UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Model-driven development and utilization of an oxygen-independent platform strain

Permalink

https://escholarship.org/uc/item/7n78s64z

Author

Portnoy, Vasiliy A.

Publication Date

2010

Supplemental Material

https://escholarship.org/uc/item/7n78s64z#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Model-driven development and utilization of an oxygen-independent platform strain.

A dissertation submitted in partial satisfaction of the requirement of

the degree Doctor of Philosophy

by

Vasiliy A. Portnoy

Committee in charge:

Professor Bernhard Ø. Palsson, Chair
Professor Douglas Bartlett
Professor Michael Heller
Professor Andrei L. Osterman
Professor Shankar Subramaniam
Professor Kun Zhang

2010

Copyright

Vasiliy A. Portnoy, 2010

All rights reserved

The dissertation of Vasiliy A. Portnoy is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2010

DEDICATION

This dissertation is dedicated to wife, Victoria, and my kids, Misha and Mia for their patience, support, and inspiration through the years that resulted in this document, and for that I will always be grateful.

TABLE OF CONTENTS

SIGNATURE PAGE iii
DEDICATION iv
TABLE OF CONTENTS v
LIST OF FIGURES x
LIST OF TABLES xiii
ACKNOWLEDGEMENT xv
VITA xix
PUBLICATIONS xix
ABSTRACT OF THE DISSERTATION xxi

CHAPTER I: INTRODUCTION 1

1.1 METABOLIC ENGINEERING 2
1.2 SYSTEMS BIOLOGY 4
1.3 METABOLIC ENGINEERING BASED ON HIGH-THROUGHPUT TECHNOLOGIES 6
1.4 UTILIZING SINGLE HIGH-THROUGHPUT TECHNOLOGIES 7
1.5 METABOLIC ENGINEERING BASED ON INTEGRATION OF HIGH-THROUGHPUT DATA SETS 9
1.6 SYNTHETIC BIOLOGY 11
1.7 CONCLUSION 12
1.8 ACKNOWLEDGEMENTS 13
1.9 REFERENCES 14

CHAPTER II: SYSTEMS BIOLOGY APPROACH TO SOLVING METABOLIC ENGINEERING CHALLENGES 23

2.1 GENOME-SCALE METABOLIC MODELS 23
2.2 METABOLIC ENGINEERING BASED ON GENOME-SCALE MODELS 25

2.3 MODEL-DRIVEN STRAIN DESIGN FOR METABOLIC ENGINEERING APPLICATIONS 26
2.4 *IN SILICO* STRAIN DESIGN ALGORITHMS AND METHODOLOGIES 27
2.5 METABOLIC ENGINEERING AIDED BY THE *IN SILICO* APPROACHES 29
2.6 ALTERNATIVE IN SILICO DEVELOPMENTS TO AID METABOLIC ENGINEERING 31
2.7 INTEGRATION OF GENOME-SCALE MODELS WITH HETEROGENEOUS DATA 33
2.8 METABOLIC ENGINEERING AIDED BY ADAPTIVE EVOLUTION 35
2.9 DISSERTATION OVERVIEW 36
2.10 ACKNOWLEDGEMENTS 39
2.11 REFERENCES 40

CHAPTER III: DEVELOPMENT OF THE OXYGEN-INDEPENDENT PLATFORM STRAIN 54

3.1 INTRODUCTION 55
3.2 RESULTS 57
3.2.1 Strain construction and growth adaptation 57
3.2.2 Initial phenotypic characterization and adaptive evolution 57
3.2.3 Phenotypic characterization of the evolved populations 59
3.2.4 Clonal Analysis 60
3.2.5 Oxygen dependency of ECOM strains 61
3.3 DISCUSSION 63
3.3.1 *In silico* analysis using a genome-scale model 63
3.3.2 Oxygen-independent platform strain 64
3.3.3 Phase plane and gene deletion analysis using metabolic model 65
3.4 CONCLUSION 67
3.5 MATERIALS AND METHODS 68
3.5.1 Strains and media 68
3.5.2 Generation of mutant strains 69

3.5.3 Adaptive evolution 69
3.5.4 Phenotype assessment 70
3.5.5 Quantitative PCR 71
3.5.6 Clonal analysis 73
3.5.7 Computational analysis 74
3.6 ACKNOWLEDGEMENTS 74
3.7 REFERENCES 76

CHAPTER IV: PHYSIOLOGY OF THE OXYGEN-INDEPENDENT PLATFORM STRAIN 96

4.1 INTRODUCTION 96
4.2 RESULTS 98
4.2.1 Strain engineering and adaptive evolution 98
4.2.2 Phenotypic characterization revealed substantial similarity between aerobic and anaerobic ECOM4 199
4.2.3 Gene expression analysis reveals a shift to anaerobic metabolism in ECOM4LA under oxic conditions 100
4.2.4 Gene expression suggests ArcA is active in aerobic ECOM4LA, while FNR is not 103
4.2.5 Quinone pool is dominated by menaquinones in ECOM4LA under oxic condition 105
4.2.6 Targeted gene expression measurements 106
4.2.7 Carbon labeling experiments 106
4.3 DISCUSSION 110
4.3.1 Aerobic ECOM4LA shows anaerobic gene expression 112
4.3.2 Aerobic ECOMLA4 uses anaerobic respiration 112
4.3.3 Anaerobic regulator ArcA is active in ECOM4LA during oxic growth 113
4.3.4 ^{13}C analysis complements gene expression data 114
4.3.5 Similar physiological behavior under oxic and anoxic conditions 115
4.4 MATERIALS AND METHODS 116

4.4.1 Strains and media 116
4.4.2 Adaptive evolution 116
4.4.3 Phenotype assessment 117
4.4.4 Clonal analysis 118
4.4.5 Transcriptome analysis 119
4.4.6 Quantitative PCR analysis 120
4.4.7 ^{13}C tracing studies 121
4.4.8 Quinone extraction 125
4.5 ACKNOWLEDGEMENTS 126
4.6 REFERENCES 127

CHAPTER V: PRODUCTION OF D-LACTATE AT pH 7.0 AND pH 4.5 146

5.1 INTRODUCTION 146
5.2 RESULTS 148
5.2.1 Production of high purity D-lactic acid under oxic and anoxic conditions at neutral pH 148
5.2.2 Growth assessment of ECOM4LA strain at different pH before and after the adaptation 149
5.2.3 Assessment of D-lactic acid production before and after adaptation to low pH 150
5.3 DISCUSSION 150
5.3.1 Succinate secretion during mid-log growth phase 151
5.3.2 Volumetric productivities associated with D-lactate production under neutral and low pH 151
5.3.3 Mechanism of adaptation and overall observations 152
5.4 MATERIALS AND METHODS 153
5.4.1 Strains and media 153
5.4.2 Adaptive evolution 153
5.4.3 Phenotype assessment 154
5.4.4 Fermentation 155
5.5 ACKNOWLEDGEMENTS 156

5.6 REFERENCES 157

CHAPTER VI: REDOX-COUPLED PRODUCTION OF ORGANIC AND AMINO ACIDS USING THE RECOMBINANT ECOM4 PLATFORM 164

6.1 INTRODUCTION 164
6.2 RESULTS 166
6.2.1 Production of high purity racemic lactate mixture by recombinant ECOM4LA strain 166
6.2.2 Production of high purity L-alanine using recombinant ECOM4A platform 167
6.3 DISCUSSION 167
6.4 MATERIALS AND METHODS 169
6.4.1 Strains and media 169
6.4.2 Cloning 169
6.4.3 Phenotype assessment 170
6.4.4 Fermentation 171
6.5 ACKNOWLEDGEMENTS 172
6.7 REFERENCES 173

CHAPTER VII: CONCLUDING REMARKS AND FUTURE DIRECTIONS 180

7.1 INTRODUCTION 180
7.2 SUMMARY OF RESULTS 181
7.3 FUTURE DIRECTIONS 185
7.4 REFERENCES 186

LIST OF FIGURES

Figure 1.1: Metabolic engineering based on systematic approaches, high-throughpt techniques and genome-scale models. 20

Figure 1.2: Relation of hig-throughput “omic” measurements to the central dogma of molecular biology 21

Figure 1.3: High-throughput methodologies underlying metabolic reconstructions and properties of the reconstructed networks 22

Figure 2.1: Uses of *E. coli* metabolic reconstructions. 48

Figure 2.2: The concept of growth-coupled strain designs. 49

Figure 2.3: Metabolic engineering of *E. coli* for improved succinic acid production, based on comparative genome analysis between *M. succiniciproducens* (A) and *E. coli* (B), and combinatorial *in silico* simulation. 51

Figure 2.4: Uses of laboratory scale adaptive evolution. 52

Figure 2.5: Various demonstrative studies of metabolic engineering based on systematic approaches including studies the use either high-throughput technologies (X-omics studies) or genome-scale models or both. 53

Figure 3.1: The overall experimental design, and nomenclature. 86

Figure 3.2: Evolutionary trajectories of ECOM3 populations. 87

Figure 3.3: Growth rate measurements for three evolved ECOM3 populations as a function of EZ Supplements addition (ml). 88

Figure 3.4: Aerobic growth and secretion profiles of three end-point populations. ... 89

Figure 3.5: Mechanism of *D*-lactate production and associated gene expression analysis. 90

Figure 3.6: D-lactate concentration from clonal analysis.....91

Figure 3.7: Anaerobic growth and secretion profile of three endpoint population……..... 92

Figure 3.8: Proposed mechanism of oxygen utilization by ECOM3 strains and corresponding gene expression**.**.....93

Figure 3.9: Comparison of experimental and computationally predicted growth rates and lactate secretion rates for the evolved ECOM3 strains.....94

Figure 3.10: Lactate secretion as a function of substrate uptake.....95

Figure 4.1: Phenotypic characteristics of ECOM4 populations during evolution. ...138

Figure 4.2: Batch fermentation of ECOM4LA strain**.**.....139

Figure 4.3: Gene expression analysis. 140

Figure 4.4: Transcriptomics analysis of the ECOM4LA and MG1655 strains under oxic and anoxic conditions. 141

Figure 4.5: Global gene expression suggests that ArcA is active in ECOM4LA in aerobic conditions.....142

Figure 4.6: Quinone pool content (nmol/gDW) under various environmental conditions. 143

Figure 4.7: Metabolic flux distribution through branching areas of the central metabolism of MG1655 and ECOM4LA cell lines. 144

Figure 4.8: Respiratory chain rearrangements in ECOM4LA strain compared to MG1655 and the effect of the content of the quinone pool on ArcA activity.....145

Figure 5.1: Batch fermentation of ECOM4LA strain.....161

Figure 5.2: Assessment of the growth rate of ECOM4LA at different pH before and after the adaptive evolution. 162

Figure 5.3: Evolutionary trajectory of the ECOM4LA strain during the 100 day adaptation to low pH environment. 163

Figure 6.1: Metabolic flux distribution in the recombinant ECOM4 strain capable of overproducing the racemic mixture of lactic acid. 174

Figure 6.2: Composition of racemic mixture depends on the amount of IPTG 175

Figure 6.3: Metabolic flux distribution in the recombinant ECOM4 strain capable of overproducing L-Alanine. 176

Figure 7.1: Used of ECOM4 strain as a redox-coupled production platform. 187

LIST OF TABLES

Table 3.1: Phenotypic characteristics of ECOM3 populations during adaptive evolution 80

Table 3.2: Phenotypic comparison of ECOM3 populations between the oxic and anoxic growth environment 81

Table 3.3: Physiological data for the ECOM3 and ECOM4 strains under oxic conditions 82

Table 3.4: List of the double reaction deletions of central metabolism in the *i*AF1260 that produce optimal phenotypes similar to the ECOM3. 83

Table 3.5: Strains and plasmids used in this study 84

Table 3.6: Sequences of the primers used for gene deletion of *cydAB*; *cyoABCD,* and *cbdAB* operons as well as sequences of primers used to confirm the deletions…… ... 85

Table 4.1: Physiological characteristics during 30 day adaptive evolution of ECOM4 strains…..... 132

Table 4.2: Phenotypic characteristics of ECOM4LA strain under oxic and anoxic conditions 133

Table 4.3: Comparison between gene expression levels in ECOM4LA and MG1655 cells grown aerobicaly and anaerobically..... 134

Table 4.4: Whole-genome expression and qPCR comparison for selected genes in the ECOM4LA strain 135

Table 4.5: Physiological parameters inferred from ^{13}C labeling data. 136

Table 4.6: DNA primers used for the deletion of *ygiN* 137

Table 5.1: Pysiological characteristics of the evolved ECOM4LA strain at pH 7.0 and pH 4.5 following a 100 day period of adaptative evolution 160

ACKNOWLEDGEMENTS

It is easy to accumulate a great deal of gratitude over the course of 5 years of graduate work! Systems biology is an incredible field and I had a great pleasure to work with a large group of incredibly talented and generous people at the University of California, San Diego.

First of all, I would like to acknowledge Dr. Bernhard Ø. Palsson as a helpful adviser and an excellent scientist. Professor Palsson, gave me the opportunity to pursue the almost every crazy idea that I came up with. He played a major role in exposing me to the "crème of the crop" of the scientific community ether by facilitating scientific collaborations or by sending me to a variety of conferences. By the way, the Metabolic Engineering conference in Jeju, was a blast.

My committee members: Dr Shankar Subramaniam, Dr. Michael Heller, Dr. Kun Zhang, Dr. Douglas Bartlett, and Dr. Andrei Osterman, were extremely generous with their time and talents, and I thank them collectively for helping me to develop as a scientist and for providing important suggestions for my work and thesis.

My thanks also goes to all the people in the Systems Biology Research Group, through the years, in particular to Dr. Markus Herrgard for years of friendship and guidance. I started working with Markus during my senior undergraduate year at UCSD, Markus taught me how to "think science", if there is a term like this. He taught me how to setup experiments, analyze the data, and ask the "right" questions in order to generate hypothesizes. I have gained a lot of knowledge from out work together and for that I will always be grateful. I am extremely grateful to my friend and colleague

Karsten Zengler, who not only provided an endless supply of coffee but also facilitated a number of scientific and not-so-scientific discussions that were highly educational. I am also thankful to Yekaterina Tarasova, who worked with me during the major part of my graduate career and helped me with almost every project that is presented in this thesis. I am grateful to her for all the time and devotion that she invested into this work, and I wish here all the best in life. BK Cho, Neema Jamshidi, Andrew Joyce, Yuri Matusov and others have been great friends and colleagues and our common interests have made this experience more enjoyable. Eric Knight for putting up with all my complains and his assistance in the experimental techniques.

I will always be indebted for the support and unconditional love from my beautiful wife, Victoria Portnoy and my kids Misha and Mia. Victoria has been an ultimate source of my mental and physical strength, and a great friend through these years. Without her support and help the completion of this work would have been almost impossible.

I don’t have words to describe my appreciation for my parent’s support and sacrifices through all these years, in particular for providing an opportunity for me to come to the United States and get educated here. My Dad and Mom have been an example of hard work and perseverance, and great role models.

The work presented in this document would not be possible without all the collaborators who put their efforts and determination into it.

The text of the chapter 1, in part or in full, is a reprint of material as it appears in Markus J. Herrgard, Vasiliy A. Portnoy, Hyun Uk Kim, Hongseok Yun, Seung Bum Sohn, Jin Sik Kim, Bernhard Palsson*, Sang Yup Lee*,. "Genome-Scale Models, Systems Biology, and Metabolic Engineering" In the Metabolic pathway engineering handbook: Tools and Applications. Editor Christina D. Smolke, CRC Press, FL, 2010. The dissertation author was the secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of the chapter 2, in part or in full, is a reprint of material as it appears in Markus J. Herrgard, Vasiliy A. Portnoy, Hyun Uk Kim, Hongseok Yun, Seung Bum Sohn, Jin Sik Kim, Bernhard Palsson*, Sang Yup Lee*,. "Genome-Scale Models, Systems Biology, and Metabolic Engineering" In the Metabolic pathway engineering handbook: Tools and Applications. Editor Christina D. Smolke, CRC Press, FL, 2010. The dissertation author was the secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

I thank Eric M. Knight and Larisa Kagan for help with generation of *cydAB E. coli* mutant strain, Ines Thiele, Karsten Zengler and Dr. Byung Kwan Cho, and Dae-Hee Lee for valuable discussions. The text of the chapter 3, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Herrgård, M.J., Palsson, B.Ø. Aerobic Fermentation of D-Glucose by an Evolved Cytochrome Oxidase Deficient Escherichia coli Strain, *Appl Environ Microbiol*. 2008 Dec;74**(**24):7561-9. The dissertation author was the primary researcher and author and the co-authors listed in

this publication directed and supervised the research which forms the basis for this chapter.

I thank Jan Schellenberger for useful discussions regarding metabolic fluxes, and Yuri Matusov, Marc Abrams, Dr. Daniel Hyduke, and Karsten Zengler for helpful discussions and critical revision of this chapter. The text of the chapter 4, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Scott D.A., Lewis, N.E., Tarasova Y., Osterman A. L., Palsson, B.Ø. Deletion of cytochrome oxidase and quinol monooxygenase blocks aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010 Oct;76(19). The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of the chapter 5, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Tarasova, Y., Palsson, B.Ø. Redox-coupled production of organic and amino acids using *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010, *In preparation*. The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of the chapter 6, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Tarasova Y., Palsson, B.Ø.Redox-coupled production of organic and amino acids using *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010, *In preparation.* The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

VITA

2003 - 2005	B.S., University of California, San Diego
2005 - 2008	M.S., University of California, San Diego
2008 - 2010	Ph.D., University of California, San Diego

PUBLICATIONS

Portnoy, V.A., Palsson, B.Ø., *Redox-coupled Production of Organic and Amino Acids in Escherichia coli*. **Applied and Environmental Microbiology**, 2010, In preparation.

Qiu, Y., Ingeniis, J., Portnoy, V.A., Cho, B.K., Lovley, D., Palsson, B.Ø., Zengler, K., *Regulation and Control of Nitrogen Assimilation in Geobacter sulfurreducens.* **The Journal of Biological Chemistry**, 2010, In preparation.

Portnoy, V.A., Palsson, B.Ø., *Chapter 7: Orders of Magnitude*. In **Systems Biology: Simulation of Dynamic Network States**., 2010.

Portnoy, V.A., Scott, D.A., Lewis, N.E., Tarasova, Y., Osterman, A.L., Palsson, B.Ø., *Deletion of cytochrome oxidases and quinol monooxygenase blocks aerobic-anaerobic shift in Escherichia coli K-12 MG1655.* **Applied and Environmental Microbiology**, 2010, Oct;76(19).

Portnoy, V.A., Herrgård, M.J., Palsson, B.Ø., *Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain.* **Applied and Environmental Microbiology**, 2008 Dec; 74(24):7561-9.

Shlomi, T., Herrgård, M.J., Portnoy, V.A., Naim, E., Palsson, B.Ø., Sharan, R., Ruppin, E., *Systematic condition-dependent annotation of metabolic genes.* **Genome Research**, 2007 Nov; 17(11):1626-33.

Herrgård, M.J., Portnoy, V.A., Kim, H.U., Yun, H., Sohn, S.B., and Kim, J.S., Lee, S.Y., Palsson, B.Ø., *Systems Biology, Genome-Scale Models, and Metabolic Engineering.* In **Metabolic Engineering**. Edited by Smolke, K., 2007.

Herrgård, M.J., Lee, B.S., Portnoy, V.A., Palsson, B.Ø., *Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae*. **Genome Research**, 2006 May; 16(5):627-35.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S.,

Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V.A., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V., *The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.* **Nucleic Acids Research**, 2005 Oct; 33(17):5691-702.

ABSTRACT OF THE DISSERTATION

Model-driven development and utilization of an oxygen-independent platform strain

by

Vasiliy A. Portnoy

Doctor of Philosophy in Bioengineering

University of California, San Diego

Professor Bernhard Ø. Palsson, Chair

The microbial production of commodity chemicals is a promising avenue for the development of sustainable processes for the utilization of renewable resources and reducing our dependency on foreign oil. In order to become cost and energy effective, the process must utilize an organism that is optimized for production of a number of reduced by-products from variety of feedstocks. *Escherichia coli* is one of the most commonly used host organisms for metabolic engineering and overproduction of metabolites due to its metabolic versatility, amenability to genetic manipulation, and the ability to produce a wide variety of reduced by-products such as bio-ethanol and organic acids. *E. coli* has also been extensively characterized with respect to its metabolic physiology. It is capable of surviving in a variety of

environmental conditions, such as oxic and anoxic; however the different growth rates and different secretion products under oxic anc anoxic conditions poses a significant challenge for metabolic engineering processes in which environmental perturbations will influence the outcome of the bio-catalytic process. Therefore, the utilization of the oxygen-independent strain for bio-catalysis eliminates the need for the stringent control over the fermentation environment with respect to oxygenation, thus significantly reducing the cost of the entire bio-catalytic process. Therefore, it is of interest to develop an *E. coli* strain incapable of oxygen utilization, to be used as a platform strain for metabolic engineering.

Here, I present the work aimed at the (i) development of an oxygen-independent platform strain, (ii) understanding of its physiological behavior, and (iii) utilization of this strain for metabolic engineering applications. Such a strain can be useful for the overproduction of commodity chemicals under various conditions independent of oxygen supply and optimization of anaerobic metabolic engineering designs using adaptive evolution under oxic conditions. The results show that upon the removal of the oxygen-utilization pathway, the ECOM4 (*Escherichia coli* Cytochrome Oxidase Mutant 4) strain was unable to undergo an aerobic-anaerobic shift and exhibited similar phenotypes under both conditions with D-lactic acid as a sole growth-associated by-product. Moreover, I show that the ECOM4 strain can be used for the overproduction of organic and amino acids from renewable resources.

CHAPTER I:

INTRODUCTION

What I cannot create, I do not understand.
Richard Feynman

The field of "biochemical engineering" first emerged in the middle of the 20th century. That was the time when increased production of penicillin was needed, and the research has lead to the development of a first aerobic submerged culture, increasing the production of penicillin, required to cure battle wounds of World War II [1]. Later, with the discovery of the DNA structure and the emergence of the genetic tools, the term "biochemical engineering" was frequently substituted by the term "metabolic engineering", which was defined by James E. Bailey in 1991as:

> Metabolic engineering is the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology.[2]

The goal of this chapter is to illustrate the transitions from biochemical to metabolic engineering and the emergence of synthetic biology, as a new cutting edge research and technology field. This chapter will present various tools available for the successful metabolic engineering and focus on some examples where these tools have lead to remarkable developments and discoveries. I will talk about metabolic engineering in the content of current genetic technology such as high-throughput approaches to gathering genomic data as well as in the content of adaptive evolution.

Furthermore, I will discuss the emergence of the systems biology, aimed at integration of individually characterized components and understanding functionality of the biological system as a whole. And lastly, I will present the fusion between the "classical" metabolic engineering and systems biology leading to an emergence of synthetic biology and advances associated with this field.

1.1 METABOLIC ENGINEERING

The concept of utilization of biological systems to produce products is a very old one. We have many examples of successful production of antibiotics [3-5], alcohols [6-8], fine chemicals[9-11], etc. Early metabolic engineering approach was highly dependent on traditional mutagenesis and selection. Although creative ideas for identification and selection of mutant strains for production purposes were employed, the overall approach was one of trial and error [12]. Regardless of acceptance and success of the resulted mutants, this "black box" approach was a random process, often without a complete understanding of the biological process.

Progress in science is often driven by the progress in technology. Molecular biology, for example, entered a new era when James Watson and Francis Crick announced in 1953 that they had "found the secret of life", and presented the double stranded helical structure of the DNA molecule [13]. This discovery has facilitated advances in the technological methods aimed at manipulation of DNA in order to introduce latent functionalities into the bacterial and mammalian organisms. In the early 1970th another remarkable and ground breaking discovery was announced.

Recombinant DNA technique [14] was a new kid on the block, made possible by the discovery, isolation and application of restriction nucleases by Werner Arber and colleagues [15]. Recombinant DNA technique allowed researchers to express the foreign DNA molecules in cells. This initiated the use of bacteria as a factory to produce non-native products. These discoveries were the foundation stones for the new era of molecular biology and genetic manipulation, and on top of these stones was emerging a new field of metabolic engineering.

In 1991, a formal definition for the field of metabolic engineering was finally proposed by James Bailey, he stated that: Metabolic engineering is the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology. [2]

One of the key features of metabolic engineering is the modification of metabolic networks. These manipulations are often archived by using recombinant DNA technologies and chemical engineering methods [12]. However, due to the complexity of metabolic networks and their regulation, modifications of metabolic pathways can have unpredictable consequences that may hamper achieving the original engineering goals. Prior to the advent of systemic approaches, metabolic engineering relied on intuition and biochemical knowledge for the selection of metabolic pathways for manipulation. Results of this approach were often unexpected and the resulting strains required extensive fine tuning to yield viable production strains. Implementation of complicated metabolic engineering designs involves genetic modifications that are associated with significant phenotypic changes of the organism. Such changes can result in slower growth rates and production of unnecessary and potentially toxic

by-products among other complications [16]. Due to these issues, classical metabolic engineering approaches are often time consuming, labor intensive, and ineffective from an economical standpoint.

Recently, metabolic engineering began to rely extensively on large-scale screening and experimentation, and computational analysis of metabolic and regulatory networks (Figure 1.1). Presently, there is significant interest among scientists and engineers to study cells and microorganisms in the context of systems biology.

1.2 SYSTEMS BIOLOGY

In recent years, metabolic engineering has shifted its paradigm towards the implementation of systemic approaches focused on integration of different components of biological network, such as: metabolism, regulation, signaling, etc., in order to improve metabolic engineering designs.

Biologists have traditionally used a bottom-up approach for systems analysis; the system is broken down to its elementary components and based on their properties, systemic behavior is inferred. Although understanding the behavior and role the individual components in a particular biological system is critical, studying each component in isolation cannot give the full picture of how the system works. In order to fully understand the behavior of the system, the various components need to be studied simultaneously in an integrative fashion [17]. Systems biology seeks to integrate existing knowledge of the biology of a particular system with quantitative high-throughput measurements in order to elucidate how different subsystems affect each other and

function as a whole [18] (Figure 1.2).

With the advances made in high-throughput techniques, information on the molecular characteristics of cells is being generated at an increasing rate. As a result, methods capable of extracting valuable information from noisy large-scale datasets are necessary. Moreover, methods that are able to link information extracted from high-throughput datasets to cellular phenotypes must also be developed. For example, genetic data may identify specific alleles that increase susceptibility to certain diseases, but the data does not reveal the biological mechanisms that cause the increased susceptibility. It is only by combining the genetic information with knowledge of metabolic, regulatory and signaling network structures that allows determining how specific genetic variants cause the observed phenotypic consequences (i.e. disease susceptibility).

Large amounts of biological information are commonly used in metabolic engineering applications; this information is acquired from high-throughput data sets such as transcriptomic, proteomic, metabolomic and phenotypic data (Figure 1.2). While each of these data sets allows studying a particular facet of the overall microbial physiology, the data sets must be analyzed together in order to maximize the value extracted from the data. Systems biology seeks to achieve this aim by generating comprehensive models of biological networks that can be used as a framework for data integration in order to facilitate scientific discovery and hypothesis generation. Mechanisms such as alternative metabolic pathways, feedback effects in transcriptional regulation and signaling cross-talk can be represented and interrogated with such models. The ability to systematically account for these complex systems level

mechanisms can significantly improve our ability to engineer bacteria to produce desired bio-products.

Moreover, the advances in high-throughput technologies that include an ever-growing collection of –omics techniques such as transcriptomics, fluxomics, sequenceomics, proteomics, metabolomics, etc. enabled the building of comprehensive databases representing our knowledge of metabolic networks in specific organisms. The KEGG (Kyoto Encyclopedia of Genes and Genomes), the EcoCyc (A comprehensive view of *Escherichia coli* biology), the SGD (Saccharomyces Genome Database) are only a few examples of organism specific databases emerged in recent years[19-21]. These types of databases contain biochemical, molecular, and genomic information that can be used to enable more systematic and efficient metabolic engineering.

1.3 METABOLIC ENGINEERING BASED ON HIGH-THROUGHPUT TECHNOLOGIES

High-throughput technologies refer to experimental methods that produce genome-scale data on any of the molecular components (genes, proteins, transcripts, metabolites) or their interactions in an organism. These technologies increasingly play a critical role in metabolic engineering as they allow researchers to strategies that take into account complex interactions between the target metabolic pathways and all other cellular processes. Herein we provide a brief introduction of high-throughput methods focusing on their use in metabolic engineering applications followed by discussion of how integration of multiple data types can further facilitate metabolic engineering.

1.4 UTILIZING SINGLE HIGH-THROUGHPUT TECHNOLOGIES

High-throughput genome sequencing has allowed determining the gene repertoire of both prokaryotic and eukaryotic organisms [22-24]. The sequencing of a novel microbial genome can now be done routinely, and multiple databases storing annotated complete genome sequences exist [25,26]. Genomics has been widely used for the identification of novel biochemical activities present in a genome and discovery of new metabolic pathways. With the increasing number of fully sequenced organisms, comparative genomics has become an extremely useful tool for research, discovery, and metabolic engineering. Similarly, targeted sequencing of production strains created by mutagenesis followed by comparison with non-mutated wild type strain sequences can be used to identify specific beneficial mutations.

The comparative genomics approach has been demonstrated for *Corynebacterium glutamicum* for the purpose of lysine overproduction [27]. Here, the specific regions of the genome of an overproduction strain created through multiple rounds of mutagenesis were compared with the wild-type strain to identify specific mutations that increase production of lysine. This approach can also help in identifying target genes for further manipulation for metabolic engineering purposes. Metabolic engineering has traditionally involved insertion or deletion of genes to obtain a strain that can be used for bioprocessing applications and with the help of a fully sequenced genome, identifying target genes for such purposes is made much easier.

Transcriptomics, which makes use of high-density DNA microarrays, allows for

the parallel study of the relative abundance of mRNAs under different conditions and in different strains [28]. Transcriptome profiles can give insight on the metabolic and regulatory state of the cells as well as explain any physiological changes in the cell, thereby providing information about active pathways in specific conditions [29]. By comparative analysis of the transcriptome profiles, genes and potential mechanisms responsible for any physiological behaviors such as alcohol tolerance or changes in growth rate can be identified [30]. Transcriptomics was successfully used to improve the production of human insulin-like growth factor I fusion protein (IGF-If) in *E. coli* [31]. Here, the profile of gene expression during the production of IGF-If was obtained and genes that were down-regulated compared to the wild type strain were selected for overexpression. An additional problem was caused by slow growth rates during high cell density culture (HCDC) conditions required for practical applications. By studying the transcriptome profiles, it was possible to eliminate this problem and increase production of IGF-If though overexpression of a small number of genes.

Proteomics allows researchers to identify and quantity the levels of proteins present in an organism in a given condition [29]. This analysis is usually done by isolating the proteins expressed using 2D electrophoresis or other separation technologies and identifying them using mass spectroscopy [32,33]. Although proteome analysis has not yet been as widely utilized in metabolic engineering as genomic or transcriptomic methods, the information from the proteome analysis can provide researchers with additional insights on the activity of metabolic pathways in engineered strains. An example where proteome analysis was used for strain enhancement is with the recombinant *E. coli* strain for overproducing the human hormone leptin [34]. When the initial strain designed

for the overproduction of leptin was studied, researchers discovered that the levels of proteins responsive to heat shock increased and the levels of the enzymes present in the amino acid biosynthetic pathways decreased. Furthermore, it was observed that the enzymes in the serine family amino acid biosynthetic pathways were expressed at lower levels compared to other amino acid biosynthesis pathways. The overexpression of a key enzyme in the serine biosynthetic pathway was found to increase leptin production indicating that this pathway was a bottleneck for producing leptin.

Metabolomics is devoted to the identification and quantification of the concentrations of metabolites in a system. Techniques involving mass spectrometry, NMR spectroscopy, and various other automated tools [35] are used to create a profile of a subset of cellular metabolites. Due to the wide range of metabolite structure/composition, no single technique is capable of detecting and quantifying all the metabolites present in a cell. Therefore, a mixture of techniques, such as liquid chromatography-mass spectroscopy (LC-MS), needs to be employed, depending on the chemical and physical properties of the metabolites of interest. In spite of the limited success in employing metabolomics due to technical difficulties, its potential in a wide range of applications including metabolic engineering has lead to significant efforts to improve and standardize experimental methods [36,37].

1.5 METABOLIC ENGINEERING BASED ON INTEGRATION OF HIGH-THROUGHPUT DATA SETS

Typically the information obtained based on a single high-throughput data type

is not enough to fully characterize the behavior of an organism due to the nonlinear relationship that relates gene expression to enzymatic activities and metabolic fluxes. For example, protein levels and their activity are not directly proportional, and the same can be said about the relationship between gene and protein expression[38-43]. Therefore, a combination of different types of high-throughput data is necessary for complete understanding of a biological system within a particular physiological state [31,39,44,45].

