1.82
	M+2	7.11	0.17	2.37

	M+3	1.09	0.14	13.08
Gly246	M+0	71.08	0.56	0.78
	M+1	18.44	0.73	3.97
	M+2	7.15	0.41	5.68
	M+3	1.25	0.20	15.85
Phe302	M+0	69.05	0.93	1.34
	M+1	20.96	0.62	2.95
	M+2	7.84	0.44	5.61
	M+3	1.32	0.29	21.81
Phe308	M+0	37.07	0.63	1.71
	M+1	35.98	0.59	1.65
	M+2	18.65	0.25	1.32
	M+3	6.48	0.39	6.09
	M+4	1.57	0.16	10.11
Phe336	M+0	37.50	0.92	2.46
	M+1	36.46	0.78	2.13
	M+2	17.73	0.91	5.11
	M+3	6.08	0.33	5.39
	M+4	1.68	0.25	15.02
Val260	M+0	4.80	0.86	18.02
	M+1	38.28	0.75	1.97
	M+2	40.38	0.19	0.48
	M+3	12.01	0.18	1.52
	M+4	3.83	0.09	2.43
Val288	M+0	15.47	0.77	4.96
	M+1	34.6	0.52	1.50
	M+2	35.04	0.30	0.86
	M+3	10.31	0.25	2.46
	M+4	3.41	0.19	5.47
Val302	M+0	53.03	0.57	1.07
	M+1	25.27	0.49	1.95
	M+2	16.27	0.74	4.57
	M+3	4.16	0.23	5.42

7.6 Estimation of Fluxome

7.6.1 Flux Observability and Variability

Perhaps the most important decision made in designing an isotopic tracer study such as the one carried out in this experiment involves the choice of labeling patterns used in substrates fed to the cultures of interest [Weichert and de Graaf, 1997]. The major goal in selecting which labeled substrates should be applied to culture growth, through either

continuous or shot-wise addition, is to maximize the mathematical observability of the fluxes present in the metabolic network, with another minor goal being to reduce the experimental cost associated with the use of these very expensive labeled compounds. In optimizing the composition of isotopically labeled substrates to maximize flux analysis precision, the three areas of concern are: (1) the precision of the measurements of metabolite isotopic distributions; (2) the degree of sensitivity of metabolite isotopic distributions to changes in flux values; and (3) the amount of substrate that has been labeled [Wittman and Heinzle, 2001].

This first area was dealt with in Table 7-2, in which the measurement errors associated with each amino acid fragment isotopic mass fraction were presented and found to be very minor, even when considered across multiple time points. In considering the second and third areas of concern, the optimal isotopomeric substrate composition for this experiment was determined by repeated computer simulations using the MNA software package, with the goal of reducing the propagation of errors in the estimations of several key flux values. These simulations demonstrated that the use of $[1\text{-}^{13}C]$-glucose and $[3\text{-}^{13}C]$-L-lactate against a background of unlabeled glucose allowed for the most accurate analysis of all of the fluxes involved in our modified metabolic network. The final optimized medium concentrations of the labeled substrates were given earlier in section 7.3.

Lactate was of course a necessary component of our medium because of our goal of investigating the effects of switching from a predominantly glucose medium to one that was primarily lactate. However, as in the work of Petersen *et al.* (2000), an additional benefit of the use of labeled lactate was as a means of producing a differently labeled intracellular pyruvate pool. This provided a way of discriminating between fluxes originating from PEP and those originating from pyruvate, because the fractional enrichments of the two chemical species would be otherwise identical. Unlike in the Petersen *et al* study, however, there was no benefit from the use of uniformly labeled $[^{13}C_6]$-glucose because of the network simplifications applied to the anaplerotic pathways involving oxaloacetate and malate.

Petersen *et al.* (2000) were also able to use metabolite balancing and ^{13}C NMR to quantify each of the individual parallel and bidirectional fluxes present in the key

anaplerotic node around oxaloacetate. However to do this they needed to disregard the possibility of flux taking place through the glyoxylate bypass pathway. They justified this simplification by noting that previous ^{13}C labeling studies had found the glyoxylate bypass to be inactive under the condition of growth on a medium with glucose as its carbon source [Marx, 1996 and 1999; Wendisch, 1997]. In their more detailed study of the oxaloacetate node, Petersen *et al.* used a medium containing 91% of its carbon in the form of glucose, allowing them to apply their simplification. In contrast, the experiment described in this chapter involves the use of a medium in which glucose is the minority carbon source, rendering this type of simplification invalid. Here, the complete estimation of all individual fluxes operating in both the glyoxylate bypass and the oxaloacetate node was not possible given the mathematical complexity of the reaction network and the number of intracellular and extracellular measurements available, and as a result an alternate model network was required.

The total set of biochemical reactions that were used to represent the functioning of the metabolic network, as well as the equation used to relate biomass measurements to the flux of central carbon metabolites and other observable amino acid products, are given in Appendix B. These reactions, which were entered into the MNA-Flux 1.0 software for use in metabolic network analysis calculations, are based in large part on work previously published by Vallino (1991), Jetten and Sinskey (1995), and Park (1997). Included in the reaction set are equations linking the formation of the compounds trehalose, glutamine, alanine, valine, acetate, and phenylalanine with intermediates of central carbon metabolism. The concentrations and isotopic distribution patterns of these compounds, which are measured by HPLC and enzymatic assays described in section 7.3, provide data that is crucial in increasing the observability and accuracy of the flux measurements. A rigorous mathematical analysis of the detailed relationship between the number and type of mass isotopomer measurements obtained and the number and type of solvable unknown fluxes is outside of the scope of this thesis but is covered very well by Klapa *et al.* (2003).

Because the objectives of this experiment involved among other areas the branch point between the tricarboxylic acid cycle and the glyoxylate bypass pathway located at isocitrate dehydrogenase, one network simplification used by the previously mentioned

studies could not be used. This was the combining of several reactions of the TCA cycle into one representative reaction converting acetyl-CoA and oxaloacetate into isocitrate. These other studies, like that of Petersen *et al.* (2000), also did not consider the two reactions of the glyoxylate bypass pathway, so two additional equations were added to the reaction set. The simplifications that were inherited from the metabolic network model were: (1) the combination of all pentose 5-phosphate compounds, such as ribose 5-phosphate and xylulose 5-phosphate, together as a single species; (2) the treatment of the reactions catalyzed by pyruvate and phosphoenolpyruvate carboxylases as being reversible, when in actuality the reverse flux of these reactions are catalyzed by the distinct enzymes, oxaloacetate decarboxylase and phosphoenolpyruvate carboxykinase, respectively; (3) the lumping of flux from the tricarboxylic acid cycle to the Embden-Meyerhof-Parnas pathway, as catalyzed by both the malic enzyme and oxaloacetate decarboxylase, into one flux acting as the reverse of pyruvate carboxylase; (4) the reduction of the four-reaction succinylation pathway of lysine biosynthesis to a two-reaction pathway involving only the starting and ending materials and one intermediate; and (5) the collapsing of the sequences of reactions involved in the formation of trehalose, acetate, valine, and phenylalanine into one reaction for each metabolite, summarizing the overall carbon balances of its anabolic pathway. With this metabolic network model, the MNA software was used to calculate a data set of fluxes that minimized the difference between simulated and experimentally determined values for all measurements of extracellular product formation, intracellular mass isotopomer fractions, the formation of biomass, and the consumption of glucose and lactate.

7.6.2 Flux Maps

The model metabolic network, and the associated flux values for each of the four cultures, is shown in Figures 7.6. In each case, the fluxes were normalized so that the total carbon uptake rate would be fixed at 600 mol carbon per hour. Together, the four overlaid flux maps represent the metabolic behaviors of each of the two strains in response to growth on each of the two carbon sources.

PEP Gluc
CO2
R1
R11
Pyr G6P Ru5P Ru5P F6P
R25 R2 R12 R14
trehal
F6P S7P GAP
R3 R13
FBP F6P E4P GAP
R4
GAP DHAP DAHP CO2
R6 R5 R39 R40 Phe
BPG PEP
R7
3PG
R8 Pyr Glu OAC
2PG R29 R31
R9 Ala AKG Asp
PEP R33
R10 Aspald Pyr
Lac R34
CO2 Pyr R28
R15 R35
CO2 CO2 H4D
R23 R24 R32 R36
AcCoA Ac m-DAP LL-DAP
R16 R37
OAC Cit R38
R22 R41 CO2
R43 Lys
Mal
R21 Glyox IsoCit
Fum R17 CO2 Pyr Pyr
R42 AKG Glu Gln R30
R26 R27 Val CO2
R18
R20 Suc SucCoA CO2
R19

21253 on 10% lactate
21253 on 90% lactate

21253(pKD7) on 90% lactate

Figure 7-6: Flux maps. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either 10% or 90% in the form of lactate, with the remainder being glucose. The reactions comprising the model metabolic network shown are described in greater detail in Appendix B. The net forward flux values present in each of the four cultures are shown for each reaction. The order in which the flux values are displayed from top to bottom for each reaction is shown in the legend.

7.7 Comparison of Flux and Transcriptional Changes

7.7.1 Definition of Comparison Metric

DNA microarrays were used to measure the mRNA concentrations within the cell samples taken during the metabolic pseudo-steady state that occurred in each culture. These concentrations were then compared to the fluxes that were calculated for the reactions catalyzed by the enzymes encoded by the corresponding mRNA sequences. To assist in this comparison, for each mRNA/flux pair, the results were normalized to a percentage of the maximum value determined from among the four cultures studied. In this way, we could see the extent to which the up-regulation and down-regulation of transcript concentrations was related to increases and decreases in the metabolite flux. In total, 28 reactions in the network were associated with genes represented on the microarray.

As an example of this normalization method and a definition of the comparison metric used to relate the transcriptional and flux values, Figure 7-7 presents data for the reaction converting acetyl-CoA and oxaloacetate to citrate through the catalysis of the enzyme citrate synthase. The abundance of the mRNA for the citrate synthase gene is highest in the case of 21253(pKD7) cultivated on the predominantly lactate medium. This same strain grown on the glucose-rich medium had approximately 35% less amount of the transcript, as is shown in the upper graph of the figure. Similarly for the flux values, the highest material flux through this reaction was seen in 21253(pKD7) grown on primarily lactate, while when the strain was cultivated on mostly glucose, a 50% lower flux value was measured. The difference of 15% between the 35% change that occurs in transcription and the 50% change that occurs in flux between these two cultures is used as the y-value of the point shown in lower graph of Figure 7-7. In an identical fashion, the x-value of 5% is determined by calculating the difference between the transcript change and flux change in the two cultures of the parental strain 21253. A mathematical representation of this comparison is shown below.

172
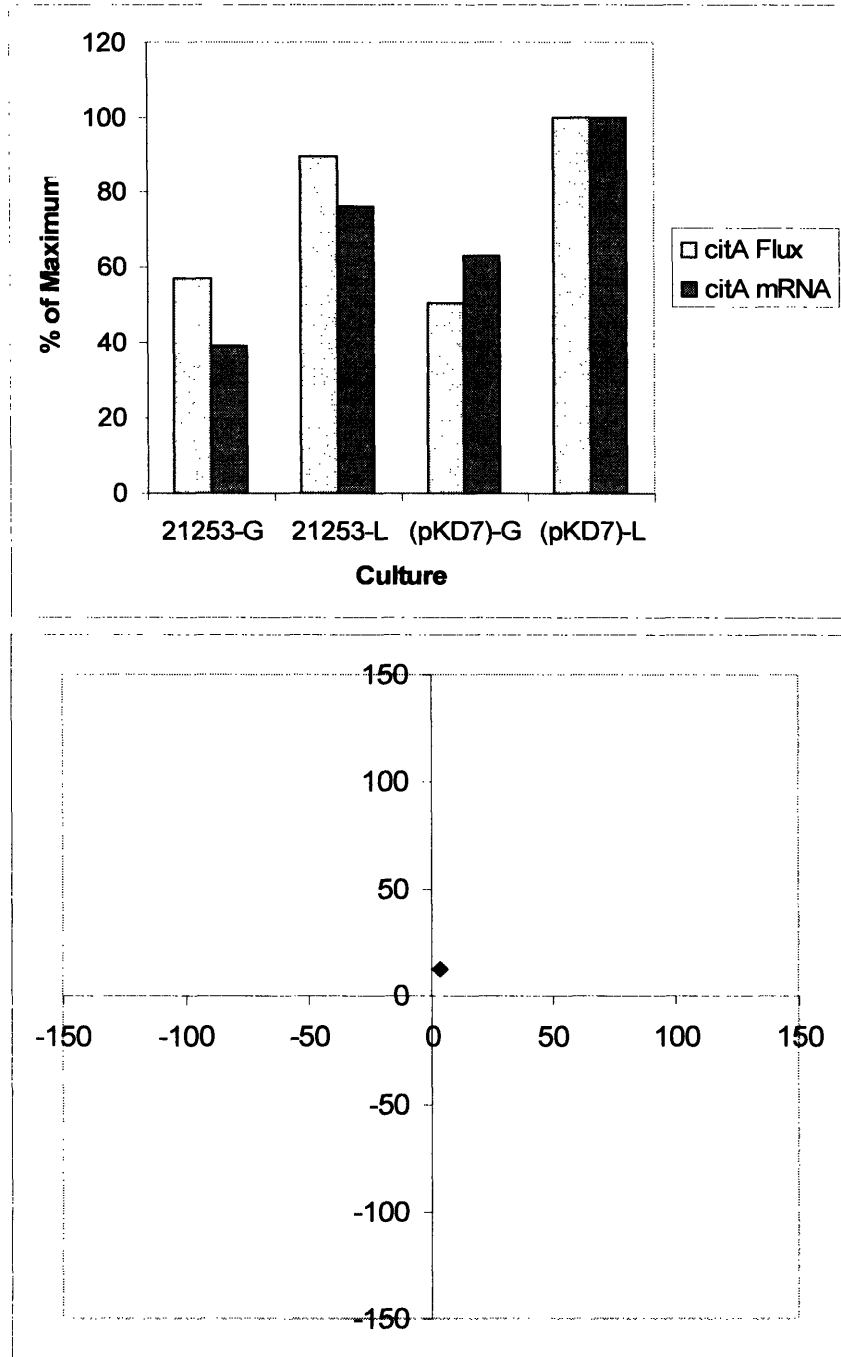
Figure 7-7: Example of comparison between transcriptional and flux responses. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The upper graph shows the relative mRNA concentrations and flux values for the reaction catalyzed by citrate synthase (citA) in the four cultures. The lower graph shows the same data in the form of a comparison metric defined in the text.