Successful integration of multiple high-throughput data types was demonstrated in the case of high cell density culture (HCDC) of *E. coli* utilized for the production of various bio-products [31]. It was observed that during HCDC, the specific production rate (g product/ gDW/ h) decreased as the cell density increased. To understand why this phenomenon occurred, an integrated transcriptome and proteome analysis was conducted [31,45-47]. The transcriptome and proteome were both were both measured as a function of time during various phases of HCDC. The genes encoding TCA cycle enzymes as well as genes for NADH dehydrogenase and ATPase were up-regulated during the exponential phase and down-regulated upon entering the stationary phase, indicating a significant reduction in aerobic respiration at high cell densities [31]. Moreover, chaperone genes were found to be up-regulated suggesting that high cell density induces cellular stress. Surprisingly, a significant reduction in the expression of genes involved in amino acid biosynthesis was observed as cell density increased. Decreased availability of amino acids may then explain the decrease in the specific productivity during HCDC. With further study, the regulation of protein production as a function of cell density may be elucidated and strategies for increasing productivity can be developed.

1.6 SYNTHETIC BIOLOGY

All of the strategies and tools presented in this chapter suggest only one thing: the complete understanding of elementary components of biological systems is necessary in order to engineer a better microbe. This paradigm has led to active progress in the field of synthetic biology. Synthetic biology includes metabolic engineering, molecular biology, and systems biology and uses these tools to address biological chemical and environmental problems that we are faced with today. The term "synthetic biology" has a history spanning the twentieth century[48]. In 1974, the Polish geneticist Waclaw Szybalski introduced the term "synthetic biology"[49], writing:

> Let me now comment on the question "what next". Up to now we are working on the descriptive phase of molecular biology. ... But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with the unlimited expansion potential and hardly any limitations to building "new better control circuits" and..... finally other "synthetic" organisms, like a "new better mouse". ... I am not concerned that we will run out exciting and novel ideas ... in the synthetic biology, in general.

It is fair to say that we have entered the synthetic biology phase: in the June, 2010 in the scientific journal Science, researchers from the J. Craig Venter Institute presented the first “synthetic” cell [50]. It took over a decade of dedicated research in order to develop protocols to create a synthetic genome of simplest bacterial parasite *Mycoplasma mycoides*; however despite the challenges scientists succeeded and were able to “boot-up” the artificial genome inside the bacterial cell forcing it to replicate and

enable cell replication. This scientific achievement has will lead the way toward engineering useful microbes or other biological systems.

1.7 CONCLUSION

Systems biology can be considered to be the bridge that allows connecting the different aspects of biological function of a specific system to address various biological problems [51] (Figure 1.3). Systems biology methodologies can also be used in metabolic engineering, consequently enabling a systemic approach to engineering for example the overproduction of a metabolite. Modern systems biology includes a number of steps that lead to systemic description of e.g. a microbial cell [52]. Generally, the first step in the systemic approach is the acquisition of large-scale quantitative and qualitative data and the identification of the components of the biological system. Genome-scale data such as genomic sequence and transcript or protein profiles can then used to construct genome-scale models of metabolic as well as other types of networks. Once these models are built and verified experimentally, they can be used for various applications including metabolic engineering.

High-throughput technologies and genome-scale metabolic models will serve as powerful tools for metabolic engineering [53-55]. The scope and depth of these tools is growing rapidly to respond to the needs created by novel metabolic engineering challenges. This chapter has provided examples of metabolic engineering studies that utilize high-throughput data [27,31,33,56] as well as hypothesis-driven metabolic engineering studies using genome-scale models[57-59]. These types of methods are now

firmly established as a part of the metabolic engineer's toolbox where they complement both classical metabolic engineering approaches and other emerging technologies such as those developed in the growing field of synthetic biology. In the near future, systems biology-enabled metabolic engineering based on high-throughput technologies and genome-scale models will give rise to advancements in various scientific fields, such as medical research allowing for the discovery of novel drug targets for specific diseases [53,55] and synthetic biology which can be used to design useful microbes.

1.8 ACKNOWLEDGEMENTS

The text of this chapter, in part or in full, is a reprint of material as it appears in Markus J. Herrgard, Vasiliy A. Portnoy, Hyun Uk Kim, Hongseok Yun, Seung Bum Sohn, Jin Sik Kim, Bernhard Palsson*, Sang Yup Lee*,. "Genome-Scale Models, Systems Biology, and Metabolic Engineering" In the Metabolic pathway engineering handbook: Tools and Applications. Editor Christina D. Smolke, CRC Press, FL, 2010. The dissertation author was the secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

1.9 REFERENCES

1. Grossman CM. The first use of penicillin in the United States. *Ann Intern Med* 2009;150:737.

2. Bailey JE. Toward a science of metabolic engineering. *Science* 1991;252:1668-75.

3. Newton GL, Buchmeier N, Fahey RC. Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiol Mol Biol Rev* 2008;72:471-94.

4. Le Roes-Hill M, Goodwin C, Burton S. Phenoxazinone synthase: what's in a name? *Trends Biotechnol* 2009;27:248-58.

5. Rassi H, Ghasemi SF. Production of novel chemotherapeutic drugs by Streptomycetes. *Lik Sprava* 2008:40-7.

6. Parawira W, Tekere M. Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review. *Crit Rev Biotechnol*.

7. Smith KM, Cho KM, Liao JC. Engineering Corynebacterium glutamicum for isobutanol production. *Appl Microbiol Biotechnol*.

8. Connor MR, Liao JC. Microbial production of advanced transportation fuels in non-natural hosts. *Curr Opin Biotechnol* 2009;20:307-15.

9. Datta R, Henry M. Lactic acid: recent advances in products, processes and technologies - a review. *Journal of Chemical Technology & Biotechnology* 2006;81:1119-1129.

10. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007;3:149.

11. Altaras NE, Cameron DC. Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli. *Appl Environ Microbiol* 1999;65:1180-5.

12. Lee SY, Papoutsakis ET. Metabolic Engineering. CRC Press, 1999.

13. Watson JD, Crick FHC. Molecular structure of nucleic acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953;171:737.

14. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* 1973;70:3240-4.

15. Roberts RJ. Restriction endonucleases. *CRC Crit Rev Biochem* 1976;4:123-64.

16. Lee SJ, Song H, Lee SY. Genome-based metabolic engineering of Mannheimia succiniciproducens for succinic acid production. *Appl Environ Microbiol* 2006;72:1939-48.

17. Kitano H. Systems biology: a brief overview. *Science* 2002;295:1662-4.

18. Witkamp RF. Genomics and systems biology--how relevant are the developments to veterinary pharmacology, toxicology and therapeutics? *J Vet Pharmacol Ther* 2005;28:235-45.

19. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 2005;102:2685-9.

20. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 2002;420:190-193.

21. Borodina I, Krabben P, Nielsen J. Genome-scale analysis of Streptomyces coelicolor A3(2) metabolism. *Genome Res* 2005;15:820-9.

22. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;268:78-94.

23. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 1998;26:544-8.

24. Salzberg SL, Pertea M, Delcher AL, Gardner MJ, Tettelin H. Interpolated Markov models for eukaryotic gene finding. *Genomics* 1999;59:24-31.

25. Fraser-Liggett CM. Insights on biology and evolution from microbial genome sequencing. *Genome Res.* 2005;15:1603-1610.

26. Mount DW, Pandey R. Using bioinformatics and genome analysis for new therapeutic interventions. *Mol Cancer Ther* 2005;4:1636-43.

27. Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K, Ikeda M. A novel methodology employing Corynebacterium glutamicum genome information to generate a new L-lysine-producing mutant. *Appl Microbiol Biotechnol.* 2002;58:217-223.

28. Mills JC, Roth KA, Cagan RL, Gordon JI. DNA microarrays and beyond: completing the journey from tissue to cell. *Nat Cell Biol* 2001;3:E175-8.

29. Joyce AR, Palsson BO. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 2006;7:198-210.

30. Gonzalez R, Tao H, Shanmugam KT, York SW, Ingram LO. Global gene expression differences associated with changes in glycolytic flux and growth rate in Escherichia coli during the fermentation of glucose and xylose. *Biotechnol Prog* 2002;18:6-20.

31. Yoon SH, Han MJ, Lee SY, Jeong KJ, Yoo JS. Combined transcriptome and proteome analysis of Escherichia coli during high cell density culture. *Biotechnol Bioeng* 2003;81:753-67.

32. Patterson SD. Proteomics: beginning to realize its promise? *Arthritis Rheum* 2004;50:3741-4.

33. Han MJ, Lee SY. Proteome profiling and its use in metabolic and cellular engineering. *Proteomics* 2003;3:2317-24.

34. Han MJ, Jeong KJ, Yoo JS, Lee SY. Engineering Escherichia coli for increased productivity of serine-rich proteins based on proteome profiling. *Appl Environ Microbiol* 2003;69:5772-81.

35. Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst* 2005;130:606-25.

36. Rochfort S. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J Nat Prod* 2005;68:1813-20.

37. Sauer U. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* 2004;15:58-63.

38. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997;18:533-7.

39. Corbin RW, Paliy O, Yang F, Shabanowitz J, Platt M, Lyons CE, Jr., Root K, McAuliffe J, Jordan MI, Kustu S, Soupene E, Hunt DF. Toward a protein profile of Escherichia coli: comparison to its transcription profile. *Proc Natl Acad Sci U S A* 2003;100:9232-7.

40. Glanemann C, Loos A, Gorret N, Willis LB, O'Brien XM, Lessard PA, Sinskey AJ. Disparity between changes in mRNA abundance and enzyme activity in Corynebacterium glutamicum: implications for DNA microarray analysis. *Appl Microbiol Biotechnol* 2003;61:61-8.

41. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in Saccharomyces cerevisiae. *Mol Cell Proteomics* 2002;1:323-33.

42. Lee PS, Shaw LB, Choe LH, Mehra A, Hatzimanikatis V, Lee KH. Insights into the relation between mrna and protein expression patterns: II. Experimental observations in Escherichia coli. *Biotechnol Bioeng* 2003;84:834-41.

43. Mehra A, Lee KH, Hatzimanikatis V. Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. *Biotechnol Bioeng* 2003;84:822-33.

44. Jurgen B, Hanschke R, Sarvas M, Hecker M, Schweder T. Proteome and transcriptome based analysis of Bacillus subtilis cells overproducing an insoluble heterologous protein. *Appl Microbiol Biotechnol* 2001;55:326-32.

45. Lee JH, Lee DE, Lee BU, Kim HS. Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. *J Bacteriol* 2003;185:5442-51.

46. Hong SH, Kim JS, Lee SY, In YH, Choi SS, Rih J-K, Kim CH, Jeong H, Hur CG, Kim JJ. The genome sequence of the capnophilic rumen bacterium Mannheimia succiniciproducens. *Nat Biotech* 2004;22:1275-1281.

47. Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers ME, Blomquist PR, Martinez EJ, Monreal AW, Feibelman TP, Mayorga ME, Maxon ME, Sykes K, Tobin JV, Cordero E, Salama SR, Trueheart J, Royer JC, Madden KT. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nat Biotechnol* 2003;21:150-6.

48. Campos L. That Was the Synthetic Biology That Was. In: Schmidt M, Kelle A, Ganguli-Mitra A, Vriend H, eds. Synthetic Biology: The Technoscience and Its Societal Consequences: Springer Academic Publishing, 2010.

49. Szybalski W. In Vivo and in Vitro Initiation of Transcription. In: Kohn A, Shatkay A, eds. Control of Gene Expression. New York: Plenum Press, 1974:405-417.

50. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*;329:52-6.

51. Tadmor B, Tidor B. Interdisciplinary research and education at the biology-engineering-computer science interface: a perspective. *Drug Discov Today* 2005;10:1183-9.

52. Bork P, Serrano L. Towards cellular systems in 4D. *Cell* 2005;121:507-9.

53. Lee SY, Lee DY, Kim TY. Systems biotechnology for strain improvement. *Trends Biotechnol* 2005;23:349-58.

54. Borodina I, Nielsen J. From genomes to in silico cells via metabolic networks. *Curr Opin Biotechnol* 2005;16:350-5.

55. Aderem A. Systems biology: its practice and challenges. *Cell* 2005;121:511-3.

56. Choi JH, Lee SJ, Lee SJ, Lee SY. Enhanced production of insulin-like growth factor I fusion protein in Escherichia coli by coexpression of the down-regulated genes identified by transcriptome profiling. *Appl Environ Microbiol* 2003;69:4737-42.

57. Bro C, Regenberg B, Forster J, Nielsen J. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab Eng* 2006;8:102-11.

58. Lee SY, Hong SH, Moon SY. In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered Escherichia coli as an example. *Genome Inform* 2002;13:214-23.

59. Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 2005;23:612-6.

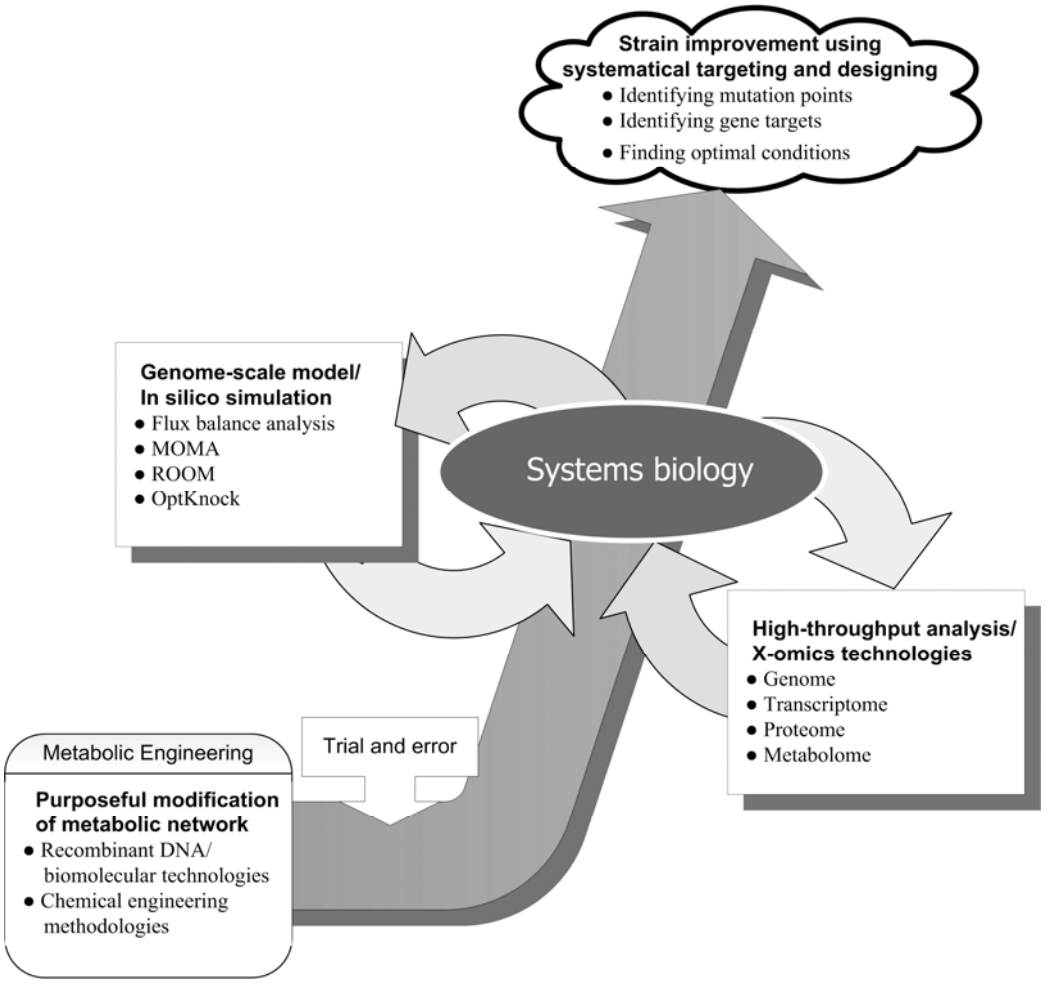


Figure 1.1: Metabolic engineering based on systematic approaches, high-throughput techniques, and genome-scale models. These approaches will contribute to more efficient metabolic engineering.

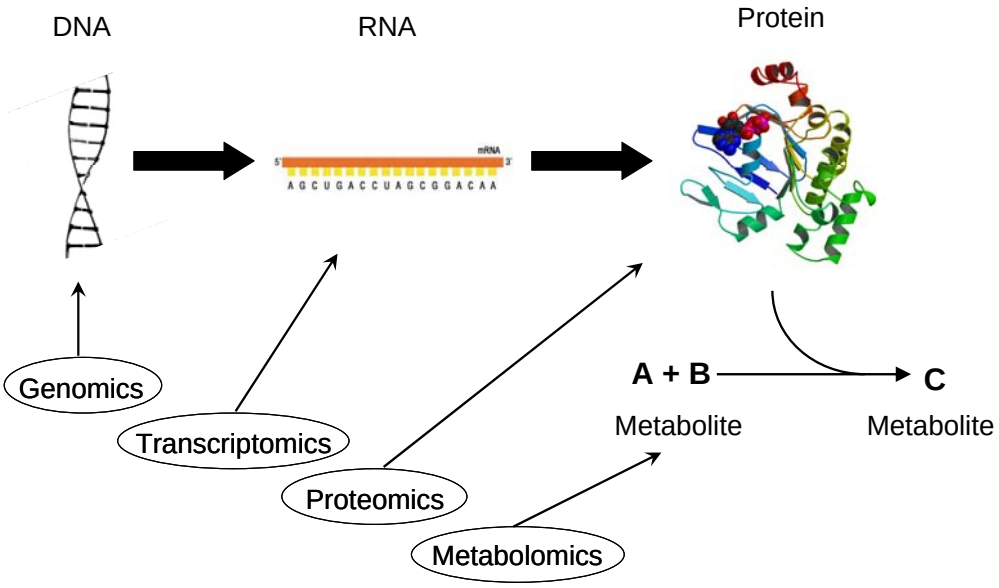


Figure 1.2: Relation of high-throughput "omic" measurements to the central dogma of molecular biology.

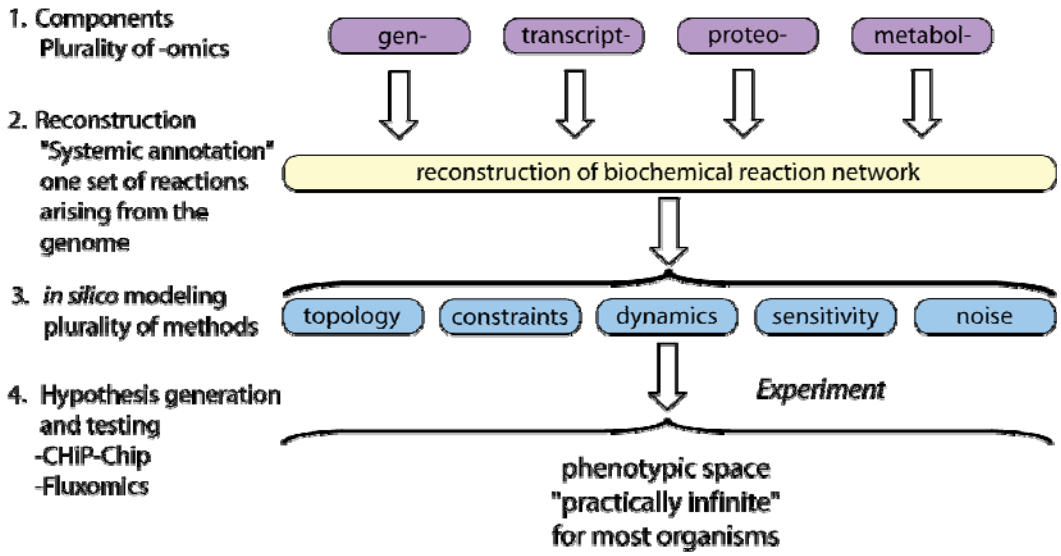


Figure 1.3: High-throughput methodologies underlying metabolic reconstructions and properties of the reconstructed networks

CHAPTER II:

SYSTEMS BIOLOGY APPROACH TO SOLVING METABOLIC ENGINEERING CHALLENGES

High-throughput experimental technologies are allowing us to obtain comprehensive data on the molecular make-up of cells and enabling us to reconstruct large-scale (and sometimes genome-scale) reaction networks. The resulting systems view of cellular functions demands the study of network-based properties. An effective approach for understanding systems biological properties is to define and explore biochemical pathways in the content of metabolic reconstruction that is a fundamental keystone for systems biology. This chapter is devoted to future prospects of metabolic engineering based on systems biology and genome-scale models with illustrations of successful case studies.

2.1 GENOME-SCALE METABOLIC MODELS

Biochemical network reconstructions provide a common denominator in systems biology (Figure 2.1). These reconstructions contain the most up-to-date information on molecular components and their interactions available for an organism,

and thus serve as both content databases and a basis for computational models. Genome-sequencing together with biochemical and physiological literature has enabled the development of genome-scale metabolic network reconstructions for industrially important micro-organisms including *E. coli*. Most applications of network reconstructions depend on the ability to convert these reconstructions to predictive models. The constraint-based reconstruction and modeling approach (COBRA) has been thoroughly reviewed previously and has been applied to a wide variety of organisms [3,4]. One of the most fundamental constraint-based applications is flux balance analysis (FBA) [9-11]. In this framework, system inputs and network stoichiometry limit the amounts of predefined objective targets that a system can produce (e.g., cellular biomass precursors and/or energetic units). COBRA approaches can be utilized to probe a number of cellular functions including the production of desirable by-products.

Applications of constraint-based modeling of *E. coli*[12,13] range from pragmatic to theoretical studies, and can be classified into general categories (Figure 2.1). Of particular relevance to this work are successful applications of constraint-based models in metabolic engineering [14-24] and in predicting phenotypic behavior [25-35]. Different types of applications of constraint-based models have also required development of novel *in silico* analysis methods [36] including a number of methods for model-driven strain design [15-19,21,23,37-41].

2.2 METABOLIC ENGINEERING BASED ON GENOME-SCALE MODELS

One of the biggest challenges in efforts to engineer overproduction of desired by-products is identifying genes that must be manipulated to successfully generate the desired phenotype. Choosing the most productive genetic manipulation strategies requires understanding how the altered metabolic pathways will function in the context of the whole system. The high-throughput data-based approaches described in Chapter I typically allow identifying potential bottlenecks in already existing engineered strains, but they are less useful when used in a prospective fashion to attempt to select genes to overexpress or delete. This prospective design phase of metabolic engineering is where genome-scale models of metabolic networks have shown great promise. As mentioned before the constraint-based modeling techniques such as flux balance analysis (FBA) can be applied to genome-scale metabolic models to gain better insight into the interplay between metabolic pathways within the *in vivo* system and to rapidly evaluate potential engineering strategies [9,45,46]. The increasing availability of high-quality genome-scale metabolic models together with the development of novel constraint-based modeling methods has lead metabolic engineers to increasingly apply such models in engineering microorganisms to overproduce commercially desirable metabolic products [38,49-51].

2.3 MODEL-DRIVEN STRAIN DESIGN FOR METABOLIC ENGINEERING APPLICATIONS

Through the use of computational methods that incorporate constraint-based modeling, it has been demonstrated that model-directed strain design can lead to increased metabolite production [14,15,17,21,23,39,40,42,52,53]. In these studies, the *E. coli* genome-scale model is principally used to analyze the metabolite production capabilities of *E. coli* and to identify metabolic interventions. *E. coli* strains have been systematically designed through *in silico* analysis to produce target metabolites such as ethanol [21], lycopene [19,42], lactic acid [20], succinic acid [38,43], L-valine [56], L-threonine[24], other amino acids [16], as well as diverse end-products from hydrogen to vanillin [17]. A growing number of metabolic engineering studies have demonstrated the use of genome-scale models to generate strain designs that are often non-intuitive and non-obvious. Genome-scale reconstructions are extremely useful because they enable the examination and simulation of metabolism as an integrated network. This circumvents the possible shortcomings of methods that rely on manual assessment of limited interactions and fail to detect non-intuitive causal interactions. With the growing availability of organism and strain specific models, methods for designing microbial strains for industrial production and other applications [57,58] are beginning to emerge.

2.4 *IN SILICO* STRAIN DESIGN ALGORITHMS AND METHODOLOGIES

Computational approaches to strain design have been developed and implemented with success. One group of methods that shows promise in implementation is those that couple metabolite production to cellular growth are introduced and briefly outlined:

1) *OptKnock*: Optknock [15] is a bi-level algorithm that suggests gene deletion strategies leading to the forced overproduction of a specified growth-coupled target metabolite. The OptKnock method relies on bi-level mixed-integer optimization that finds the optimal combination of gene deletions that allows maximizing the production rate of a bioproduct simultaneously with maximizing the biomass formation rate. The basic action of OptKnock algorithm is shown in Figure 2.2. The phase plane analysis indicates the solution space accounting for the production of the desired metabolite and growth of the bacterial host; however, it is clear that the living system will choose to grow at the maximum efficiency and not produce the desired metabolite (Figure 2.2A). An alternative solution (Figure 2.2B) is found where the metabolite production and the bacterial growth are balanced and the solution space is minimized due to the modifications of the network (i.e. gene deletions). This solution couples metabolite production to the bacterial growth forcing these two objectives to be satisfied at the same time. OptKnock has been computationally examined [14,15] and suggested strain designs have been experimentally verified with success [23].

2) *OptGene:* OptGene [59] is based on a genetic algorithm that can also produce growth-coupled strain designs. Its advantages include the potential for running at a higher speed than *OptKnock* and utilizing non-linear objectives. It has been tested using a genome-scale model of yeast, but has yet to be applied to engineer *E. coli* designs [60].
3) *OptStrain:* OptStrain [17] is a hierarchical computational framework incorporating mixed integer programming that identifies pathways that are targets for recombination of non-native pathways to host organisms. It is effectively similar to Optknock with the added feature that additional reactions can be added to the model to simulate a genetic addition to a cell (i.e., a gene insertion). For recombinant pathways, it chooses both the pathway that will produce the greatest potential yield and require the smallest number of genetic additions.

One of the most promising and widely used computational methods for designing metabolic overproducing strains is OptKnock. The advantage of using OptKnock is that resulting *in vivo* gene deletion strain could then be subjected to evolutionary engineering (i.e. continuous adaptation) and increased product formation would happen as by-product of increasing growth rate. This framework was applied to predict optimal gene deletions that allow growth-coupled overproduction of succinate, lactate, 1.3-propanediol [15], as well as amino acids [16], and the computational results were found to be in good agreement with experimental data collected from the literature.

2.5 METABOLIC ENGINEERING AIDED BY THE *IN SILICO* APPROACHES

Results obtained from genome-scale *in silico* model analysis frequently suggest metabolic engineering strategies that differ from those derived from simple inspection of the target pathways. For instance, *in silico* analysis may suggest overexpression of genes in pathways that were not considered as initial targets for engineering. This usually occurs due to the cofactor induced high level of interconnectivity between metabolic pathways [15]. Also, intracellular flux distributions obtained from simulating genome-scale models may suggest limiting factors for successful metabolic engineering, such as an inadequate supply of reducing agents or cofactors, or a pathway bottleneck [49,51].

So far, metabolic engineering with *in silico* experiments has mainly focused on the overproduction of bioproducts. Lee et al. [38] provide a typical example in which researchers identified the optimal combination of gene knockout targets that would improve succinic acid-production capability of an organism that does not naturally produce succinic acid in sufficient quantities (Figure 2.3). Researchers first used comparative genomics to identify genes that are present in *E. coli*, but are missing in the *natural* succinic acid producer *M. succiniciproducens*. These candidate gene targets for gene deletion were further investigated using combinatorial *in silico* knockout simulations with FBA of a genome-scale *E. coli* metabolic model. *In silico* analysis allowed identifying a set of multiple gene deletions that was predicted to result in succinic acid overproduction. The suggested genetic modifications were implemented in *E. coli* and fermentation data showed that the genetically modified

strain had significantly increased succinic acid production. The ability to quickly simulate the outcome of a particular genetic modification and monitor its secretion or production profile *in silico* is one of the powerful tools that systems biology provides to researchers.

It is likely that *in silico* experiments will play a key role in metabolic engineering applications beyond overproduction of metabolites that are native to the host organism. These applications include the production of new biologics that are not native to the host organism, broadening the substrate utilization range of an organism, designing novel biodegradation pathways, and modification of general cellular properties (e.g. stress tolerance) for the facilitation of bioprocess applications. For instance, in the case of production of non-native biological by a host organism, the corresponding novel metabolic pathways can be easily added *in silico* to the reconstructed metabolic network of the wild-type organism. The properties of the organism expressing a heterologous pathway can then at least in principle be investigated *in silico* using standard constraint-based modeling methods. Whether such models that include heterologous pathways can be successfully validated and employed for the metabolic engineering of the target organism remains to be seen. In addition to applications to microbial metabolic engineering, the use of genome-scale models has expanded to the field of biomedical engineering, where drug discovery can be aided by genome-scale models [61,62].

2.6 ALTERNATIVE *IN SILICO* DEVELOPMENTS TO AID METABOLIC ENGINEERING

Although FBA and methods based on it such as OptKnock have proven to be very useful in the analysis of metabolic networks, these methods have a number of shortcomings. The fluxes obtained from FBA are those that support maximal growth of the cell (or the maximization of some other cellular objective such as ATP) and thus do not necessarily reflect its true intracellular fluxes. The flux distributions obtained from FBA are typically also not unique as there may be multiple alternative pathways that the cell can utilize even in the optimal growth state. These issues become even more obvious in the case of simulations of gene deletion or over expression strain phenotypes as it is not expected that metabolism would operate optimally in such strains. For this reason a number of new algorithms for metabolic network analysis have been developed that relax the optimality assumption.

One such method is the Minimization of Metabolic Adjustment (MOMA) algorithm to simulate the phenotype of the gene knockout mutant, which is based on the hypothesis that mutants generated in the laboratory have not undergone enough adaptation to achieve the optimal growth phenotype predicted by FBA [63]. MOMA employs quadratic programming to search the point in the altered solution space of a mutant strain that is closest to the optimal point in the solution space of the wild-type strain. In this way, MOMA tries to minimize the flux redistribution of the mutant and thereby more realistically capture the phenotypic characteristics of the mutant. MOMA was employed in the work of Alper, et al., in which researchers identified a number of gene deletion targets that result in increased biosynthesis of lycopene in *E.*

coli [18]. When implemented *in vivo* these deletions contributed to increasing production of lycopene by 40% compared to the lycopene-producing parental strain.

Shlomi, et al. [64] reported another new algorithm called regulatory on/off minimization (ROOM) for the search of realistic flux distributions of the gene knockout mutant. They specifically focus on the metabolic steady-state of the cell after knockout mutants are introduced. Previous experiments showed that global regulatory changes due to the knockout mutations eventually converge to a steady-state that is close to the metabolic state of the wild-type [63]. Based on these findings, researchers attempted to minimize the total number of significant flux changes after gene knockout compared to the wild-type flux distribution. Thorough investigation of the three algorithms for deletion strain phenotype prediction (FBA, MOMA and ROOM) show that MOMA more accurately describes the initial transient growth rates after gene knockout, whereas ROOM and FBA provide more accurate predictions of the growth of the gene knockout mutant at the final optimal metabolic steady state. In terms of the flux distribution at the final metabolic steady state, ROOM is shown to be superior to FBA and MOMA. Although there has been no report yet on the utilization of ROOM in metabolic engineering applications, it will contribute to efficient metabolic engineering through more accurate analysis of intracellular flux distributions.

2.7 INTEGRATION OF GENOME-SCALE MODELS WITH HETEROGENEOUS DATA

It has been demonstrated that genome-scale metabolic models are able to predict growth phenotypes for gene deletion strains with 65-80% accuracy depending on the organism and growth condition [6,9,46]. However, it has been observed that the accuracy of the prediction decreases and observed growth characteristics deviate from the experimental data in cases when multiple gene knockouts are required [18,49]. Aside from the incompleteness of the metabolic network, this discrepancy is often thought to be caused by the regulatory effects that the genome-scale stoichiometric metabolic model fails to capture. Therefore, approaches have been developed to incorporate regulatory information *in silico* into the metabolic network modeling process to improve prediction accuracy.

Covert, et al incorporated data on transcriptional regulation into the metabolic model of *E. coli* in the form of a Boolean model of the known transcriptional regulatory network in *E. coli* [65]. In this regulatory network, genes can have only two states, either expressed or not expressed. Consequently, the reactions associated with particular genes that are inactivated under certain culture conditions can be constrained to zero flux and the metabolic model can be used to make predictions of metabolic phenotypes in the presence of regulatory constraints. Covert, et al demonstrated that the prediction capability of the genome-scale metabolic model was improved when it was combined with the regulatory network model. However, the predictive power of combined metabolic/regulatory networks is limited by our incomplete understanding of transcriptional regulatory network structures even in

well-characterized organisms such as *E. coli.* For this reason, it is crucial that the network structures are continuously updated based on both high-throughput data and targeted experimentation.

Another example where the combination of high-throughput data into together with *in silico* model predictions results in improved metabolic engineering strategies has been demonstrated by Alper et al. [19]. Alper et al. generated a transposon mutagenesis library of *E. coli in vivo* and screened the library to identify deletion strains with enhanced lycopene production. This screening approach allowed discovering additional gene knockout targets that may be strongly affected by the regulatory mechanism in the parental lycopene producing strain and thus would not have been identified based on the metabolic model. Implementing the additional knockouts that had not been predicted by the metabolic model *in vivo* resulted in further increase in lycopene production.

The two studies above clearly demonstrate the usefulness of integration of genome-scale models with high-throughput data for metabolic engineering purposes. It is expected that metabolic engineering will incorporate both *in silico* approaches and high-throughput experimentation as powerful and sophisticated tools for strain design. Metabolic engineering strain designs produced using i*n silico* methodologies are usually exhibit a sub-optimal growth and therefore have a low production rates then predicted from calculations. Due to the growth-coupled nature of these designs, the increase in growth rate should improve production properties of these strains. The improvement of growth can be archived using adaptive evolution.

2.8 METABOLIC ENGINEERING AIDED BY ADAPTIVE EVOLUTION

Laboratory scale adaptive evolution has been used for basic biological discoveries as well as for biotechnological application; some of these applications are presented in Figure 2.4.

Some of the biological applications include understanding of the bases of adaptation, unraveling effects of beneficial gene mutations on phenotype, and establishment of dominant physiological traits through direct bacterial competition. Biotechnological applications include: activation of dormant pathways [1], increasing the arsenal of utilizable sugars [66,67], and well as adaptation of bacterial strains to different environmental conditions [8]. Laboratory adaptive evolution has emerged as a valuable tool for strain optimization in metabolic engineering [68,69]. Adaptive evolution strategies allow for the combination of rational engineering of the starting point strain with the efficient generation of genetic variation and selection of beneficial mutations in an unbiased fashion. These adaptive evolution strategies have been shown to efficiently optimize a production strain [66,70]. In contrast to rational engineering and directed modification of specific enzymes, adaptive evolution has the advantage of allowing non-intuitive beneficial mutations to occur in many different genes. After resequencing strains evolved under constant conditions, mutations have been found in both genes that code for metabolic enzymes such as glycerol kinase (*glpK*) and in global regulators such as sigma factor σ^S (*rpoS*) [27]. Furthermore, mutations can activate latent metabolic pathways, bypassing deleted reactions to increase the growth rate of a strain [71]. So far, the majority of applications of adaptive evolution in metabolic engineering have used traditional well-characterized platform organisms

such as yeast [66,67,70,72,73] and *E. coli* [37,74,75]. There have been 4 distinct classes of metabolic engineering applications that have utilized adaptive evolution: 1) Increasing the uptake of one desirable substrate or a combination of substrates [66,67,70]; 2) Increasing tolerance for stresses that commonly occur in industrial processing [72,74]; 3) Adaptation of genetically modified strains to bypass bottlenecks such as cofactor balancing or production of essential biomass precursors [73,75,76]; 4) Adaptation of genetically modified strains to an un-natural growth condition (i.e. low pH, high temperature) or removal of media supplementation[8]. The last two applications are of a particular interest as the rational genetic modifications introduced in the parental strain typically redirect metabolic flux from biomass production to a desirable by-product that results in low growth rates, reduction of total product yields and productivities, and in many cases will lead to the requirement of media supplementation. Adaptive evolution of these types of strains under selection for higher growth rates allows strains to attain an optimal balance between by-product secretion and biomass generation.

2.9 DISSERTATION OVERVIEW

As mentioned in this chapter, the metabolic engineering has shifted its paradigm towards systematic approach: where the high-throughput data can be used to suggest new ways to improve native production capabilities of the specific organism. Thus, the genome-scale metabolic models are becoming an ultimate platform for the analysis of such data and understanding physiological behavior on the systems-level

(Figure 2.5). Metabolic models were actively used to predict the growth-coupled metabolic engineering designs for overproduction of commodity chemicals and bio-fuels[15,16,20,52,77,78]. As was mentioned earlier, one of the biggest advantages in using growth coupled metabolic engineering strain designs is the ability to adaptively evolve the resulted stain to further improve production capabilities. One of the most important requirements of adaptation is consistency in the environmental condition throughout the adaptation period[1,8,20,52,76,79-82]. Environmental inconsistency will introduce alternative adaptive pressure and might results in optimization of undesired physiological traits such as: increased oxygen uptake, etc. As the majority of metabolic engineering designs rely in anaerobic growth conditions, adaptation of such strains is a challenging task and complicates the process of adaptive evolution. The goal of this work is to utilize model-driven approach to develop a platform strain capable of exhibiting a similar phenotype under oxic and anoxic growth conditions to be used for implementation and adaptation of the successful metabolic engineering designs to further improve production. This broad goal can be further disseminated into sub aims in the following way: Aim 1: identification of the successful strain design with no oxygen uptake, constructing the mutant strain and adaptation if necessary (Chapter 3); Aim 2: phenotypic characterization and understanding of the physiological behavior and underlying biochemistry of the mutant strain (Chapter 4); Aim 3: implementation of the metabolic designs for the production of organic acid (*D*/*L* lactic acid) and amino acid (*L*-alanine) and adaptation in order to improve production (Chapter 5).

Biology has many interesting and mysterious aspects to it that are not fully

understood, so it was unavoidable that in the midst of conducting experiments in pursuit of these main goals numerous interesting observations and hypotheses were generated. Details of these observations are provided bellow. Chapter 3 will focus on the development of the oxygen-independent platform strain and challenges and findings associated with this. One of the most significant findings were the ability of the mutant strain to secret *D*-lactic acid at nearly theoretical maximum yield without any pathway engineering and nearly diminished oxygen uptake rate. Chapter 4 will be devoted to the understanding of the physiology associated with the resulted phenotype, using high-throughput methodologies such as transcriptomics analysis and carbon tracing (fluxomics) experiments. Building upon the findings of these chapters, Chapter 5 will describe a study aimed at adaptation of ECOM4 strain to production of lactic acid in low pH environment. Chapter 6 will describe development of the organic and amino acid overproducing strains based on the platform strain developed earlier. In the final chapter, will be devoted to summarizing the main results of the three major aims of this dissertation: i) development of the oxygen-independent platform strain, ii) understanding physiological behavior of this strain and underlying genetic changes, and iii) development of the overproducing strains using the platform strain. Moreover, the future directions and implementations of the developed technology will be discussed and some general commentary associated with the field will be given.

2.10 ACKNOWLEDGEMENTS

The text of this chapter, in part or in full, is a reprint of material as it appears in Markus J. Herrgard, Vasiliy A. Portnoy, Hyun Uk Kim, Hongseok Yun, Seung Bum Sohn, Jin Sik Kim, Bernhard Palsson*, Sang Yup Lee*,. "Genome-Scale Models, Systems Biology, and Metabolic Engineering" In the Metabolic pathway engineering handbook: Tools and Applications. Editor Christina D. Smolke, CRC Press, FL, 2010. The dissertation author was the secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

2.11 REFERENCES

1. Lee DH, Palsson BO. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. *Appl Environ Microbiol* 2010;76:4158-68.

2. Applebee MK, Herrgard MJ, Palsson BO. Impact of individual mutations on increased fitness in adaptively evolved strains of Escherichia coli. *J Bacteriol* 2008;190:5087-94.

3. Price ND, Papin JA, Schilling CH, Palsson B. Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology* 2003;21:162-169.

4. Reed JL, Famili I, Thiele I, Palsson BO. Towards multidimensional genome annotation. *Nat Rev Genet* 2006;7:130-41.

5. Fong SS, Marciniak JY, Palsson BØ. Description and Interpretation of Adaptive Evolution of Escherichia coli K-12 MG1655 Using a Genome-scale in silico Metabolic Model. *Journal of Bacteriology* 2003;185:6400-8.

6. Fong SS, Palsson BO. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* 2004;36:1056-58.

7. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 2002;420:186-9.

8. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain. *Appl Environ Microbiol* 2008;74:7561-9.

9. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 2001;19:125-130.

10. Varma A, Palsson BO. Metabolic Flux Balancing: Basic concepts, Scientific and Practical Use. *Nat Biotechnol* 1994;12:994-998.

11. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict

growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology* 1994;60:3724-3731.

12. Edwards JS, and Palsson, B.O. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics* 2000;1.

13. Price ND, Reed JL, Papin JA, Famili I, Palsson BO. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophys J* 2003;84:794-804.

14. Burgard AP, Maranas CD. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 2003;82:670-7.

15. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;84:647-57.

16. Pharkya P, Burgard AP, Maranas CD. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* 2003;84:887-99.

17. Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* 2004;14:2367-76.

18. Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* 2005;7:155-64.

19. Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 2005;23:612-6.

20. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-648.

21. Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial

systems. *Metab Eng* 2006;8:1-13.

22. Lee SY, Lee DY, Kim TY. Systems biotechnology for strain improvement. *Trends Biotechnol* 2005;23:349-58.

23. Wang Q, Chen X, Yang Y, Zhao X. Genome-scale in silico aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. *Appl Microbiol Biotechnol* 2006;V73:887-894.

24. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007;3:149.

25. Ghim CM, Goh KI, Kahng B. Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*. *J Theor Biol* 2005;237:401-11.

26. Imielinski M, Belta C, Halasz A, Rubin H. Investigating metabolite essentiality through genome-scale analysis of *Escherichia coli* production capabilities. *Bioinformatics* 2005;21:2008-16.

27. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 2006;38:1406-1412.

28. Samal A, Singh S, Giri V, Krishna S, Raghuram N, Jain S. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinformatics* 2006;7:118.

29. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys J* 2006;90:1453-61.

30. Datta R, Henry M. Lactic acid: recent advances in products, processes and technologies - a review. *Journal of Chemical Technology & Biotechnology* 2006;81:1119-1129.

31. Beard DA, Liang SD, Qian H. Energy balance for analysis of complex metabolic networks. *Biophys J* 2002;83:79-86.

32. Kümmel A, Panke S, Heinemann M. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics* 2006;7.

33. Kümmel A, Panke S, Heinemann M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* 2006;2:2006.0034.

34. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121.

35. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2007;2:727-738.

36. Feist AM, Orth JD, Zielinski DC, Palsson BO. Construction and evolution of E. coli production strains designed through a model-driven analysis. *In preparation* 2008.

37. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-8.

38. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol* 2005;71:7880-7887.

39. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* 2007;446:572-6.

40. Choi HS, Kim TY, Lee DY, Lee SY. Incorporating metabolic flux ratios into constraint-based flux analysis by using artificial metabolites and converging ratio determinants. *J Biotechnol* 2007;129:696-705.

41. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 2005;102:2685-9.

42. Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* 2005;102:12678-83.

43. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ. A bottom-up approach to gene regulation. *Nature* 2006;439:856-60.

44. Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY. Systems biology as a foundation for genome-scale synthetic biology. *Curr Opin Biotechnol* 2006;17:488-492.

45. Hong SH, Kim JS, Lee SY, In YH, Choi SS, Rih J-K, Kim CH, Jeong H, Hur CG, Kim JJ. The genome sequence of the capnophilic rumen bacterium Mannheimia succiniciproducens. *Nat Biotech* 2004;22:1275-1281.

46. Famili I, Forster J, Nielsen J, Palsson BO. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci U S A* 2003;100:13134-9.

47. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. *Genome. Res.* 2004;14:301-12.

48. Ko J, Kim I, Yoo S, Min B, Kim K, Park C. Conversion of methylglyoxal to acetol by Escherichia coli aldo-keto reductases. *J Bacteriol* 2005;187:5782-9.

49. Bro C, Regenberg B, Forster J, Nielsen J. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab Eng* 2006;8:102-11.

50. Hong SH, Park SJ, Moon SY, Park JP, Lee SY. In silico prediction and validation of the importance of the Entner-Doudoroff pathway in poly(3-hydroxybutyrate) production by metabolically engineered Escherichia coli. *Biotechnol Bioeng* 2003;83:854-63.

51. Lee SY, Hong SH, Moon SY. In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered Escherichia coli as an example. *Genome Inform* 2002;13:214-23.

52. Fong SS, Joyce AR, Palsson BO. Parallel adaptive evolution cultures of

Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res* 2005;15:1365-72.

53. Lee SY, Woo HM, Lee D-Y, Choi HS, Kim TY, Yun H. Systems-level analysis of genome-scale *in silico* metabolic models using MetaFluxNet. ***Biotechnol. Bioproc. Eng.*** 2005;10:425-431.

54. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO. Systems approach to refining genome annotation. *Proc Natl Acad Sci U S A* 2006;103:17480-4.

55. Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM. Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics* 2006;7.

56. Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A* 2007;104:7797-802.

57. Mahadevan R, Bond DR, Butler JE, Esteve-Nunez A, Coppi MV, Palsson BO, Schilling CH, Lovley DR. Characterization of Metabolism in the Fe(III)-Reducing Organism Geobacter sulfurreducens by Constraint-Based Modeling. ***Appl. Environ. Microbiol.*** 2006;72:1558-1568.

58. Feist AM, Scholten JCM, Palsson BO, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Mol Syst Biol* 2006;2:1-14.

59. Bailey JE. Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. *Metab Eng* 2001;3:111-4.

60. Bailey JE. Complex biology with no parameters. *Nature Biotechnology* 2001;19:503-4.

61. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB. Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res* 2004;14:917-24.

62. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid

pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* 2005;1:e46.

63. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 2002;99:15112-7.

64. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* 2005;102:7695-700.

65. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004;429:92

66. Sonderegger M, Sauer U. Evolutionary engineering of Saccharomyces cerevisiae for anaerobic growth on xylose. *Appl Environ Microbiol* 2003;69:1990-8.

67. Wisselink HW, Toirkens MJ, del Rosario Franco Berriel M, Winkler AA, van Dijken JP, Pronk JT, van Maris AJ. Engineering of Saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose. *Appl Environ Microbiol* 2007;73:4881-91.

68. Sauer U. Evolutionary engineering of industrially important microbial phenotypes. *Adv Biochem Eng Biotechnol* 2001;73:129-69.

69. Chatterjee R, Yuan L. Directed evolution of metabolic pathways. *Trends Biotechnol* 2006;24:28-38.

70. Kuyper M, Toirkens MJ, Diderich JA, Winkler AA, van Dijken JP, Pronk JT. Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting Saccharomyces cerevisiae strain. *FEMS Yeast Res* 2005;5:925-34.

71. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* 2006;281:8024-8033.

72. Cakar ZP, Seker UO, Tamerler C, Sonderegger M, Sauer U. Evolutionary engineering of multiple-stress resistant Saccharomyces cerevisiae. *FEMS Yeast Res* 2005;5:569-78.

73. van Maris AJ, Geertman JM, Vermeulen A, Groothuizen MK, Winkler AA, Piper MD, van Dijken JP, Pronk JT. Directed evolution of pyruvate decarboxylase-negative Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol* 2004;70:159-66.

74. Yomano LP, York SW, Ingram LO. Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production. *J Ind Microbiol Biotechnol* 1998;20:132-8.

75. Meynial Salles I, Forchhammer N, Croux C, Girbal L, Soucaille P. Evolution of a Saccharomyces cerevisiae metabolic pathway in Escherichia coli. *Metab Eng* 2007;9:152-9.

76. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable Escherichia Coli adaptation to loss of key metabolic enzymes. *J Biol Chem* 2005.

77. Hua Q, Joyce AR, Fong SS, Palsson BO. Metabolic analysis of adaptive evolution for in silico designed lactate-producing strains. *Biotechnol Bioeng* 2006;95:992-1002.

78. Aristidou A, Penttilä M. Metabolic engineering applications to renewable resource utilization. *Current Opinion in Biotechnology* 2000;11:187-98.

79. Lenski RE, Travisano M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci U S A* 1994;91:6808-14.

80. Pal C, Papp B, Lercher MJ. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* 2005;37:1372-5.

81. Perrenoud A, Fong SS, Sauer U, Palsson BO. Adaptive evolution of *Escherichia coli* gene deletion mutants results in novel flux patterns. *Journal of Biological Chemistry* Submitted.

82. Wichman HA, Millstein J, Bull JJ. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 2005;170:19-31.

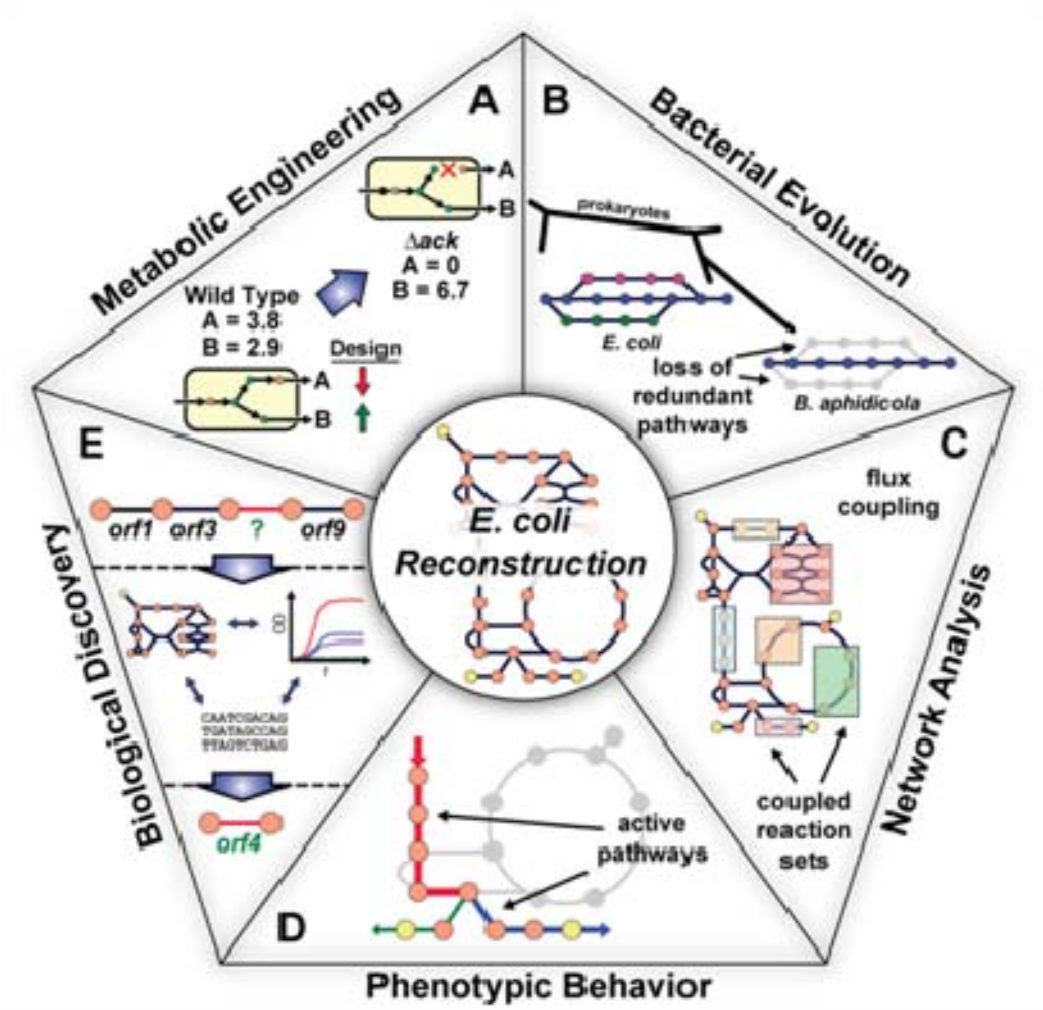


Figure 2.1: Uses of *E. coli* metabolic reconstructions. (A) *In silico* strain design using *E. coli* to overproduce commercially valuable products [14,15,17,20,21,38,42,43]. (B) Addressing questions in evolutionary biology such as reductive evolution of metabolic capabilities [44]. (C) Analysis of intrinsic properties of biological networks, such as coupled reaction activities [47,48] across different growth conditions. (D) *In silico* studies of gene [25], metabolite [26,27] and reaction [28,29] essentiality as well as predictions of flux distributions using stoichiometric and thermodynamic constraints [29,31-35]. (E) Biological discovery such as predicting genes encoding for orphan reactions [54,55].

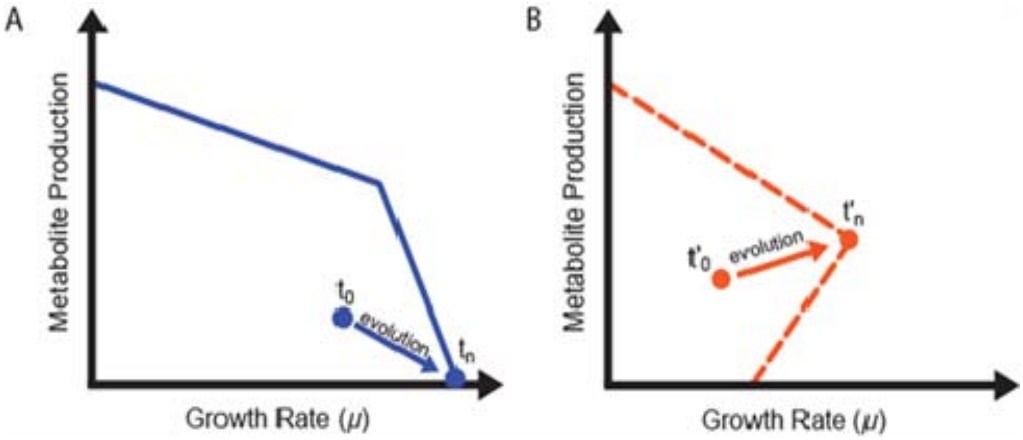
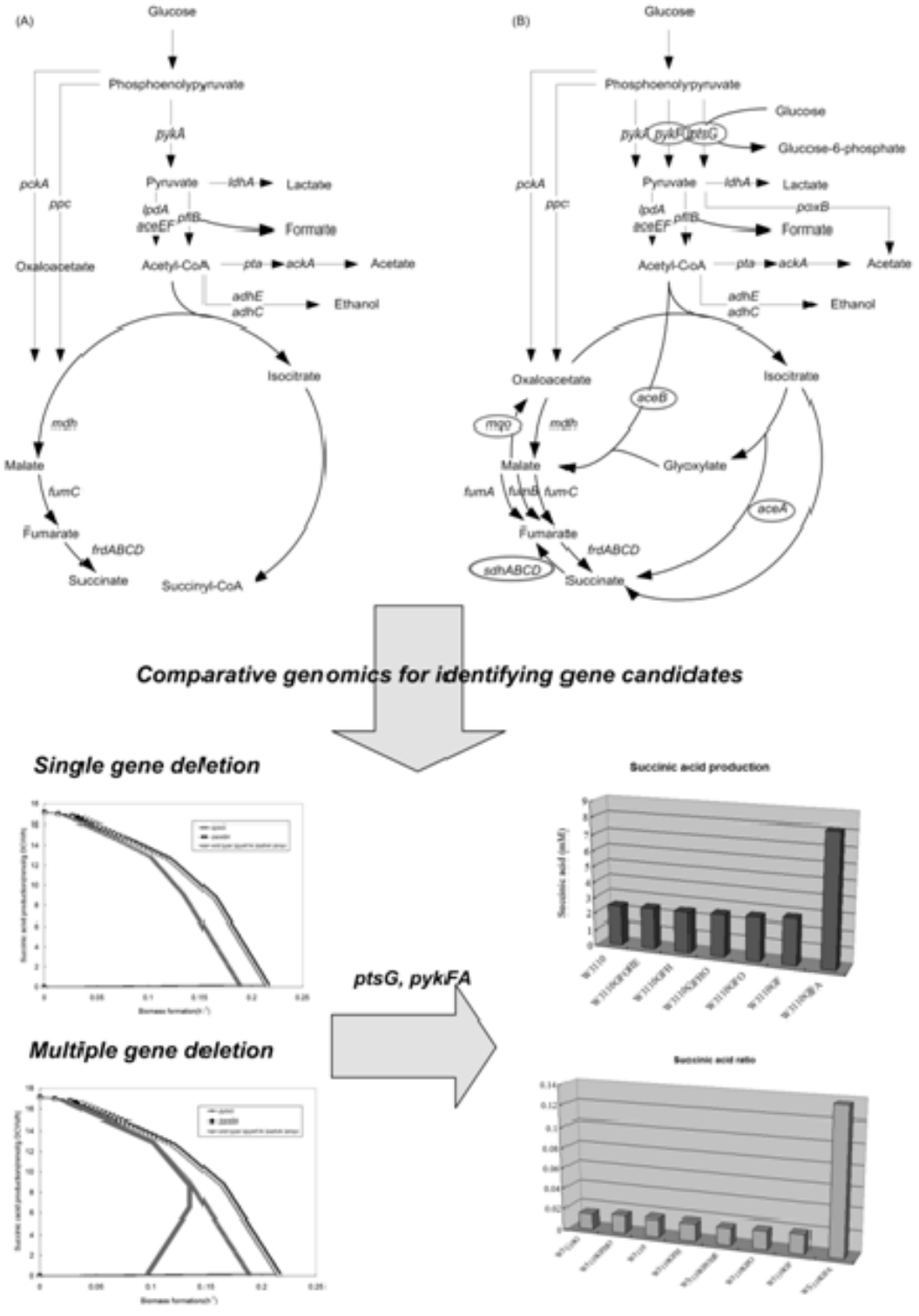


Figure 2.2: The concept of growth-coupled strain designs. Shown are two metabolite production envelopes for both a wild type strain (A) and a growth-coupled designed strain (B). The envelopes are defined by the governing physio-chemical constrains and are determined using FBA and a metabolic model. Any combination of growth rate and production rate inside an envelope is possible. Through the process of adaptive evolution, a strain is predicted to evolve from an initial physiological state (t_0 or t'_0), to an optimal state corresponding with the maximum possible growth rate of the strain (t_n or t'_n) [5-7]. Therefore, by designing a stain that couples the production rate for a desired metabolite to growth rate (i.e., the 'designed strain' in B), overproducing strains can be generated.

Figure 2.3: Metabolic engineering of *E. coli* for improved succinic acid production, based on comparative genome analysis between *M. succiniciproducens* (A) and *E. coli* (B), and combinatorial *in silico* simulation. Genes present only in *E. coli* were first identified as gene knockout targets. Subsequently, all possible combinations of those genes were simulated so as to find the mutant strain that produces the maximal production rate of succinic acid. In this study, mutant strain whose *ptsG* and *pykFA* genes were knocked-out showed the highest value among the candidates. W3110 is the wild-type strain and the mutant strains are named according to the corresponding gene disrupted; H for *sdhA*, O for *mqo*, E for *aceBA*, G for *ptsG*, and FA for *pykFA*.


(A)
Glucose
Phosphoenolypyruvate
pykA
pckA
ppc
Pyruvate
ldhA
Lactate
lpdA
aceEF
pflB
Formate
Oxaloacetate
Acetyl-CoA
pta
ackA
Acetate
adhE
adhC
Ethanol
Isocitrate
mdh
Malate
fumC
Fumarate
frdABCD
Succinate
Succinyl-CoA
(B)
Glucose
Phosphoenolypyruvate
pykA
pykF
ptsG
Glucose-6-phosphate
pckA
ppc
Pyruvate
ldhA
Lactate
poxB
lpdA
aceEF
pflB
Formate
Acetyl-CoA
pta
ackA
Acetate
adhE
adhC
Ethanol
Oxaloacetate
aceB
Isocitrate
mqo
mdh
Malate
Glyoxylate
fumA
fumB
fumC
aceA
Fumarate
frdABCD
sdhABCD
Succinate
Comparative genomics for identifying gene candidates
Single gene deletion
Succinic acid production
ptsG, pykFA
Multiple gene deletion
Succinic acid ratio

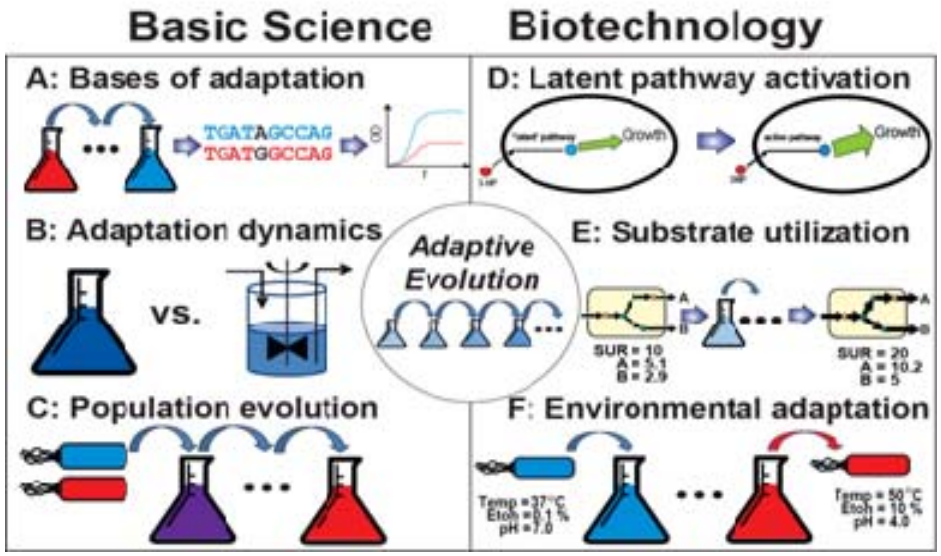


Figure 2.4: Uses of laboratory scale adaptive evolution. (A) utilization of adaptive evolution to identify genetic bases underlying bacterial adaptation[1,2]. (B) Adaptive evolution can undergo different evolutionary trajectories depending on the environment and selection pressure. (C) Adaptive evolution can be used to isolate the most superior strain. (D) and (E) Activation of "latent" pathways by forcing the bacteria to utilize non-native substrates leads to valuable biological discoveries[1]. (F) Adaptation of the production strains to alternative growth environment [8].

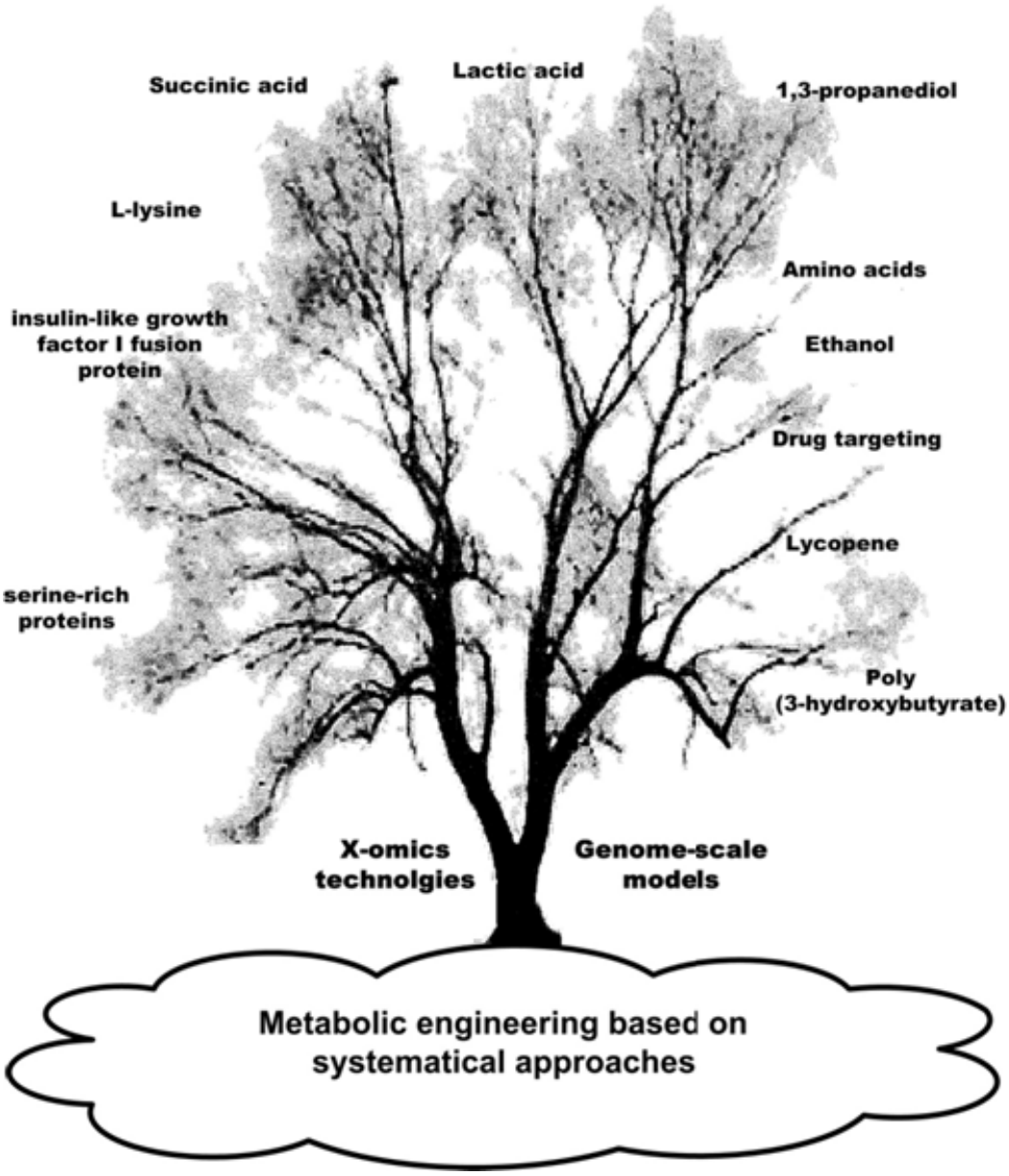


Figure 2.5: Various demonstrative studies of metabolic engineering based on systematic approaches including studies the use either high throughput technologies (X-omics studies) or genome-scale models or both.