$$\Delta F_{12} = \text{Flux 1} - \text{Flux 2}$$

$$\Delta T_{12} = \text{Transcript 1} - \text{Transcript 2}$$

$$x = \frac{\Delta F_{12} \left| \Delta T_{12} \right|}{\Delta T_{12} \left| \Delta F_{12} \right|} \left| \Delta F_{12} - \Delta T_{12} \right|$$

$$y = \frac{\Delta F_{34} \left| \Delta T_{34} \right|}{\Delta T_{34} \left| \Delta F_{34} \right|} \left| \Delta F_{34} - \Delta T_{34} \right|$$

$$1 = 21253 \text{ on glucose}$$

$$2 = 21253 \text{ on lactate}$$

$$3 = 21253(\text{pKD7}) \text{ on glucose}$$

$$4 = 21253(\text{pKD7}) \text{ on lactate}$$

The choice of defining the flux and transcript differentials as measuring the changes between identical strains grown on different media is arbitrary, and could easily have been written as measuring instead the changes between different strains grown on identical media. In this way, the metric would be directly examining the effect of pyruvate carboxylase over-expression rather than the effect of carbon source availability. The fractions contained within the equations for the graph x- and y-values are present to indicate the similarity or difference between the directionality of changes that take place. For example, if the flux values for a given reaction increase from one culture to another, while the related mRNA abundances decrease, the value calculated by this equation will be negative in sign. In contrast, if both the flux and mRNA simultaneously increase or decrease, the calculated metric value will be positive. Because of this sign convention, flux/mRNA pairs that rise or fall in tandem will plot within the upper right quadrant of a graph of the type shown in Figure 7-7. Also, the closer the correlation is between transcriptional and flux responses, the closer the data point for the pair will be to the origin.
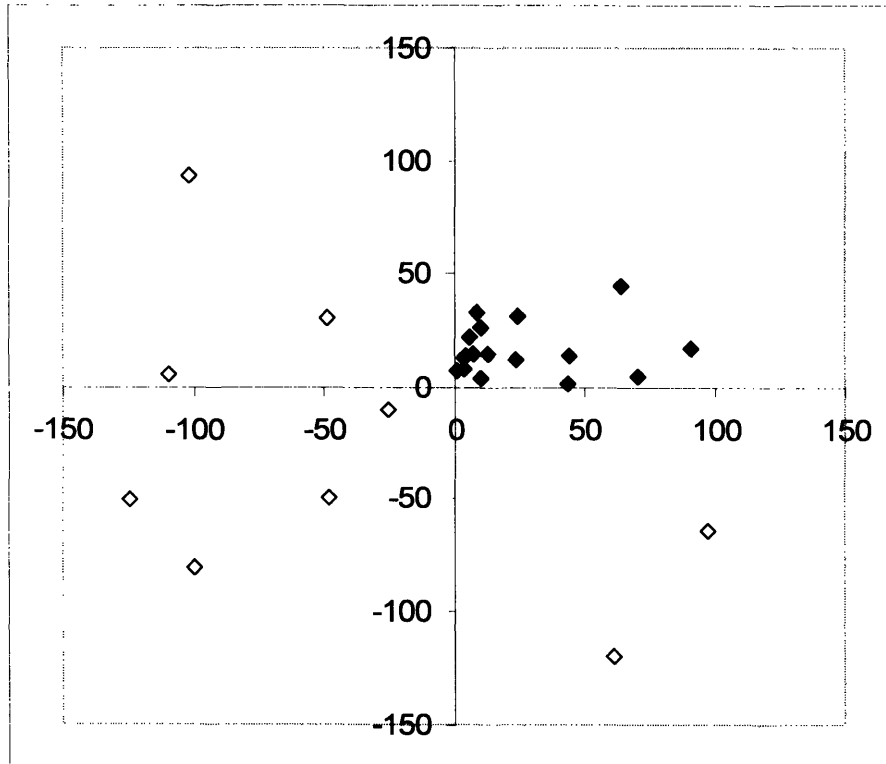
Figure 7-8: Comparison of all measured flux/mRNA pairs. The data for each related pair of flux and transcriptional data sets are plotted according to the comparison metric described in section 7.7.1. The points with closed symbols have good correlation between flux and transcriptional responses, while those with open symbols do not possess such a correlation.

7.7.2 Genes with Weak Correlation between Flux and Transcript Concentration

Of the 28 pairs of flux magnitudes and mRNA levels quantified by this experiment, nine of them, or approximately 1/3 of the total, were found to not have significant correlations between the two types of measurements. These sets of data were found to have large discrepancies and opposite directionalities in the behavior of the transcriptome and the fluxome to changes in the medium carbon source for one or both of the tested *C. glutamicum* strains. In the graph of Figure 7-8, which depicts the comparison between the

Figure 7-9: Weak correlations between transcriptional and flux responses I. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by phosphoglucoisomerase (pgi), phosphoglyceromutase (gpm), and fructose bisphosphate aldolase (fba) in the four cultures.

flux and transcript changes as described in section 7.7.1, the pairs with weak correlations are shown in the three quadrants other than the upper right.

The data for three of these nine cases—reactions catalyzed by phosphoglucoisomerase (pgi), phosphoglyceromutase (gpm), and fructose bisphosphate aldolase (fba)—are shown in Figure 7-9. These reactions are each involved in glycolysis and gluconeogenesis. As was expected, the fluxes decreased for both strains as the glucose carbon source was replaced with lactate. This is because, as was discussed in section 4.3, the importance of the Embden-Meyerhof-Parnas pathway to the cells decreases as the lower glucose availability forces a shift from a glycolytic mode to a gluconeogenic mode. The mRNA concentrations for these genes, though, stayed relatively constant among all

176

four cases tested. In the experiment described in chapter 4 studying carbon source effects on gene transcription in the *C. glutamicum* E12 strain, mRNA levels for these genes decreased in the case of a lactate medium, although the decreases were the smallest measured for the Embden-Meyerhof-Parnas pathway genes, as can be seen from Table 4-3. In the results from the experiment of chapter 5, shown in Figures 5-8, 5-9, and 5-10, as the percentage of lactate relative to glucose in the growth medium increased, the transcription concentrations of gpm dropped relatively steadily. The mRNA levels of pgi, though, increased in response to identical changes, while those of fba showed no significant trend. Together, the results from these three experiments suggest that the flux through the reactions catalyzed by these three genes is not likely to be controlled by the cells through manipulations of gene transcription.

Another three of the nine genes—α-ketogluterate dehydrogenase (sucA), succinyl-coA synthase (sucC), and succinate dehydrogenase (sdhA)—are for tricarboxylic acid cycle enzymes that catalyze sequentially related reactions. Data for these genes and their reactions are given in Figure 7-10. In strain E12, of the these three tricarboxylic acid genes, only sdhA showed significant up-regulation for growth on lactate in the data of Table 4-3, and its up-regulation was the least of the significant genes of that experiment. In Figures 5-12 and 5-13, the mRNA levels of sucC increased with increasing lactate content in the growth medium, while those of sucA and sdhA did not.

It should be emphasized that the results for these nine pairs of genes and reactions should not be viewed negatively, and do not imply that the quality of the data provided by either the DNA microarrays or the metabolic flux analysis are suspect. Instead, these findings provide information about the behavior of the metabolism of the cells that is of equal importance to that given by results in which the transcription and flux correlate more closely. The data from these nine flux/mRNA pairs clearly imply that the metabolite flows through the reactions catalyzed by the genes of the glycolytic pathway and tricarboxylic acid cycle discussed above, as well as those involving pyruvate kinase, aspartate aminotransferase, and dihydrodipicolinate reductase, as shown in Figure 7-11, may be primarily regulated at a level other than transcript concentration, such as that of mRNA translation, protein stability, or enzyme activity.
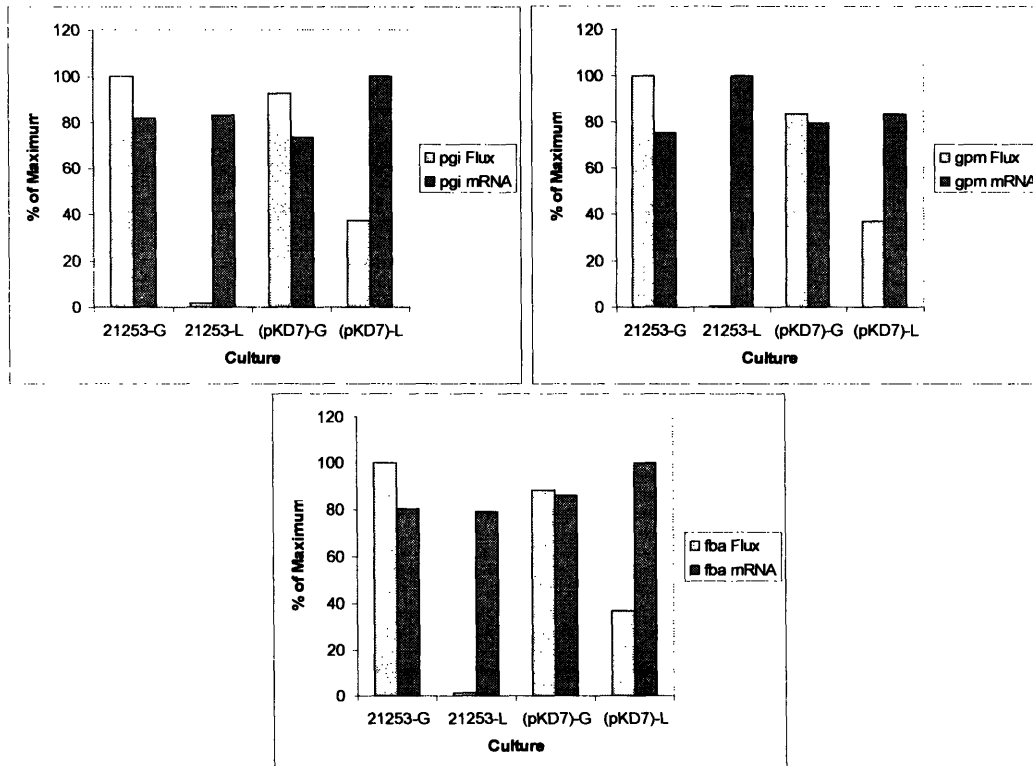
Figure 7-10: Weak correlations between transcriptional and flux responses II. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by α-ketogluterate dehydrogenase (sucA), succinyl-CoA synthase (sucC), and succinate dehydrogenase (sdhA) in the four cultures.

7.7.3 Embden-Meyerhof-Parnas Genes

Although three of the glycolytic reactions did not show a close similarity between mRNA concentration and metabolic flux, the remaining five—phosphofructokinase (pfk), triosephosphate isomerase (tpi), glyceraldehyde 3-phosphate dehydrogenase (gap), phosphoglycerate kinase (pgk), and enolase (eno)—very much did so. For these reactions, the data of which are given by Figure 7-12, again the flux magnitudes showed very significant drops for both strains when growth was based primarily on lactate. Here, however, the drops in flux magnitude were matched by those for the values of the related transcript amounts, although the flux changes were more dramatic than those of the
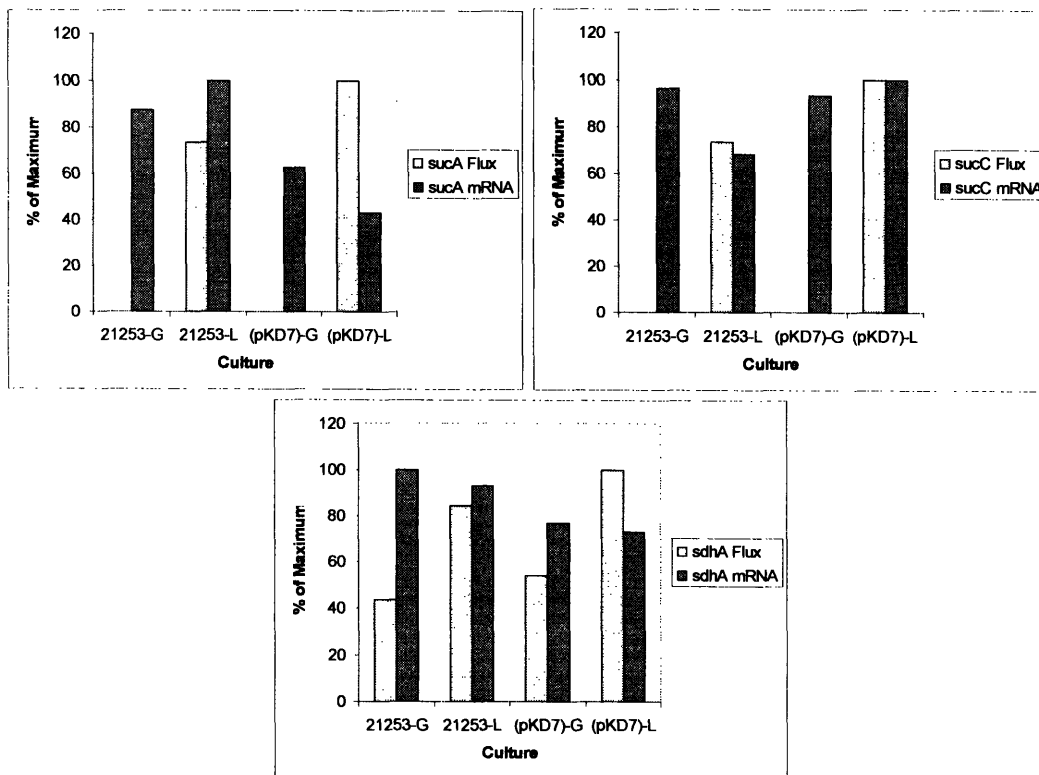
178
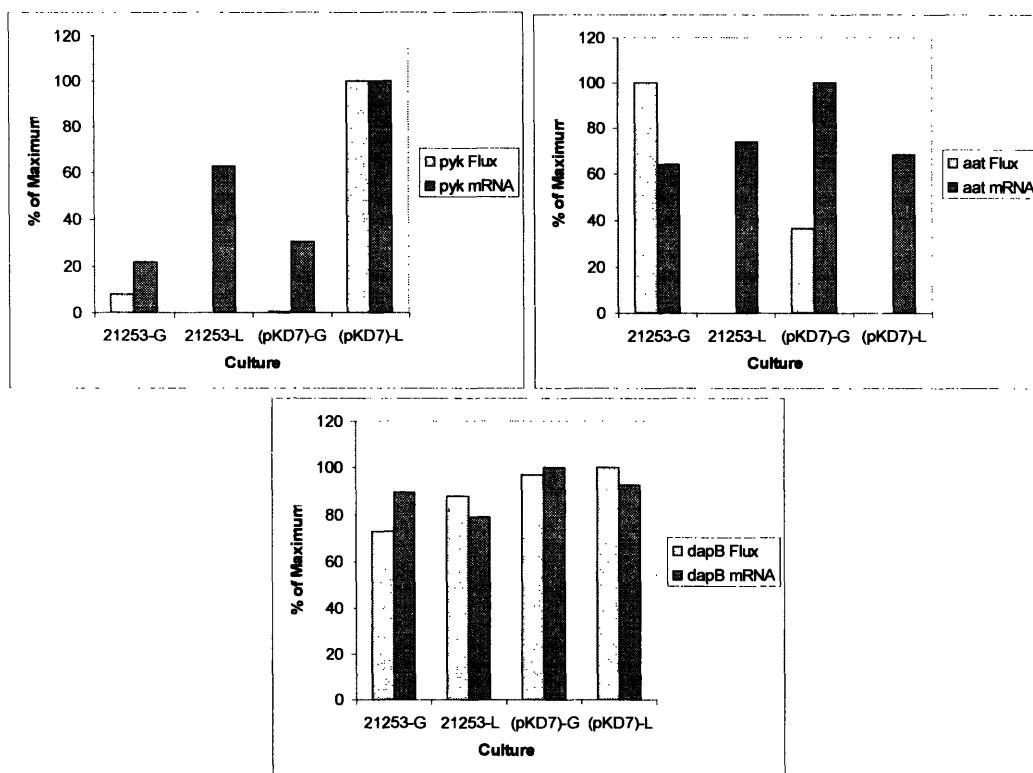
Figure 7-11: Weak correlations between transcriptional and flux responses III. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by pyruvate kinase (pyk), aspartate aminotransferase (aat), and dihydrodipicolinate reductase (dapB) in the four cultures.