CHAPTER III:

DEVELOPMENT OF THE OXYGEN-INDEPENDENT PLATFORM STRAIN

In an effort to develop an *E. coli* strain that would exhibit the similar phenotypic behavior under both oxic and anoxic culture conditions we selected three operons directly involved in oxygen utilization for removal. Cytochrome oxidases (*cydAB, cyoABCD, cbdAB*) were removed from the *E. coli* K12 MG1655 genome resulting in the ECOM3 (*E. coli* Cytochrome Oxidase Mutant 3) strain. Removal of cytochrome oxidases reduced the oxygen uptake rate of the knock-out strain by nearly 85%. Moreover, the knock-out strain was initially incapable of growing on M9 minimal media. After subjecting the ECOM3 strain to adaptive evolution on glucose M9 medium for 60 days, the growth rate equivalent to anaerobic wild type *E. coli* was achieved. My findings demonstrate that three independently adaptively evolved ECOM3 populations acquired different phenotypes: one produced lactate as a sole fermentation product while the other two strains exhibited a mixed acid fermentation under oxic growth conditions with lactate remaining as the major product. The homofermenting strain showed the *D*-lactate yield of 0.8 g/g from glucose. Gene expression and *in silico* model-based analysis was employed to identify perturbed

pathways and explain phenotypic behavior. Significant upregulation of *ygiN* and *sodAB* explain the remaining oxygen uptake that was observed in evolved ECOM3 strains. *E. coli* strains presented herein showed the ability to produce lactate as a fermentation product from glucose as well as undergo mixed-acid fermentation during aerobic growth.

3.1 INTRODUCTION

Escherichia coli is one of the most commonly used host organisms for metabolic engineering and overproduction of metabolites due to its fast growth rate, amenability to genetic manipulation, and its ability to produce wide variety of anaerobic fermentation products such as organic acids. *E. coli* has also been extensively characterized with respect to its metabolic physiology [1] enabling the utilization of rational model-based engineering strategies [5-8]. Rational model-based approach to engineering *E. coli* that aims to couple specific metabolite overproduction to growth combined with adaptive evolution have shown promise for strain optimization [6,9]. For engineered strains that couple desirable byproduct secretion to growth, adaptation to higher growth rates has been shown to lead to the increased production of the product [10].

The anaerobic growth of *E. coli* is characterized by the formation of a number of reduced byproducts as a result of mixed-acid fermentation; the majority of the metabolic engineering designs rely on anoxic growth conditions. Maintenance of a

strict anoxic condition is a challenging task and complicates the procedure of experimental adaptation.

The goal of this study was to develop an *E. coli* strain that could show similar phenotypic behavior under both oxic and anoxic growth conditions. The resulting strain could be used as a platform strain in evolutionary engineering where long term laboratory evolution in aerobic conditions is used to optimize desirable phenotypic traits [11]. It has been reported that simultaneous deletion of *cyd* and *cyo* genes has resulted in a significant decrease, but not elimination, of oxygen uptake [12]. However, no secretion analysis or other physiological characterization of this mutant strain has been reported [12,13]. Based on these results, we hypothesized that removal of all of three cytochrome oxidases would result in anaerobic growth characteristics even under oxic conditions. Moreover, we expected activation of anaerobic pathways responsible for mixed acid fermentation as a means of NADH recycling, leading to the production of fermentation products aerobically. Wild type *E. coli* strain (MG1655) was subjected to genetic manipulation and all active cytochrome oxidases were removed from its genome. The resulting ECOM3 (*E. coli* Cytochrome Oxidase Mutant) strain was subjected to adaptive evolution and phenotypic characterization throughout the course of evolution. The experimental setup for the adaptive evolution of ECOM3 and the nomenclature used in this study are presented in Figure 3.1. The three evolved populations were extensively characterized and results are presented.

3.2 RESULTS

3.2.1 Strain construction and growth adaptation

In an effort to develop an *E. coli* strain that would exhibit the similar phenotypic behavior under both oxic and anoxic culture conditions, we constructed a triple mutant of wild-type *E. coli* K-12 MG1655 strain (ECOM3 or *E. coli* Cytochrome Oxidase Mutant) that had the genes coding for the cytochrome oxidase *bd* (*cydAB*, b0733-b0734), cytochrome oxidase *bo* (*cyoABCD*, b0432-b0429) and putative cytochrome oxidase (*cbdAB*, b0979-b0978) completely removed from the genome. The *cbd* locus is annotated as a putative cytochrome oxidase and studies indicate that its gene products do not form a fully functional terminal oxidase [12,14]. Deletions of terminal cytochrome oxidases, encoded by the *cydAB* (cytochrome *bd* complex), *cyoABCD* (cytochrome *bo* complex), and *cbdAB* (also known as *appBC*) operons, have been previously reported in *E. coli* [14-18].

3.2.2 Initial phenotypic characterization and adaptive evolution

The resulting ECOM3 strain was initially incapable of growing on M9 minimal medium supplemented with 2 g/l glucose and also demonstrated slow growth on LB media (0.034 +/- 0.002 1/h). Significant growth (0.25 +/- 0.02 1/h) was only observed on Rich Defined Media (Technova) and on M9 minimal media supplemented with a full mixture of amino acids (Supplement EZ, Technova). Three parallel

adaptive evolutions (denoted by ECOM31, ECOM32, ECOM33) were conducted to adapt the initial ECOM3 strain to growth on M9 minimal media with glucose as the sole carbon source (Figure 3.1). Initially, a rapid decrease in growth rate was observed upon reduction of the amino acid supplements (Figure 3.2), but the growth rate increased once the supplement volume was reduced to 60 µl. The cells were then allowed to adapt to the new environment with the amount of supplement remaining unchanged for an additional six days. At day 14 of adaptive evolution, amino acid supplements were further reduced to 30 µl resulting in a significant change in growth rate. Further reduction in amino acid supplements had no effect on cell growth rate, and starting on day 17, amino acid supplements were no longer added to the media (Figure 3.3). A rapid increase in growth rate followed complete removal of supplements and the growth rate reached a maximum of 0.44 +/- 0.01 1/h on day 56 of evolution. The evolutions were continued for an additional four days with no further observed growth rate increases (Figure 3.2). Number of cell divisions that occurred during the term of evolution was estimated based on the amount of cells passed each day and the doubling time. The total number of cell divisions assuming a small death rate was 5.00×10^{12} +/- 0.5×10^{12} (average is reported) (Figure 3.2).

The three independently evolved end-point populations (eECOM31, eECOM32, eECOM33) showed similar growth rate gains and acquired the ability to grow on glucose minimal medium without amino acid supplementation. Evolutions were stopped once the observed growth rates for the three end-point populations (0.42 +/- 0.02 1/h – average reported) became equivalent to the growth rate of wild type *E. coli* cultivated under anoxic conditions (0.45 +/- 0.02 1/h) indicating a similarity of the

evolved ECOM3 populations to the anaerobic phenotype of the wild type strain. In order to further probe the metabolic phenotypes of the populations during and after evolutions, growth rates, oxygen uptake rates, sugar uptake rates, and product secretion rates were measured for each of the three on day 1, day 30, and day 60.

3.2.3 Phenotypic characterization of the evolved populations

Phenotypic characterization of the evolved populations revealed that the three evolutionary endpoints had slightly different metabolic phenotypes (Table 3.1). As a common feature, a two-fold increase in substrate uptake rate was observed for all three populations from 9.98 +/- 2.0 mmol/gDW/h to 20.80 +/- 0.7 mmol/gDW/h on average within the first 30 days of evolution. Similarly, the *D*-lactate secretion rate increased more than two-fold by day 30 of adaptive evolution (from 17.00 +/- 2.9 mmol/gDW/h to 35.44 +/- 5.5 mmol/gDW/h on average), but the endpoint populations showed higher variance in lactate secretion than glucose uptake.

Acetate, one of the major fermentation products of wild type *E. coli*, was not a major growth byproduct of the ECOM3 parental strain prior to evolution. However, the three evolutionary endpoint populations had significantly different acetate secretion rates. Acetate secretion was strongly reduced in the eECOM31; while only a moderate reduction was seen in the eECOM32. In contrast, the acetate secretion rate of eECOM33 increased more than two-fold by day 60 (Table 3.1).

Detailed secretion analysis of the ECOM3 end-point populations under oxic conditions (Figure 3.4) indicated that the strains acquired the ability to secrete *D*-

lactate as a dominant fermentation product at yields of 0.76, 0.73 and 0.65 g lactate/g glucose for eECOM31, eECOM32, and eECOM33, respectively. Compared to the wild type strain, the lactate secretion rates have increased by 94.5, 92.3 and 77.2 fold, while the substrate uptake rates have increased by 2.3, 2.2, and 2.3 fold, respectively, for the three evolved ECOM3 populations. These results are consistent with gene expression analysis using RT-PCR that showed that the *ldhA* gene, encoding the lactate dehydrogenase protein, was up-regulated by 21, 8, and 23 fold, respectively, in the three evolved strains compared with the wild type strain (Figure 3.5).

The increase in glucose uptake and lactate secretion, without the production of any other major fermentation products, results in a significantly elevated flux through the glycolytic pathway. Consistent with the physiological characteristics of the end point strains, we did not observe a significant increase in the expression of pyruvate dehydrogenase gene (*aceF*) or pyruvate formate lyase (*pflA*) except in the case of the eECOM32 population (Figure 3.5).

3.2.4 Clonal analysis

In order to characterize and study heterogeneity in the evolved populations, we used clonal analysis. Evolved populations were plated on solid media at day 60 of evolution and ten random colonies from each population were selected for the analysis. We observed a higher level of heterogeneity within the eECOM33 population. The clone with the highest lactate yield (eECOM31LA) was identified within the eECOM31 population based on the highest final lactate concentration

(Figure 3.6). Phenotypic assessment of eECOM31LA showed compatible growth rates to eECOM31 population under oxic conditions, with slightly higher glucose uptake and lactate secretion rates (GUR: 21.61 +/- 0.16 mmol/gDW/h; LactSR: 42.32 +/- 2.52 mmol/gDW/h). The final lactate concentration for eECOM31LA strain was 1.58 +/- 0.1 g/l, which is equivalent to 80% conversion of glucose to lactate (0.8 g lactate/g glucose). The growth rate of the eECOM31LA (0.39 +/- 0.01 1/h) was slightly lower than that of the eECOM31 population. Moreover, it was observed that eECOM31LA mutant had significantly lower oxygen uptake rate (2.44 mmol O_2/gDW/h) then eECOM31 population, demonstrating a wide range of oxygen requirements established during adaptive evolution.

3.2.5 Oxygen dependency of ECOM strains

In order to determine whether the presence or absence of oxygen affects the phenotype of ECOM3 strain we subjected three evolved populations to anaerobic growth on M9-glucose minimal media. We observed a slight decrease in growth rate, 12%, 5%, and 7% for three populations respectively as well as radically different secretion profiles compared to the aerobic phenotype (Table 3.2). Secretion analysis demonstrated that eECOM31 population preserved its aerobic phenotype and produced *D*-lactate as the sole byproduct at a concentration similar to the one observed under oxic growth conditions. In contrast, the eECOM32 and eECOM33 lost their ability to secrete *D*-lactate as a sole byproduct and presented with a phenotype similar to the wild-type *E. coli* under anoxic growth conditions with formate, acetate, and

ethanol as byproducts (Figure 3.7). Gene expression analysis of the *ldhA, pflA*, and *aceF* genes under anoxic growth conditions failed to reveal a clear mechanistic basis for the observed physiological differences between the three strains. We originally hypothesized that deletion of cytochrome oxidases would completely eliminate oxygen consumption by the ECOM3 strain. However, while oxygen consumption was significantly reduced in the evolved strains, some residual oxygen uptake remained. Prior to evolution (day 0), the oxygen uptake rate of the ECOM3 strain was 6.89 +/- 1.61 mmol O_2/gDW/h, which is almost three times lower than the oxygen uptake rate of the wild type strains. The oxygen consumption was further reduced to 3.84 +/- 1.7 mmol O_2/gDW/h (average reported for all evolved populations) by the end of the adaptive evolution (Table 3.1). Evolved populations showed significantly different oxygen uptake rates: 5.62 +/- 0.34, 3.69 +/- 0.66, and 2.21 +/- 0.38 mmol O_2/gDW/h for the eECOM31, eECOM32, and eECOM33, respectively. Gene expression analysis together with *in silico* phenotypic modeling (see Discussion) revealed that oxygen uptake levels were consistent with observed levels of expression of *ygiN* gene (annotated as quinol monooxygenase [2]; however since the proposed biochemical reaction [2] does not involve incorporation of oxygen into an electron donor the term monooxygenase should not be used). The eECOM31 population showed the highest OUR consistent with the highest *ygiN* expression. In order to determine if YgiN is involved into oxygen uptake we conducted an additional gene deletion and removed *ygiN* from the original unevolved ECOM3 strain. The removal of *ygiN* almost completely eliminated oxygen uptake (Table 3.3).

3.3 DISCUSSION

To our knowledge, we describe here the first *E. coli* strain that is able to homoferment glucose to lactic acid under aerobic growth conditions. This strain (ECOM3) was engineered by removing all active cytochrome oxidases. Genes were removed using homologous recombination techniques and the resulted strain was evolved to achieve growth on M9 minimal media with trace elements and 2 g/l glucose. The observed growth rate after 60 days of evolution was equivalent to the growth rate of wild type *E. coli* under anoxic conditions. Adaptive evolution produced three end-point populations that exhibited similar behavior aerobically and had radically different phenotypic characteristics anaerobically. Lactic acid was identified as a major product of aerobic fermentation for all there end-point populations. The best representative of eECOM31 population exhibited the highest lactate secretion and glucose uptake rate. The yield of lactate from glucose was close to 80%. We used gene expression analysis to investigate genetic perturbations that underlined secretion of lactic acid, and remaining oxygen uptake rate. We also utilized a genome-scale metabolic model of *E. coli* (*i*AF1260) to understand the mechanism of oxygen utilization in the ECOM3 phenotype.

3.3.1 *In silico* analysis using a genome-scale model

In order to identify potential metabolic fates of oxygen in ECOM3 strain, we employed a genome-scale metabolic model of *E. coli* (*i*AF1260) [19]. The *in vivo* genotype of the ECOM3 strain was implemented computationally though the removal

of reactions catalyzed by the deleted genes. The *in silico* model was further constrained using experimental data to set the glucose and oxygen uptake rates and acetate secretion rates (with experimental error accounted for by allowing a range of uptake/secretion rates). Analysis of the computationally predicted flux distributions utilizing these constraints provided insights into the observed residual oxygen utilization. The model predicted that the mechanism that could account for residual oxygen uptake at the observed level would be through the activity of the *ygiN*. The simulation of the ECOM3 phenotype with computational model and scientific evidence showed flux coupling between the NADH dehydrogenase (*nuo/ndh* operons) and the reaction catalized by the *ygiN* gene forming a ubiquinone cycle. It has been shown that ubiquinone is the electron acceptor for the NADH dehydrogenase (*ndh/nuo*) [4] and a preferred electron carrier for *E. coli* during aerobic growth [3]. Furthermore, Adams and Jia [2] indicate that *ygiN* can potentially react with ubiquinol molecule and oxidize it to the ubiquinone form, through coupling of this oxidation reaction with reduction of the molecular oxygen. Based on these findings we proposed the mechanism for oxygen utilization (Figure 3.8).

3.3.2 Oxygen-independent platform strain

The removal of the *ygiN* gene *in silico*, predicted elimination of the oxygen uptake in ECOM3 strain. In order to determine if *ygiN* accounts for the oxygen uptake we removed it from the parental ECOM3 strain and observed nearly complete elimination of oxygen uptake: 0.03 +/- 0.04 mmol O_2/gDW/h. This experimental

evidence demonstrates that observed oxygen uptake can be attributed to the activity of YgiN. Consistent with the role of this pathway in residual oxygen utilization, we found that levels of aerobic expression of all the genes in this pathway were increased in the ECOM3 populations compared to the wild type strain (Figure 3.8). In particular, the expression of *ygiN* gene was increased nearly three times in the eECOM31.

3.3.3 Phase plane and gene deletion analysis using metabolic model

When the experimentally derived uptake/secretion rate constraints were imposed on the model, it predicted no lactate fermentation when a particular flux distribution was determined by maximizing biomass production using flux balance analysis (FBA, Figure 3.9A). We used the model and experimental data obtained for ECOM3 strains to map the observed phenotypes to the predicted space of allowed lactate secretion rates as a function of growth rate (Figure 3.9B). Phenotypes of the three end point strains lie close to the corner point of this space (growth rate 0.42 1/h; average lactate production 33.9 mmol/gDW/h) that correspond to trade-off between growth and lactate production. The inability of biomass maximization to predict observed phenotypes was interesting as in previous studies metabolic phenotypes of evolved strains could be accurately predicted using FBA[6,20]. We assumed that the reason for the inability of FBA to correctly predict the lactate secretion phenotype was that there are additional constraints or metabolic bottlenecks that limit the ability of the evolved ECOM3 strains to grow optimally (Figure 3.9B).

In order to identify candidate bottleneck reactions, we used the computational model to determine what additional gene deletions could explain observed secretion of lactate as an optimal phenotype. By systematically enumerating all single and double reaction deletions of central metabolic and amino acid biosynthetic reactions in the *i*AF1260 model, we identified a number of double reaction deletions in the acetate and ethanol secretion pathways that predicted phenotypes (i.e. growth and lactate secretion) that were almost identical to the observed ECOM3 phenotype (Figure 3.9C and Table 3.4). For instance, the removal of the pyruvate dehydrogenase (PDH) and pyruvate formate lyase (PFL) reactions from the model resulted in a predicted flux solution space where lactate secretion is coupled to biomass formation (growth). This prediction was generated by constraining the oxygen uptake and glucose uptake rate to the experimentally measured values (with experimental error included as lower/upper bounds). Predicted value for the lactate secretion at maximum growth rate matched the experimental measurements closely (Figure 3.9C). In order to check that this prediction was not a result of the applied constrains we varied the value of glucose uptake within the range of 10-30 mmol/gDW/h, and obtained similar values for lactate secretion to growth ratio (Figure 3.10).

Experimentally, we observed a significant increase in expression of *ldhA* gene but no downregulation of the *pflA* and *aceF* genes (corresponding to the PFL and PDH reactions) compared to the wild type strain. It is known that pyruvate formate lyase is inactivated aerobically by post-transcriptional mechanisms [21-23] and thus the PFL reaction is inactive in ECOM3 strain. Since the glycolytic flux in the ECOM3 strains was significantly increased, the lack of upregulation of *aceF* gene and inactivation

PFL would effectively result in decreased flux through PDH and PFL reactions relative to the glycolytic flux. This in turn would result in a bottleneck effect in the acetate and ethanol secretion pathway, thus forcing secretion of lactate in accordance with the model predictions.

3.4 CONCLUSION

In summary, we have engineered an *E. coli* strain (ECOM3) capable of homofermenting glucose to lactate in both aerobic and anaerobic conditions by deleting all cytochrome oxidase genes and adapting the strain to grow on minimal media without amino acid supplementation. Clonal analysis allowed identifying the best lactic acid producer from the eECOM31 population with lactate yields close to 80% from glucose. Interestingly, cell populations derived from the adaptive evolution showed significant residual oxygen uptake. We identified the mechanisms accounting for and the observed residual oxygen uptake using a combination of genome-scale metabolic model of *Escherichia coli* and gene expression analysis of specific pathways. The resulting ECOM3 populations have been shown to be amenable to genetic manipulation (results not shown) and thus can be used as a platform strain for further metabolic engineering that redirect lactate flux into other desirable byproducts. However, complete reduction of oxygen uptake was archived through removal of the quinol monooxygenase enzyme encoded by ygiN gene. The direct involvement of the YgiN into oxygen uptake has only been suggested based on the *in vitro* experiments

by Adams et al. [2]. Our findings present *in vivo* evidence of the biochemical function of quinol monooxygenase.

This quadruple mutant was named the ECOM4 strain (*E. coli* Cytochrome Oxidase Mutant 4). As mentioned in section 2.3.2 the oxygen uptake was nearly abolished upon removal of YgiN. The next chapter will be devoted to the physiological characterization of the developed strain using high-throughput gene expression profiling and carbon tracing experiments.

3.5 MATERIALS AND METHODS

3.5.1 Strains and media

E. coli K12 MG1655 (ATCC: 700926), obtained from the American Type Culture Collection (Manassas, VA), was used as a parent strain for all gene deletions in this study. During the gene deletion process the strains were cultured on Luria-Bertani medium supplemented with 50μg/ml kanamycin and 100 μg/ml ampicillin when necessary. Evolution and phenotype assessments of the mutant strain were carried out using M9 minimal media [24] with glucose (2 g/l) as the carbon source containing $Na_2HPO_4 \cdot 7H_2O$ (6.8g), KH_2PO_4 (3g), NaCl (0.5g), NH_4Cl (1g), $MgSO_4$ (2 mM), and $CaCl_2$ (0.1 mM) and trace elements [25]. During the early stage of adaptive evolution minimal media was additionally supplemented with EZ supplements (Technova) containing the mixture of L-amino acids at following concentrations (numbers indicate milimolarity): Ala (0.8), Arg (5.2), Asn (0.4), Asp (0.4), Cys (0.1), Glu (0.6), Gln (0.6), Gly (0.8), His (0.2), Ile (0.4), Leu (0.8), Lys (0.4), Met (0.2), Phe

(0.4), Pro (0.4), Ser (10.0), Thr (0.4), Trp (0.1), Tyr (0.2), Val (0.6), adenine, guanine, cytosine, and uracil (0.2 each), and thiamine, calcium pantothanate, p-aminobenzoic acid, p-hydroxybenzoic acid, and 2,3-dihydroxybenzoic acid (0.01 each) [26].

3.5.2 Generation of mutant strains

All strains and plasmids used in this study are listed in Table 3.5. The *Escherichia coli* K-12 MG1655 (ECOM3 or *E. coli* Cytochrome Oxidase Mutant: Δ*cydAB-cyoABCD-cbdAB*) strain lacking three known cytochrome oxidases was generated by homologous recombination using the lambda Red recombinase system [27,28] with primers listed in Supplementary Table 1. In short, the gene to be deleted was replaced by a kanamycin gene flanked by FRT sites and the insert was removed with a FLP recombinase. The *cydAB* operon was removed first, followed by the *cyoABCD* operon and then the *cbdAB*. For the *cbdAB* operon deletion the resistance cassette was not removed with a FLP recombinase. In order to verify the genotype of all evolved mutants, colonies were isolated from solid media and tested with PCR. Primers used for deletion verification are presented in Table 3.6. Wild type *E. coli* colonies were tested in parallel as a negative control.

3.5.3 Adaptive evolution

The mutant strain was adapted though continuous passage in M9 minimal medium supplemented with 2g/l *D*-glucose and trace elements as reported earlier [9,25]. To initiate evolutions, an ECOM3 mutant was plated on the solid M9 minimal media

containing 2 g/l glucose, trace elements, EZ supplements and kanamycin antibiotic and incubated overnight at 37°C. A single colony was selected from the ECOM3 plate, re-suspended in 10µl of sterile water and inoculated into three 500 ml Erlenmeyer flasks containing 250 ml of M9 minimal media supplemented with 20 ml of 5X Supplement EZ (Teknova). Flasks were incubated at 37°C using a stir bar for mixing and aeration (~1000 rpm). Every day, optical density measurements (OD at 600 nm) were taken and cells were passed into a fresh medium. The volume of the inoculum for each passage was adjusted to account for changes in growth rate, and ensure that cultures would not enter the stationary phase before next passage. The amount of EZ supplements added to the media was reduced exponentially within the first two weeks of evolution. The evolutions were propagated under oxic condition for 60 days [~ 700 generations] until a stable growth rate was reached. Cultures were screened every other day for contamination using PCR. The evolutions were also supplemented with 50µg/ml kanamycin once a week in order to prevent contamination. Each evolved cell population was sampled to investigate the effects of adaptive evolution on cellular metabolism at day 1, day 30 [~ 268 doublings], and day 60 [~ 700 doublings]. Samples were also frozen on day 1 and every 2 days throughout the evolution.

3.5.4 Phenotype assessment

To measure growth rate and byproduct secretion, each population was grown in batch culture at 37°C under oxic and anoxic conditions. Aerobic cultivation was conducted in 500 ml Erlenmeyer flasks containing 250 ml of M9 minimal media with

trace elements and 2 g/l glucose as a sole carbon source. Temperature was controlled at 37°C by a circulating water bath, mixing and aeration was controlled with a stir bar at ~1000 rpm. Anaerobic cultivation was conducted in 250 ml Erlenmeyer flasks with 200 ml of medium, sealed with rubber stoppers containing necessary inlet tubing. Anoxic conditions were achieved by continuous flashing of cultures with 95% N_2 5% CO_2 gas mixture at flow rate of 1 ml/min. The temperature was controlled by using a circulating water bath; the mixing was controlled with a stir speed of ~200 rpm. Samples were taken from the batch cultures regularly (every 30 min), filtered through a 0.2 μm filter and stored at -20°C for byproduct secretion analysis. Glucose concentration in the media was assessed using an enzymatic assay kit (R-Biopharm), while *D*-lactate secretion was measured using RI (refractive index) detection by HPLC (Waters, Milford, MA) with a Bio-Rad Aminex HPX87-H ion exclusion column (injection volume, 50 μl) and 5mM H_2SO_4 as the mobile phase (0.6 ml/min, 65°C). The identities of metabolites and organic acids in the fermentation broth were further verified with enzymatic kits (R-Biopharm). The oxygen uptake rate of each aerobic culture was determined by measuring the rate of dissolved oxygen depletion in an enclosed respirometer chamber using a polarographic dissolved oxygen probe (Cole-Parmer Instruments, Vernon Hills, IL)**.**

3.5.5 Quantitative PCR

RNA samples were taken from exponentially growing cells and added to two volumes of RNA protect (Qiagen, Valencia, CA). Total RNA was isolated using an

RNeasy mini kit (Qiagen, Valencia, CA). Reverse transcription was performed on 10 μg of total RNA. The reverse transcription mixture (60 μL) contained 10 μg total RNA, 75 μg random primers, 1X 1st Strand Buffer, 10 mM DTT, 0.5 mM dNTP's, 30 U of Superase, and 1500 U of Superscript II. The mixture was incubated in a thermocycler (Bio-Rad, Hercules, CA) at 25 °C for 10 min, 37 °C for 1 hr and then 42 °C for 1 hr. The reaction was followed by an incubating at 70 °C for 10 min to inactivate the superscript. The RNA was then degraded by adding 20 μL of 1 N NaOH and incubating at 65 °C for 30 min. After the incubation, 20 μL of 1 N HCl was added to neutralize the solution. QIAquick PCR Purification Kits were used to clean up the cDNA synthesis product. Following the purification, the cDNA was quantified and then directly used in qPCR reactions. The 50 μl of qPCR reaction contained 25 μl of SYBR Green Tag master mix (Qiagen), 0.2 μM forward primer, 0.2 μM reverse primer, and cDNA as a template. Each qPCR reaction was run in triplicates in the Bio-Rad thermocycler (Bio-Rad, Hercules, CA) with the following settings: 95°C for 15 min, 94°C for 15 s, 52°C for 30 s, 72°C for 30 s; the denaturation, annealing and extension steps were repeated for 40 cycles. Gene expression of evolved ECOM3 strains was analyzed under oxic and anoxic growth conditions and compared to the wild type strain under similar growth conditions. In order to determine the binding affinity of each primer set, a standard curve was calculated for each primer and reaction efficiency obtained from it. Using the standard curve, the relative cDNA quantity was obtained for each gene by normalizing it to the quantity of *acpP* (acyl carrier protein) cDNA in the same sample. *acpP* was chosen as the internal control

gene since it is constitutively expressed in wild type and mutants under both aerobic and anaerobic conditions [29].

3.5.6 Clonal analysis

Evolved ECOM3 populations (day 60) were cultured overnight on solid M9 media with 2 g/l glucose and 50 μg/ml kanamycin. Ten random individual colonies were selected from each plate and grown overnight in M9 minimal media with 2g/l glucose. Cells were harvested at 5000 rpm in the centrifuge (Thermo CR3i), washed three times with M9 minimal media without carbon source and loaded on a Bioscreen C machine (Growth Curves USA). Cultures were inoculated into 300 μl wells containing M9 minimal media with 2 g/l glucose and trace elements, the initial OD of each well was kept bellow the 0.05. Cells were grown for 8 hours at 37°C with continuous shaking to ensure good mixing and aeration and optical density (OD at 600 nm) measurements were taken every 15 min. Once the cells reached stationary phase, the assay was stopped and the final *D*-lactate concentration (g/l) was assessed by HPLC. Strains with the highest production yield were identified and subjected to aerobic batch cultivation. Strains were grown in 500 ml Erlenmeyer flasks with 250 ml of M9 minimal media for 8 hours at 37°C with continuous agitation as described above. Samples were taken every 30 min, filtered and analyzed using the HPLC (Waters, Milford, MA). Growth rate (1/h), oxygen uptake rate (mmol/gDW/h), sugar uptake rate (mmol/gDW/h), and product secretion rates (mmol/gDW/h) were measured as described above.

3.5.7 Computational analysis

The computational analyses were done using the *i*AF1260 genome-scale metabolic model of *E. coli* K-12 MG1655 [19]. The simulations were performed using either the Simpheny software platform (Genomatica, San Diego, CA) and the Matlab COBRA Toolbox [30] using established methods for gene deletion [31] and robustness analysis [32]. The metabolic reconstruction has been examined with flux balance analysis (FBA). FBA provides a solution space that contains all of the possible steady-state flux distributions satisfying given constrains [33,34]. Phase plane analysis [31,35] was used to calculate the range of characteristic phenotypes that a network can display as a function of variations in the activity of two reactions, such as LDH (lactate dehydrogenase) and the biomass function (growth). The model was constrained by experimental data by setting lower/upper bounds of uptake/secretion fluxes to the experimentally measured values. In order to allow for experimental uncertainty the bounds were set to within one experimental standard deviation of the experimentally measured mean value.

3.6 ACKNOWLEDGEMENTS

We thank Eric M. Knight and Larisa Kagan for help with generation of *cydAB* *E. coli* mutant strain, Ines Thiele, Karsten Zengler and Dr. Byung Kwan Cho, and Dae-Hee Lee for valuable discussions. The text of this chapter, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Herrgård, M.J., Palsson, B.Ø. Aerobic Fermentation of D-Glucose by an Evolved Cytochrome Oxidase Deficient

Escherichia coli Strain, *Appl Environ Microbiol*. 2008 Dec;**74**(24):7561-9. The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

3.7 REFERENCES

1. Neidhardt FC. *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington, D.C.: ASM Press, 1996:2 v. (xx, 2822 , lxxvii).

2. Adams MA, Jia Z. Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. *J Biol Chem* 2005;280:8358-8363.

3. Bekker M, Kramer G, Hartog AF, Wagner MJ, de Koster CG, Hellingwerf KJ, de Mattos MJ. Changes in the redox state and composition of the quinone pool of *Escherichia coli* during aerobic batch-culture growth. *Microbiology* 2007;153:1974-1980.

4. Jaworowski A, Campbell HD, Poulis MI, Young IG. Genetic Identification and Purification of the Respiratory NADH Dehydrogenase of *Escherichia coli*. *Biochemistry* 1981;20:2041-2047.

5. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007;3:149.

6. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-648.

7. Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol* 2008;26:659-667.

8. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol* 2005;71:7880-7887.

9. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* 2006;281:8024-8033.

10. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-8.

11. Fong SS, Joyce AR, Palsson BO. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res* 2005;15:1365-72.

12. Calhoun MW, Oden KL, Gennis RB, de Mattos MJ, Neijssel OM. Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. *J Bacteriol* 1993;175:3020-3025.

13. Shioi J, Tribhuwan RC, Berg ST, Taylor BL. Signal transduction in chemotaxis to oxygen in *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* 1988;170:5507-5511.

14. Sturr MG, Krulwich TA, Hicks DB. Purification of a cytochrome bd terminal oxidase encoded by the *Escherichia coli* app locus from a delta cyo delta cyd strain complemented by genes from *Bacillus firmus* OF4. *J Bacteriol* 1996;178:1742-1749.