transcript, possibly suggesting that there is a minimal baseline amount of mRNA of the enzyme genes always present in the cells, even in the case of negligible flux through the pathway. For all five reactions, the flux values were highest in the case of 21253 grown on primarily glucose and lowest for the same strain grown on primarily lactate. The cultures of 21253(pKD7) also saw declining Embden-Meyerhof-Parnas pathway flux associated with the change from a predominantly glucose to a predominantly lactate growth medium. While the transcriptional changes did not exactly match the flux changes associated in contrasting the behavior of the two strains cultivated on the same medium, the two types of data sets were in closer agreement in describing the more significant

Figure 7-12: Correlations between transcriptional and flux responses for the Embden-Meyerhof-Parnas pathway. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by phosphofructokinase (pfk), triosephosphate isomerase (tpi), glyceraldehyde 3-phosphate dehydrogenase (gap), phosphoglycerate kinase (pgk), and enolase (eno) in the four cultures.

changes associated with the differing behavior of the same strain cultivated on different media.
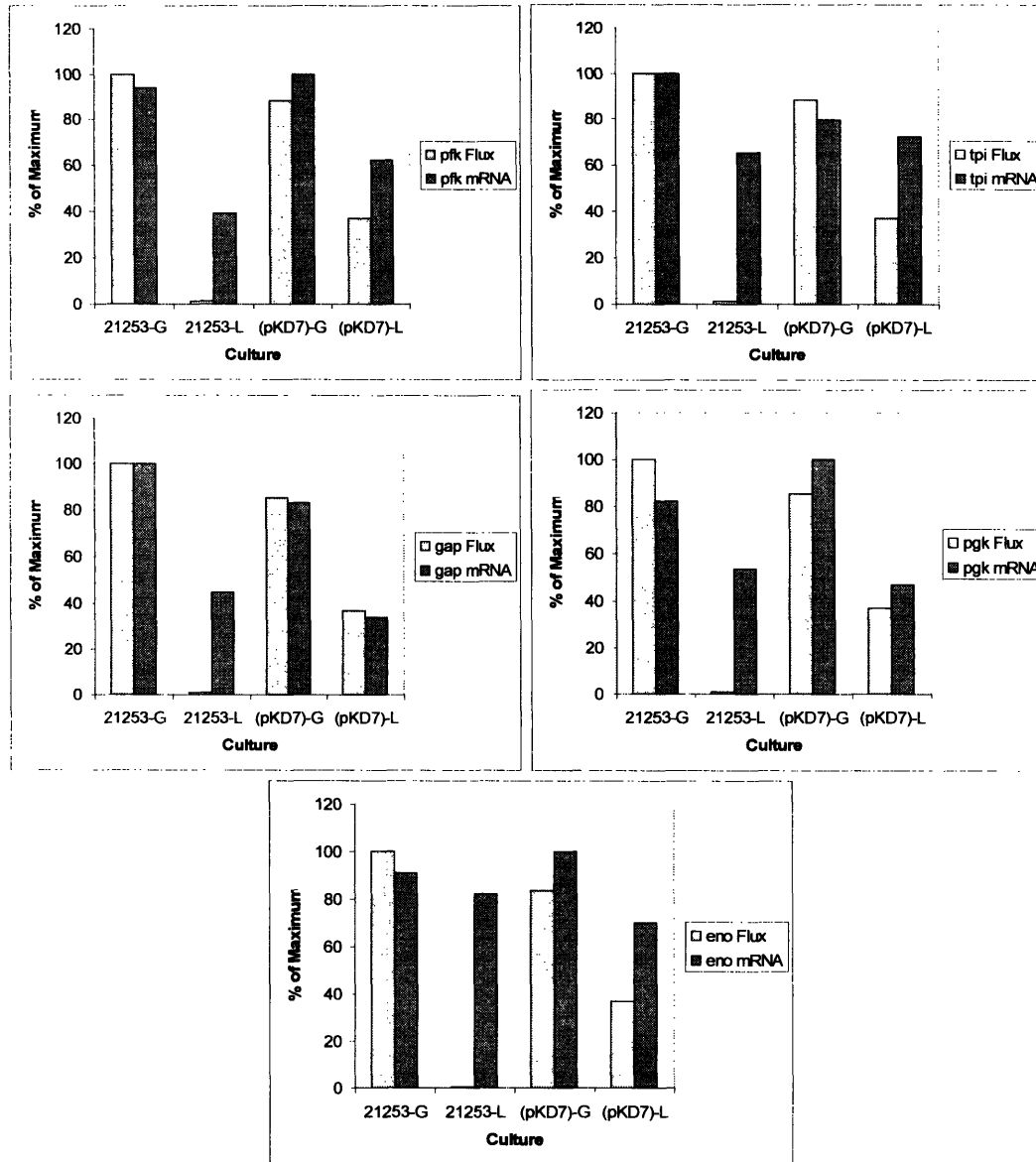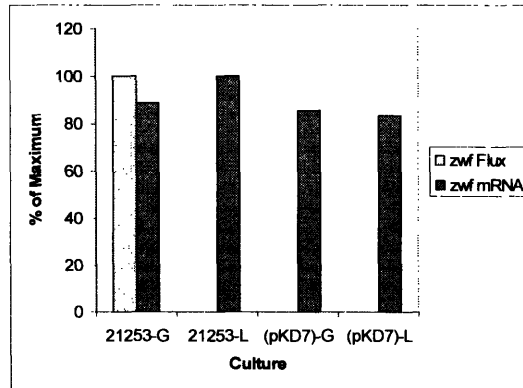
Figure 7-13: Correlations between transcriptional and flux responses for the pentose phosphate pathway. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reaction catalyzed by glucose 6-phosphate dehydrogenase in the four cultures.

7.7.4 Pentose Phosphate Pathway Genes

The pentose phosphate pathway, which branches off of the glycolytic pathway, would be expected to also show a drop in carbon flow as glucose is replaced with lactate, which enters the metabolic network downstream of either pathway. Instead, as is shown in Figure 7-6, the flow entering the pentose phosphate pathway remains relatively unchanged among the four cultures. This constant flow is matched by an equally steady level of the transcript of the glucose 6-phosphate dehydrogenase gene, shown in Figure 7-13, the product of which catalyzes the reaction at the entry point into the pathway. This matches the general pattern of results relating to the pentose phosphate pathway in the experiments previously described in sections 4.6.2 and 5.4.3, in which little or no change is observed in the transcript levels of the enzyme genes for the pathway.

7.7.5 Tricarboxylic Acid Cycle and Glyoxylate Bypass Genes

In contrast to the declining fluxes and transcript levels of glycolysis, for four of the genes involved in the tricarboxylic acid cycle—aconitase (aco), citrate synthase (citA), isocitrate dehydrogenase (icd), and fumarase (fum)—increases were observed as lactate was used as the available carbon source. These patterns are presented in Figure 7-14. In these cases, a shift from glucose to lactate as the major carbon source available in the medium coincides with an increased need for the cells to rely on the tricarboxylic acid cycle to produce energy and reducing equivalents that could otherwise be generated from glycolysis and the pentose phosphate pathway. A more detailed discussion of this metabolic behavior is given in section 4.6.3. Also, the increases in flux associated with the change in growth medium were magnified in the case of cultures of the pyruvate carboxylase over-expressing strain. Other genes for which closely agreeing increases in flux and transcript levels could be seen for both strains upon growth on lactate were isocitrate lyase (aceA) and malate synthase (aceB) of the glyoxylate bypass, also shown in Figure 7-14. This pathway experiences increases in flux and gene transcription because of its use in replenishing intracellular pools of malate and oxaloacetate that are consumed in the increased TCA cycle flux.

7.7.6 Pyruvate/Oxaloacetate Node Genes

Other genes for which closely agreeing increases in flux and transcript levels could be seen for both strains upon growth on lactate were pyruvate dehydrogenase (pdh) and pyruvate carboxylase (pyc), both of which direct the flow of carbon away from the pyruvate entry point of lactate into central carbon metabolism, as well as phosphoenolpyruvate carboxykinase (pck), which allows carbon to flow from the tricarboxylic acid cycle into the gluconeogenic pathway for the purpose of forming sugar phosphates that are used as cellular building blocks. The transcriptional and flux behavior of these three sets of genes and reactions are shown in Figure 7-15. Similar increases for
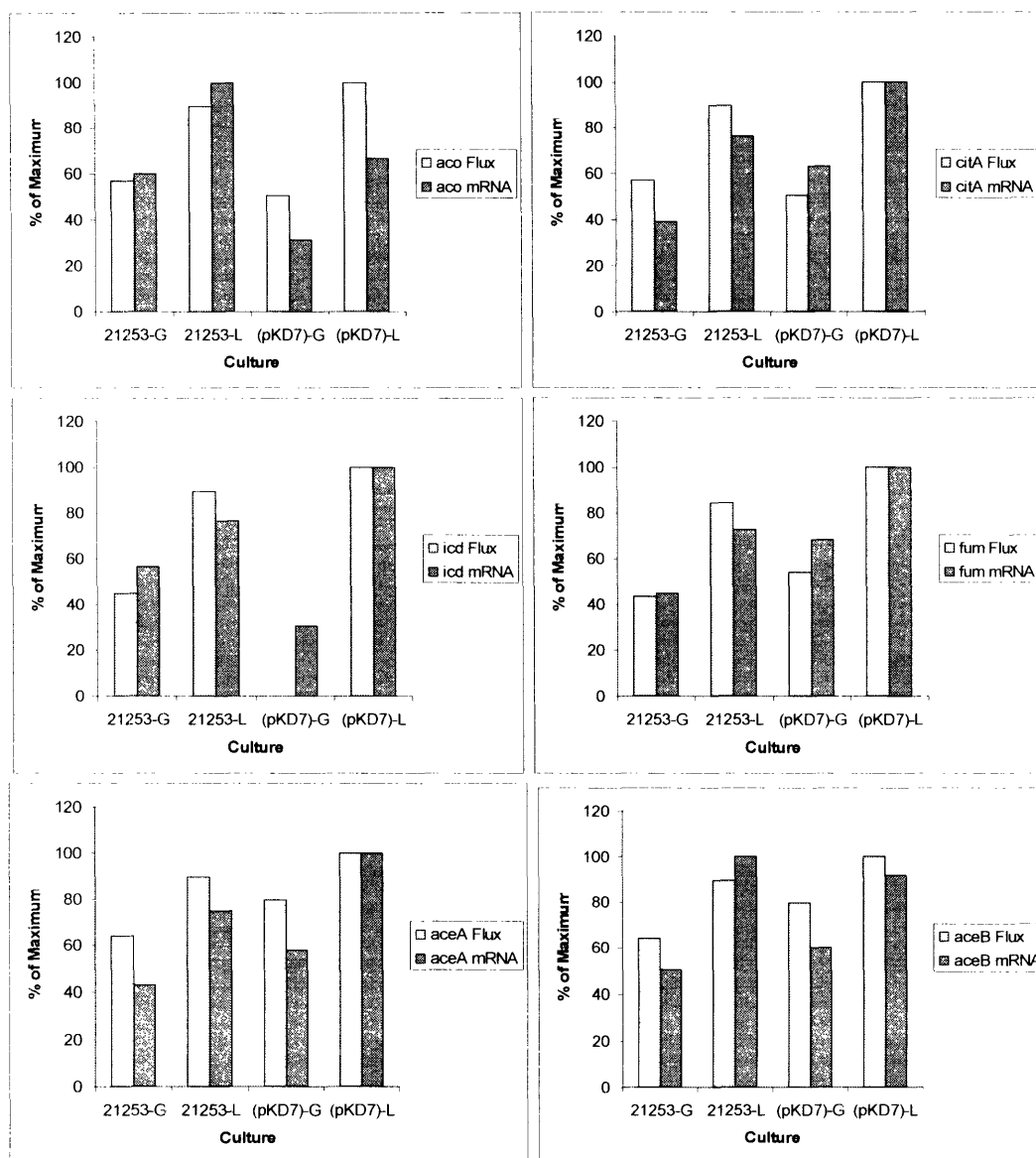
Figure 7-14: Correlations between transcriptional and flux responses for the tricarboxylic acid cycle and glyoxylate bypass pathway. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by aconitase (aco), citrate synthase (citA), isocitrate dehydrogenase (icd), fumarase (fum), isocitrate lyase (aceA), and malate synthase (aceB) in the four cultures.

the mRNA abundances of pck in cultures grown on lactate were seen in Table 4-3 and Figure 5-16, and of pdh were seen in Figure 5-17.
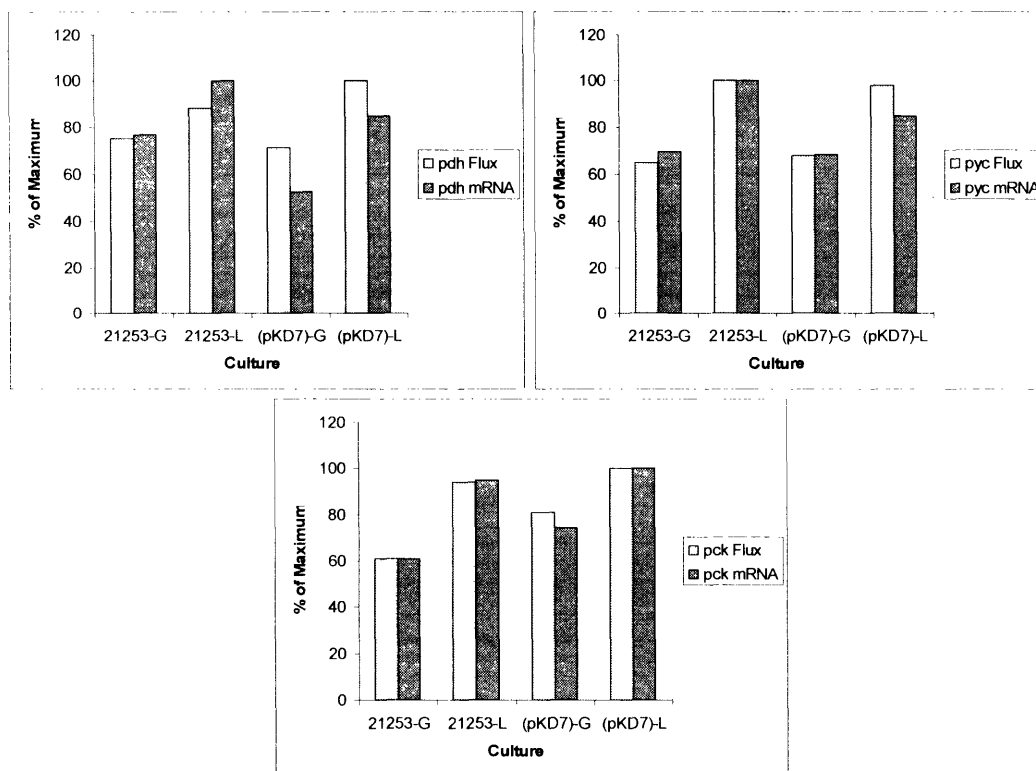
Figure 7-15: Correlations between transcriptional and flux responses for the pyruvate/oxaloacetate node. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by pyruvate dehydrogenase (pdh), pyruvate carboxylase (pyc), and phosphoenolpyruvate carboxykinase (pck) in the four cultures.

7.7.7 Lysine Biosynthesis Genes

Four of the five steps in the reaction network that leads from oxaloacetate to lysine show trends, depicted in Figure 7-16, that were among those for which the metabolite fluxes and gene transcript concentrations agreed with one another very well. These include the first and last reactions of the lysine biosynthetic pathway, catalyzed by aspartokinase (ask) and diaminopimelate decarboxylase (lysA), respectively, and reactions in parallel branches of the split portion of the pathway, catalyzed by succinyl-diaminopimelate decarboxylase (dapE) and meso-diaminopimelate dehydrogenase (ddh).
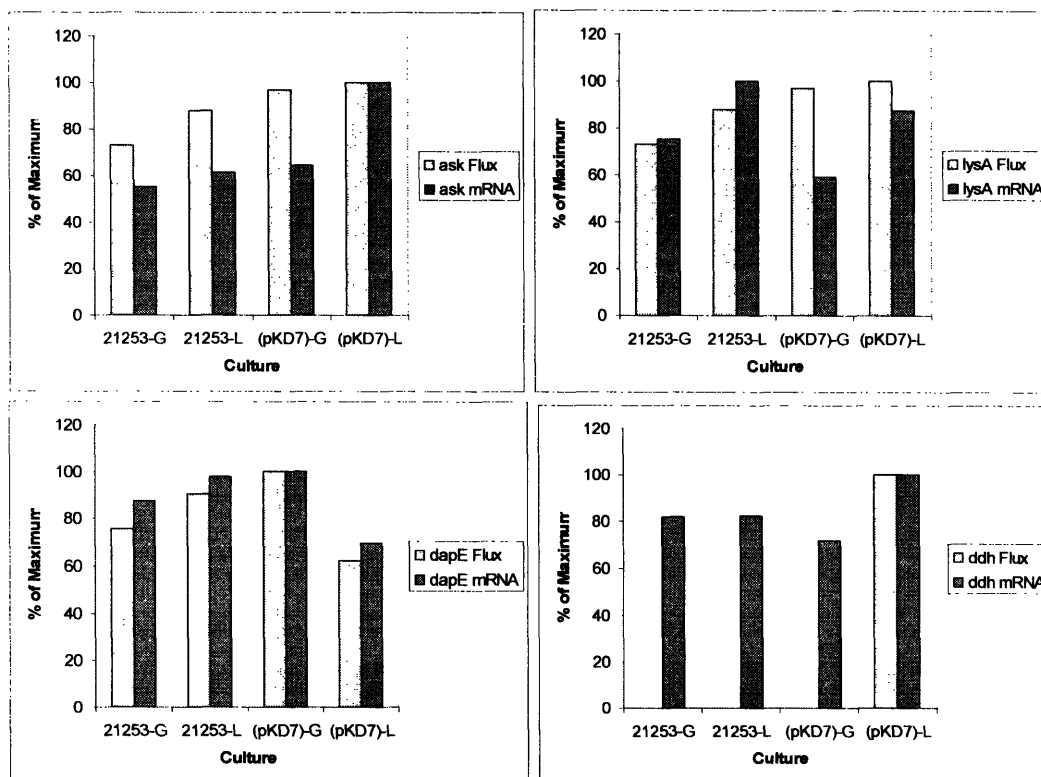
Figure 7-16: Correlations between transcriptional and flux responses for the lysine biosynthesis pathway. Strains 21253 and 21253(pKD7) were grown on minimal media containing 20 g/l carbon source that was either predominantly glucose or lactate. The graphs show the relative mRNA concentrations and flux values for the reactions catalyzed by aspartokinase (ask), diaminopimelate decarboxylase (lysA), succinyl-diaminopimelate desuccinylase (dapE), and meso-diaminopimelate dehydrogenase (ddh) in the four cultures.

For the reactions of ask and lysA, growth on lactate caused the flux and enzyme gene transcription to increase for both strains. This coincided with increased lysine productivities that were seen under this condition and described in Table 7-1. However, in the previous experimental results of Figure 5-20, the mRNA levels of ask, and of the transcriptionally linked gene for aspartate semialdehyde dehydrogenase (asd), decreased steadily as the growth media contained an increasing percentage of its carbon source in the form of lactate. That ask was up-regulated, rather than down-regulated, in the predominantly lactate cultures could offer an explanation for the otherwise inconsistent lysine yield data of Table 7-1. Typically, when the strain that has been engineered for the over-expression of pyruvate carboxylase is cultivated on lactate, the resulting yield of

lysine on biomass is less than that seen with the parental strain cultivated under identical conditions. This was the case in, for example, Koffas et al. (2002) and section 5.2.2 of this thesis. The likely explanation for this is that an increased supply of oxaloacetate from pyruvate through the action of the pyruvate carboxylase can not be sufficiently used as a substrate for lysine synthesis because of the low native activity of the feedback inhibited ask enzyme for which oxaloacetate is a substrate. In this case, the oxaloacetate is instead consumed by the tricarboxylic acid cycle, where it serves to support additional growth at the expense of additional lysine production. This bottleneck at the point of ask was relieved in another study by the coordinated over-expression of both the pyc and ask genes, resulting in enhanced lysine yields [Koffas et al., 2003]. A similar phenomenon may be occurring here, with the ask expression levels and flux throughput increasing unexpectedly along with an equally unexpected increase in lysine productivity.

The other very interesting result from the measurements of the lysine biosynthesis flux/mRNA pairs involved the branched portion of the pathway. In the succinylation route, catalysis is carried out by the product of the dapE gene. In the dehydrogenation route, the relevant gene is ddh. For three of the four cultures, the fluxes and mRNA abundances of these two pairs of reactions and genes remain fairly unchanged. For the culture of 21253(pKD7), though, an increase in both flux and mRNA titer is seen in the ddh reaction, along with a corresponding decreases in the alternative dapE pathway. One possible cause of this may be that the supply of the succinyl-CoA required by the succinylation route is low in the culture in which 21253(pKD7) is grown on mostly lactate. The reason for this is that this culture has the highest tricarboxylic acid cycle flux among the four grown for this experiment, as was shown in Figure 7-16, possibly consuming the succinyl-CoA in that pathway before it can be applied to lysine synthesis. This suggests both some conditions for which a switch in the flow of metabolites within the network occurs, and that the switch takes place as a result of regulation at the level of gene transcription. This type of regulatory action is a good example of how DNA microarray analyses provide metabolically meaningful insights into the operation of intracellular networks
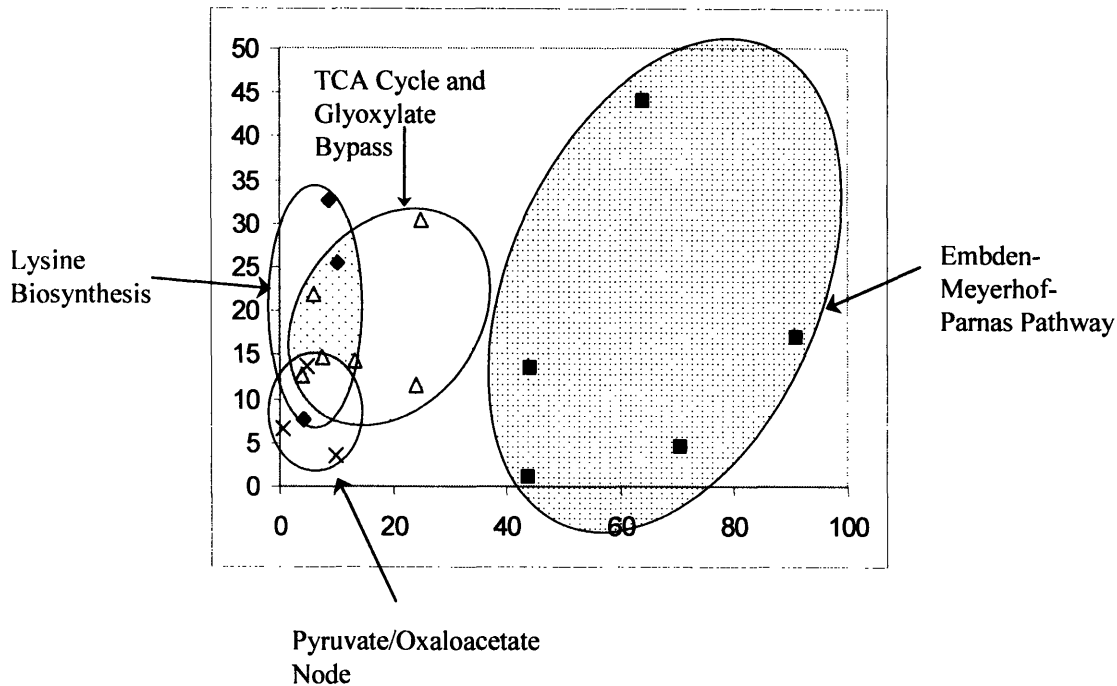
Figure 7-17: Comparison of flux/mRNA pairs with good correlation to one another. The data for each related pair of flux and transcriptional data sets with good correlation are plotted according to the comparison metric described in section 7.7.1. The clusters indicated on the graph each consist of reactions comprising one individual pathway of the metabolic network.

7.7.8 Implications for Transcriptional Control

By returning to the concept of the comparison metric that was introduced in section 7.7.1, further insight can be gained into the flux and transcriptional behavior of the cells on a pathway-by-pathway basis. The data for the flux/mRNA pairs for which a good correlation existed between the two types of measurements have been plotted using the metric in Figure 7-17. Additionally, all points associated with a particular pair of reaction and enzyme have been marked identically, so that any clustering of the data becomes evident. Examination of the graph does indeed show such clustering to occur, suggesting that regulatory strategies used by the cells to control flux are to some extent conserved within an individual pathway. These clusters are located at different positions relative to the origin, which represents the case in which transcriptional and flux responses to environmental changes are identical in both directionality and magnitude. In this way, the

187

chart can be thought of as a measure of the degree to which the fluxes associated with a particular pathway are determined by the amounts of transcript present for the genes of the enzymes catalyzing the reactions within it.

That the clusters do not entirely overlap indicates that there are different means of regulation at work in different pathways. From the plot it can be seen that of the four clusters present, the Embden-Meyerhof-Parnas pathway cluster is furthest removed from the origin, indicating that the metabolic flux through the glycolytic and gluconeogenic reactions are more likely to be regulated at the post-transcriptional level. In contrast, the genes associated with the tricarboxylic acid cycle and the anaplerotic pathways show a very strong link in general between changes in transcript concentrations and resulting changes in metabolic flux. These types of findings are very useful from a metabolic engineering point of view, because they suggest that reactions of certain pathways may be much more attractive targets for influencing the up-regulation or down-regulation of gene expression as a means of altering the flow of metabolites within a reaction network.

Chapter 8: Conclusions and Recommendations

8.1 Summary

The work of this thesis was carried out in an attempt to gain a more complete understanding of the functioning of the network of reactions used by microorganisms to convert available substrates into the components that are required for cell culture growth and maintenance. The goal of metabolic engineering is to use the knowledge gained about these metabolic networks in altering the fundamental chemical operations that take place within the cells, thereby providing them with novel properties or altering those already in existence in ways that are beneficial to research or industrial applications. To do this successfully, one needs as detailed and quantitative a description of the metabolic network as can be attained. Global measurement techniques, such as that of DNA microarrays, are then of particular benefit to the field of metabolic engineering.

By applying DNA microarray technology to the study of various strains and growth conditions for the lysine-producing bacteria *C. glutamicum*, we have been able to gain a more complete understanding of metabolic adaptations that take place within the cells to compensate for shifts in enzyme expression levels and medium composition. These adaptations manifested themselves extracellularly as macroscopic changes in the physiology of cultures that could be measured in terms of yields of both biomass and lysine, and intracellularly as changes in the profiles of transcript and metabolite concentrations and flux rates present in the reaction network of the cells.

Partial genome DNA microarrays were constructed through the printing of DNA sequences representing 57 *C. glutamicum* genes onto amino-silane-coated glass slides. These genes encode enzymes that catalyze each of the reactions in the pathways that directly constitute the metabolic network used by the cells in converting glucose and lactate substrates to lysine product. As such, the interconnected pathways represented are those of glycolysis and gluconeogenesis, pentose phosphates, tricarboxylic acids, the

glyoxylate bypass, central carbon anaplerosis, biotin biosynthesis, and lysine biosynthesis.

In optimizing the protocols used in the creation and application of the DNA microarrays several areas were found to be significant in determining the quality of the final results. Key among these were the method used for isolating and purifying total RNA samples obtained from cultures, and the parameters of the reverse transcription reaction used to generate fluorescently labeled cDNA from the RNA templates. It was shown that the use of DNase and repeated RNA filtration purification steps improved the quality of the RNA significantly, and that by using gene specific primers to initiate the reverse transcription of this RNA, the amount of nonspecific hybridization occurring on the arrays and obscuring the true data was greatly minimized.

Initial experiments with the novel microarrays demonstrated that the transcriptional data that they provided was very low in error and high in reproducibility. Also, distinct microarray spots that represented the same gene using DNA sequences of different lengths were shown to generate nearly identical results. In contrast, multiple microarray spots that represented the same gene using DNA sequences originating from the genomes of two different but phylogentically related species—*C. glutamicum* and *M. tuberculosis*—yielded data sets that were somewhat similar to another, but not nearly as interchangeable. The microarrays were then applied to studies of three different *C. glutamicum* strains that were grown on six different medium compositions and sampled during three different culture phases. In this way the transcriptional levels of all of the key genes related to central carbon metabolism and lysine biosynthesis in the cells and their response to changes in available carbon sources and gene over-expression could be monitored.

By summarizing the changes in transcript determined by the DNA microarray results of all experiments of this work, we can make more informed predictions about the activities of enzymes and flows of metabolites. From this combined data, in general those cells that are grown on glucose rely strongly on glycolysis for the production of energy and the pentose phosphate pathway for the production of reducing equivalents. The tricarboxylic acid cycle operates at a moderate level, and the glyoxylate bypass is not significant. Cells that are cultivated instead on lactate have increased demand for the tricarboxylic acid

cycle, the glyoxylate bypass, and malic enzyme. Expression of the aspartokinase and aspartate semialdehyde genes declines. As pyruvate carboxylase is over-expressed, cells grown on lactate have a further increased demand for both the tricarboxylic acid cycle and malic enzyme, and a stronger decline in the levels of aspartokinase and aspartate semialdehyde. Other genes whose transcript levels are up-regulated for the case of growth on lactate include the acyl-CoA carboxylases dtsR1, dtsR2, accC, and accD, as well as the genes of the biotin biosynthesis pathway.