15. Newton G, Gennis RB. In vivo assembly of the cytochrome d terminal oxidase complex of *Escherichia coli* from genes encoding the two subunits expressed on separate plasmids. *Biochim Biophys Acta* 1991;1089:8-12.

16. Puustinen A, Finel M, Haltia T, Gennis RB, Wikstrom M. Properties of the two terminal oxidases of *Escherichia coli*. *Biochemistry* 1991;30:3936-3942.

17. Kita K, Konishi K, Anraku Y. Terminal oxidases of *Escherichia coli* aerobic respiratory chain. II. Purification and properties of cytochrome b558-d complex from cells grown with limited oxygen and evidence of branched electron-carrying systems. *J Biol Chem* 1984;259:3375-3381.

18. Dassa J, Fsihi H, Marck C, Dion M, Kieffer-Bontemps M, Boquet PL. A new oxygen-regulated operon in *Escherichia coli* comprises the genes for a putative third cytochrome oxidase and for pH 2.5 acid phosphatase (appA). *Mol Gen Genet* 1991;229:341-352.

19. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for

Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121.

20. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S. Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* 2006;188:8259-71.

21. Conradt H, Hohmann-Berger M, Hohmann HP, Blaschkowski HP, Knappe J. Pyruvate formate-lyase (inactive form) and pyruvate formate-lyase activating enzyme of Escherichia coli: isolation and structural properties. *Arch Biochem Biophys* 1984;228:133-42.

22. Knappe J, Sawers G. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of Escherichia coli. *FEMS Microbiol Rev* 1990;6:383-98.

23. Rasmussen LJ, Moller PL, Atlung T. Carbon metabolism regulates expression of the pfl (pyruvate formate-lyase) gene in Escherichia coli. *J Bacteriol* 1991;173:6390-7.

24. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual New York: Cold Spring Harbor Laboratory Press, 2001:2.

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

26. Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *J Bacteriol* 1974;119:736-747.

27. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000;97:6640-6645.

28. Cherepanov PP, Wackernagel W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 1995;158:9-14.

29. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004;429:92-96.

30. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2007;2:727-738.

31. Edwards JS, and Palsson, B.O. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics* 2000;1.

32. Edwards JS, Palsson BO. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog* 2000;16:927-939.

33. Edwards JS, Ibarra RU, Palsson BO. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 2001;19:125-130.

34. Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology* 2000;203:229-48.

35. Bell SL, Palsson BO. Phenotype phase plane analysis using interior point methods. *Computers & Chemical Engineering* 2004;29:481-486.

Table 3.1: Phenotypic characteristics of ECOM3 populations during adaptive evolution

	MG1655	ECOM3	ECOM31			ECOM32			ECOM33		
			day 1	day 30	day 60	day 1	day 30	day 60	day 1	day 30	day 60
SUR mmol/gDW/h	9.02 +/- 0.23	11.88 +/- 2.01	7.67 +/- 0.60	20.88 +/- 0.89	19.89 +/- 0.94	11.07 +/- 0.78	20.53 +/- 1.01	21.15 +/- 1.50	11.23 +/- 1.20	16.90 +/- 2.16	21.28 +/- 0.97
OUR mmol/gDW/h	14.92 +/- 0.21	6.9 +/- 1.52	7.90 +/- 1.00	3.81 +/- 0.45	5.61 +/- 0.30	7.75 +/- 0.42	5.26 +/- 0.24	3.69 +/- 0.60	5.04 +/- 0.50	4.29 +/- 0.21	2.22 +/- 0.38
LactSR mmol/gDW/h	0.40 +/- 0.01	17.60 +/- 0.65	16.63 +/- 0.70	35.09 +/- 0.36	36.36 +/- 0.25	15.94 +/- 0.98	35.15 +/- 2.64	35.63 +/- 1.56	17.24 +/- 0.98	25.84 +/- 2.19	29.98 +/- 2.81
AcSR mmol/gDW/h	3.4 +/- 0.02	3.84 +/- 0.45	3.75 +/- 0.49	0.33 +/- 0.05	0.50 +/- 0.02	4.32 +/- 0.58	2.73 +/- 0.47	2.31 +/- 0.31	3.43 +/- 0.68	3.48 +/- 0.22	8.17 +/- 0.91

Table 3.2: Phenotypic comparison of ECOM3 populations between the oxic and anoxic growth environment

Strain	Aerobic GR (1/h)	Anaerobic GR (1/h)	Aerobic Lactate Titer (g/L)	Anaerobic Lactate Titer (g/L)
Unevolved wild-type	0.71 +/- 0.01	0.45 +/- 0.02	0	0.04 +/- 0.01
eECOM31	0.42 +/- 0.01	0.37 +/- 0.01	1.51 +/- 0.01	1.47 +/- 0.01
eECOM32	0.40 +/- 0.02	0.38 +/- 0.02	1.45 +/- 0.04	0.29 +/- 0.02
eECOM33	0.44 +/- 0.02	0.42 +/- 0.01	1.30 +/- 0.06	0.31 +/- 0.03

Table 3.3: Physiological data for the ECOM3 and ECOM4 strains under oxic conditions

Name	Growth rate 1/h	SUR	LactSR	OUR	AcSR	Lactate Yield (%)
			mmol/gDW/h			
ECOM3	0.42 +/- 0.02	21.34 +/- 0.3	35.44 +/- 0.3	3.84 +/- 0.5	3.66 +/- 1.3	80
ECOM4	0.36 +/- 0.01	18.92 +/- 0.6	36.38 +/- 1.0	0.03+/- 0.04	0	97

Table 3.4: List of the double reaction deletions of central metabolism in the iAF1260 that produce optimal phenotypes similar to the ECOM3. Minimal and Maximum D-lactate secretion flux (mmol/gDW/h) and Maximum growth rate (1/h) are shown for WT (wild type) and ECOM3 strains. Experimentally measured fluxes were used as constrains for the model. * Reaction Abbreviations are taken from the *E. coli* computational model *i*AF1260.

Strain	Reactions Abreviation*	Reaction Name	Min D-lactate production (mmol/gDW/h)	Max D-lactate production (mmol/gDW/h)	Max growth rate (1/h)
WT					
	ALCD2x FHL	Alcohol dehydrogenase / Formate-hydrogen lyase	10.33	10.33	0.22
	ACALD PFL	Acetaldehyde dehydrogenase / Pyruvate formate lyase	13.22	13.22	0.20
	ALCD2x PFL	Alcohol dehydrogenase / Pyruvate formate lyase	13.22	13.22	0.20
	PHD PFL	Pyruvate dehydrogenase / Pyruvate formate lyase	13.88	13.88	0.18
*i*ECOM31					
	ALCD2x FHL	Alcohol dehydrogenase / Formate-hydrogen lyase	29.28	29.28	0.39
	ACALD PFL	Acetaldehyde dehydrogenase / Pyruvate formate lyase	32.64	32.64	0.38
	ALCD2x PHL	Alcohol dehydrogenase / Pyruvate formate lyase	32.64	32.64	0.38
	PHD PFL	Pyruvate dehydrogenase / Pyruvate formate lyase	33.98	33.98	0.35
*i*ECOM32					
	ALCD2x FHL	Alcohol dehydrogenase / Formate-hydrogen lyase	31.39	31.39	0.43
	ACALD PFL	Acetaldehyde dehydrogenase / Pyruvate formate lyase	34.82	34.82	0.41
	ALCD2x PHL	Alcohol dehydrogenase / Pyruvate formate lyase	34.82	34.82	0.41
	PHD PFL	Pyruvate dehydrogenase / Pyruvate formate lyase	36.15	36.15	0.38
*i*ECOM33					
	ALCD2x FHL	Alcohol dehydrogenase / Formate-hydrogen lyase	32.19	32.19	0.43
	ACALD PFL	Acetaldehyde dehydrogenase / Pyruvate formate lyase	35.63	35.63	0.42
	ALCD2x PHL	Alcohol dehydrogenase / Pyruvate formate lyase	35.63	35.63	0.42
	PHD PFL	Pyruvate dehydrogenase / Pyruvate formate lyase	36.96	36.96	0.39

Table 3.5: Strains and plasmids used in this study

Strains and Plasmids:	Relevant Characteristics	Source or Reference
Strains		
MG1655	*Escherichia coli* (Wild Type)	ATCC: 47076
ECOM3	MG1655, Δ(*cydAB-appBC-cyoABCD)* ::*FRT-kan-FRT*	This study
ECOM3*ygiN*	MG1655, Δ(*cydAB-appBC-cyoABCD-ygiN)* ::*FRT-kan-FRT*	This study
eECOM31	evolved ECOM3 strain 1 (60 days)	This study
eECOM32	evolved ECOM3 strain 2 (60 days)	This study
eECOM33	evolved ECOM3 strain 3 (60 days)	This study
eECOM31LA	Best Lactate producer isolated from eECOM31 culture	This study
Plasmids		
pKD46	*bla γβ exo (*Red recombinase), temp. conditional pSC101 operon	27
pKD13	Template plasmid with *FRT-kan-FRT* (kanamycin cassett)	27
pCP20	*FLP+ , λ cI857+ , λ pr RepTS, ApR, CmR*	28

Table 3.6: Sequences of the primers used for gene deletion of *cydAB*; *cyoABCD,* and *cbdAB* operons as well as sequences of primers used to confirm the deletions.

Deletion Primers	Forward Primer (5'-->3')	Reverse Primer (5' --> 3')
cydAB	TGTGCCACTGACGCTCGGTATGGCGTTCCT GCTGGCCATTATGGAAACGGTCTACGTCCT GTGTAGGCTGGAGCTGCTTC	TACAGAGAGTGGGTGTTACGTTCAATATCT TCTTTGGTGATACGACCGAACATTTTCCAG ATTCCGGGGATCCGTCGACC
cyoABCD	ATGAGACTCAGGAAATACAATAAAAGTTT GGGATGGTTGTCATTATTTGCAGGCACTGT AGTGTAGGCTGGAGCTGCTTC	TTAGTGCATCATCATGTTGTAGTTGAGGTT CCACATAATCCAGATGGAGCCTACAACCAG ATTCCGGGGATCCGTCGACC
cbdAB	ATGTGGGATGTCATTGATTTATCGCGCTGG CAGTTTGCTCTGACCGCGCTGTATCACTTT GTGTAGGCTGGAGCTGCTTC	TTAGTACAACTCGTTTTCGTTACGGCGGAG AGTTTCTGTTGTCATGCGCCCCCACATTTT ATTCCGGGGATCCGTCGACC
Confirmation primers	Forward Primer (5'-->3')	Reverse Primer (5' --> 3')
cydAB	AAAGAATTAAGGTCAACCG	CGCCCGCAGGGGGCGCTTGTCCATT
cyoABCD	ATAACGCCCTTTTGCAACAG	GTTAAACACACAACCCGACGCCACA
cbdAB	GCTTAGCGAGGTATGTCAGT	TGTCAGATATGAAAAGCGGAAACAT

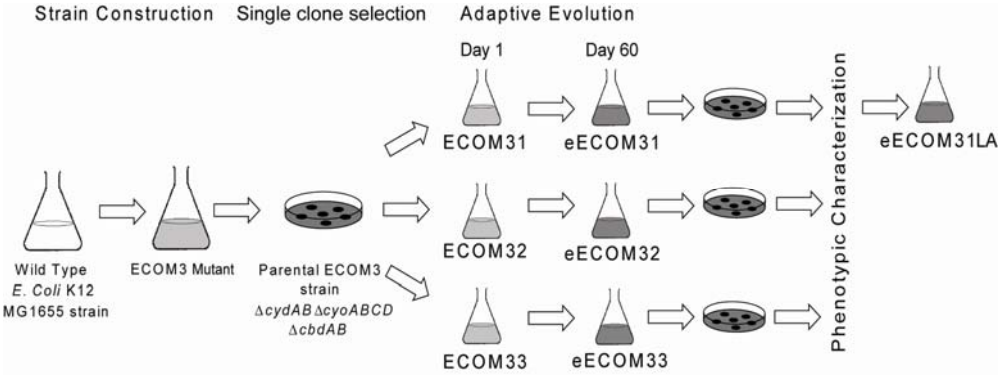


Figure 3.1: The overall experimental design, and nomenclature. Wild type *E. coli* K12 MG1655 strain was converted to parental ECOM3 strain through a series of genetic manipulations. A single clone (colony) of ECOM3 was isolated from the solid media, and subjected to adaptive evolution (ECOM31, ECOM32, ECOM33) resulting in three end-point populations (eECOM31, eECOM32, eECOM33; 'e'-indicates the evolved strain). Following the phenotypic characterization the strain with the highest lactate yield was identified (eECOM31LA).

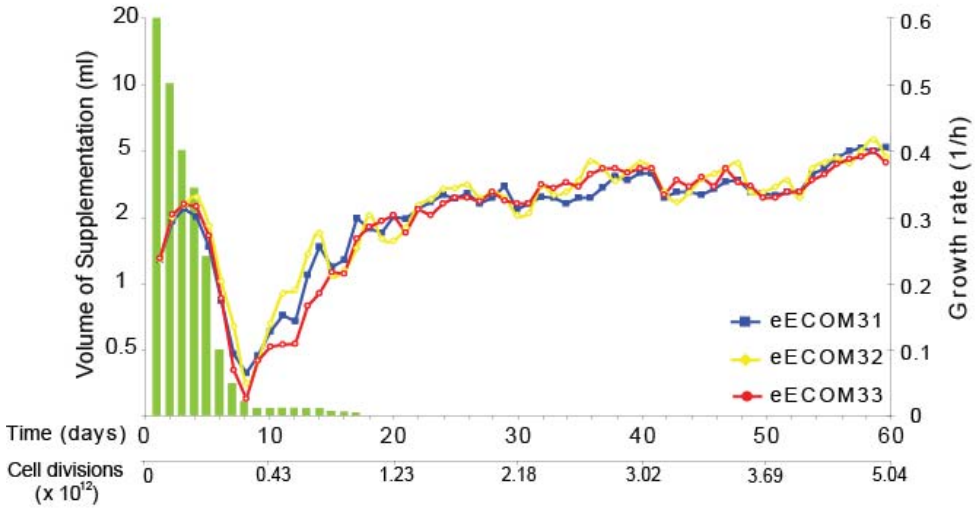


Figure 3.2: Evolutionary trajectories of ECOM3 populations. Growth rate measurements for three evolved ECOM3 strains and cell divisions are shown as a function of time of evolution. Final average growth rate was 0.42 +/- 0.02 1/h. Anaerobic growth rate of wild-type *E. coli* is 0.45 +/- 0.02 1/h. The EZ amino acid supplement amount (in ml) is shown with black bars. The total number of cell divisions for the entire period of adaptation is presented on a secondary abscissa.

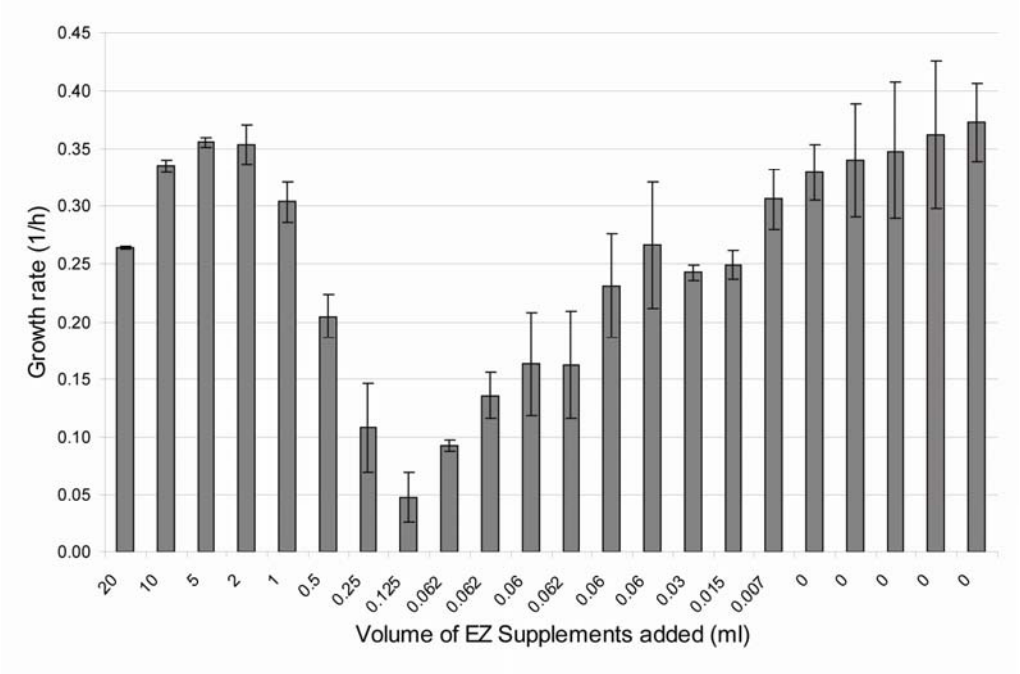


Figure 3.3: Growth rate measurements for three evolved ECOM3 populations as a function of EZ Supplements addition (ml). Drop in growth rate was observed as supplementation was decreases during the first two weeks of evolution. No such response was observed after day 15 of evolution (EZ supplements amount is 30 μl).

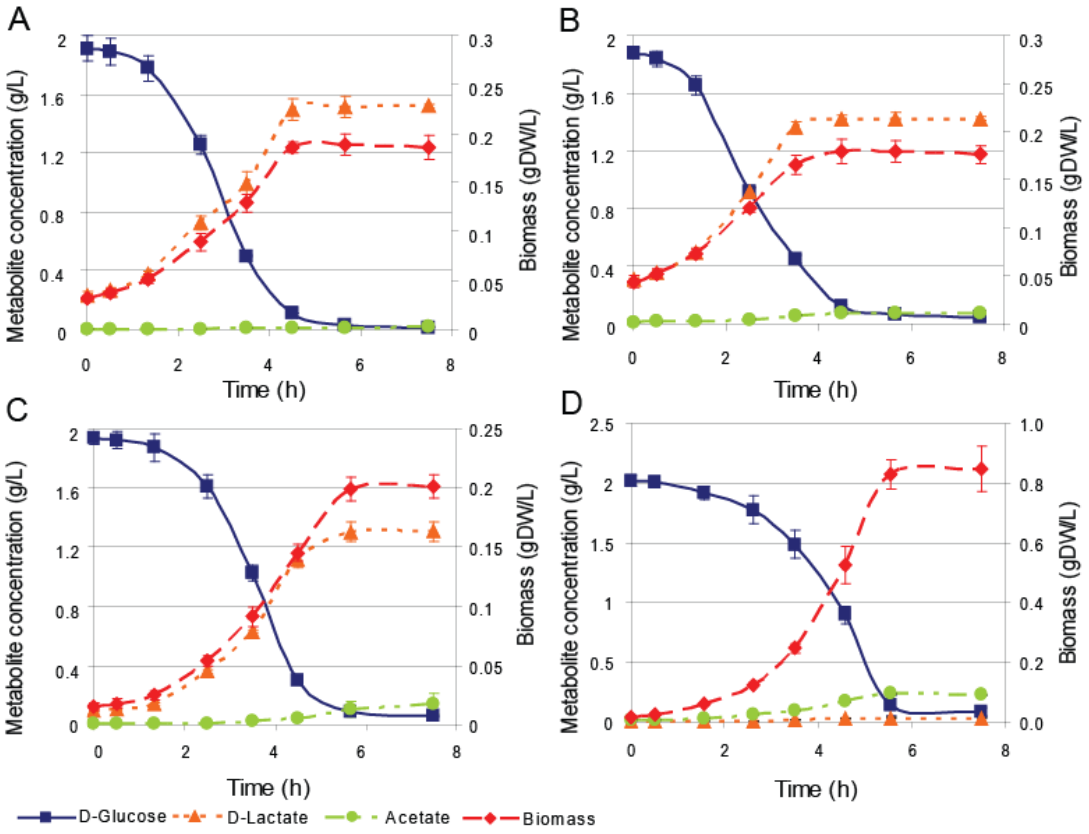


Figure 3.4: Aerobic growth and secretion profiles of three end-point populations. A) eECOM31, B) eECOM32, C) eECOM33, and D) Wild Type strains. Data taken on day 60 of evolutions, all measurements were done in triplicate. (■) indicates the concentration of glucose remaining in the culture; (Δ) indicates the amount of *D*-lactate produced; (O) – concentration of acetate; (◆) – cell density (gDW/L).

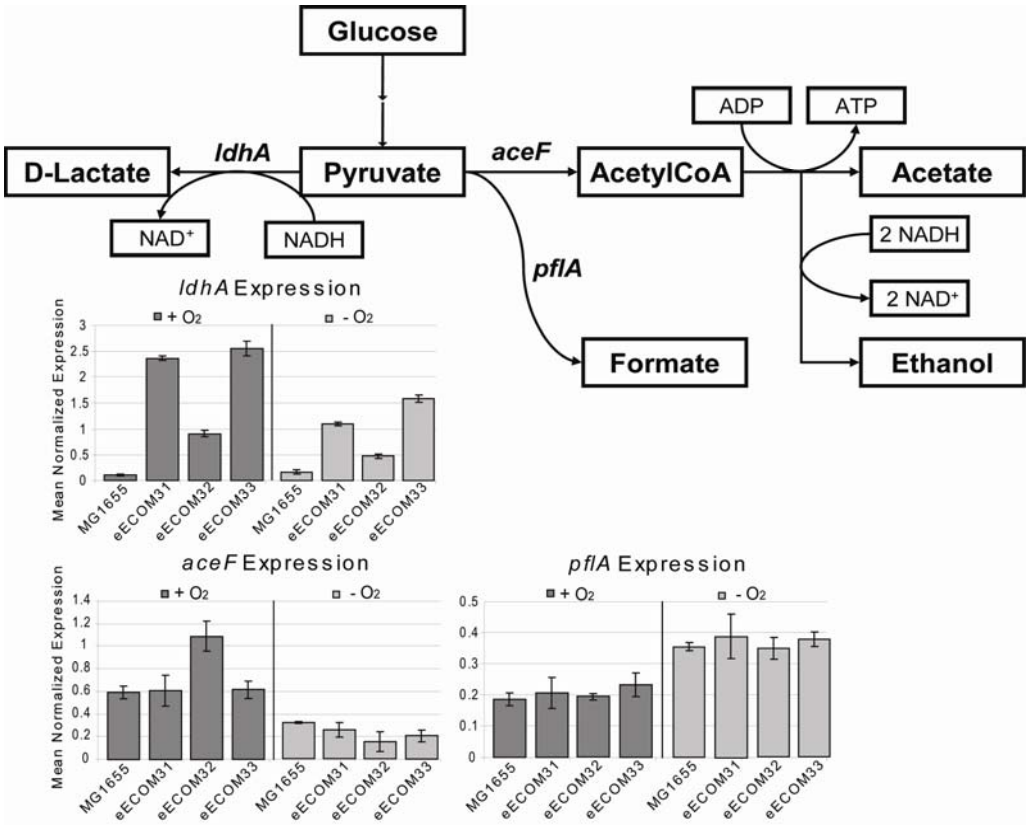


Figure 3.5: Mechanism of *D*-lactate production and associated gene expression analysis. Pathways of conversion of pyruvate to common organic acids are presented with corresponding enzyme names. Gene expression of *ldhA, aceF,* and *pflA* genes was measured and presented by the bar diagrams. Gene expression was measured under oxic (dark grey bars) and anoxic (light grey bar) conditions. *ldhA* showed a significant upregulation while no upregulation was observed for *aceF* and *pflA* genes.

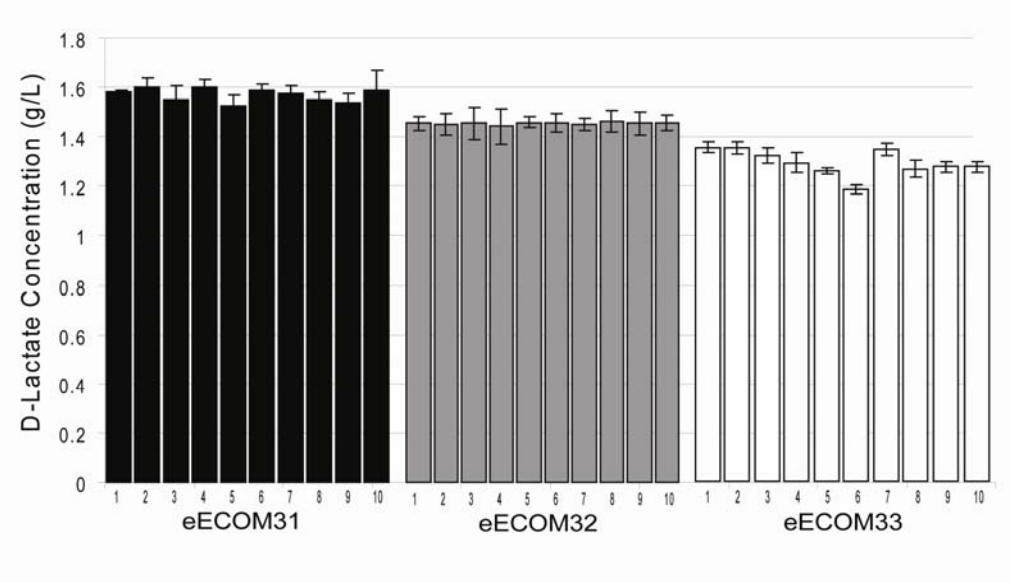


Figure 3.6: D-lactate concentration from clonal analysis. Concentration of D-lactate produced by 30 clones isolated from three end point strains after 8 hours of growth. eECOM31 – black bar; e ECOM32 – grey bar, eECOM33 – white bar.

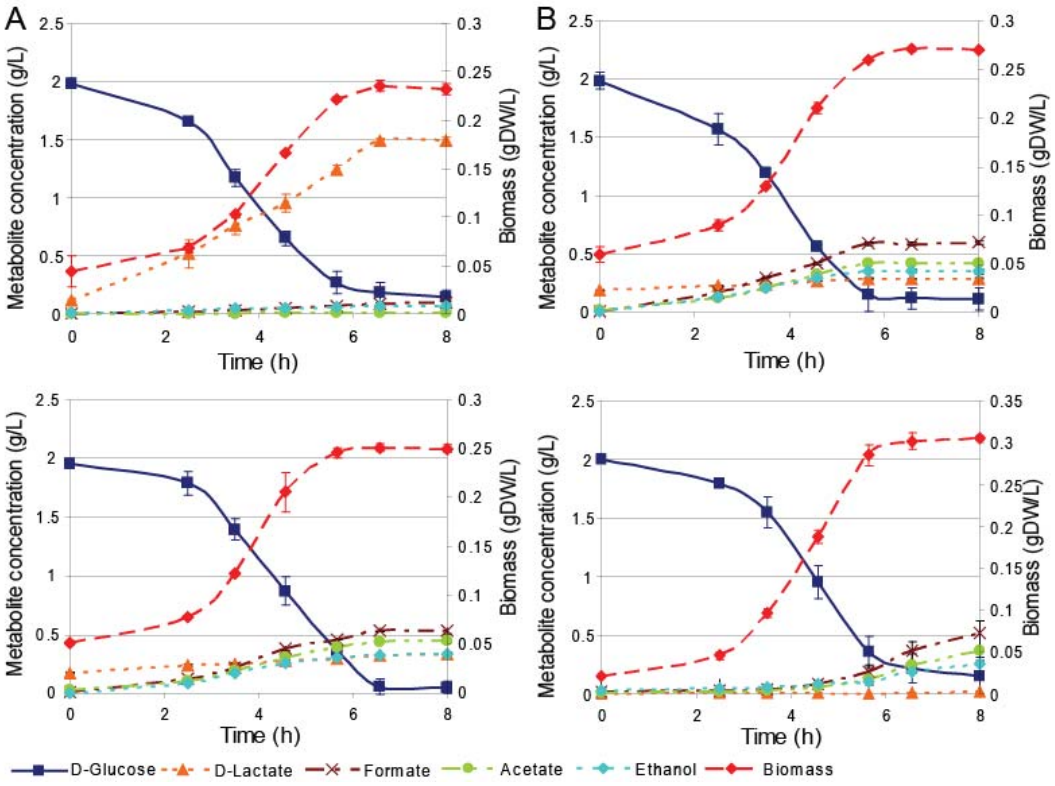


Figure 3.7: Anaerobic growth and secretion profile of three endpoint populations. A - eECOM31, B - eECOM32, C - eECOM33, D - Wild Type strain. Solid black line (■) indicates the decrease in concentration of glucose remaining in the culture; dashed black line (Δ) indicates the amount of *D*-lactate produced; (O) – acetate concentration; (X) – formate concentration; (◇) – ethanol concentration and (◆) – cell density (gDW/L).

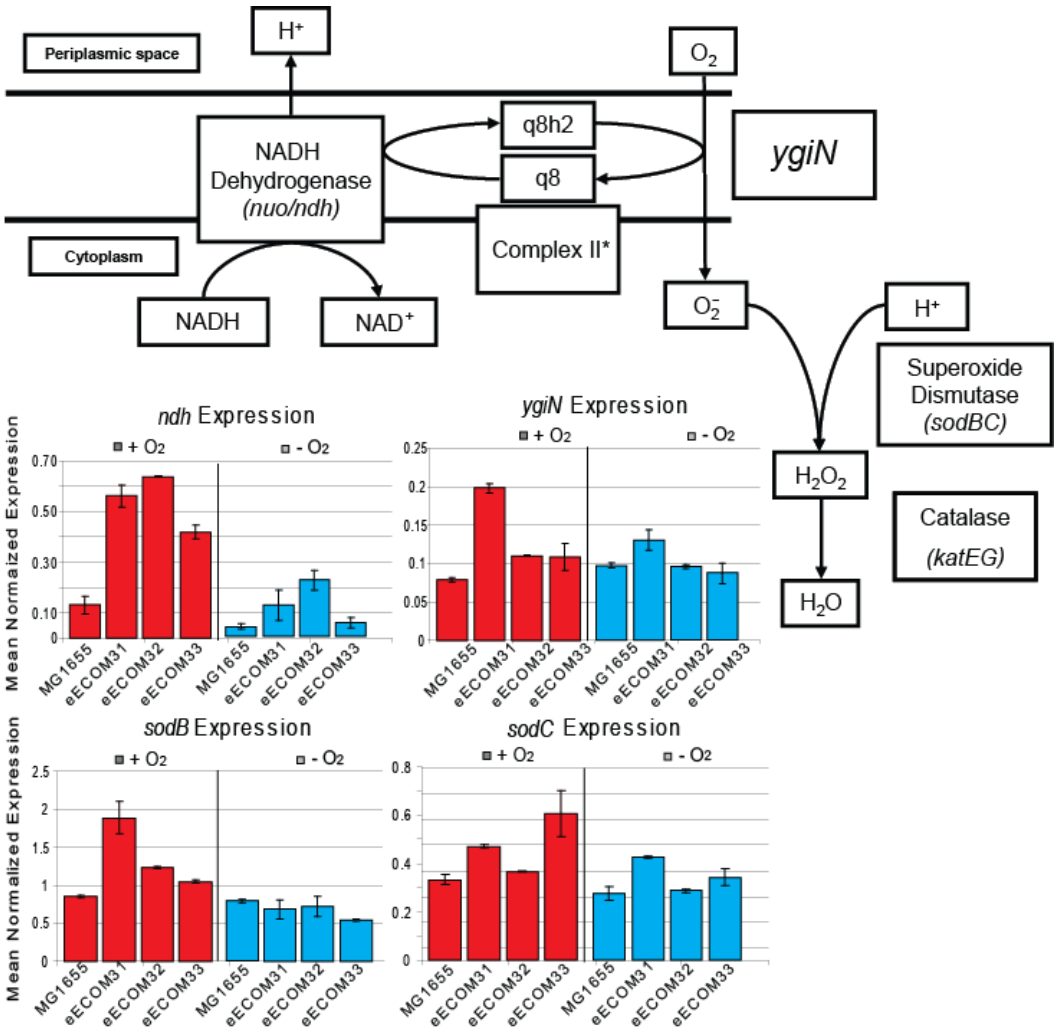


Figure 3.8: Proposed mechanism of oxygen utilization by ECOM3 strains and corresponding gene expression**.** Possible mechanism of oxygen utilization by ECOM3 strains is presented. This mechanism was elucidated based on gene expression analysis (bar diagrams) and scientific evidence [2-4]. Gene expression was measured under oxic (red bars) and anoxic (blue bar) conditions. Expression for the catalase is not shown.

* Complex II was not included in this figure due to lack of evidence of its involvement.

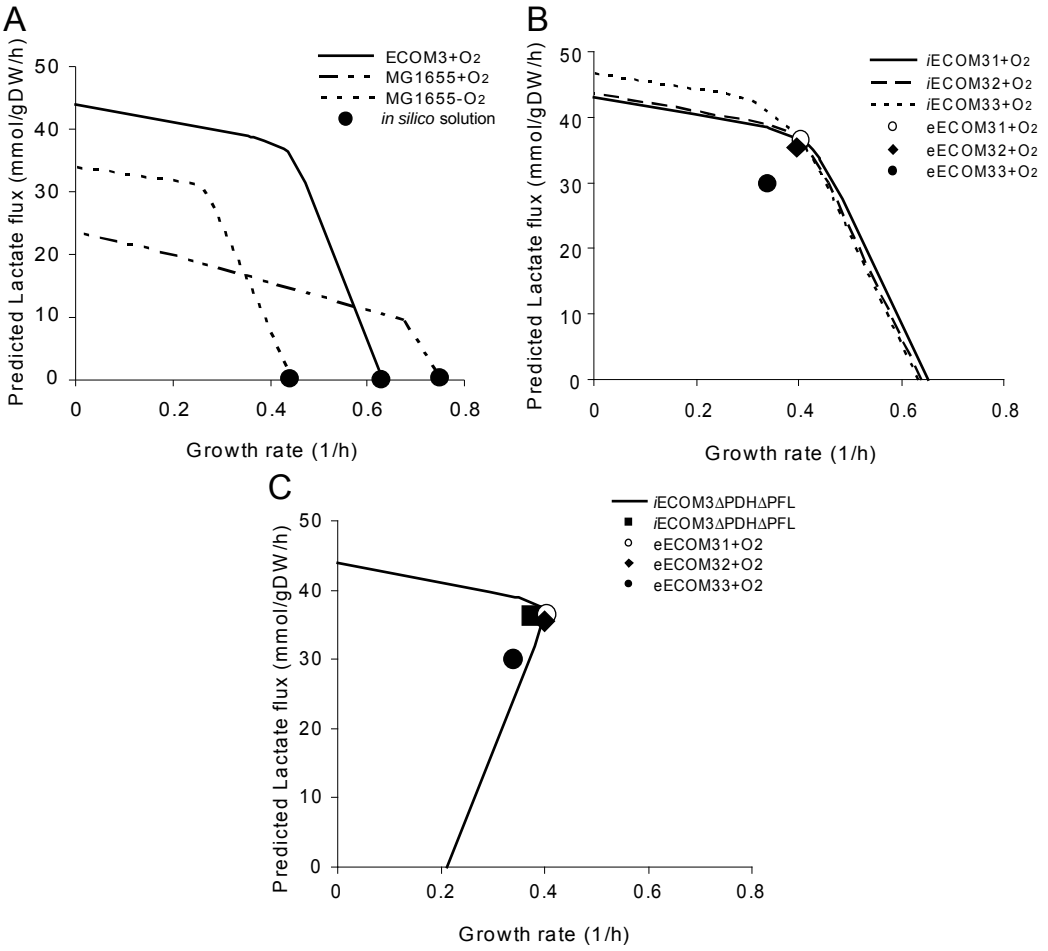


Figure 3.9: Comparison of experimental and computationally predicted growth rates and lactate secretion rates for the evolved ECOM3 strains. A. Two-dimensional computational solution envelopes (lines) containing all possible flux distributions for the given constrains and computationally predicted flux values (circles) are shown for aerobic and anaerobic wild type (*i*MG1655) strain and for aerobic *i*ECOM3 strain (using average values for the three evolved strains as constraints). The model predicts that the experimental flux values need to reside inside the solution envelope and that the optimal flux value to maximize biomass formation would reside at the maximum allowed growth rate point of the flux space (circles). B. Computational solution envelopes (lines) are shown for each of three end-point strains (solid line - ECOM31, dashed line - ECOM32, dotted line - ECOM33) using experimental data as constrains. eECOM31 (○), eECOM32 (◆), eECOM33 (●) solutions are based on experimental results and are lying within the solution envelope; C. The modified solution envelope after the removal of PHD (pyruvate dehydrogenase) and PFL (pyruvate formate lyase) reactions is shown (line). The optimal computational solution (*i*ECOM3ΔPDHΔPFL) as well as experimental data points (eECOM31, eECOM32, and eECOM33) are shown. Note: *'i'*- indicates an *in silico* solution.

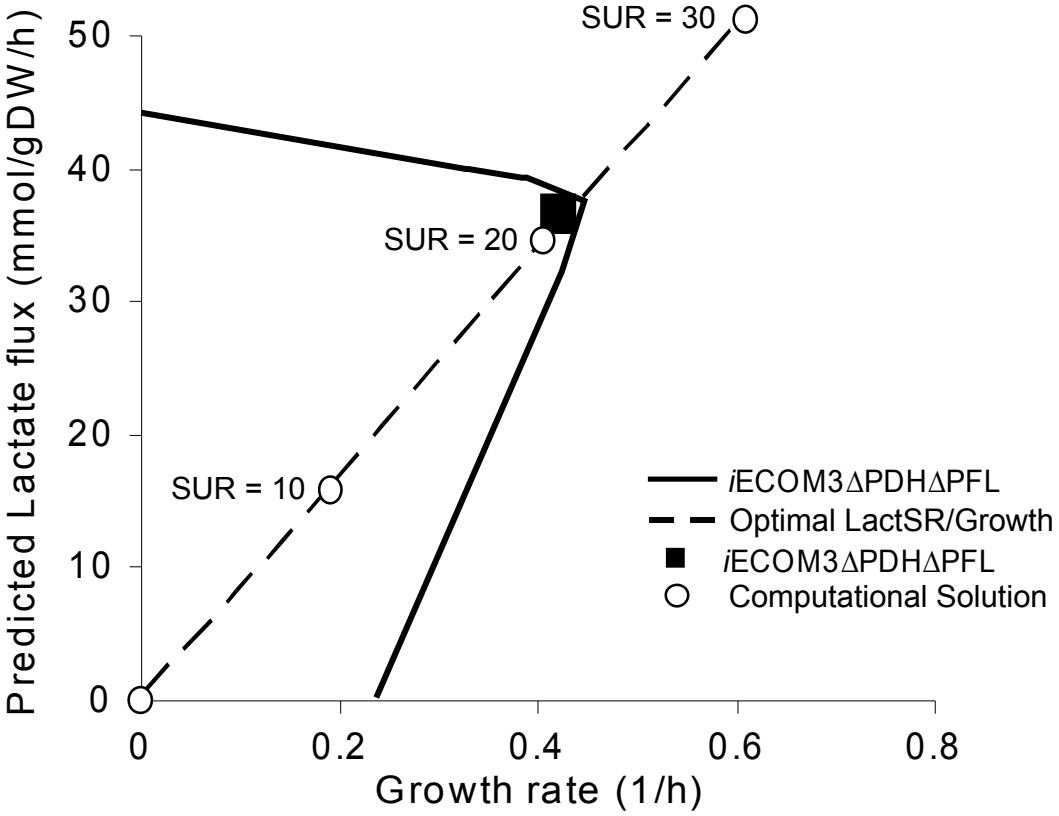


Figure 3.10: Lactate secretion as a function of substrate uptake. Correlations between predicted lactate flux and growth rate with respect to different substrate uptake rates (SUR) are presented. With variable SUR the model predict different lactate/growth ratio but values are located on the same line as the original prediction.

CHAPTER IV:

PHYSIOLOGY OF THE OXYGEN-INDEPENDENT PLATFORM STRAIN

Removal of three terminal cytochrome oxidases (Δ*cydAB,* Δ*cyoABCD,* Δ*cbdAB*) and quinol monooxygenase (Δ*ygiN*) from the *E. coli* K12 MG1655 genome resulted in the activation of ArcA aerobically. These mutations resulted in reduction of oxygen uptake rate by nearly 98% and production of *D*-lactate as a sole by-product under oxic and anoxic conditions. The knock-out strain exhibited nearly identical physiological behavior under both conditions, suggesting that mutations have led to significant metabolic and regulatory perturbations. In order to fully understand the unique physiology of this mutant and identify underlying metabolic and regulatory reasons that prevent transition from aerobic to anaerobic phenotype we utilized whole-genome transcriptome analysis, combined with ^{13}C tracing experiments and physiological characterization. Our analysis showed that the deletions resulted in the activation of anaerobic respiration under oxic conditions and a consequential shift in the content of the quinone pool from the ubiquinones to menaquinones as electron carriers. Increase in menaquinone concentration results in activation of ArcA. The activation of ArcB/ArcA regulatory system has led to a major shift in the metabolic flux distribution through the central metabolism of the mutant strain. Flux analysis

indicated that the mutant strain had undetectable fluxes through the TCA cycle and the oxidative pentose phosphate pathway, and elevated flux through the glycolysis and the anaplerotic reaction. Flux and transcriptomics data were highly correlated and showed similar patterns.

4.1 INTRODUCTION

Escherichia coli has been studied extensively with respect to its physiology, genetics, and metabolism. One of the unique features of its metabolism is the ability to support robust growth under both oxic and anoxic conditions [1]. During aerobic growth, when oxygen is used as a terminal electron acceptor, *E. coli* divides rapidly and produces carbon dioxide and acetate as major growth by-products [1] representing an efficient form of energy metabolism. In the absence of oxygen, *E. coli* and other microorganisms rely on anaerobic respiration and fermentation in order to oxidize substrates, recycle electron carriers, and generate energy [1]. This metabolic versatility of *E. coli* allows it to survive and thrive over a wide range of conditions.

Since the ability to produce a number of reduced by-products such as organic acids and ethanol is of importance in the field of metabolic engineering, the majority of the metabolic engineering designs rely on anaerobic conditions [5-7]. It has also been shown that *E. coli* strains developed for overproduction of commodity chemicals can be further improved using adaptive evolution strategies [5]. Adaptive evolution is often performed through a series of dilutions allowing cells to remain within the exponential phase; the environmental condition is hereby kept similar from passage to passage as

environmental perturbations might result in an incorrect evolutionary trajectory [5]. Implementation of adaptive evolution for strains required strict anoxic growth conditions is a challenging task; therefore a development of a platform strain that would be insensitive to oxygen and would exhibit similar physiological behavior under oxic and anoxic conditions would be beneficial and will significantly simplify the process of adaptation of anaerobic strain designs.

We have shown in the previous chapter that the simultaneous deletion of three terminal cytochrome oxidases (*cydAB, cyoABCD and cbdAB*) and quinol monooxygenase (*ygiN*) results in nearly complete abolishment of oxygen uptake rate [8]. The strain harboring these four mutations was named ECOM4 (*Escherichia coli* Cytochrome Oxidase Mutant 4). The ECOM4 strain was unable to undergo aerobic-anaerobic shift and exhibited similar phenotype under both conditions, making it suitable to be used as a platform strain for the implementation and adaptation of the strain designs. Comprehensive understanding of the metabolism and physiology of the platform strain is important as it provides insights for further engineering. Therefore, it is imperative to understand how the deletions reprogram the metabolic network of *E. coli.*

4.2 RESULTS

4.2.1 Strain engineering and adaptive evolution

The ECOM4 strain was constructed from the unevolved ECOM3 strain previously described by us [8]. The ECOM4 strain was initially incapable of growing on M9 minimal medium and required amino acid supplementation for the robust growth.

ECOM4 was adapted in culture to grow in unsupplemented M9 minimal media with glucose as the sole carbon source. Evolutionary trajectories for three populations are presented in Figure 4.1A. After 30 days of adaptation, the average growth rate of the three evolved populations (ECOM41, ECOM42, and ECOM43) in minimal medium was nearly identical (within 5%) to the growth rate in supplemented medium prior to evolution (Figure 4.1B). Lactate yield decreased slightly during growth adaptation. The evolved oxygen uptake rate (OUR) was nearly identical to the unevolved strain, at a level of 0.25 +/- 0.12 mmol/g-dwt/hr, and nearly 60 times lower than the OUR of WT *E. coli*. Detailed phenotypic data such as growth rate, glucose and oxygen uptake rates, and by-product secretion rates measured during evolution are presented in Table 4.1. In order to characterize and study heterogeneity in the evolved populations, we used clonal analysis. A single clone (ECOM4LA) was selected from one of the three evolved populations (ECOM41) based on the highest lactate yield. The following experiments were performed in triplicates using the ECOM4LA clone.

4.2.2 Phenotypic characterization revealed substantial similarity between aerobic and anaerobic ECOM4

The ECOM4LA strain was predicted to have a similar growth rate irrespective of the oxygen supply. Growth rates were comparable for aerobic and anaerobic conditions, being 0.32 +/- 0.02 1/hr and 0.27 +/- 0.06 1/hr, respectively. Similarly, the conversion of glucose to *D*-lactate was only slightly affected by oxygen supply, with yields of 98% and 92% for aerobic and anaerobic growth conditions (Table 4.2). Lactate was produced with 70% yield during the exponential phase and with nearly

100% yield during the stationary phase (Figure 4.2); illustrating that 30% of carbon was directed towards biomass formation. Succinic acid was present in a low amount during the exponential growth phase and was metabolized in stationary phase. The highest measured concentration of succinate during the exponential phase was on the order of 30.0 +/- 15.0 mg/L; however, this measurement was highly variable due to re-uptake of succinate. Based on these results we hypothesized that the ECOM4LA strain might be using the anaerobic respiratory chain in order to remove the excess electrons during aerobic growth (see Discussion), while the majority of the electrons are removed by means of *D*-lactate production. Gene expression analysis was used to determine metabolic changes, and to decipher possible regulatory alterations that underlie the inability of ECOM4LA to undergo the normal aerobic-anaerobic shift.

4.2.3 Gene expression analysis reveals a shift to anaerobic metabolism in ECOM4LA under oxic conditions

Genome wide gene expression profiles were determined for the WT and ECOM4LA strains under aerobic and anaerobic conditions. Expressed genes were selected based on criteria described earlier (see Materials and methods). The gene expression comparison between aerobic and anaerobic growth in WT *E. coli* revealed that 564 genes (13% of the genome – based on 4468 total genes in the *E. coli* genome [3,4]) had significant changes in expression (Figure 4.3). Comparison of mRNA transcript levels between ECOM4LA and WT under oxic growth conditions revealed that 538 genes were significantly affected, accounting for nearly 13% of the genome, similar to the previous comparison. Interestingly, we observed that only ~6% of the

genome (250 genes) was affected by an aerobic-anaerobic shift in ECOM4LA cell line (Figure 4.3). This observation suggested that the inability to utilize oxygen has a significant effect on global gene expression and regulation, which contributes significantly to the inability of ECOM4LA strain to undergo an aerobic-anaerobic shift.

Gene Ontology (GO) term enrichment was employed to identify biological processes that are enriched within differentially expressed genes between various experimental conditions. Interestingly, the WT aerobic-anaerobic shift and the aerobic WT/ECOM4LA comparison shared several enriched metabolic GO biological processes such as "aerobic respiration," "anaerobic respiration," "tricarboxylic acid cycle," "oxidation reduction," and "glycolysis" (Supplemental Table 4.1). Moreover, in the comparison between WT and ECOM4LA under oxic conditions, most of the significantly enriched GO terms in the down-regulated genes were similar to the enriched terms in the WT aerobic-anaerobic shift (Supplemental Table 4.2), and these were dominated by metabolic processes.

Since metabolic terms dominated in the differentially expressed genes, we mapped the transcriptomic data onto the *E. coli* metabolic network reconstruction [9]. Central metabolism was analyzed in detail (i.e., glycolysis, TCA cycle, pentose phosphate pathway (PPP), and fermentative pathways). We considered the gene expression pattern acquired from aerobic and anaerobic growth in WT *E. coli* as a benchmark to which we compared gene expression in the ECOM4LA strain under similar conditions. We observed that during the aerobic-anaerobic shift, WT downregulated the TCA cycle and upregulated expression of certain enzymes involved

in glycolysis and fermentative pathways such as formate, acetate, and succinate production (Figure 4.4A). When examining differences between gene expression of the ECOM4LA strain and WT *E. coli* grown in an oxic environment, we noticed that the majority of genes involved in glycolysis were significantly upregulated, while genes involved in TCA cycle were downregulated in ECOM4LA (Table 4.3, Figure 4.4B).

Similar expression patterns were observed between WT (anaerobic) and ECOM4LA (aerobic) compared to WT (aerobic) (Figure 4.4A/B). These results suggest that the ECOM4LA strain relies on glycolysis under oxic growth conditions for energy generation through substrate level phosphorylation. This result might be attributed to deletions in respiratory chain genes and an inability to build a sufficient proton gradient to produce energy by ATP-synthase. Also upregulation of the anaplerotic reaction (from phosphoenolpyruvate to oxaloacetate) was observed in the ECOM4LA strain. We also noticed that the lactate dehydrogenase (*ldhA*) gene was upregulated over 3 fold similarly to what was observed earlier for the parent strain [8]. Comparable, but less profound expression changes were observed in anaerobic ECOM4LA compared to aerobic WT (Figure 4.4C). We observed significant downregulation of the TCA cycle and upregulation of some glycolytic enzymes, suggesting that similar regulatory mechanisms are active in this strain under both environmental conditions.

The most interesting result was observed when we mapped the gene expression of ECOM4LA strain during an aerobic-anaerobic shift. This comparison indicated no changes to central metabolism in ECOM4LA under oxic and anoxic conditions

(Figure 4.4D). The only gene that had almost a 2-fold increase in expression was fumarate reductase (*frdABCD*). Other significantly expressed genes mapped sparsely, without any definite pattern, onto the entire metabolic map in the *i*AF1260 metabolic model. The lack of more significant changes in gene expression between aerobic and anaerobic profiles for the ECOM4LA strain illustrates that functionalities of the central metabolism of respiratory deficient ECOM4LA strain has been reduced to perform similar functions under both studied growth conditions.

4.2.4 Gene expression suggests ArcA is active in aerobic ECOM4LA, while FNR is not

Since differential expression of central metabolic genes is similar in ECOM4LA in comparison to anaerobically grown WT *E. coli*, we asked if this anaerobic behavior in ECOM4LA extended beyond its metabolism. *E. coli* has two different regulators that control expression of genes involved in the aerobic-anaerobic shift, consisting of the ArcB/ArcA two-component system and FNR. FNR is a transcriptional regulator whose activity is regulated directly by oxygen [10]; therefore, it is expected that under aerobic conditions, the FNR regulon would not be significantly differentially expressed between the WT and ECOM4LA. Consistent with this, the FNR regulon (from RegulonDB v 6.0) does not have more differentially expressed genes in the microarray data than expected by chance ($p = 0.11$, hypergeometric test). Moreover, in the comparison between aerobic WT and ECOM4LA, less than 4% of the FNR regulon (excluding ArcA/FNR co-regulated genes) is differentially expressed in the direction consistent with FNR activity (Figure 4.5A). In like manner, the

comparison between WT and ECOM4LA under anoxic conditions also shows little difference. However, for both strains, the shift from aerobic to anaerobic conditions clearly causes gene expression changes in the FNR regulon consistent with known FNR activity of activation or repression (Figure 4.5A), suggesting that FNR activity changes in the aerobic-anaerobic shift, but not between these two strains.

Conversely, the analysis of the ArcA regulon revealed significant differences between ECOM4LA and WT, but little ArcA-associated change in the ECOM4LA aerobic-anaerobic shift. When WT and ECOM4LA are compared, the ArcA regulon is enriched among the differentially expressed genes ($p = 4.5 \times 10^{-12}$, hypergeometric test), which is the second most significantly enriched regulon for this condition (Supplemental Table 4.3). Moreover, in the WT aerobic-anaerobic shift, 71% of the differentially expressed ArcA regulon genes that are consistent with reported ArcA function (as an activator/repressor), also show the same consistency when comparing WT and ECOM4LA under aerobic conditions (Figure 4.5B). Furthermore, differentially expressed genes between WT and ECOM4LA under aerobic conditions were consistent with known functions of ArcA ($p = 0.03$, Fischer's exact test). Conversely, few genes in the ArcA regulon are significantly differentially expressed between anaerobic WT and ECOM4LA or between aerobic and anaerobic ECOM4LA, suggesting that ArcA activity is similar in these three scenarios. Together, these results show that gene expression changes are consistent with ArcA being active in ECOM4LA under aerobic conditions.

4.2.5 Quinone pool is dominated by menaquinones in ECOM4LA under oxic condition

The activation of ArcA under aerobic conditions may be due to the fact that ArcA is a part of a two-component regulatory system that responds to the redox state of the quinone pool [11]. Since the aerobic respiratory chain cannot be utilized in ECOM4LA, the ubiquinone pool is diminished and complemented by menaquinones that are involved in the anaerobic respiration mechanism (from NADH to fumarate). Relative amounts of ubiquinone and menaquinone species present in actively growing ECOM4LA and WT under oxic and anoxic conditions were measured (Figure 4.6). Consistent with previous reports, the quinone pool of the WT is dominated by ubiquinones during aerobic growth (800 +/- 60 nmol/g-dwt) and by menaquinones during anaerobic growth (650 +/- 80 nmol/g-dwt) [11]. For the ECOM4LA strain we observed that the ubiquinone content varied between 150.0 and 200.0 nmol/g-dwt for anaerobic and aerobic conditions respectively, while menaquinones were present at a much higher concentration for both conditions (500.0 +/- 150.0 nmol/g-dwt for anaerobic and 450.0 +/- 100.0 nmol/g-dwt in aerobic conditions) (Figure 4.6). Since ECOM4LA cannot utilize molecular oxygen, the quinone pool in the mutant strain has a completely different content with respect to WT. The ECOM4LA quinone pool is dominated primarily by menaquinones under both conditions. It has been shown that presence of ubiquinones inhibits activation of the ArcB/ArcA system, while abundance of menaquinones alleviates this inhibition leading to activation of ArcA [11].

These results suggest that the anaerobic phenotype of aerobically growing ECOM4LA is due to the activation of the ArcA regulon by a disruption in the cellular redox balance. Under oxic conditions, phosphorylated ArcA activates numerous operons involved in fermentative metabolism [12,13] and represses operons involved in respiratory metabolism [14].

4.2.6 Targeted gene expression measurements

In order to validate observed levels of gene expression, we used qPCR. Genes selected for qPCR analysis under oxic condition, in ECOM4LA as compared to WT, included: NADH: menaquinone oxidoreductase (*yieF, wrbA*), fumarate reductase (*frdABCD*), and succinate dehydrogenase (*sdhABCD*). qPCR analysis confirmed that *yieF* was upregulated nearly 10 fold, *wrbA* was upregulated over 40 fold and the *frdABCD* operon was upregulated over 50 fold in the aerobic ECOM4LA strain. We observed a significant downregulation (over 25 fold down) of *sdh* operon in aerobic ECOM4LA, which is similar to WT under anoxic conditions (over 30 fold down) (Table 4.4). These findings, together with observed downregulation of the TCA cycle, imply that the regulation in aerobic ECOM4LA cell line is similar to anaerobic WT *E. coli.*

4.2.7 Carbon labeling experiments

Gene expression analysis indicated major differences in the metabolism of ECOM4LA versus its WT *E. coli* parent when grown aerobically. To confirm this,

metabolism was assayed directly by ^{13}C labeling of both strains under aerobic conditions. ^{13}C labeling was used to infer relative flux through different sections of central metabolism, particularly the pentose phosphate pathway (PPP), glycolysis, and the TCA cycle (Table 4.5).

Pentose phosphate pathway and glycolysis

PPP versus glycolytic flux was calculated in two ways. First, it was estimated from labeling patterns of alanine produced from U-13C-glucose. Using the calculation of Szyperski (11) for the reassortment of intermediates in the non-oxidative branch of the PPP (leading to a reassortment of C1-C2 in pyruvate and therefore alanine), in WT a maximum 13% of pyruvate was formed from PPP. In ECOM4LA, the calculated percentage was -2% (or effectively zero; Table 4.5).

Similar values were found for flux through the PPP versus glycolysis using 1-^{13}C-glucose-generated data. Here, alanine labeling patterns were analyzed for loss of ^{13}C label consequent upon the loss of the labeled 1-carbon of glucose as CO_2 during transit through the oxidative branch of the PPP. This analysis yielded a value for PPP flux of 15% in WT and 2% in ECOM4LA, relative to glycolysis (Table 4.5, Figure 4.7B). If it is assumed that all glucose taken up was channeled to glycolysis or the PPP, then the relative PPP flux can be converted to an absolute flux by multiplying these percentages by the measured glucose uptake rates (Table 4.2). This calculation gives values for PPP flux of 1.4 and 0.5 mmol/g-dwt/hr for WT and ECOM4LA, respectively.

A different perspective on the PPP was provided by analyzing histidine labeling from U-13C-glucose cultures. From histidine labeling patterns, it was

possible to calculate relative inputs to the P5P pool (including ribose-5-phophate needed for RNA and DNA synthesis) from oxidative or non-oxidative PPP. In WT and ECOM4LA, respectively 19% and 13% of input to P5P was from oxidative PPP, with the balance from non-oxidative PPP (Table 4.5). The slightly stronger preference for non-oxidative PPP in ECOM4LA versus WT corresponds to generally enhanced expression of non-oxidative PPP genes in ECOM4LA versus WT, while expression of most of the non-oxidative PPP genes shows no difference (Figure 4.4A/B).

Functioning of TCA cycle

Amino acid labeling data from U-^{13}C-glucose-grown *E. coli* were used to determine the relative input of anaplerosis (via PEPC or malic enzyme) versus the TCA cycle to OAA (aspartate) (Figure 4.7C, Table 4.5). WT *E. coli* were found to have a split input to OAA, ~40% from TCA cycle and 60% from anaplerotic reactions. This result is similar to that found before for WT *E. coli* strains growing in glucose minimal medium in aerated flasks [15,16]. In contrast, in the ECOM4LA strain, OAA was (within limits of error) exclusively synthesized by anaplerosis. This indicated that the TCA cycle was non-functional somewhere between oxoglutarate (glutamate) and OAA. The labeling patterns of aspartate fragments in ECOM4LA did indicate some recycling of oxaloacetate through the symmetrical TCA intermediate fumarate (and possibly also succinate) (Table 4.5), which would indicate a break in the cycle closer to oxoglutarate. These measurements corresponded to the gene expression results, which showed that in the ECOM4LA relative to WT, expression of almost all of the TCA cycle enzymes was lower; while expression of the anaplerotic enzyme phosphoenolpyruvate carboxykinase was greater (Figure 4.4B).

Other pathways

The ED pathway was evaluated as an alternate route to pyruvate from glucose. Although expression of genes encoding the ED pathway are usually weak in *E. coli* grown on glucose [17], it was previously shown that *E. coli* mutants which were disabled in components of the TCA cycle (Sdh/Mdh or FumA) produced ~20% of their pyruvate via the ED pathway [15]. Calculating the ED flux (versus glycolysis plus PPP) [15], we found that it was insignificant (Table 4.5) and certainly was not relatively more important in the TCA-non-functional ECOM4LA mutant. This corresponded to a lack of enhancement in gene expression for enzymes in the ED pathway (Figure 4.4B).

Using 1-^{13}C- or 6-^{13}C-glucose data, the degree of labeling of the 1-C pool was calculated from aspartate and methionine labeling (the latter being equivalent in its origins to aspartate plus a 1-C unit). For both *E. coli* strains cultured with 6-^{13}C-glucose, 1-C pool labeling was slightly below 50% (Table 4.5), but was less with 1-^{13}C-glucose, reduced (relative to 6-^{13}C-glucose labeling) by 20% for WT and 3% for ECOM4LA. These reduced labeling levels reflected loss of label from glucose routed through the oxidative PPP before conversion to serine and thence into the 1-C pool, and corresponded roughly to the relative flux through this pathway calculated from 1-^{13}C-glucose-labeled alanine data (Table 4.5). From fragment data for serine and glycine, the percent ^{13}C labeling at serine-3 or glycine-2 was calculated, and from this (assuming that these were the only two sources for the 1-C pool) the contribution of each to the 1-C pool. In all cases, serine-3 was the predominant precursor (Table 4.5).

Endogenous sources of CO_2

Data from U-^{13}C-glucose labeling experiments indicated that most of the CO_2/bicarbonate used in anaplerotic reactions was derived from glucose and not from atmospheric CO_2 (Table 4.5). In ECOM4LA, 70% of CO_2 was from glucose versus 90% in WT. The lesser figure in ECOM4LA is not unexpected, as the mutant was lacking in any flux through two major CO_2-evolving steps in the TCA cycle which would be expected to increase the proportion of CO_2 from internal sources. CO_2 labeling in both strains with 1-^{13}C was similar (Table 4.5), accounting for ~10% endogenous CO_2 in both cases. The lesser PPP flux in ECOM4LA (albeit relative to glycolysis) suggested that in this strain the 1-carbon of glucose might be converted to CO_2 via additional pathways. As the 1-position of glucose is equivalent to the 6-position after conversion by glycolysis to 3-carbon metabolites, labeling was also performed with 6-^{13}C glucose. This yielded no CO_2 labeling in either WT or ECOM4LA, demonstrating that in both strains oxidative PPP was the only route to convert the 1-carbon of glucose into CO_2.

4.3 DISCUSSION

The aim of this study was to gain insights into physiology of the ECOM4 strain and understand what metabolic and regulatory changes have led to the inability of to shift between aerobic and anaerobic growth. Three active cytochrome oxidases and quinol monooxygenase were completely removed in order to produce a phenotype almost incapable of oxygen utilization. The oxygen uptake rate of the resultant mutant was reduced by nearly 60 times compared to un-mutated *E. coli.* As a consequence of

these deletions, the mutant strain was unable to undergo an aerobic-anaerobic shift and presented fermentative behavior under oxic and anoxic conditions. In order to understand metabolic changes that underlie the unique physiology of the mutant strain, we conducted whole genome transcriptomics analysis coupled with ^{13}C tracing experiments and physiological characterization during aerobic and anaerobic growth.

The transition between oxic and anoxic environments has been studied extensively in *E. coli* [14,18,19]. In particular, the "shift" between aerobic and anaerobic modes of metabolism is regulated by two distinct systems of transcription factors: FNR and ArcB/ArcA [20-25]. It has been reported that FNR is able to sense oxygen directly [26], while the ArcB/ArcA system responds to the content of the quinone pool [11,27,28], to switch on the expression of fermentation genes and represses the aerobic pathways when *E. coli* encounters low oxygen growth conditions [29,30].

Here we hypothesized that oxygen uptake-mediated regulation (ArcB/ArcA) will be significantly perturbed as a result of the inability to utilize oxygen, while oxygen sensing regulation (FNR) should exhibit similar behavior as in the wild type. In order to investigate the behavior of the mutant strain, we employed whole-genome transcriptome analysis, combined with ^{13}C tracing experiments and physiological characterization. Consistent with our hypothesis we observed activation of ArcA regulator under oxic conditions, and consequently activation of fermentative metabolism during aerobic growth, while the regulatory action of the FNR regulator remained similar to wild type. We also show that the resulting mutant strain demonstrated negligible oxygen uptake (oxygen uptake rate is reduced by 98%). Physiological behavior of the mutant strain was similar to the so-called Warburg effect

found in most cancer cells [31,32], where energy is produced predominantly by glycolysis followed by lactic acid fermentation, rather than by oxidation of pyruvate through the TCA cycle, even when oxygen is plentiful [32-34].

4.3.1 Aerobic ECOM4LA shows anaerobic gene expression

As expected, the deletion of the respiratory chain components had a greater effect on metabolism in an oxic environment. Comparable gene expression patterns between the aerobic ECOM4LA and anaerobic WT not only indicate similar regulation but also suggest similar metabolic functions. In particular, high flux (based on uptake and secretion rates) and increased expression of glycolytic enzymes suggest that glycolysis is a main energy producing pathway in ECOM4LA during aerobic and anaerobic growth, similar to that seen in WT anaerobic growth. It is possible that ECOM4LA is unable to build a sufficient proton gradient due to mutations in cytochrome oxidases, thus requiring the production of ATP molecules by substrate-level phosphorylation under oxic growth conditions. The similarity of growth rates between aerobically grown ECOM4LA and anaerobically grown WT *E.coli* (Table 4.2) suggests that energy requirements are similar in both strains under given conditions, unlike that of aerobically and anaerobically grown WT *E. coli*.

4.3.2 Aerobic ECOMLA4 uses anaerobic respiration

It is well known that *E. coli* has a highly versatile respiratory chain that allows it to adapt to conditions that vary with respect to oxygen availability and the redox

state (Figure 4.8A,B) [11]. It is possible that mutations introduced in the ECOM4LA strain, together with adaptive evolution, resulted in the rearrangement of the respiratory chain and a shift in the content of the quinone pool (Figure 4.8C). During anaerobic growth, *E. coli* uses different respiratory pathways as compared to during aerobic growth [1]. In the respiratory chain formed by NADH menaquinone oxidoreductase (*yieF* and *wrbA*) and fumarate reductase (*frdABCD*) electrons are transferred from the NADH to fumarate by a menaquinone pool [35], resulting in the formation of succinate (Figure 4.8C) [36]. Based on our gene expression results and physiological observations we conclude that anaerobic respiration consisting of *yieF/wrbA* and *frdABCD* is active and used is to remove excess electrons during exponential growth of the ECOM4LA strain.