One of the major results of the thesis was an indication of the importance of the malic enzyme as a source of the reducing equivalents required for both biomass and lysine production. The lower flux that exists in the pentose phosphate pathway as the cultures are grown on increasing proportions of lactate rather than glucose requires that the cells generate NADPH elsewhere in the network. The increased utilization of the tricarboxylic acid cycle under these conditions can replace through the activity of isocitrate dehydrogenase some of the NADPH lost due to lower pentose phosphate pathway flux, but this change alone can not supply all of the reducing equivalents necessary for cell growth and amino acid synthesis. Additionally, some of the flow through isocitrate dehydrogenase is diverted into the glyoxylate bypass pathway in the case of growth medium containing only lactate as the available carbon source, further reducing NADPH production. The activity of malic enzyme provides a third means that can be used by the cells for forming the reducing equivalents needed to meet the anabolic requirements of the cells. For growth on glucose, flux mapping experiments have confirmed that lowering the flow of carbon from the tricarboxylic acid cycle back to the Embden-Meyerhof-Parnas pathway yields increases in the productivities of lysine, most likely by reducing what is essentially a futile cycle under the given growth conditions [Drysch et al., 2003], but for growth on lactate, increasing this flow instead may lead to enhanced lysine production as NADPH availability will be improved.

Although the large majority of our findings are compatible with physiological observations and proven or proposed metabolic networks operating in C. glutamicum, others do not fit as well. For example, not all genes associated with lysine biosynthesis are up-regulated during growth on lactate although lysine yields increase under these conditions. Also, some genes, such as those of the glyoxylate bypass, show significant

differences in transcript concentration found in the *C. glutamicum* strains 21253 and 21253(pKD7) for growth on glucose even though the two strains have very similar growth and lysine yields when cultivated in that medium. Rather than being explained merely by experimental errors, these counterintuitive results may be signs that the regulation of these genes and pathways occurs at a level other than transcription.

To investigate the relationship that exists between the profiles of transcription and flux in the metabolism of the *C. glutamicum* cultures, DNA microarray results were compared with those obtained through metabolic network analysis. It was found that in the majority of cases there was a good correlation between the transcriptional responses and the flux responses, in terms of both magnitude and directionality, of the cells to perturbations in growth media and gene over-expression. Using a novel metric to describe the degree to which the flux through a reaction is related to the mRNA abundance of the gene for the enzyme catalyzing that reaction, it was shown that many of the reactions of the network tended to cluster together with other reactions functioning in the same pathway. Comparisons between these pathway-specific clusters revealed that in general some pathways, such as the Embden-Meyerhof-Parnas pathway, were more likely to be regulated by the cells at a level other than that of transcription. Other pathways, such as the tricarboxylic acid cycle and several central carbon anaplerotic reactions, appeared to have metabolite fluxes whose changes were in closer coordination with changes in mRNA levels. These results have implications both for the application of DNA microarray results as predictors of network reaction behavior, and for the selection of which genes within the network are the best candidates for metabolic engineering through transcriptional modifications such as over-expression or repression.

8.2 Conclusions

The major conclusions of this thesis work were the following:

(1) The two most significant factors in ensuring that microarray results of high quality were obtained were identified as being the purity of the total RNA isolated

from cell culture samples and the specificity of the reverse transcription reactions used to produce labeled cDNA from the isolated mRNA templates. As a result, it was found that the microarray protocols should be modified to include DNase treatment of RNA samples, additional RNA filtration purification steps, and the use of gene specific primers in the reverse transcription.

(2) The results obtained from a DNA microarray spot are independent of the size of the probe sequence printed within the spot in the range from 400-4000 base pairs.

(3) The use of DNA sequences originating from one species on a microarray as a probe used to assay the transcriptional profile of a second species provides results that have some qualitative value, but can not reliably be used for accurate quantification.

(4) The malic enzyme and pyruvate carboxylase are regulated by the cells in a coordinated manner at the level of transcription, most likely for the shared purpose of generating reducing equivalents in a transhydrogenation cycle or modified tricarboxylic acid cycle.

(5) Higher mRNA abundances of the genes encoding the acyl-CoA carboxylases dtsR1, dtsR2, accC, and accD correlate well with conditions that promote increased lysine productivity.

(6) The transcript levels of the genes for aspartokinase and aspartate semialdehyde dehydrogenase decrease as the proportion of lactate carbon source relative to glucose in the growth medium increases. This causes a limitation on the potential flow of material through the lysine biosynthetic pathway, resulting in decreased lysine yields.

(7) An increase in the mRNA concentrations and flux magnitudes associated with the dehydrogenation route relative to the succinylation route of the branched lysine biosynthetic pathway is observed when pyruvate carboxylase is over-expressed and lactate is the available carbon source.

(8) The transcription of the enzyme genes for the incomplete pathway of biotin biosynthesis is up-regulated in the case of growth on lactate.

(9) There is a fair, but incomplete correlation between the transcriptional profile and the flux profile of the metabolic network, with different pathways of the network

being controlled via transcription in varying degrees. On average, the Embden-Meyerhof-Parnas pathway was shown to be less likely to be regulated though transcription than the pathways of the tricarboxylic acid cycle and central carbon anaplerosis.

8.3 Recommendations

As was described in the previous sections, there were several modifications made to the original DNA microarray protocols that brought about necessary improvements to the accuracy and reliability of the data. However, even those techniques that appeared to be clearly beneficial to the microarray protocols may contain themselves new sources of error when applied to other systems. For this reason future work that seeks to apply the findings of this thesis to different types of arrays or investigations of other species should seek to reconfirm the presented protocol results. One example of this can be found in the use of gene specific primers, which here greatly reduced the degree of problematic fluorescence associated with each array that was attributable to nonspecific hybridization. Another work that tested the use of specific primers against the use of random hexamers primers in carrying out *Escherichia coli* microarray experiments found that the specific primers actually reduced the accuracy of their gene expression measurements [Arfin *et al.*, 2000]. The researchers of that study attributed this finding to two effects—the unequal efficiency with which the different primer sequences bind to the target mRNA during the reverse transcription reaction, and the differing degradation rates of the 25-mers complementary to the primer sequences that could influence the abilities of isolated mRNA sequences to serve as templates for reverse transcription. In that work, though, there were several gene sequences for which no significant fluorescence was detected on the corresponding microarray spots when specific primers were used. That was not a phenomenon observed in any of the experiments of this thesis. Also, the specific primers used in the study were all derived from the 3'-sequences of the open reading frames. Since bacterial mRNA decay occurs in the 3' to 5' direction, the sequences

complementary to these primers would be in general prone to more rapid degradation than the sequences complementary to the specific primers of this thesis, which were randomly located throughout the gene sequences.

One of the most significant benefits of the application of DNA microarrays to metabolic engineering is that the ability to examine the simultaneous transcriptional behavior of large numbers of genes allows for a better identification of particularly promising enzyme candidates for modification or manipulation. In this work, aspartokinase, aspartate semialdehyde dehydrogenase, and malic enzyme were found to be excellent targets for coordinated alteration or over-expression along with pyruvate carboxylase to assist the cells in balancing the demands of producing energy, biomass, reducing equivalents, and amino acid byproducts. The coordinated over-expression of aspartokinase and pyruvate carboxylase has already proved to be beneficial for the production of lysine [Koffas et al., 2003]. The further addition of aspartate semialdehyde dehydrogenase or malic enzyme to this over-expression strategy may yield still further improvements either limiting bottlenecks that exist in anabolic pathways, or allowing for the increased generation of NADPH that is required in large quantities for lysine synthesis.

Another interesting discovery made in this work was that all of the genes associated with the partial pathway for biotin biosynthesis in C. glutamicum exhibit up-regulated expression for the case of growth on lactate. Because the species lacks one enzyme of this pathway, biotin must be supplemented in the extracellular growth medium. There may then be a benefit realized from engineering a strain to possess the ability to generate its own supply of this vital cofactor. This is especially true given the evidence presented here and elsewhere [Tosaka et al., 1979] that an increase in the availability of biotin is positively correlated with the production of lysine. Further evidence that C. glutamicum should be capable of supporting biotin production after some genetic modification comes from the work of Ogino et al. (1974a and 1974b), who reported the successful conversion of the biotin analog DL-cis-tertrahydro-2-oxo-4-n-pentyl-thieno-(3,4-D)-imidazoline by the species.

Equally important to the understanding and successful manipulation of the metabolic network of the organism is an overview of not only what the gene expression profile

associated with a given phenotype is, but also how that profile is realized by the cells. To achieve this goal, more information will be needed regarding the characteristics of, for example, different promoters and other regulatory sequences at work within the cells. Investigations done to date in this area are summarized in the recent review by Pátek *et al.* (2003). It should be noted that genes that appear to be co-regulated by the cells are not necessarily controlled by the same regulatory element. Instead, the genes could be controlled by distinct regulators, each of which responds in a similar fashion to the particular stimulation being studied. Semantically, this scheme has been labeled a stimulon, referring to a clustering of genes sharing a behavior that correlates strongly with a given extracellular or intracellular stimulation, in contrast to a regulon, referring to a cluster of genes directly affected by a given regulator molecule [Wendisch, 2003].

Chapter 9: References

Abe, S.K. and K. Takayama. 1967. Taxonomical studies on glutamic-acid producing bacteria. J. Gen. Appl. Microbiol. **13**: 279-301.

Albrecht, M., N. Misawa, and G. Sandmann. 1999. Metabolic engineering of the terpanoid biosynthetic pathway of *E. coli* for production of the crotenoids β-carotene and zeazanthin. Biotechnol. Lett. **21**: 791-795.

Alwine, J.C., D.J. Kemp, and G.R. Stark. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. USA. **74**: 5350-5354.

Arfin, S.M., A.D. Long, E.T. Ito, L. Tolleri, M.M. Riehle, E.S. Paegle, and G.W. Hatfield. Global gene expression profiling in *Eschericia coli* K12: The effects of integration host factor. J. Biol. Chem. **275**: 29672-29684.

Bailey, J.E. 1998. Mathematical modeling and analysis in biochemical engineering. Past accomplishments future opportunities. Biotechnol. Prog. **14**: 8-20.

Baliga, N.S., M. Pan, Y.A. Goo, E.C. Yi, D.R. Goodlett, K. Dimitrov, P. Shannon, R. Aebersold, W.V. Ng, and L. Hood. 2002. Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global approach. Proc. Natl. Acad. Sci. U.S.A. **99**: 14913-14918.

Baronofsky, J.J., W.J.A. Schreurs, and E.R. Kashket. 1984. Uncoupling by acetic acid limits growth of and acetogenesis by *Clostridium thermoaceticum*. Appl. Environ. Microbiol. **48**: 1134-1139.

Behr, M.A., M.A. Wilson, W.P. Gill, H. Salamon, G.K. Schoolnik, S. Rane, and P.M. Small. 1999. Comparitive genomics of BCG vaccines by whole-genome DNA microarray. Science. **284**: 1520-1523.

Chen, G., T.G. Gharib, C.C. Huang, J.M. Taylor, D.E. Misek, S.L. Kardia, T.J. Giordano, M.D. Iannettoni, M.B. Orringer, S.M. Hanash, and D.G. Beer. 2002. Discordant protein and mRNA expression in lung adenocarcinomas. Mol. Cell. Proteomics. **1**: 304-313.

Christensen, B. and J. Nielsen. 1999. Isotopomer analysis using GC-MS. Metab. Eng. **1**: 282-290.

Christensen, B. and J. Nielsen. 2000. Metabolic network analysis of *Penicillium chrysogenum* using 13C-labeled glucose. Biotechnol. Bioeng. **68**: 652-659.

Christiansen, T., B. Christenesen, and J. Nielsen. 2002. Metabolic network analysis of *Bacillus clausii* on mimimal and semirich medium using 13C-labeled glucose. Metab. Eng. **4**: 159-169.

Cocaign-Bousquet, M., C. Monnet, and N.D. Lindley. 1993. Batch kinetics of *Corynebacteruim glutamicum* during growth on various substrates: Use of substrate mixtures to localize metabolic bottlenecks. Appl. Microbiol. Biotechnol. **40**: 526-530.

Cocaign-Bousquet, M. and N.D. Lindley. 1995. Pyruvate overflow and carbon flux with the central metabolic pathways of *Corynebacterium glutamicum* during growth on lactate. *Enz. Microbiol. Technol.* **17**: 260-267.

Cole, S.T., R. Brosch, Pakhill, J., T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, S. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, S. Skelton, S. Squares, R. Squares, J.E. Sulston, K. Taylor, S. Whitehead, and B.G. Barrell. 1998. Deciphering the biology of *Mycobacterium* tuberculosis from the complete genome sequence. Nature. **393**: 537-544.

Collins, M.D. and C.S. Cummings. 1986. "Genus *Corynebacterium* Lehmann and Neumann 1896." **Bergey's Manual of Systematic Biology II**. P.H.A. Sneath, N.S. Mair, M.E. Sharpe, and J.G. Holt, eds. pp. 1266-1276.

Colon, G.E., T.T. Nguyen, M.S.M. Jetten, A.J. Sinskey, and G. Stephanopoulos. 1995. Production of isoleucine by overexpression of ilvA in a *Corynebacterium lactofermentum* threonine producer. Appl. Microbiol. Biotechnol. **43**: 482-488.

Cremer, J., L. Eggeling, and H. Sahm. 1991. Control of the lysine biosynthetic sequence in *Corynebacterium glutamicum* as analyzed by overexpression of the individual corresponding genes. Appl. Environ. Microbiol. **57**: 1746-1752.

Dauner, M. M .Sonderegger., M. Hochuli, T. Szyperski, K. Wuthrich, H.P. Hohmann, U. Sauer, and J.E. Bailey. 2002. Intracellular carbon fluxes in riboflavin-producing *Bacillus subtilitis* during growth on two-carbon substrate mixtures. Appl. Environ. Microbiol. **68**: 1760-1771.

de Graaf, A.A., M. Mahle, M. Mollney, W. Wiechert, P. Stahmann, and H. Sahm. 2000. Determination of the full 13-C isotopomer distributions for metabolic flux analysis using heteronuclear spin echo difference NMR spectroscopy. J. Biotechnol. **77**: 25-35.

de Graaf, A.A., L. Eggeling, and H. Sahm. 2001. Metabolic engineering for L-lysine production by *Corynebacterium glutamicum*. Adv. Biochem. Eng. Biotechnol. **73**: 9-29.

Debabov, V.G. 2003. The threonine story. Adv. Biochem. Eng. Biotechnol. **79**: 113-136.

Delaunay, S., P. Daran-Lapujade, J.-M. Engasser, and J.-L. Goergen. 2004. Glutamate as an inhibitor of phosphoenolpyruvate carboxylase activity in *Corynebacterium glutamicum*. J. Ind. Microbiol. Biotechnol. **31**: 183-188.

Demain, A.L. and J. Birnbaum. 1968. Alteration of permeability for the release of metabolites from the microbial cell. Curr. Top. Microbiol. Immunol. **46**: 1-25.

DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nature Genet. **14**: 457-460.

De Saizieu, A., U. Certa, J. Warrington, C. Gray, W. Keck, and J. Mous. 1998. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide array. Nat. Biotechnol. **16**: 45-48.

Dominguez, H., M. Cocaign-Bousquet, and N.D. Lindley. 1993. Simultaneous consumption of glucose and fructose from sugar mixtures during batch growth of *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. **47**: 600-603.

Dominguez, H., C. Rollin, A. Guyonvarch, J.-L. Guerquin-Kern, M. Cocaign-Bousquet, and N.D. Lindley. 1998. Carbon-flux distribution in the central metabolic pathways of *Corynebacterium glutamicum* during growth on fructose. Eur. J. Biochem. **254**: 96-102.

Donato, L.D., C.D. Rosiers, J.A. Montgomery, F. David, M. Garneau., and H. Brunengraber. 193. Rates of gluconeogenesis and citric acid cycle in perfused livers, assessed from the mass spectrometric assay of the 13C labeling pattern of glutamate. J. Biol. Chem. **268**: 4170-4180.

Drysch, A., M.E. Massaoudi, W. Wiechert, A.A. de Graaf, and R. Takors. Serial flux mapping of *Corynebacterium glutamicum* during fed-batch L-lysine production using the sensor reactor approach. Biotechnol. Bioeng. **85**: 497-505.

Duperray, F., D. Jezequel, A. Ghazi, L. Letelier, and E. Shechter. 1992. Excretion of glutamate by *Corynebacterium glutamicum* triggered by surfactants. Biochim. Biophys. Acta. **1103**: 250-258.

Eggeling, L., S. Oberle, and H. Sahm. 1998. Improved L-lysine yield with *Corynebacterium glutamicum*: Use of dapA resulting in increased flux combined with growth limitation. Appl. Microbiol. Biotechnol. **49**: 24-30.

Eggeling, L. and H. Sahm. 1999. L-glutamate and L-lysine: traditional products with tempestuous developments. Appl. Microbiol. Biotechnol. **52**: 146-153.

Eikmanns, B.J. 1992. Identification, sequence analysis, and expression of a *Corynebacterium glutamicum* gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomerase. J. Bacteriol. **174**: 6076-6086.

Eikmanns, B.J., N. Thum-Schmitz, L. Eggeling, K. Ludtke, and H. Sahm. 1994. Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum gltA* gene encoding citrate synthase. Microbiol. **140**: 1817-1828.

Fischer, E. and U. Sauer. 2003. Metabolic flux profiling of *Eschericia coli* mutants in central carbon metabolism using GC-MS. Eur. J. Biochem. **270**: 880-891.

Follettie, M.T. and A.J. Sinskey. 1986. Food Technol. **40**: 88-94.

Follettie, M.T. O.P. Peoples, C. Agoropoulou, and A.J. Sinskey. 1993. Gene structure and expression of the *Corynebacterium flavum* N13 ask-asd operon. J. Bactriol. **175**: 4096-4103.

Gerstmeir, R., V.F. Wendisch, S. Schnicke, H. Ruan, M. Farwick, D. Reinscheid, and B.J. Eikmanns. 2003. Acetate metabolism and its regulation in *Corynebacterium glutamicum*. J. Biotechnol. **104**: 99-122.

Glannemann, C., A. Loos, N. Gorret, L.B. Willis, X.M. O'Brien, P.A. Lessard, and A.J. Sinskey. 2003. Disparity between changes in mRNA abundance and enzyme activity in *Corynebacterium glutamicum*: implications for DNA microarray analysis. Appl. Microbiol. Biotechnol. **61**: 61-68.

Gourdon, P., M.-F. Baucher, N.D. Lindley, and A. Guyonvarch. 2000. Cloning of the malic enzyme gene from *Corynebacterium glutamicum* and role of the enzyme in lactate metabolism. Appl. Environ. Microbiol. **66**: 2981-2987.

Griffin, T.J., S.P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood, and R. Aebersold. 2002. Complemetary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cereisiae*. Mol. Cell. Proteomics. **1**: 323-333.

Gubler, M.E., M.S.M. Jetten, S.H. Lee, and A.J. Sinskey. 1994. Effects of phosphoenol pyruvate carboxylase deficiency on metabolism and lysine production in *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. **40**: 857-863.

Hartmann, M., A. Tauch, L. Eggeling, B. Bathe, B. Mockel, A. Puhler, and J. Kalinowski. 2003. Identification and characterization of the last two unknown genes,

dapC and *dapF*, in the succinylase branch of the L-lysine biosynthesis of *Corynebacterium glutamicum*. J. Biotechnol. **104**: 199-211.

Hatakeyama, K., K. Kohama, A.A. Vertes, M. Kobayashi, Y. Kurusu, and H. Yukawa. 1993. Genomic organization of the biotin biosynthetic genes of coryneform bacteria: cloning and sequencing of the *bioA-bioD* genes from *Brevibacterium flavum*. DNA Seq. **4**: 177-184.

Hayashi, M., H. Mizoguchi, N. Shiraishi, M. Obayashi, S. Nakagawa, J. Imai, S. Watanabe, T. Ota, and M. Ikeda. 2002. Transcriptome analysis of acetate metabolism in *Corynebacterium glutamicum* using a newly developed metabolic array. Biosci. Biotechnol. Biochem. **66**: 1337-1344.

Hermann, T. 2003. Industrial production of amino acids by coryneform bacteria. J. Biotechnol. **104**: 155-172.

Hochuli, M., H. Patzelt, D. Oesterhelt, K. Wuthrich, and T. Szyperski. 1999. Amino acid biosynthesis in the halophilic archaeon *Haloarcula hispanica*. J. Bacteriol. **181**: 3226-3237.

Hua, Q., C. Yang, and K. Shimizu. 2000. Metabolic control analysis for lysine synthesis using *Corynebacterium glutamicum* and experimental verifications. J. Biosci. Bioeng. **20**: 184-192.

Hüser, A.T., A. Becker, I. Brune, M. Dondrup, J. Kalinowski, J. Plassmeier, A. Pühler, I. Wiegräbe, and A. Tauch. 2003. Development of a *Corynebacterium glutamicum* DNA microarray and validation by genome-wide expression profiling during growth with propionate as carbon source. J. Biotechnol. **106**: 269-286.

Ideker, T., V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, and L. Hood. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science. **292**: 929-934.

Ikeda, M. 2003. Amino acid production processes. Adv. Biochem. Eng. Biotechnol. **79**: 1-35.

Inbar, L., Z.E. Kahana, and A. Lapidot. 1985. Natural-abundance 13C nuclear magnetic resonance studies of regulation and overproduction by L-lysine by *Brevibacterium flavum*. Eur. J. Biochem. **149**: 601-607.

Inui, M., S. Morakami, S. Okino, H. Kawaguchi, A.A. Vertes, and H. Yukawa. 2004. Metabolic analysis of *Corynebacterium glutamicum* during lactate and succinate productions under oxygen depravation conditions. J. Mol. Microbiol. Biotechnol. **7**: 182-196.

Ishige, T., M. Krause, M. Bott, V.F. Wendisch, and H. Sahm. 2003. The phosphate starvation stimulon of *Corynebacterium glutamicum* as determined by DNA microarray analysis. J. Bacteriol. **185**: 4519-4529.

Jäger, W., P.G. Peters-Wendisch, J. Kalinowski, and A. Pühler. 1996. A *Corynebacterium glutamicum* gene encoding a two-domain protein similar to biotin carboxylase and biotin-carboxyl-carrier proteins. Arch. Microbiol. **166**: 76-82.

Jetten, M.S.M. and A.J. Sinskey. 1993. Characterization of phosphoenolpyruvate carboxykinase from *Corynebacterium glutamicum*. FEMS Microbiol. Lett. **111**: 183-188.

Jetten, M.S., M.E. Gubler, S.H. Lee, and A.J. Sinskey. 1994. Structural and functional analysis of pyruvate kinase from *Corynebacterium glutamicum*. Appl. Environ. Microbiol. **60**: 2501-2507.

Jetten, M.S., M.T. Follettie, and A.J. Sinskey. 1995. Effect of different levels of aspartokinase on the lysine production by *Corynebacterium lactofermentum*. Appl. Microbiol. Biotechnol. **43**: 76-82.

Jetten, M.S.M. and A.J. Sinsky. 1995. Recent advances in the physiology and genetics of amino acid producing bacteria. Crit. Rev. Biotechnol. **15**: 73-103.

Jonsbu, E., T.E. Ellingsen, and J. Nielsen. 2000. Effects of nitrogen sources on cell growth and production of nystatin by *Streptomyces noursei*. J. Antibiot. **53**: 1354-1362.

Jucker, B.M., J.Y. Lee, and R.G. Shulman. *In Vivo* ^{13}C NMR measurements of hepatocellular tricarboxylic acid cycle flux. 1998. J. Biol. Chem. **273**: 12187-12194.

Kalinowski, J., B. Bachmann, G. Thierbach, and A. Puhler. 1990. Aspartokinase genes lysCα and lysCβ overlap and are adjacent to the aspartate semialdehyde dehydrogenase gene *asd* in *Corynebacterium glutamicum*. Mol. Microbiol. **5**: 1197-1204.

Kalinowski, J., J. Cremer, B. Bachmann, L. Eggeling, and A. Pühler. 1991. Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. Mol. Microbiol. **5**: 1197-1204.

Kalinowski, J. B. Bathe, D. Bartels, N. Bischoff, M. Bott, A. Burkovski, N. Dusch, L. Eggeling, B.J. Eikmanns, L. Gaigalat, A. Goesmann, M. Hartmann, K. Huthmacher, R. Kramer, B. Linke, A.C. McHardy, F. Meyer, B. Mockel, W. Pfefferle, A. Puhler, D.A. Rey, C. Ruckert, O. Rupp, H. Sahm, V.F. Wendisch, I. Wiegrabe, and A. Tauch. 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J. Biotechnol. **104**: 5-25.

Kamada, N., A. Yasuhara, Y. Takano, T. Nakano, and M. Ikeda. 2001. Effect of transketolase modifications on carbon flow to the purine-nucleotide pathway in *Corynebacterium ammoniagenes*. Appl. Microbiol. Biotechnol. **56**: 710-717.

Kascer, H. and J.A. Burns. 1973. Control of enzyme flux. Symp. Soc. Exp. Biol. **27**: 64-104.

Kascer, H. and J.A. Burns. 1995. The control of flux. Biochem. Soc. Trans. **23**: 387-391.

Katsumata, R., A. Ozaki, T. Oka, and A. Furuya. 1984. Protoplast transformation of glutamate-producing bacteria with plasmid DNA. J. Bacteriol. **159**: 306-311.

Kawahara, Y., K. Takahasi-Fuke, E. Shimizu, T. Nakamatsu, and S. Nakamori. 1997. Relationship between the glutamate production and activity of 2-oxogluterate dehydrogenase in *Brevibacterium lactofermentum*. Biosci. Biotechnol. Biochem. **61**: 1109-1112.

Kiefer, P., E. Heinzle, O. Zelder, and C. Wittmann. 2003. Comparitive metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* culture on glucose or fructose. Appl. Environ. Microbiol. **70**: 229-239.

Kim, H.-J., T.-H. Kim, Y. Kim, and H.-S. Lee. 2004. Identification and characterization of glxR, a gene involved in regulation of glyoxylate bypass in *Corynebacterium glutamicum*. J. Bacteriol. **186**: 3453-3460.

Kimura, E., C. Abe, Y. Kawahara, and T. Nakamatsu. 1996. Molecular cloning of a novel gene, DtsR, which rescues the detergent sensitivity of a mutant derive from *Brevibacterium lactofermentum*. Biosci. Biotechnol. Biochem. **60**: 1565-1570.

Kimura, E., C. Abe, Y. Kawahara, T. Nakamatsu, and H. Tokuda. 1997. A dtsR gene-disrupted mutant of *Brevibacterium lactofermentum* requires fatty acids for growth and efficiently produces L-glutamate in the presence of an excess of biotin. Biochem. Biophys. Res. Commun. **234**: 157-161.

Kimura, E., C. Yagoshi, Y. Kawahara, T. Ohsumi, and T. Nakamatsu. 1999. Glutamate overproduction in *Corynebacterium glutamicum* triggered by a decrease in the level of a complex comprising DtsR and a biotin-containing subunit. Biosci. Biotech. Biochem. **63**: 1274-1278.

Kimura, E. 2002. Triggering mechanism of L-glutamate overproduction by DtsR1 in coryneform bacteria. J. Biosci. Bioeng. **94**: 545-551.

Kimura, E. 2003. Metabolic engineering of glutamate production. Adv. Biochem. Eng. Biotechnol. **79**: 37-57.

Kinoshita, S., S. Udoka, and M. Shimono. 1957. Studies on the amino acid fermentation. I. Production of L-glutamic acid by various microorganisms. J. Gen. Appl. Microbiol. **3**: 193-205.

Kiss, A.D. and G. Stephanopoulos. 1991. Metabolic activity control of the L-lysine fermentation by restrained growth fed-batch strategies. Biotechnol. Prog. **7**: 501-509.

Klapa, M.I. 2001. High resolution metabolic flux quantification using stable isotopes and mass spectroscopy. Ph.D. Thesis. MIT, Cambridge.

Klapa, M.I., J.-C. Aon, and G. Stephanopoulos. 2003. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectroscopy. Eur. J. Biochem. **270**: 3525-3542.

Koffas, M.A.m R. Ramamoorthi, W.A. Pine, A.J. Sinskey, and G. Stephanopoulos. 1998. Sequence of the *Corynebacterium glutamicum* pyruvate carboxylase gene. Appl. Microbiol. Biotechnol. **50**: 346-352.

Koffas, M.A.G. 2000. Metabolic engineering of *Corynebacterium glutamicum* for amino acid production improvements. Ph.D. Thesis. MIT, Cambridge.

Koffas, M.A.G., G.Y. Jung, J.C. Aon, and G. Stephanopoulos. 2002. Effect of pyruvate carboxylase overexpression on the physiology of *Corynebacterium glutamicum*. Appl. Environ. Microbiol. **68**: 5422-5428.

Koffas, M.A.G., G.Y. Jung, and G. Stephanopoulos. 2003. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. Metab. Eng. **5**: 32-41.

Krömer, J.O., O. Sorgenfrei, K. Klopprogge, E. Heinzle, and C. Wittmann. 2004. In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J. Bacteriology. **186**: 1769-1784.

Lange, C., D. Rittman, V.F. Wendisch, M. Bott, and H. Sahm. 2003. Global expression profiling and physiological characterization of *Coryebacterium glutamicum* grown in the presence of L-valine. Appl. Environ. Microbiol. **69**: 2521-2532.

Lipshutz, R.J., S.P.A. Fodor, T.R. Gingeras, and D.J. Lockhart. 1999. High density synthetic oligonucleotide araays. Nat. Gen. **21**: 20-24.

Loos, A., C. Glanemann, L.B. Willis, X.M. O'Brien, P.A. Lessard, R. Gerstmeir, S. Guillouet, and A.J. Sinskey. Development and validation of *Corynebacterium* DNA microarrays. 2001. Appl. And Environ. Eng. **67**: 2310-2318.

Lucchini, S., A. Thompson, and J.C.D. Hinton. 2001. Microarrays for microbiologists. Microbiology. **147**: 1403-1414.

Lynch, T. J., E.A. Tallant, and W.Y. Cheung. 1975. *Brevibacterium liquefaciens* adenylate cyclase and its in vivo stimulation by pyruvate. J. Bacteriol. **124**: 1106-1112.

Malumbres, M. and J.F. Martin. 1996. Molecular control mechanism of lysine and threonine biosynthesis in amino acid-producing Corynebacteria: Redirecing carbon flow. FEMS Microbiol. Lett. **143**: 103-114.

Marx, A., A.A. de Graaf, W. Weichert, L. Eggeling, and H. Sahm. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. 1996. Biotechnol. Bioeng. **49**: 111-129.