4.3.3 Anaerobic regulator ArcA is active in ECOM4LA during oxic growth

E. coli has two distinct regulators that control expression of the many genes involved in the aerobic-anaerobic shift: the ArcB/ArcA two-component system and FNR. In the expression data, we saw that ArcA activation in ECOM4LA is likely responsible for the anaerobic phenotype under aerobic conditions. However, there was a small number of ArcA targets (19 out of 143; see Appendix Table 4.3) that were further changed in the anaerobic ECOM4LA, consistent with known ArcA activity. Thus, it seems that ArcA still increases its level of activity slightly in the ECOM4LA aerobic-anaerobic shift. To further validate activation of ArcA in ECOM4LA during aerobic growth we looked at genes previously identified as direct targets of ArcA

regulation. We observed significant downregulation of succinate dehydrogenase (*shdABCD*) and fad operon under oxic conditions, which is known to be repressed by the ArcA. These operons were also repressed significantly under anoxic conditions, indicating activity of ArcA during anaerobic growth consistent with WT *E. coli*. Our results suggest that the action mode of one of the global transcription regulators (ArcA) has been altered as a result of major metabolic adjustment, which affected the gene expression in a non-intuitive way. In particular, the inability to utilize oxygen has led to a decrease in ubiquinone content and an increase in menaquinone content (Figure 4.6) leading to the activation of the ArcB/ArcA regulatory system [27]. Thus, the change in a composition of the quinone pool (Figure 4.8C) leads to activation of ArcA and subsequent activation of fermentative metabolism during aerobic and anaerobic growth of the ECOM4LA cell line.

4.3.4 ^{13}C analysis complements gene expression data

Metabolic flux calculations based on ^{13}C glucose labeling data were highly consistent with the gene expression data. Most notably, glycolysis was upregulated in ECOM4LA compared with the PPP, and flux through the TCA cycle was not detectable (Figure 4.7). Flux analysis indicated reduced PPP flux between glucose and pyruvate not just relative to glycolysis, but also with conversion to absolute flux using glucose uptake rates. This contrasted with gene expression data (Figure 4.4B) which showed greater expression of genes for the non-oxidative PPP in ECOM4LA. However, the reversibility of the reactions catalyzed by the enzymes in the non-

oxidative PPP should be noted. Data for input into the P5P pool via the two branches of the PPP showed a greater contribution of the non-oxidative PPP in ECOM4LA (Table 4.5), which might account for the increased expression of the genes for this branch of the pathway here. Producing P5P via a non-reducing route might help ECOM4LA maintain its redox balance in the absence of the ability to utilize oxygen.

4.3.5 Similar physiological behavior under oxic and anoxic conditions

Even though we observed a nearly 15% difference in growth rate of ECOM4LA between oxic and anoxic conditions, the overall physiological behavior was similar (Table 4.2). Lower lactate yield observed during anaerobic growth can be attributed to a higher cell density. ECOM4LA strain grew to a 20% higher cell density anaerobically than aerobically (data not shown). Oxygen uptake rate measured after gene deletions was nearly 60 times lower than in wild type. We were unable to identify the metabolic function accounting for the remaining oxygen uptake; however, since no major physiological differences were observed under oxic and anoxic conditions, we can conclude that oxygen does not have a significant metabolic function in the ECOM4LA strain.

4.4 MATERIALS AND METHODS

4.4.1 Strains and media

The strain described in this study was generated from the cytochrome oxidase mutant strain (ECOM3) presented before [37]. The quinol monooxygenase (*ygiN*) [38] was removed from the unevolved ECOM3; the resulting strain harbored the following mutations: *cydAB, cyoABCD, cbdAB, ygiN*, and was named ECOM4 (*Escherichia Coli* Oxidase Mutant 4). The deletion of the *ygiN* gene was conducted using homologous recombination of a PCR-amplified linear fragment using lambda Red recombinase system [39]. In short, the gene to be deleted was replaced by a kanamycin gene flanked by FRT sites and the insert was removed with a FLP recombinase. In order to verify the genotype of the mutant, single colonies were isolated from the solid media and tested with PCR. Primers used for the deletion and verification are presented in Table 4.6. Wild type (WT) *E. coli* colonies were tested in parallel as a negative control. Bacterial strains were cultured at 37°C in M9 minimal liquid medium containing 4 gram/L glucose, except as noted.

4.4.2 Adaptive evolution

Mutant strains were adaptively evolved using the technique described earlier [5,40]. In short, a colony off a fresh agar plate was inoculated in 250 ml M9 medium containing EZ supplements (Teknova), grown overnight, and passed into a new flask containing fresh medium. The volume of inoculum was adjusted on a daily basis in

order to maintain exponential phase growth. The amount of EZ supplements added to the medium was reduced exponentially during the first two weeks of evolution. Cells were propagated aerobically for 30 days (>500 generations) following the protocol reported by Fong et al. 2005 [5]. Evolving cultures were also supplemented with 50 μg/ml kanamycin once a week and screened daily with PCR to prevent contamination. Samples were frozen every 2 days throughout the evolution.

4.4.3 Phenotype assessment

To assess phenotypic characteristics of evolved and isolated strains, growth rates and byproduct secretion profiles were measured. Each strain was grown in batch culture under oxic, and anoxic conditions. Aerobic cultivation was conducted in 500 ml Erlenmeyer flasks containing 250 ml M9 medium. Temperature was controlled by a circulating water bath, mixing and aeration was controlled with a stir bar at ~1000 rpm. Anaerobic cultivation was conducted in 250 ml Erlenmeyer flasks with 200 ml medium, sealed with rubber stoppers containing necessary inlet tubing. Anoxic conditions were achieved by continuously flushing of cultures with a 95% N_2 / 5% CO_2 gas mixture at a flow rate of 1 ml/min. The temperature was controlled by using a circulating water bath; mixing was controlled with a stir speed of ~200 rpm. Samples were taken from batch cultures periodically (every 30 min), filtered through a 0.2 μm filter and stored at -20°C for by-product analysis. Glucose concentration in the media was assessed using an enzymatic assay kit (R-Biopharm), while *D*-lactate secretion was measured using RI (refractive index) detection by HPLC (Waters) with a Bio-Rad

Aminex HPX87-H ion exclusion column (injection volume, 10 μl) and 5mM H_2SO_4 as the mobile phase (0.5 ml/min, 45°C). The identities of metabolites and organic acids in the fermentation broth were further verified with enzymatic kits (R-Biopharm). The oxygen uptake rate of each aerobic culture was determined by measuring the rate of dissolved oxygen depletion in an enclosed respirometer chamber using a polarographic dissolved oxygen probe (YSI).

4.4.4 Clonal analysis

ECOM4 populations evolved for 30 days were cultured overnight on solid M9 media with 4 gram/l glucose and 50 μg/ml kanamycin. Ten random individual colonies were selected from each plate and grown overnight in M9 liquid medium. Cells were harvested by centrifugation, washed three times with medium without a carbon source and loaded on a Bioscreen C machine (Growth Curves USA). Cultures were inoculated into 300 μl wells containing medium; the initial OD of each well was less than 0.05. Cells were grown for 8 hours with continuous shaking to ensure good mixing and aeration and optical density (OD at 600 nm) measurements were taken every 15 min. Once the cells reached stationary phase, the assay was stopped and *D*-lactate concentration was assessed by HPLC. Strains with the highest production yield were identified and subjected to aerobic batch cultivation. Strains were grown in 500 ml Erlenmeyer flasks with 250 ml medium for 8 hours with continuous agitation as described before. Samples were taken every 30 min, filtered and analyzed using HPLC (Waters). Growth rate (1/hr), oxygen uptake rate (mmol/g-dwt/hr), sugar uptake rate

(mmol/g-dwt/hr), and product secretion rates (mmol/g-dwt/hr) were measured as described before.

4.4.5 Transcriptome analysis

Cultures were grown to mid-log growth phase aerobically, and anaerobically (OD A600 ~ 0.6 for WT and OD A600 ~ 0.25 for ECOM4LA). The cultures (3 mL of WT and 7 ml of ECOM4LA) were then added to 2 volumes of RNAprotect Bacteria Reagent (Qiagen) and total RNA was isolated by using RNeasy columns (Qiagen) with DNaseI treatment. Total RNA yields were measured by using a spectrophotometer (A260) and quality was checked by visualization on agarose gels and by measuring the sample A260/A280 ratio (>1.8). cDNA preparation was performed as described in Cho *et al*. [41]. Affymetrix GeneChip *E. coli* Genome 2.0 arrays were used for genome-scale transcriptional analyses. cDNA synthesis, fragmentation, end-terminus biotin labeling, and array hybridization were performed as recommended by Affymetrix standard protocols. Differentially expressed genes were selected by using fold-change threshold and student t-test with false discovery rate (FDR) correction as implemented in ArrayStar 3 software (DNAStar). Genes with at least two-fold expression level change and FDR-adjusted P-value of less than 0.05 were considered significant and were used for strain analysis. Transcriptome data was mapped to *i*AF1260 metabolic reconstruction of *E. coli* [9], by using the Simpheny software platform (Genomatica).

The probability of regulon and GO term enrichment among differentially expressed genes was computed using the hypergeometric distribution. Regulons were obtained from RegulonDB v6.0 [42] and GO terms from Ecocyc v12.0 [3,4]. Correction for multiple hypotheses was done as reported by Storey, et al. [43] (FDR = 0.01). Consistency of differential expression with ArcA and FNR activity in their respective regulons was determined by comparing differential expression (up or down) with increased ArcA or FNR activity (activator or repressor) as reported by RegulonDB.

4.4.6 Quantitative PCR analysis

RNA purification and cDNA synthesis were conducted following the same protocol as described for the gene expression analysis. The 50 µl qPCR reaction contained 25 µl of SYBR Green Tag master mix (Qiagen), 0.2 µM forward primer, 0.2 µM reverse primer, and cDNA as a template. Each qPCR reaction was run in triplicates in a Bio-Rad thermocycler (Bio-Rad, Hercules) with the following settings: 95°C for 15 min, 94°C for 15 s, 52°C for 30 s, 72°C for 30 s; the denaturation, annealing and extension steps were repeated for 40 cycles. Targeted gene expression of the mutant strain was analyzed under oxic and anoxic growth and compared to WT. Using a standard curve for each primer set, the relative cDNA quantity was obtained for each gene by normalizing it to the quantity of *acpP* (acyl carrier protein) cDNA in the same sample. *acpP* was chosen as the internal control gene since it is constitutively expressed in WT and mutant under both aerobic and anaerobic conditions [18].

4.4.7 ^{13}C tracing studies

Culture labeling: Prior to labeling, single colonies were selected from stock plates and inoculated directly into 250 ml M9 medium in 500 Erlenmeyer flasks aerated by stirring at 1000 rpm. Cells were grown overnight, harvested, washed twice with water and used to inoculate 50 ml flasks containing 25 ml medium with 2 g/L ^{13}C-labeled D-glucose, with initial OD_{600} 0.005-0.01. Glucose was supplied as either 100% 1-^{13}C-labeled, 100% 6-^{13}C-labeled, or a mixture of 20% uniformly (U-^{13}C-) labeled with 80% natural glucose (which is randomly 1% ^{13}C). Cells were grown to mid-log phase, corresponding to OD600 of 0.6 (WT) or 0.25 (mutant). 3 ml (WT) or 10 ml (mutant) of each culture was harvested by centrifugation at 4°C. Media were aspirated and analyzed with HPLC to determine the remaining glucose concentration. Cell pellets were placed at -80°C prior to further analysis.

Derivatization and GC-MS analysis: Cells were resuspended in 0.1 ml 6 M HCl, transferred to glass vials and protein was digested into amino acids under a nitrogen atmosphere for 18 hr at 105°C in an Eldex H/D Work Station. Digested samples were dried to remove residual HCl, resuspended with 75 µl each tetrahydrofuran and *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide (Aldrich), and incubated for 1 hr at 80°C to derivatize amino acids. Samples were filtered through 0.2 µm PVDF filters, and injected into a Shimadzu QP2010 Plus GC-MS (0.5 µl with 1:50 split ratio). GC injection temperature was 250°C and the GC oven temperature was initially 130°C for 4 min, rising to 230°C at 4°C/min and to 280°C at 20°C/min with a final hold at this temperature for 2 min. GC flow rate with helium carrier gas was 50 cm/s. The GC column used was a 15 m x 0.25 mm x 0.25 µm

SHRXI-5ms (Shimadzu). GC-MS interface temperature was 300°C and (electron impact) ion source temperature was 200°C, with 70 eV ionization voltage. The mass spectrometer was set to scan m/z range 50-600.

Processing of GC-MS data: Mass data were retrieved from the GC-MS for fragments of 14 derivatized amino acids: cysteine and tryptophan were degraded during amino acid hydrolysis; asparagine and glutamine were converted respectively to aspartate and glutamate; arginine was not stable to the derivatization procedure. For each fragment, these data comprised mass intensities for the base isotopomer (without any heavy isotopes, M+0), and isotopomers with increasing unit mass (up to M+6) relative to M+0. These mass distributions were normalized by dividing by the sum of M+0 to M+6, and corrected for naturally-occurring heavy isotopes of the elements H, N, O, Si, S, and (in moieties from the derivatizing reagent) C, using matrix-based probabilistic methods as described [44,45] implemented in Microsoft Excel. Data were also corrected for carry-over of unlabeled inoculum [44].

Corrected mass distributions for amino acid fragments from U-^{13}C-glucose-labeled cells were used to infer the trafficking and reassortment through metabolism of linked chains of carbons derived from glucose, while mass distribution data from 1- or 6-^{13}C-glucose-labeled cells were used to track the fate of individual carbon atoms. The analysis is summarized here and is described in more detail in Supplemental Methods. Amino acid labeling data originating from U-^{13}C-glucose was used to estimate two aspects of pentose phosphate pathway (PPP) flux. The mass distribution data for alanine (as a marker for pyruvate) were used to calculate the fraction of alanine originating from the PPP versus glycolysis. Flux from glucose via glucose-6-

phosphate to pentose-5-phosphates (P5P; ribose-5-phosphate, xyulose-5-phosphate and ribulose-5-phosphate – all assumed to be in equilibrium) in the oxidative PPP and back to glycolytic intermediates in non-oxidative PPP ultimately yields 5 pyruvate molecules per 3 input glucose. Of these 5 pyruvate molecules, 3 are composed of 3-carbon units linked as they were in glucose (same as pyruvate produced via glycolysis); 2 are re-assorted such that C1 has a different origin from the rest of the molecule [16]. The fraction of pyruvate split across the C1-C2 bond was calculated from the mass distributions of alanine fragments.

Secondly, histidine labeling from U-^{13}C-glucose was used to calculate relative input to P5P from oxidative or non-oxidative PPP. The carbon backbone of histidine is equivalent to P5P plus one carbon from the tetrahydrofolate-linked one-carbon (1-C) pool. Input to P5P from oxidative PPP removes the C1 carbon from glucose but otherwise the carbon backbone remains intact (giving an M+5 P5P fraction). In contrast, inputs from non-oxidative PPP necessarily yield re-assorted P5P, with the split between different source molecules being largely across the C2-C3 bond (yielding M+2 or M+3 P5P).

Data from 1-^{13}C-glucose labeling experiments were used to provide another measure of flux through PPP versus glycolysis. As noted above, glucose routed through the oxidative branch of the PPP loses carbon from position 1 as CO_2. Therefore, by measuring the degree of loss of ^{13}C-label in alanine (pyruvate) in 1-^{13}C-glucose-labeled cells, relative flux through glycolysis versus PPP was calculated.

Mass data for U-^{13}C-glucose labeling of aspartate, which was assumed to be in equilibrium with oxaloacetate (OAA), were used to assess the relative inputs to OAA

from the TCA cycle versus the anaplerotic reactions phosphoenolpyruvate carboxylase (PEPC) and malic enzyme. In broad terms, input from anaplerosis was apparent as +3 mass units labeling of aspartate, indicative of incorporation of linked [^{13}C] 3-carbon units arising from PEP or pyruvate, while input from the TCA cycle appeared as +2 mass units labeling indicative of input of 2-carbon units originating as acetyl-CoA. Data from various fragments of aspartate were used to calculate the backflux in the TCA cycle from oxaloacetate to symmetrical metabolites (i.e., fumarate), and the ^{13}C labeling of cellular CO_2/bicarbonate. The ^{13}C labeling pattern of anaplerotic input to oxaloacetate was then modeled as the product of CO_2 labeling and ^{13}C labeling of alanine C1-3 (as a surrogate for pyruvate or PEP), while the input to oxaloacetate from α-ketoglutarate in the TCA cycle was assumed to correspond to the labeling of glutamate (C2-C5 fragment). The relative contributions of these inputs to oxaloacetate were then calculated using least-squares fit in MATLAB. These results were checked with alternate amino acid fragments providing the inputs (See Supplemental Methods). 1- or 6-^{13}C-glucose data were used to calculate relative flux from glucose to pyruvate through the Entner-Doudoroff (ED) pathway versus glycolysis or the PPP. The ED, in contrast to the latter pathways, converts 1-^{13}C-glucose to 1-^{13}C-pyruvate, and not 3-^{13}C-pyruvate. Flux through the ED was therefore estimated by comparing labeling of C1-3 and C2-3 fragments of alanine [15].

Positionally-labeled glucose data were also used to determine ^{13}C-labeling of the 1-C pool, utilizing methionine and aspartate labeling data, as methionine is produced from aspartate plus a 1-C unit. Furthermore, the relative contributions of

serine or glycine to the 1-C pool were determined, based on the labeling of the 3-position of serine and 2-position of glycine.

4.4.8 Quinone extraction

The ubiquinone-8 (UQ) and menaquinone (MQ) extraction was conducted according to the protocol outlined previously [11,27,46]. In short, 2 ml of WT culture and 4 ml of ECOM4LA culture were quenched with 6 ml of ice cold methanol. Next, 6 ml of petroleum ether were added rapidly and mixture was vortexed for 1 min. Following centrifugation of the mixture (900xG, for 2 min), the top phase was transferred into a new tube. Another 3 ml of petroleum ether were added and the vortexing and centrifugation steps were repeated. The upper phases were combined and allowed to evaporate to dryness. Dried extracted quinones were resuspended in 100 µl of ethanol and analyzed using HPLC (Waters) system fitted with Pursuit XRs (Varian) C18 reverse phase column with methanol as a mobile phase and flow rate of 1.0 ml/min at ambient temperature. Detection of quinones was conducted using a dual-wavelength UV detector (Waters) with 290 nm for UQ and 248 nm for MQ [27,46]. Ubiquinone-10 and menaquinone-4 were used as standards. The total amount of each species was calculated using the relevant peak area, plotted against the molar absorption coefficient as described by Shestopalov *et al.* [46]. Analytical grade methanol, petroleum ether, and ethanol were acquired from Sigma Aldrich.

4.5 ACKNOWLEDGEMENTS

We thank Jan Schellenberger for useful discussions regarding metabolic fluxes, and Yuri Matusov, Marc Abrams, Dr. Daniel Hyduke, and Karsten Zengler for helpful discussions and critical revision of this manuscript.

The text of this chapter, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Scott D.A., Lewis, N.E., Tarasova Y., Osterman A. L., Palsson, B.Ø. Deletion of cytochrome oxidase and quinol monooxygenase blocks aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010, date pending. The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

4.6 REFERENCES

1. Neidhardt F. *Escherichia coli* and *Salmonella*. In: Bock A. SG, ed. Fermentaion. Washington, DC: ASM Press 1996.

2. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010;38:D355-360.

3. Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I, Collado-Vides J, Gama-Castro S, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Bonavides-Martinez C, Ingraham J. Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res* 2007;35:7577-7590.

4. Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP, Johnson DA, Krummenacker M, Nolan LM, Paley S, Paulsen IT, Peralta-Gil M, Santos-Zavaleta A, Shearer AG, Karp PD. EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res* 2009;37:D464-470.

5. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-648.

6. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007;3:149.

7. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol* 2005;71:7880-7887.

8. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain. *Appl Environ Microbiol* 2008;74:7561-9.

9. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for

Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121.

10. Jervis AJ, Crack JC, White G, Artymiuk PJ, Cheesman MR, Thomson AJ, Le Brun NE, Green J. The O2 sensitivity of the transcription factor FNR is controlled by Ser24 modulating the kinetics of [4Fe-4S] to [2Fe-2S] conversion. *Proc Natl Acad Sci U S A* 2009;106:4659-4664.

11. Bekker M, Kramer G, Hartog AF, Wagner MJ, de Koster CG, Hellingwerf KJ, de Mattos MJ. Changes in the redox state and composition of the quinone pool of *Escherichia coli* during aerobic batch-culture growth. *Microbiology* 2007;153:1974-1980.

12. Brondsted L, Atlung T. Anaerobic regulation of the hydrogenase 1 (hya) operon of *Escherichia coli*. *J Bacteriol* 1994;176:5423-5428.

13. Lynch AS, Lin EC. Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters. *J Bacteriol* 1996;178:6238-6249.

14. Gunsalus RP, Park SJ. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. *Res Microbiol* 1994;145:437-450.

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

16. Szyperski T. Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. *Eur J Biochem* 1995;232:433-448.

17. Murray EL, Conway T. Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of *Escherichia coli*. *J Bacteriol* 2005;187:991-1000.

18. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004;429:92-96.

19. Spiro S, Guest JR. Adaptive responses to oxygen limitation in *Escherichia coli*. *Trends Biochem Sci* 1991;16:310-314.

20. Salmon K, Hung SP, Mekjian K, Baldi P, Hatfield GW, Gunsalus RP. Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J Biol Chem* 2003;278:29837-29855.

21. Scott C, Partridge JD, Stephenson JR, Green J. DNA target sequence and FNR-dependent gene expression. *FEBS Lett* 2003;541:97-101.

22. Spiro S, Guest JR. FNR and its role in oxygen-regulated gene expression in *Escherichia coli*. *FEMS Microbiol Rev* 1990;6:399-428.

23. Kang Y, Weber KD, Qiu Y, Kiley PJ, Blattner FR. Genome-wide expression analysis indicates that FNR of *Escherichia coli* K-12 regulates a large number of genes of unknown function. *J Bacteriol* 2005;187:1135-1160.

24. Alexeeva S, Hellingwerf KJ, Teixeira de Mattos MJ. Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *J Bacteriol* 2003;185:204-209.

25. Compan I, Touati D. Anaerobic activation of arcA transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol Microbiol* 1994;11:955-964.

26. Sutton VR, Mettert EL, Beinert H, Kiley PJ. Kinetic analysis of the oxidative conversion of the [4Fe-4S]2+ cluster of FNR to a [2Fe-2S]2+ Cluster. *J Bacteriol* 2004;186:8018-8025.

27. Bekker M, Alexeeva S, Laan W, Sawers G, Teixeira de Mattos J, Hellingwerf K. The ArcBA two-component system of *Escherichia coli* is regulated by the redox state of both the ubiquinone and the menaquinone pool. *J Bacteriol* 2010;192:746-754.

28. Georgellis D, Kwon O, Lin EC. Quinones as the redox signal for the arc two-component system of bacteria. *Science* 2001;292:2314-2316.

29. Iuchi S, Lin EC. arcA (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc Natl Acad Sci U S A* 1988;85:1888-1892.

30. Iuchi S, Lin EC. Adaptation of *Escherichia coli* to redox environments by gene expression. *Mol Microbiol* 1993;9:9-15.

31. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009;324:1029-1033.

32. Warburg O. On the origin of cancer cells. *Science* 1956;123:309-314.

33. Bertram JS. The molecular biology of cancer. *Mol Aspects Med* 2000;21:167-223.

34. Kim JW, Dang CV. Cancer's molecular sweet tooth and the Warburg effect. *Cancer Res* 2006;66:8927-8930.

35. Yagi T, Matsuno-Yagi A. The proton-translocating NADH-quinone oxidoreductase in the respiratory chain: the secret unlocked. *Biochemistry* 2003;42:2266-2274.

36. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC. Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* 1999;284:1961-1966.

37. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. *Appl Environ Microbiol* 2008;74:7561-7569.

38. Adams MA, Jia Z. Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. *J Biol Chem* 2005;280:8358-8363.

39. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000;97:6640-6645.

40. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 2006;38:1406-1412.

41. Cho BK, Knight EM, Palsson BO. PCR-based tandem epitope tagging system for *Escherichia coli* genome engineering. *Biotechniques* 2006;40:67-72.

42. Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muniz-Rascado L, Martinez-Flores I, Salgado H, Bonavides-Martinez C, Abreu-Goodger C, Rodriguez-Penagos C, Miranda-Rios J, Morett E, Merino E, Huerta AM, Trevino-Quintanilla L, Collado-Vides J. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 2008;36:D120-124.

43. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440-9445.

44. Nanchen A, Fuhrer T, Sauer U. Determination of metabolic flux ratios from 13C-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol Biol* 2007;358:177-197.

45. van Winden WA, Wittmann C, Heinzle E, Heijnen JJ. Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol Bioeng* 2002;80:477-479.

46. Shestopalov AI, Bogachev AV, Murtazina RA, Viryasov MB, Skulachev VP. Aeration-dependent changes in composition of the quinone pool in *Escherichia coli*. Evidence of post-transcriptional regulation of the quinone biosynthesis. *FEBS Lett* 1997;404:272-284.

	ECOM4 + O_2	ECOM4 - O_2	ECOM41	ECOM42	ECOM43	ECOM41	ECOM42	ECOM43	ECOM41	ECOM42	ECOM43	ECOM41	ECOM42	ECOM43
	day 0	day 0	day 1			day 10			day 20			day 30		
GR	0.38+/-0.02	0.44+/-0.01	0.36 +/- 0.01	0.36 +/- 0.02	0.35 +/- 0.08	0.06 +/- 0.01	0.07 +/- 0.01	0.08 +/- 0.02	0.25 +/- 0.05	0.21 +/- 0.03	0.26 +/- 0.01	0.33 +/- 0.05	0.32 +/- 0.03	0.38 +/- 0.01
SUR	19.79+/-0.6	16.49+/-1.53	18.73 +/- 0.51	18.5 +/- 0.35	19.5 +/- 0.24	11.42 +/- 0.04	9.61 +/- 0.12	10.35 +/- 0.09	24.7 +/- 0.15	18.34 +/- 0.21	14.61 +/- 2.64	26.51 +/- 0.50	20.71 +/- 0.21	17.26 +/- 0.09
LactSR	37.61+/-0.9	1.75+/-0.61	36.28 +/- 0.90	35.52 +/- 1.24	37.35 +/- 0.98	22.21 +/- 0.46	18.11 +/- 0.58	20.74 +/- 0.32	41.71 +/- .10	25.71 +/- 1.24	7.47 +/- 3.61	43.78 +/- 0.12	26.58 +/- .47	9.58 +/- 2.40
Lact/Gluc	0.80	0.06	0.96	0.95	0.98	0.92	0.93	0.90	0.89	0.73	0.31	0.85	0.65	0.48
AcSR	0.00	21.12+/-0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.82	0.00	0.00	12.54
OUR	0.31+/-0.15	NA	0.27+/-0.1	0.35 +/-0.14	0.36 +/-0.09	0.15+/-0.02	0.17+/-0.08	0.12+/-0.05	0.24+/-0.1	0.22+/-0.08	0.28+/-0.12	0.25+/-0.12	0.30+/-0.1	0.38+/-0.15
EZ	+	+	+	+	+	+/-	+/-	+/-	-	-	-	-	-	-

Table 4.1: Physiological characteristics during 30 day adaptive evolution of ECOM4 strains. GR – growth rate; SUR – substrate uptake rate; LactSR – lactate secretion rate; AcSR – acetate secretion rate; OUR – oxygen uptake rate; EZ- addition of EZ supplements. Values are given in mmol/gDW/h where gDW – gram dry weigh.

Table 4.2: Phenotypic characteristics of ECOM4LA strain under oxic and anoxic conditions.

	ECOM4LA $+O_2$	ECOM4LA $-O_2$	MG1655 $+O_2$	MG1655 $-O_2$
Growth Rate (1/hr)	0.32 +/- 0.005	0.27 +/- 0.006	0.71 +/- 0.01	0.45 +/- 0.02
Glucose Uptake Rate (mmol/g-dwt/hr)	26.4 +/- 0.09	24.9 +/- 0.72	9.02 +/- 0.23	17.3 +/- 0.17
Lactate Secretion Rate (mmol/g-dwt/hr)	48.6 +/- 0.76	41.58 +/- 1.58	0	0.95 +/- .008
Acetate Secretion Rate (mmol/g-dwt/hr)	0	0	3.37 +/- 0.9	10.3 +/- 0.60
Oxygen Uptake Rate (mmol/g-dwt/hr)	0.21 +/- 0.16	0	16.49 +/- 0.67	0
Lactate/Glucose (gram/gram)	0.98 +/- 0.07	0.92 +/- 0.04	NA	NA

Table 4.3: Comparison between gene expression levels in ECOM4LA and MG1655 cells grown aerobicaly and anaerobically. † Benjamini-Hochberg false discovery rate-adjuster P value. Fold changes <2 fold were considered as no change (NC); * average expression value is presented for large operons. E4 – ECOM4LA strain, WT – wild type MG1655.