Marx, A., K. Striegel, A.A. de Graaf, H. Sahm, and L. Eggeling. Response of the central metabolism of *Corynebacterium glutamicum* to different flux burdens. 1997. Biotechnol. Bioeng. **56**: 168-180.

Marx, A., B.J. Eikmanns, H. Sahm, A.A. de Graaf, and L. Eggeling. 1999. Response of the central metabolism in *Corynebacterium glutamicum* to the use of an NADH-dependent glutamate dehydrogenase. Metab. Eng. **1**: 35-48.

Marx, A., S. Hans, B. Mockel, B. Bathe, and A.A. de Graaf. 2003. Metabolic phenotype of phosphoglucose isomerase mutants of *Corynebacterium glutamicum*. J. Biotechnol. **104**: 185-197.

Monod, J. 1949. The growth of bacterial cultures. Ann. Rev. Microbiol. **3**: 371-394.

Moody, D.E., Z. Zou, and L. McIntyre. 2002. Cross-species hybridization of pig RNA to human nylon microarrays. BMG Genomics, **3**: 27.

Morbach, S., H. Sahm, and L. Eggeling. 1996. Use of feedback-resistant threonine dehdratases of *Corynebacterium glutamicum* to increase carbon flux towards L-isoleucine. Appl. Environ. Microbiol. **61**: 4315-4320.

Mori, M. and I. Shiio. 1995. Synergistic inhibition of phosphoenolpyruvate carboxylase by aspartate and 2-oxogluterate in Brevibacterium flavum. J. Biochem. **98**: 1621-1630.

Moritz, B. K. Streigel, A.A. de Graaf, and H. Sahm. 2000. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux *in vivo*. Eur. J. Biochem. **267**: 3442-3452.

Muffler, A., S. Bettermann, M. Haushalter, A. Hörlein, U. Neveling, M. Schramm, and O. Sorgenfrei. 2002. Genome-wide transcription profiling of *Corynebacterium glutamicum* after heat shock and during growth on acetate and glucose. J. Biotechnol. **98**: 255-268.

Nakagawa. 2002. Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032. Unpublished. NCBI Refseq: NC_003450.

Nakao, Y., M. Kikuchi, M. Suzuki, and M. Doi. 1970. Microbial production of L-glutamic acid from n-paraffin by glycerol auxotrophs. Agric. Biol. Chem. **34**: 1875-1876.

Nielsen, J. 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. Biotechnol. Bioeng. **58**: 125-132.

Nunheimer, T.D., J. Birnbaum, E.D. Inhen, and A.L. Demian. 1970. Product inhibition of the fermentative formation of glutamic acid. Appl. Microbiol. **20**: 215-217.

Ogino, S. and S. Fujimito. 1974a. Coordination of dl-cis-tetrahydro-2-oxo-4-n-pentyl-thieno-(3,4-d)-imidazoline (dl-TOPTI) by soil isolates of the genus *Corynebacterium*. Agric. Biol. Chem. **38**: 275-278.

Ogino, S. and S. Fujimoto. 1974b. Production of biotin by microbial transformation of dl-cis-tetrahydro-2-oxo-4-n-pentyl-thieno-(3,4-d)-imidazoline (dl-TOPTI). Agric. Biol. Chem. **38**: 707-712.

Ozaki, A., R. Katsumata, T. Oka, and A. Furuya. 1984. Functional expression of genes of *Escherichia coli* in gram-positive *Corynebacterium glutamicum*. Mol. Gen. Genet. **96**: 175-178.

Ozaki, H., and I. Shiio. 1969. Regulation of the TCA and glyoxylate cycles in *Brevibacterium flavum*. J. Biochem. **66**: 297-311.

Paegle, L. and M. Ruklisha. 2003. Lysine synthesis control in *Corynebacterium glutamicum RC 115* in mixed substrate (glucose-acetate) medium. J. Biotechnol. **104**: 123-128.

Park, S.M. 1996. Investigation of carbon fluxes in central metabolic pathways of *Corynebacterium glutamicum*. Ph.D. Thesis. MIT, Cambridge.

Park, S.M., A.J. Sinskey, and G. Stephanopoulos. 1997. Metabolic and physiological studies of *Corynebacterium glutamicum* mutants. Biotechnol. Bioeng. **55**: 864-879.

Patek, M., M. Bilic, K. Krumbach, B. Eikmanns, H. Sahm, and L. Eggeling. 1997. Isolation and transcriptional analysis of the *dapB*-orf2-*dapA*-orf4 operon of

Corynebacterium glutamicum encoding two enzymes involved in L-lysnie synthesis. Biotechnol. Lett. **19**: 1113-1117.

Pátek, M., J. Nesvera, A. Guyonvarch, O. Reyes, and G. Leblon. 2003. Promoters of *Corynebacterium glutamicum*. J. Biotechnol. **104**: 311-323.

Payne, J. and J.G. Morris. 1969. Pyruvate carboxylase in *Rhodopseudomonas sphaeroides*. J. Gen. Microbiol. **59**: 97-101.

Pedersen, H., B. Christensen, C. Hjort, and J. Nielsen. 2000. Construction and characterization of an oxalic acid non-producing strain of *Aspergillus niger*. Metabol. Eng. **2**: 34-41.

Peters, E.P., A.F. Wilderspin, S.P. Wood, M.J.J.M. Zvelebil, O. Sezer, and A. Danchin. 1991. A pyruvate-stimulated adenylate cyclase has a sequence related to the *fes/fps* oncogenes and to eukaryotic cyclases. Mol. Microbiol. **5**: 1175-1181.

Petersen, S., A.A. de Graaf, L. Eggeling, M. Mollney, W. Weichert, and H. Sahm. 2000. *In vivo* quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacetrium glutamicum*. J. Biol. Chem. **275**: 35932-35941.

Peters-Wendisch, P.G., B.J. Eikmanns, G. Thierbach, B. Bachmann, and H. Sahm. 1993. Phosphoenolpyruvate carboxylase in *Corynebacterium glutamicum* is dispensable for growth and lysine production. FEMS Microbiol. Lett. **112**: 269-274.

Peters-Wendisch, P.G. 1996. Anaplerotische reaktionen in *Corynebacterium glutamicum*: untersuchungen zur bedeutung der PEP-carboxylase und der pyruvate-carboxylase im zentralstoffwechsel und bei der aminosäure-produktion. Ph.D. Thesis. University of Düsseldorf.

Peters-Wendisch, P.G., V.F. Wendisch, S. Paul, B.J. Eikmanns, and H. Sahm. 1997. Pyruvate carboxylase as an anaplerotic enzyme in *Corynebacterium glutamicum*. Microbiol. **143**: 1095-1103.

Peters-Wendisch, P.G., C. Kreutzer, J. Kalinowski, M. Patek, H. Sahm, and B.J. Eikmanns. 1998. Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. Microbiol. **144**: 915-927.

Pfefferle, W., B. Mockel, B. Bathe, and A. Marx. 2003. Biotechnological manufacture of lysine. Adv. Biochem. Eng. Biotechnol. **79**: 59-112.

Reidel, C., D. Rittman, P. Dangel, B. Mockel, H. Sahm, and B.J. Eikmanns. 2001. Characterization, expression, and inactivation of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the

enzyme for growth and amino acid production. J. Mol. Microbiol. Biotechnol. **3**: 573-583.

Reinscheid, D.J., B.J. Eikmanns, and H. Samm. 1994a. Characterization of the isocitrate lyase gene from *Corynebacterium glutamicum* and biochemical analysis of the enzyme. J. Bacteriology. **176**: 3474-3483.

Reinscheid, D.J., B.J.Eikmanns, and H. Samm. 1994b. Malate synthase from *Corynebacterium glutamicum*: sequence analysis of the gene and biochemical characterization of the enzyme. Microbiology. **140**: 3099-3108.

Roe, A.J., C. O'Byrne, D. McLaggan, and I.R. Booth. 2002. Inihibition of *Escherichia coli* growth by acetic acid: A problem with methionine synthesis and homocysteine toxicity. Microbiol. **148**: 2215-2222.

Rozen, S. and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S. and S. Misener (eds.) **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Humana Press, Totowa, NJ, pp. 365-386.

Ruklisha, M., R. Ionina, L. Paegle, and G. Petrovica. 2001. Metabolism and lysine biosynthesis control in *Brevibacterium flavum*: impact of stringent response. In: Simon, J.-P. and A. Durieux (Eds.), **Applied Microbiology, vol. 2**. Dordrecht: Kluwer Academic Publishers, 51-57.

Sahm, H., L. Eggeling, and A.A. de Graaf. 2000. Pathway analysis and metabolic engineering in *Corynebacterium glutamicum*. Biol. Chem. **381**: 899-910.

Sambrook, J. and D.W. Russel. 2001. **Molecular Cloning: A Laboratory Manual**, 3rd ed.Cold Spring Harbor, NY.

Santamaria, R., J.A. Gil, J.M. Mesas, and J.F. Martin. 1984. Characterization of an endogenous plasmid and development of cloning vectors and a transformation system in *Brevibacterium lactofermentum*. J. Gen. Microbiol. **130**: 2237-2246.

Sauer, U., D.R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyperski, K. Wuthrich, and J.E. Bailey. 1999. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. J. Bacteriol. **181**: 6679-6688.

Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**: 467-470.

Schneeberger, C., P. Speiser, F. Kury, and R. Zeillinger. 1994. Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. PCR Methods Appl. **4**: 234-238.

Schrumpf, B., A. Schwarzer, J. Kalinowski, A. Pühler, L. Eggeling, and H. Sahm. 1991. A functionally split pathway for lysine synthesis in *Corynebacterium glutamicum*. J. Bacteriol. **173**: 4510-4516.

Schoolnik, G.K. 2002. Functional and comparative genomics of pathogenic bacteria. Curr. Opin. Microbiol. **5**: 20-26.

Scrutton, M.C. and B.L. Taylor. 1974. Isolation and characterization of pyruvate carboxylase from *Azotobacter vinelandii*. Arch. Biochem. Biophys. **164**: 641-654.

Shiio, I., S. Otsuka, and M. Takahashi. 1962. Effect of biotin on the bacterial formation of glutamic acid. I. Glutamate formation and cellular permeability of amino acids. J. Biochem. **51**: 56-62.

Shiio, I. and H. Ozaki. 1968. Concerted inhibition of isocitrate dehydrogenase by glyoxylate plus oxaloacetate. J. Biochem. **64**: 45-53.

Shiio, I. and R. Miyajima. 1969. Concerted inhibition and its reversal by end products of aspartate kinase in *Brevibacterium flavum*. J. Biochem. **65**: 849-859.

Shiio, I., H. Momose, and A. Oyama. 1969. Genetic and biochemical studies on bacterial formation of L-glutamate. I. Relationship between isocitrate lyase, acetate kinase, and phosphate acetyltransferase levels and glutamate production in *Brevibacterium flavum*. J. Gen. Appl. Microbiol. **15**: 27-40.

Shiio, I., R. Miyajima, and S. Konosuke. 1970. Genetically desensitized aspartate kinase to the concerted feedback inhibition in *Brevibacterium flavum*. J. Biochem. **68**: 701-710.

Shiio, I., H. Ozaki, and K. Ujigawa-Takeda. 1977. Regulation of citrate synthase in *Brevibacterium flavum*, a glutamate-producing bacterium. J. Biochem. **82**: 395-405.

Shiio, I. and K. Ujigawa. 1978. Enzymes of the glutamate and aspartate synthetic pathways in a glutamate-producing bacterium, *Brevibacterium flavum*. J. Biochem. **84**: 647-657.

Shiio, I. and K. Ujigawa-Takeda. 1980. Presence and regulation of α-ketoglutarate dehydrogenase complex in a glutamate-producing bacterium, *Brevibacterium flavum*. Agruc. Biol. Chem. **44**: 1897-1904.

Shiio, I., Y. Toride, and S. Sugimoto. 1984. Production of lysine by pyruvate dehydrogenase mutants of *Brevibacterium flavum*. Agric. Biol. Chem. **48**: 3091-3098.

Shimizu, H., H. Tanaka, A. Nakato, K. Nagahisa, E. Kimura, and S. Shioya. 2003. Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxogluterate branch in glutamate production by *Corynebacterium glutamicum*. Bioprocess. Biosyst. Eng. **25**: 291-298.

Shingu, H. and G. Terui. 1971. Studies on the process of glutamic acid fermentation at the enzyme level. I. On the changes of α-ketoglutaric acid dehydrogenase in the course of culture. J. Ferment. Technol. **49**: 400-405.

Sonntag, K., L. Eggeling, A.A. de Graaf, and H. Sahm. 1993. Flux partitioning in the split pathway of lysine synthesis in *Corynebacterium glutamicum*: Quantification by 13C- and 1H-NMR spectroscopy. Eur. J. Biochem. **213**: 1325-1331.

Sonntag, K., J. Schwinde, A.A. de Graaf, A. Marx, B.J. Eikmanns, W. Weichert, and H. Sahm. 1995. ^{13}C NMR studies of the fluxes in the central metabolism of *Corynebacterium glutamicum* during growth and overproduction of amino acids in batch cultures. Appl. Microbiol Biotechnol. **44**: 489-495.

Soual-Hoebeke, E., C. de Sousa-D'Auria, M. Chami, M.-F. Baucher, A. Guyonvarch, N. Bayan, K. Salim, and G. Leblon. 1999. S-layer protein production by *Corynebacterium* strains is dependent on the the carbon source. Microbiology. **145**: 3399-3408.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503-517.

Stephanolpoulos, G. and J.J. Vallino. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science. **252**: 1675-1681.

Stephanopoulos, G. 1998. Metabolic fluxes and metabolic engineering. Metab. Eng. **1**: 1-10.

Sugimoto, S.-I. and I. Shiio. 1987. Regulation of glucose-6-phosphate dehydrogenase in *Brevibacterium flavum*. Agric. Biol. Chem. **51**: 101-108.

Sugimoto, S.-I., and I. Shiio. 1989. Regulation of enzymes for erythrose-4-phosphate synthesis in *Brevibacterium flavum*. Agric. Biol. Chem. **53**: 2081-2087.

Szyperski, T. 1995. Biosynthetically directed 13C-fractional labeling of proteinogenic amino acids. Eur. J. Biochem. **232**: 433-448.

Takinami, K., Y. Yamada, and H. Okada. 1966. Biochemical effects of fatty acids and its derivatives on L-glutamic acid fermentation. Agric. Biol. Chem. **30**: 674.

Talaat, A.M., P. Hunter, and S.A. Johnston. 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. Nat. Biotechnol. **18**: 679-682.

Tausova, T.A. and T.L. Madden. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. **174**: 247-250.

Tian, Q., S.B. Stepaniants, M. Mao, L. Weng, M.C. Feetham, M.J. Doyle, E.C. Yi, H. Dai, V. Thorsson, J. Eng, D. Goodlett, J.P. Berger, B, Gunter, P.S. Linseley, R.B. Stoughton, R. Aebersold, S.J. Collins, W.A. Hanlon, and L.E. Hood. 2004. Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. Mol. and Cell. Proteomics. **3**: 960-969.

Tosaka, O., K. Takinami, and Y. Hirose. 1978. L-lysine production by S-(2-aminoethyl) L-cysteine and α-amino-β-hydroxyvaleric acid resistant mutants of *Brevibacterium lactofermentum*. Agric. Biol. Chem. **42**: 745-752.

Tosaka, O., H. Hirakawa, K. Takinami, and Y. Hirose. 1978b. Regulation of lysine biosynthesis by leucine in *Brevibacterium lactofermentum*. Agric. Biol. Chem. **42**: 1501-1506.

Tosaka, O., H. Hirakawa, and K. Takinami. 1979. Effects of biotin levels on L-lysine formation in *Brevibacterium lactofermentum*. Agric. Biol. Chem. **43**: 491-495.

Treadway, S.L., K.S. Yanagimachi, E. Lankeneau., P.A. Lessard, G. Stephanopoulos, and A.J. Sinskey. 1999. Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain 124. Appl. Microbiol. Biotechnol. **51**: 786-793.

Tsuchida, T. and H. Momose. Genetic changes of regulatory mechanisms occurred in leucine and valine producing mutants derived from *Brevibacterium lactofermentum* 2256. Agric. Biol. Chem. **39**: 2193-2198.

Vallino, J.J. 1991. Identification of branch-point restrictions in microbial metabolism through metabolic flux analysis and local network perturbations. Ph.D. Thesis. MIT, Cambridge.

Vallino, J.J. and G. Stephanolpoulos. 1993. Metabolic flux distribution in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. **41**: 633-646.

Vallino, J., and G. Stephanolopoulos. 1994. Carbon flux distributions at the glucose 6-phosphate branch point in *Corynebacterium glutamicum* during lysine overproduction. Biotechnol. Prog. **10**: 327-334.

Van Zyl, L., S. von Arnold, P. Bozhkov, Y.Z. Chen, U. Egertsdotter, J. MacKay, R.R. Sederoff, J. Shen, L. Zelena, and D.H. Clapham. 2002. Heterologous array analysis in

Pinaceae: hybridization of *Pinus taeda* cDNA arrays with cDNA from needles and embryogenic cultures of *P. taeda*, *P. sylvestris* or *Picea abies*. Comp. and Func. Gen. **3**: 306-318.

Vrljic, M., H. Sahm, and L. Eggeling. 1996. A new type of transporter with a new type of function: L-lysine export from *Corynebacterium glutamicum*. Mol. Microbiol. **22**: 815-826.

Walker, T.E., C.H. Han, V.H. Kollman, R.E. London, and N. Matwiyoff. 1982. 13C nuclear magnetic resonance studies of the biosynthesis by *Microbacterium ammoniaphilum* of L-glutamate selectively enriched with carbon-13. J. Biol. Chem. **257**: 1189-1195.

Wehrmann, A., B. Phillipp, H. Sahm, and L. Eggeling. 1998. Different modes of diaminopimelate synthesis and their role in cell wall integrity: a study with *Corynebacterium glutamicum*. J. Bacteriol. **180**: 3159-3165.

Weichert, W. and A.A. de Graaf. 1997. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol. Bioeng. **62**: 739-750.

Wendisch, V.F., M. Spies, D.J. Reinscheid, S. Schnicke, H. Sahm, and B.J. Eikmanns. 1997. Regulation of acetate metabolism in *Corynebacterium glutamicum*: transcriptional control of the isocitrate lyase and malate synthase genes. Arch. Microbiol. **168**: 262-269.

Wendisch, V.F., A.A. de Graaf, H. Sahm, and B.J. Eikmanns. 2000. Quantitative determination of metabolic fluxes during coutilization of two carbon sources: comparative analyses with *Corynebacterium glutamicum* during growth on acetate and/or glucose. J. Bacteriol. **182**: 3088-3096.

Wendisch, V.F. 2003. Genome-wide expression analysis in *Corynebacterium glutamicum* using DNA microarrays. J. Biotechnol. **104**: 273-285.

Willison, J.C. 1998. Pyruvate and acetate metabolism in the photosynthetic bacterium *Rhodobacter capsulatas*. J. Gen. Microbiol. **134**: 2429-2439.

Wiseman, S., R. Sathiyaa, E.H. Jorgensen, and M.M. Viyajan. 2003. PCB impact on two wild populations of Arctic char in northern Norway. 42nd Annual Meeting of the Canadian Society of Zoologists.

Wittmann, C. and E. Heinzle. 2001. Modeling and experimental design for metabolic flux analysis of lysine-producing *Corynebacteria* by mass spectroscopy. Metabol. Eng. **3**: 173-191.

Wittmann, C., H.M. Kim, and E. Heinzle. 2004. Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. Biotechnol. Bioeng. **87**: 1-6.

Wolfgang, J., P.G. Peters-Wendisch, J. Kalinowski, and A. Pühler. A *Corynebacterium glutamicum* gene encoding a two-domain protein similar to biotin carboxylase and biotin-carboxyl-carrier proteins. Arch. Microbiol. **166**: 76-82.

Young, M.R., B. Tolbert, and M.F. Utter. 1969. Pyruvate carboxulase from *Saccharomyces cereivisae*. Methods Enzymol. **13**: 250-258.

Zou, J., S. Young, F. Zhu, F. Gheyas, S. Skeans, Y. Wan, L. Wang, W. Ding, M. Billah, T. McClanahan, R.L. Coffman, R. Egan, and S. Umland. 2002. Microarray profile of differentially expressed genes in a monkey model of allergic asthma. Genome Biol. **3**: 1-13.

Zupke, C. and G. Stephanopoulos. 1995. Intracellular flux analysis in hybridomas using mass balances and in vitro 13C NMR. Biotechnol. Bioeng. **45**: 292-303.

Appendix

A: Abbreviations

2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
AAT	aspartate aminotransferase
ACCA	acetyl-CoA carboxylase beta subunit
ACCC	acyl-CoA carboxylase
ACCD	acetyl-CoA carboxylase beta subunit
ACCOA	acetyl-CoA
ACEA	isocitrate lyase
ACEB	malate synthase
ACEE	pyruvate dehydrogenase E1 component
ACET	acetone
ACK	acetate kinase
ACO	aconitase
ADP	adenosine diphosphate
AKG	α-ketoglutarate
ALA	alanine
AMP	adenosine monophosphate
AROA	EPSP synthase
AROB	dehydroquinate synthase
AROE	shikimate dehydrogenase
AROG	DHAB synthase
AROH	chorismate mutase
AROQ	dehydroquinate dehydratase
ASD	aspartate semialdehyde dehydrogenase
ASK	aspartokinase
ASP	aspartate
ASPALD	aspartate semialdehyde
BPG	bisphosphoglycerate
BIOA	7,8-diaminopelargonic acid aminotransferase
BIOB	biotin synthase
BIOD	dethiobiotin synthase
BIOF	8-amino-7-oxononanoate synthase
BIRA	biotin-(acetyl-coA carboxylase) ligase
BP	base pair
CAMP	cyclic adenosine monophosphate
CIT	citrate

CITA	citrate synthase
CITEA	citrate lyase subunit
CDNA	complementary deoxyribonucleic acid
COA	coenzyme A
DAPA	dihydrodipicolinate synthase
DAPB	dihydrodipicolinate reductase
DAPD	tetrahydropicolinate succinylase
DAPE	succinyl diaminopimelate desuccinylase
DAPF	diaminopimelate epimerase
DATP	deoxyadenosine triphosphate
DCTP	deoxycytidine triphosphate
DDH	meso-diaminopimelate dehydrogenase
DEVB	glucose-6-phosphate 1-dehydrogenase
DGTP	deoxyguanosine triphosphate
DHAP	dihydroxyacetone phosphate
DMF	dimethyl formamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNTP	deoxyribonucleotide triphosphate
DTSR	propionyl-coA carboxylase
DTTP	deoxythimidine triphosphate
DUTP	deoxyuridine triphosphate
E4P	erythrose 4-phosphate
EMP	Embden-Meyerhof-Parnas pathway
ENO	enolase
F6P	fructose 6-phosphate
FBA	fructose bisphosphate aldolase
FBP	fructose-1,6-bisphosphate
FBPC	fructose-1,6-bisphosphatase
FUM	fumarase
FUMT	fumarate
G3P	glyceraldehyde 3-phosphate
G6P	glucose 6-phosphate
GALU	UDP glucose pyrophosphorylase
GAP	glyceraldehyde 3-phosphate dehydrogenase
GC	gas chromatography
GDH	glutamate dehydrogenase
GND	phosphogluconate dehydrogenase
GPC	glucose phosphatase
GPI	phosphoglucoisomerase
GPM	phosphoglycerate mutase
GPT	alanine aminotransferase
GLN	glutamine
GLNA	glutamine synthetase
GLTA	citrate synthase
GLU	glutamate

GLUC	glucose
GLXR	glyoxylate bypass regulator
GLYOX	glyoxylate
GND	phosphogluconate dehydrogenase
H4D	tetrahydrodipicolinate
HCL	hydrochloric acid
HPLC	high pressure liquid chromatography
HR	hour(s)
ICD	isocitrate dehydrogenase
ISOCIT	isocitrate
ILV	acetohydroxyacid synthase
ILVB	transaminase B
ILVC	acetohydroxyacid isomeroreductase
ILVD	acetohydroxyacid dehydratase
LACT	lactate
LDH	lactate dehydrogenase
LL-DAP	succinyl-LL-diaminopimelate
LPDA	dihydrolipoamide dehydrogenase
LPDB	dihydrolipoamide dehydrogenase
LYS	lysine
LYSA	diaminopimelate decarboxylase
LYSE	lysine exporter
LYSG	lysine export regulator
MAL	malate
MALE	malic enzyme
M-DAP	meso-diaminopimelate
MDH	malate dehydrogenase
MFA	metabolic flux analysis
MNA	metabolic network analysis
MRNA	messenger ribonucleic acid
MS	mass spectrometry
MTBSTFA	N-methyl-N-(tert-butyldimethyl-silyl)trifluoroacetamide
NAOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NADP	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
OAC	oxaloacetate
OD	optical density
OPA	ortho-phthaldialdehyde
OTSA	trehalose phosphate synthase
OTSB	trehalose phosphatase
P5P	pentose 5-phosphate
PCK	phosphoenolpyruvate carboxykinase
PDHA	pyruvate dehydrogenase E1 component subunit
PDHB	pyruvate dehydrogenase E1 component subunit

PDHC	dihydrolipoamide acetyltransferase
PCR	polymerase chain reaction
PDH	pyruvate dehydrogenase
PEP	phosphoenolpyruvate
PFK	phosphofructokinase
PFKA	phosphofructokinase
PFKB	phosphofructokinase
PGK	phosphoglycerate kinase
PGI	glucose-6-phosphate isomerase
PGL	phosphogluconolactonase
PHE	phenylalanine
PHEA	prephenate dehydratase
PPC	phosphoenolpyruvate carboxylase
PPP	pentose phosphate pathway
PTA	phosphate acetyltransferase
PVDF	polyvinylidene fluoride
PYC	pyruvate carboxylase
PYK	pyruvate kinase
PYR	pyruvate
RBSK	ribokinase
RNA	ribonucleic acid
RPE	ribulose-phosphate 3-epimerase
RPI	ribose-5-phosphate isomerase
RRNA	ribosomal ribonucleic acid
S7P	sedoheptulose 7-phosphate
SDHA	succinate dehydrogenase subunit A
SDHB	succinate dehydrogenase subunit B
SDHC	succinate dehydrogenase subunit C
SDHD	succinate dehydrogenase subunit D
SSC	saline/sodium citrate buffer
SUC	succinate
SUCA	α-ketoglutarate dehydrogenase
SUCB	dihydrolipoamide succinyltransferase
SUCC	succinyl-CoA synthetase alpha subunit
SUCCOA	succinyl-CoA
SUCD	succinyl-CoA synthetase beta subunit
TAL	transaldolase
TBDMCS	tert-butyldimethyl-chlorosilane
TCA	tricarboxylic acid
THRA	homoserine dehydrogenase
THRB	homoserine kinase
THRC	homoserine synthase
TKT	transketolase
TPI	triosephosphate isomerase
TREHAL	trehalose
VAL	valine

B: Metabolic Network Analysis Reactions

Listed below are the reactions that were entered into the MNA flux analysis software package to define the metabolic network for the purposes of computer simulation. The reactions as written below include only those species that contain carbon originating from the medium substrate, as the algorithms of the software consider only the material balances and atomic positions of carbon, and not the production and consumption of other species such as energy-storing ATP and the reducing equivalent NADPH. Also, to reduce the complexity of the network and allow for the flux quantifications to be observable and error to be reduced, some network approximations are contained in the reaction list. As an example, the compounds ribose 5-phosphate and xylulose 5-phosphate have been combined into a single species labeled pentose 5-phosphate (P5P) in reactions 11. 12, and 14. Also, the four-reaction succinylation pathway of lysine biosynthesis has been reduced to a two-reaction pathway involving only the starting and ending materials, tetrahydropicolinate (H4D) and meso-diaminopimelate (M-DAP) respectively, and one intermediate, succinyl-LL-diaminopimelate (LL-DAP). In this way the number of distinct reactions and compounds associated with the pathway are made as small as possible while still distinguishing it from the alternative, parallel, dehydrogenation route of lysine synthesis that shares the same starting and ending species. For the complete description of all reactions directly involved in the metabolic network of the pathways considered in this work, see the Tables of section 2.

Also included is the equation used to relate biomass measurements to the flux of central carbon metabolites and other observable amino acid products. This equation has been adapted from Park *et al.* (1997).

1) Gluc + PEP → G6P + Pyr

2) G6P ↔ F6P

3) F6P → FBP

4) FPB ↔ DHAP + G3P

5) DHAB ↔ G3P

6) G3P ↔ BPG

7) BPG ↔ 3PG

8) 3PG ↔ 2PG

9) 2PG ↔ PEP

10) PEP ↔ Pyr

11) G6P → P5P + CO2

12) P5P + P5P ↔ S7P + G3P

13) S7P + G3P ↔ F6P + E4P

14) P5P + E4P ↔ F6P + G3P

15) Pyr → AcCoA + CO2

16) AcCoA + OAC → Cit

17) Isocit → AKG + CO2

18) AKG → SucCoA + CO2

19) SucCoA → Suc

20) Suc ↔ Fumt

21) Fumt ↔ Mal

22) Mal ↔ OAC

23) PEP + CO2 ↔ OAC

24) Pyr + CO2 ↔ OAC

25) G6P → Trehal

26) AKG ↔ Glu

27) Glu → Gln

28) Lact → Pyr

29) Pyr + Glu → Ala + AKG

30) Pyr + Pyr → Val + CO2

31) AcCoA → Acet

32) OAC + Glu ↔ Asp + AKG

33) Asp ↔ Aspald

34) Aspald + Pyr → H4D

35) H4D → LL-DAP

36) H4D → M-DAP

37) LL-DAP → M-DAP

38) M-DAP → Lys + CO2

39) PEP + E4P → DHAP

40) PEP + DHAP → Phe + CO2

41) Cit → Isocit

42) Isocit → Glyox + Suc

43) Glyox + AcCoA → Mal

0.021 G6P + 0.007 F6P +0.09 P5P +0.036 E4P + 0.013 G3P + 0.15 3PG + 0.052 PEP + 0.03 Pyr + 0.332 ACCOA + 0.08 Asp + 0.033 Lys + 0.446 Glu + 0.025 Gln + 0.054 Ala + 0.04 Val → Biomass + 0.364 AKG + 0.143 CO2