Gene Name	Locus Number	Product	Fold Change (adj. P-value)†			
			E4/WT +O2	E4/WT -O2	WT -O2/+O2	E4 +O2/-O2
pgi	b4025	Phosphoglucose Isomerase	3.87 (0.017)	NC	NC	NC
*pfkAB**	b3916	6-Phosphofructo Kinase	2.04 (0.003)	NC	2.29 (0)	NC
fbaA	b2925	Fructose Bisphosphate Aldolase	3.00 (0.01)	NC	NC	NC
fbaB	b2097	Fructose Bisphosphate Aldolase	8.77 (0)	3.30 (0)	3.23 (0)	NC
tpiA	b3919	Triose Phosphate Isomerase	3.33 (0.005)	NC	NC	NC
gapA	b1779	Glyceraldehyde 3-Phosphate	4.53 (0.02)	NC	2.53 (0.05)	NC
pgk	b2926	Phosphoglycerate Kinase	2.10 (0.02)	NC	NC	NC
eno	b2779	Enolase	3.31 (0.008)	NC	NC	NC
ldhA	b1380	D-Lactate Dehydrogenase	3.36 (0.005)	NC	NC	NC
yieF	b3713	NADH: Menaquinone Oxydoreductase	2.47 (0.014)	2.09 (0.001)	NC	1.60 (0.023)
wrbA	b1004	NADH: Menaquinone Oxydoreductase	9.10 (0.002)	3.22 (0.001)	5.07 (0.003)	NC
*frdABCD**	b4151 - b4154	Fumarate Reductase	3.60 (0.02)	NC	4.03 (0.002)	1.90 (0.02)
*sdhABCD**	b0721 - b0724	Succinase Dehydrogenase	-25.35 (0)	NC	-33.50 (0)	NC
*nuoA - N**	b2288 - b2276	NADH: Ubiquinone Oxidoreductase	-2.10 (0.005)	NC	-2.30 (0.001)	NC

Table 4.4: Whole-genome expression and qPCR comparison for selected genes in the ECOM4LA strain

Gene Name	Gene Abreviation	b-number	Affimetrix Fold	qPCR Fold
NADH menaquinone oxidoreductase	*yieF*	b3717	2.47 up	9.5 up
	wrbA	b1004	9.1 up	47.5 up
Fumarate reductase	*frdABCD*	b4150-b4154	3.6 up	110.0 up
Succinate dehydrogenase	*shdABCD*	b0721-b0724	25.35 down	27.8 down

Table 4.5: Physiological parameters inferred from ^{13}C labeling data. Data are mean +/- standard deviation from determinations on amino acids prepared from 3 separate cultures of WT and ECOM4LA. [1]Anaplerosis via phospho-enol pyruvate carboxylase or malic enzyme.[2]This is a measure of re-orientation of oxaloacetate after cycling through symmetrical intermediates (succinate and fumarate) and does not include oxaloacetate derived from TCA cycle as this portion of the oxaloacetate pool is by default randomly oriented as it is derived from succinate and fumarate. [3]For U-^{13}C-glucose, maximum possible is 21% (from 20% U-^{13}C-glucose and 1% natural label). For 1- or 6-^{13}C, potential maximum 100% from 100% labeled glucose as sole carbon source. [4]Relative to glycolysis. [5]Measured by loss of label into alanine (pyruvate) relative to glycolysis. [6]Flux through Entner-Doudoroff pathway relative to other routes from glucose to pyruvate/ alanine.

	glucose substrates labeled amino acids		WT	ECOM4LA
Input to oxaloacetate: anaplerosis versus TCA cycle[1]	U-^{13}C-glucose ASP, ALA, GLU	Anaplerosis	62.4 +/- 0.5%	99.5 +/- 1.4%
		TCA cycle	37.6 +/- 0.5%	0.5 +/- 1.4%
Recycling of oxaloacetate to/from fumarate/ succinate[2]	U-^{13}C-glucose ASP		86.7 +/- 4.9%	39.2 +/- 0.6%
CO_2 labeling[3]	U-, 1- or 6-^{13}C-glucose ASP	U-^{13}C-glucose	19.1 +/- 0.2%	14.5 +/- 1.3%
		1-^{13}C-glucose	11.2 +/- 1.8%	9.3 +/- 0.4%
		6-^{13}C-glucose	1.7 +/- 2.4%	-0.2 +/- 0.5%
PPP to pyruvate/ alanine[4]	U-^{13}C-glucose ALA		13.2 +/- 1.7%	-2.8 +/- 1.3%
Oxidative PPP[5]	1-^{13}C-glucose ALA		15.1 +/- 1.2%	1.7 +/- 0.5%
Oxidative/ Non-oxidative PPP into ribose	U-^{13}C-glucose HIS	Oxidative PPP	19.2 +/- 1.6%	13.2 +/- 0.6%
		Non-oxidative PPP	80.8 +/- 1.6%	86.8 +/- 0.6%
1-C pool labeling	1- or 6-^{13}C-glucose MET, ASP	1-^{13}C-glucose	37.3 +/- 0.5%	46.8 +/- 0.4%
		6-^{13}C-glucose	46.8 +/- 0.2%	48.2 +/- 0.1%
Origins of 1-C pool	1- or 6-^{13}C-glucose SER, GLY, MET, ASP	From serine (1-^{13}C glucose)	93.6 +/- 2.4%	97.5 +/- 1.9%
		From serine (6-^{13}C glucose)	91.9 +/- 0.1%	98.8 +/- 0.4%
		From glycine (1-^{13}C glucose)	6.4 +/- 2.4%	2.5 +/- 1.9%
		From glycine (6-^{13}C glucose)	8.1 +/- 0.1%	1.2 +/- 0.4%
Entner-Doudoroff[6]	1- or 6-^{13}C-glucose ALA	1-^{13}C-glucose	1.4 +/- 1.0%	0.1 +/- 0.7%
		6-^{13}C-glucose	1.0 +/- 0.2%	0.3 +/- 1.0%

Table 4.6: DNA primers used for the deletion of *ygiN*

Deletion Primers	Forward Primer (5'-->3')	Reverse Primer (5' --> 3')
ygiN	ATGCTTACCGTAATCGC AGAAATCCGTACTCG TCCTGGTCGTGTAGGCTGGAG CTGCTTC	TTAAATCCCTGGCTGCA GAATACGGATA TTCATCTCCAGCATTCC GGGGATCCGTCGACC
Conformation primers	Forward Primer (5'-->3')	Reverse Primer (5' --> 3')
ygiN	CCGACATTTATCGCTAATGA	GTTGCAAGAGAAAGGCGACA

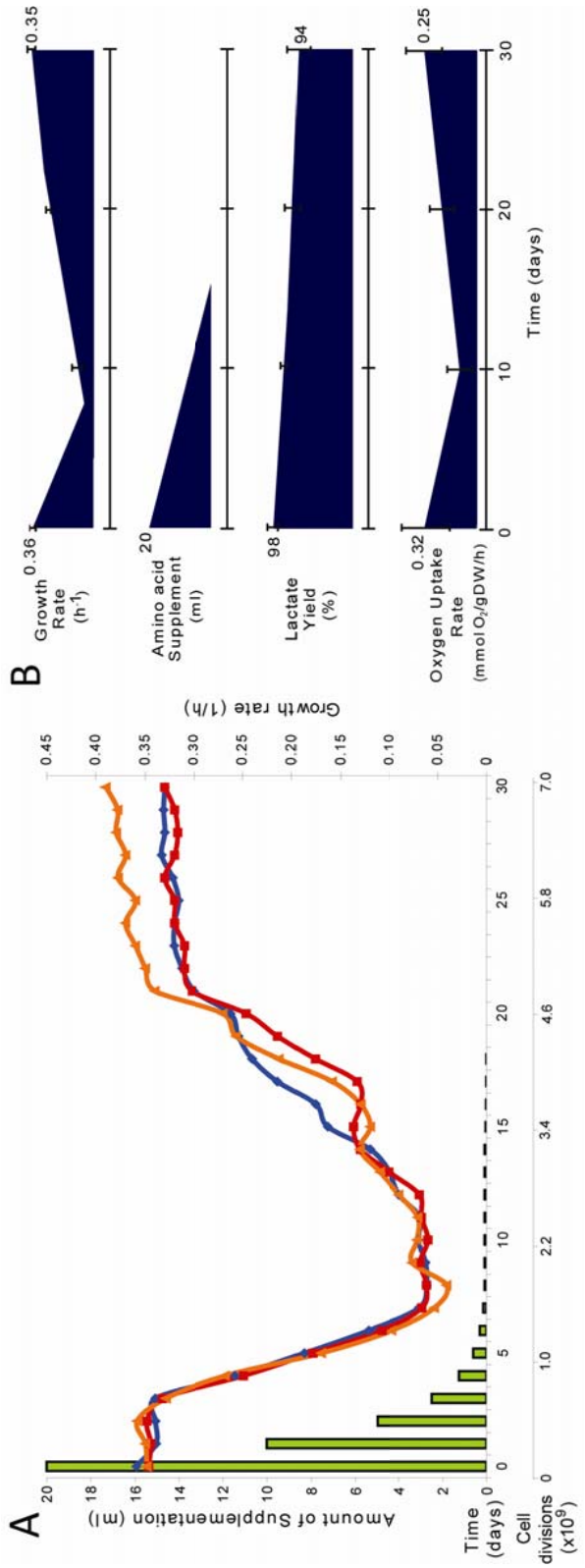


Figure 4.1: Phenotypic characteristics of ECOM4 populations during evolution. Growth rate measurements (A) and phenotypic characteristics (B) for three evolved ECOM4 populations are shown as a function of time of evolution. The EZ amino acid supplement amount (in ml) is shown with green bars. The total number of cell divisions for the entire period of adaptation is presented on a secondary abscissa.

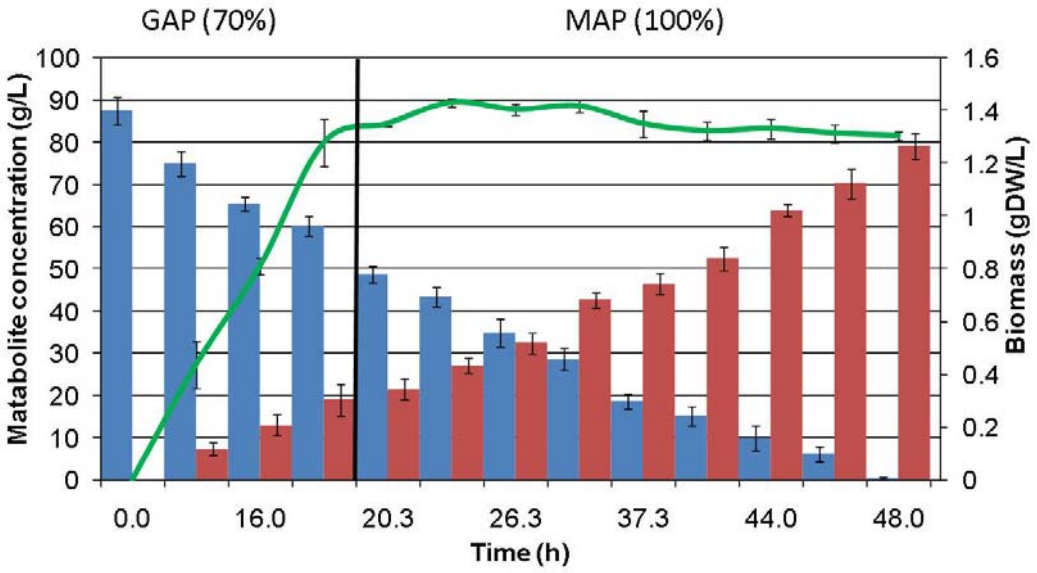


Figure 4.2: Batch fermentation of ECOM4LA strain**.** During batch fermentation the ECOM4LA strain produced D-lactic acid at 70% efficiency during the growth phase (GAP – Growth Associated Production) and at 100% efficiency during the stationary phase (MAP – Maintenance Associated Production).

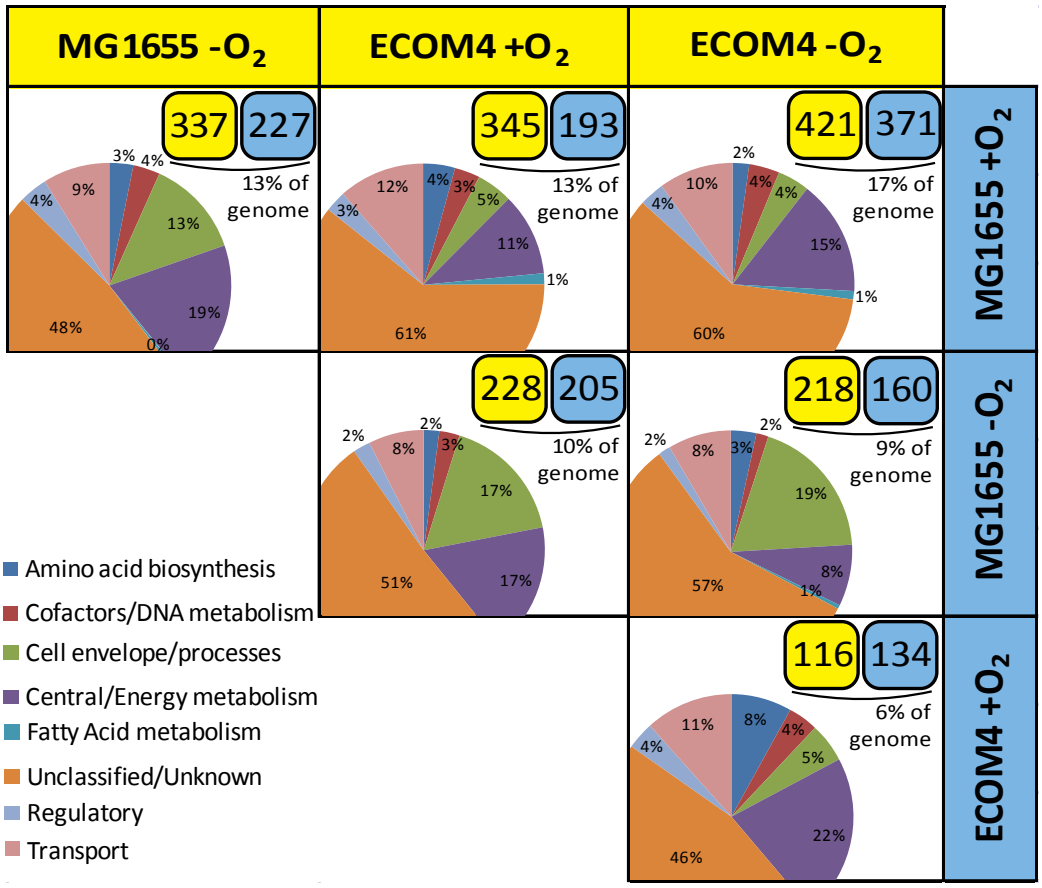


Figure 4.3: Gene expression analysis. Yellow number – genes that showed over 2-fold upregulation. Blue number – genes that showed over 2-fold downregulation. Metabolic function was assigned based on GO annotation

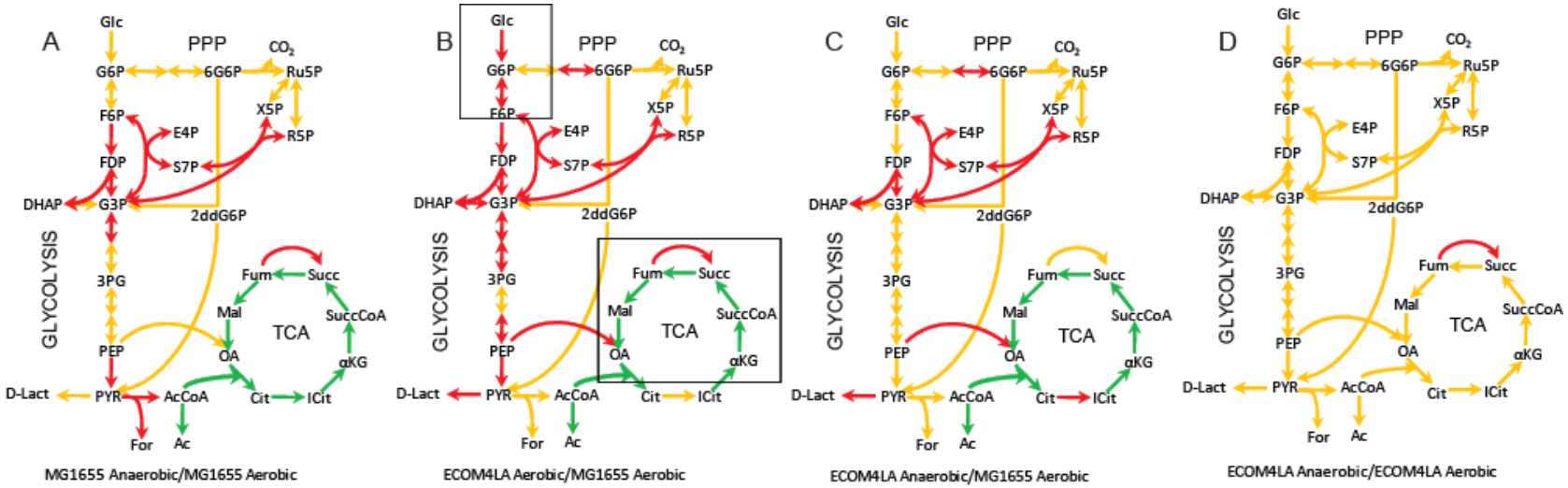


Figure 4.4: Transcriptomics analysis of the ECOM4LA and MG1655 strains under oxic and anoxic conditions. Gene expression values of 2 fold and higher ($p < 0.05$) were mapped onto the metabolic map of central metabolism. Red – at least 2-fold upregulation; Green – at least 2-fold downregulation; Yellow – no change. Panels: A - gene expression in anaerobic WT is compared to aerobic WT; B – Aerobic ECOM4LA is compared to aerobic WT; C – anaerobic ECOM4LA is compared to aerobic WT; D – anaerobic ECOM4LA is compared to aerobic ECOM4LA. Boxes in panel B enclose branching areas of metabolism for which relative metabolic flux are illustrated in Figure.4

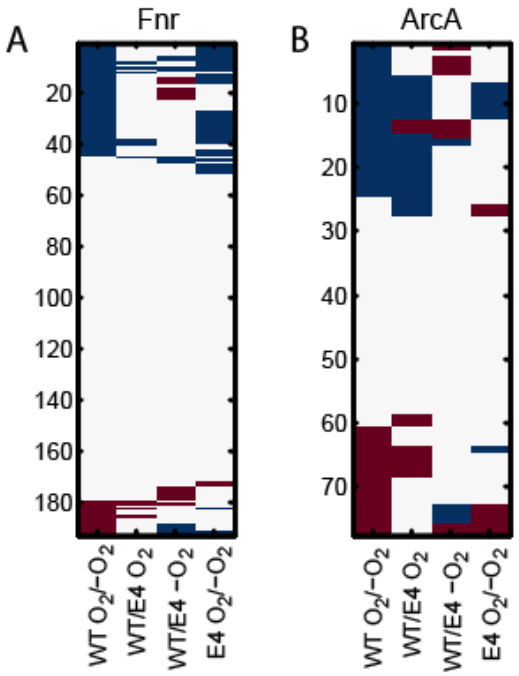


Figure 4.5: Global gene expression suggests that ArcA is active in ECOM4LA in aerobic conditions. Gene expression changes for all genes that are known to be regulated by the anerobic regulators (a) FNR or (b) ArcA according to RegulonDB were compared to the reported function of the regulators (activator or repressor). If a gene was differentially expressed in the direction of known regulatory activity, then the expression change is said to be consistent (blue). If the gene expression change is in the opposite direction of regulator activity, it is shown in red. Since RegulonDB regulatory logic is partially inferred from microarray data for these transcription factors, it is expected that some normal ArcA and FNR activity will be in the direction opposite from RegulonDB assignments. A: The FNR regulon showed greater differential expression and greater consistency (blue) for the WT aerobic-anaerobic shift, and the ECOM4LA aerobic-anaerobic shift. B: The ArcA regulon shows greater differential expression and greater consistency for the WT aerobic/anaerobic shift and in the comparison between aerobic WT and ECOM4LA; thereby suggesting that ArcA is active in ECOM4LA, even under aerobic growth.

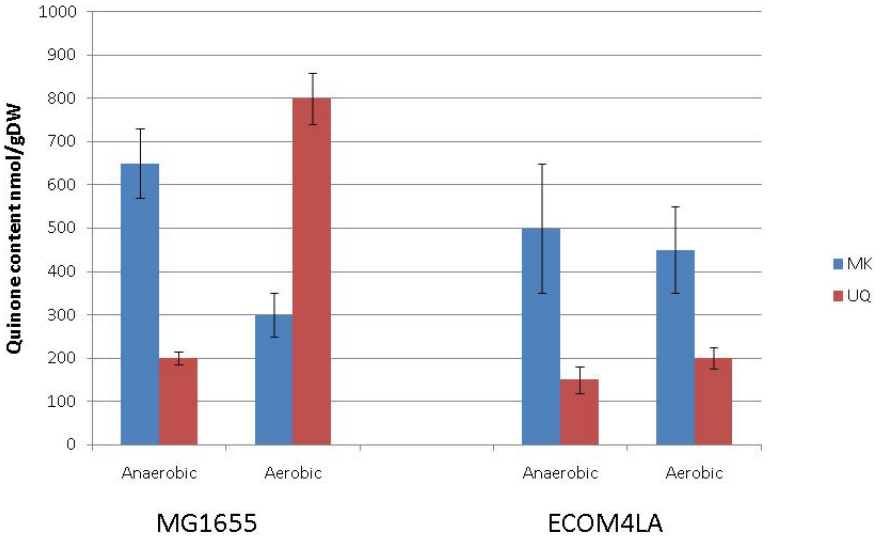


Figure 4.6: Quinone pool content (nmol/gDW) under various environmental conditions. Each measurement is a result of three samples

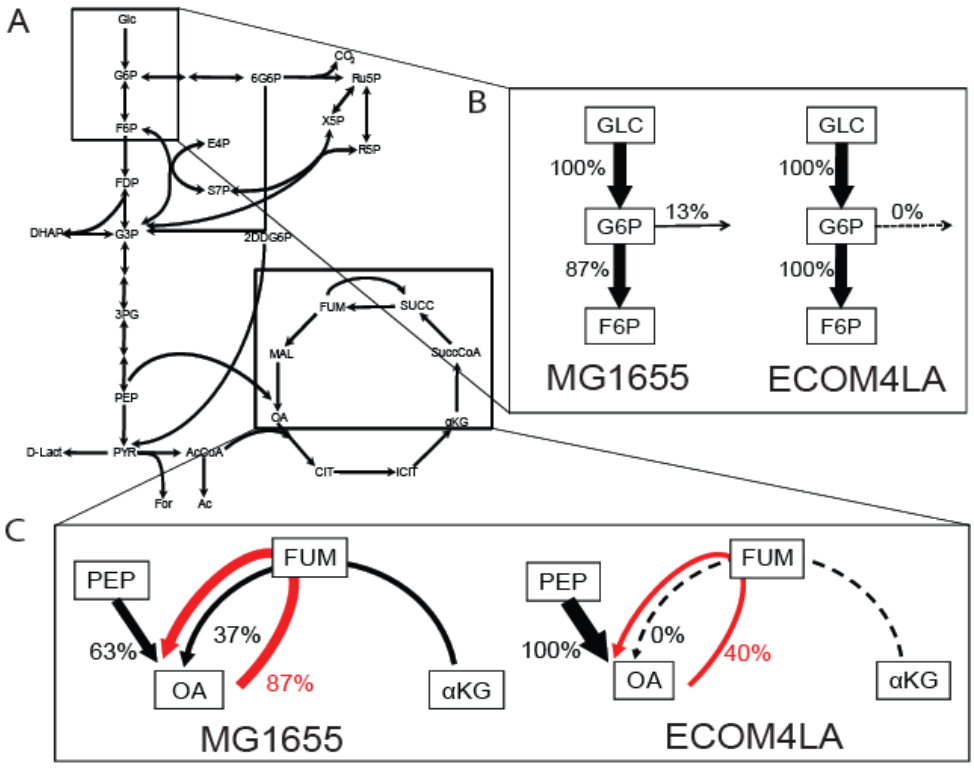


Figure 4.7: Metabolic flux distribution through branching areas of the central metabolism of MG1655 and ECOM4LA cell lines. A – Overall metabolic map of central metabolism with branching areas of metabolism boxed. Corresponding gene expression is presented in Fig. 2. B – Relative input to glycolysis (F6P) from glucose-6-phosphate (G6P) vs. pentose phosphate pathway in each strain; C - Relative input to oxaloacetate (OA) from α-ketoglutarate (αKG) in the TCA cycle or from PEP via PEP carboxylase. Also shown in red is the fraction of OA formed from PEP that recycled through fumarate (FUM) in each strain.

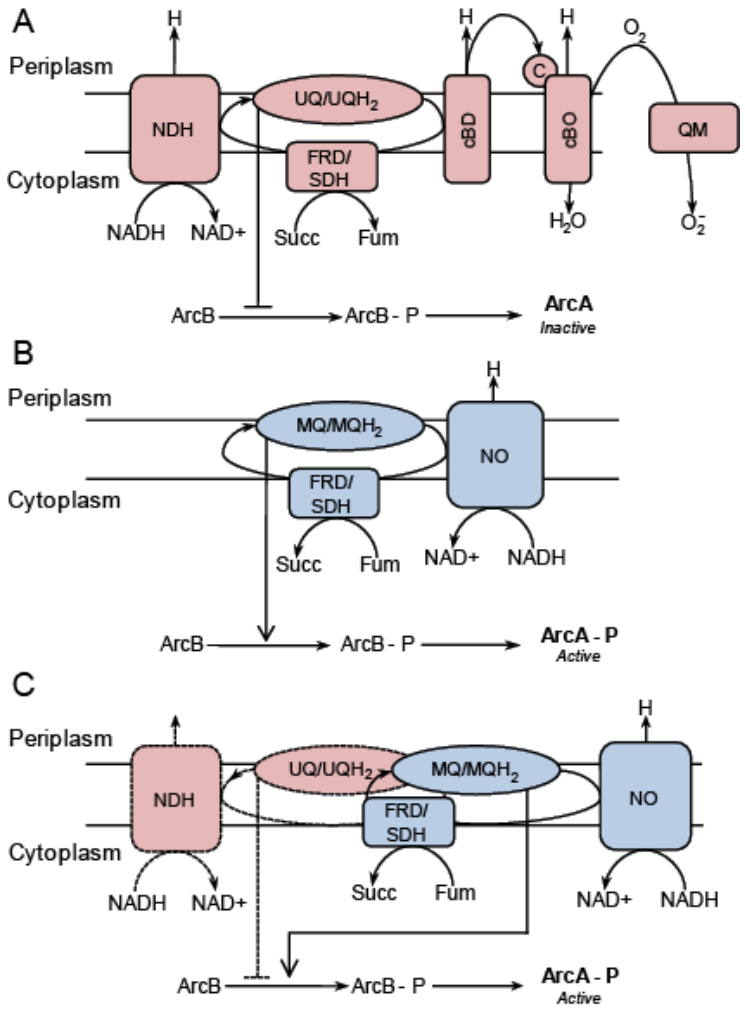


Figure 4.8: Respiratory chain rearrangements in ECOM4LA strain compared to MG1655 and the effect of the content of the quinone pool on ArcA activity. A – A classical view on aerobic respiratory chain (red) in wild type *E. coli* [2-4]. Electrons are transferred from NADH to oxygen via ubiquinone pool. An alternative oxygen utilization system through quinone monooxygenase (YgiN) is presented. High content of ubiquinones (UQs) represses the activity of ArcA regulator. B- A classical view of anaerobic respiratory chain (blue) in wild type *E. coli*. Electrons are transferred from NADH to succinate via menaquinone pool. High content of menaquinones (MQs) activates AcrA activity. C- A rearrangement in respiratory chain resulted from gene deletions. Lack of significant activity of aerobic respiration (dashed line) and increase in activity of anaerobic respiration (solid line) resulted in re-formulation of the quinone pool. Shift from UQs to MQs resulted in activation of ArcA regulator under oxic and anoxic conditions.

CHAPTER V:

PRODUCTION OF D-LACTATE AT pH 7.0 AND pH 4.5

The triple cytochrome oxidase (*cydAB, cyoABCD, cbdAB*) and quinol monooxygenase (*ygiN*) mutant *E. coli* strain (ECOM4) exhibited similar physiological behavior under oxic and anoxic conditions [1,2]. One of the key features of the ECOM4 phenotype was the secretion of D-lactic acid at nearly maximum theoretical yield, close to two molecules of lactate per one molecule of glucose. This unique feature was growth-coupled and triggered by the need to oxidize the global electron carrier NADH (Chapter 4). In this chapter we will present an application of the ECOM4 strain for the industrial level D-lactate overproduction at neutral and acidic pH.

5.1 INTRODUCTION

Lactic acid, also known as a 2-hydroxypropinoic acid, is usually produced by fermentation performed by Lactobacillus bacteria among others [3-5]. Lactic acid is a chiral compound and has two stereo isomers: D- (dextrorotatory) and L- (levorotatory) lactic acid. One of the major uses of lactic acid is in the food industry; however, since two molecules of lactic acid can be dehydrated into lactide and then polymerized to

make biodegraded polyesters there is a increased interest in using lactic acid for production of "green" plastics[6]. More and more frequently lactic acid produced using microbial fermentation is being used for the production of polylactic acid [7-10] – a plastic that can be used to substitute some of the conventional plastics produced from the petroleum oil.

Low cost, chiral and chemical purity are the key requirements of the successful commercialization of the bacterially produced lactic acid. The biggest challenge of the bacterial lactate production lies in the post-processing step, the separation of the final product from the by-products of fermentation[9]. Therefore, the utilization of the homo-fermenting strains has an advantage and leads to a significant cost reduction during the product separation. Another approach to reduce the cost of biocatalysis of lactic acid is lowering the pH of the fermentation environment. One of the key features of low pH fermentation is the production of lactate at the pH below the pKa value of 3.68 [11], which significantly simplifies the downstream separation and increasing the purity of the final product. However, engineering *E. coli* strain to sustain an active growth and lactate production at low pH conditions is not a trivial undertaking.

The laboratory scale adaptive evolution is increasingly being used to study the dynamics of bacterial adaptation, its underlying genetic basis and the altered biochemical mechanisms[12-14]. It has been shown that laboratory adaptive evolution can be used to gradually modify the cultivation environment, which results in altering the physiological behavior of the cells resulting in the improved and more robust phenotype[1,15]. It has been shown that during the adaptation, specific mutations occur

and result in phenotypic attributes that are responsible for improved fitness in a selective environment[16].

A number of studies have been aimed at engineering *E. coli* to produce lactic acid on the industrial level [17-23]. However, all of them are operating at pH in the range between 7.0 and 6.5. The ECOM4 strain is comparable to already published lactate producing *E. coli* [1,17,18,21,22,24] with respect to the lactic acid production.

Phenotypic analysis showed that the ECOM4 strain produced D-lactic acid under oxic and anoxic growth conditions at nearly theoretical maximum yield of two molecules of D-lactate per molecule of glucose[1,2] at neutral pH. In order to engineer the ECOM4 strain to produce lactic acid at low pH we used adaptive evolution. Cells were cultivated on minimal media with glucose as a carbon source and yeast extract supplementation for 100 days during which he pH was gradually adjusted from 7.0 to 4.5 and growth rate improvements were observed. The resulted strain was physiologically characterized and presented below.

5.2 RESULTS

5.2.1 Production of high purity D-lactic acid under oxic and anoxic conditions at neutral pH

The ECOM4 strain was used to produce high purity *D*-lactic acid under oxic and anoxic conditions. The conversion of glucose to *D*-lactate was only slightly affected by oxygen supply, with yields of 98% and 92% for aerobic and anaerobic growth conditions (chapter 4). Batch fermentation with glucose concentration of 88

g/L was performed for 48 hours and resulted in production of 79 g/L of D-lactic acid (over 90 % yield). The specific production rate of D-lactate was on the order of 40-45 mmol/gDW/h under oxic growth conditions.

5.2.2 Growth assessment of ECOM4LA strain at different pH before and after the adaptation

Doubling time of the pre-evolved ECOM4 strain was assessed at different pH between neutral (7.0) and acidic (4.5). The doubling time varied between 2.5 hours and 6.3 hours respectively for pH 7.0 and 4.5 (Figure 5.2). The significant increase in doubling time was observed at pH 5.5, which correlated well with decrease in the lactate secretion rate.

The ECOM4 strain was adaptively evolved for 100 days in order to achieve active growth at pH 4.5. The evolutionary trajectory of the ECOM4 strain is presented in Figure 5.3. We observed significant increase in doubling time upon increase of acidity of the media. After 56 days of adaptation the acidity was increased to pH 4.5 and the cell doubling time was close to 7 hours, similarly to the doubling time observed before adaptation at pH 4.5 (Figure 5.2). The N-Methyl-N'-nitro-N-nitrosoguanidine (NTG) mutagen was added to the culture at low level (see MATERIALS AND METHODS) and the ECOM4 was propagated at pH 4.5 with mutagen in the media for additional 7 days until the increase in growth rate was observed. Adaptation was continued for additional 20 days in order to acquire a stable growth at pH 4.5 (Figure 5.3). The final number of cell divisions acquired before

achieving the stable growth at desired conditions was close to 2 x 10^{11}, similarly to the number of cell doublings reported in similar studies [1,25]. Adaptation was stopped at pH 4.5 and physiological characterization of the adaptively evolved ECOM4 strain was performed at different pH ranging from 7.0 to 4.5.

The growth rate of the adapted strain at pH 4.5 was nearly twice that of the pre-adapted ECOM4 with doubling time of 3.8 hours (Figure 5.2) on minimal media with 4 g/L glucose and 2 g/L yeast extract as a supplementation. This improvement in doubling time accounts for nearly 60 % increase in growth rate after adaptation.

5.2.3 Assessment of D-lactic acid production before and after adaptation to low pH

Similarly as for the growth rate the lactic acid production was assessed in the ECOM4 strain before and after the adaptation. The lactate secretion rate varied between 48.6 +/- 0.76 and 5.21 +/- 0.48 mmol/gDW/h respectively for pH 7.0 and pH4.5 in the pre-evolved ECOM4 strain (Figure 5.2). The evolved strain was able to produce D-lactate at pH 4.5 with specific production rate of 20 mmol/gDW/h and yield of nearly 80%.

5.3 DISCUSSION

We present here the adaptively evolved metabolic engineering mutant strain that is capable of overproduction of the D-lactate under neutral and low pH conditions under oxic conditions. The ECOM4 strain has a natural ability to overproduce D-

lactate at nearly maximum theoretical efficiency. We adaptively evolved this strain to grow and produce D-lactic acid at low pH, in ordered to reduce the cost associated with the separation of lactate from the fermentation broth.

5.3.1 Succinate secretion during mid-log growth phase

We have observed that the ECOM4 strain produced traces amounts of succinate during the exponential growth stage. Lactate was produced with 70% yield during the exponential phase and with nearly 100% yield during the stationary phase (Figure 5.1); illustrating that 30% of carbon was directed towards the biomass formation. Succinic acid was present in a low amount during the exponential growth phase and was metabolized in stationary phase. The highest measured concentration of succinate during the exponential phase was on the order of 30.0 +/- 15.0 mg/L; however, this measurement was highly variable due to re-uptake of succinate. The succinate production during the active growth is triggered by the removal of the excess of electrons. It has been shown that the ECOM4 strain relies on anaerobic respiration as means of electron removal under oxic conditions. In the respiratory chain formed by NADH menaquinone oxidoreductase (*yieF* and *wrbA*) and fumarate reductase (*frdABCD*) electrons are transferred from the NADH to fumarate by a menaquinone pool [26], resulting in the formation of succinate[27].

5.3.2 Volumetric productivities associated with D-lactate production under neutral and low pH

During the batch fermentation of nearly 88 g/L glucose the ECOM4 produced 79 g/L D-lactate in 48 hours at neutral pH. The resulted volumetric productivity associated with this bio-conversion is 1.6 g/L/h. The evolved ECOM4 strain produced close to 70 gram of D-lactate from 90 gram of glucose in 90 hours at pH 4.5, resulting in volumetric productivity of 0.7 g/L/h (Table 5.1). The decrease in growth rate at low pH results in significant reduction of volumetric productivity.

5.3.3 Mechanism of adaptation and overall observations

The mechanism of this adaptation is unclear as well as the mechanism of survival of E. coli in highly acidic environment. However, the fact that ECOM4 was able to survive and adapt to low pH environment is remarkable. As was shown before that the basic strategy of survival at low pH involves increasing internal pH by translocating protons (H^+) to the outside of the cell and achieving would could be described as a "dormant" physiology where the metabolic activity is reduced to a minimum [28,29]. The reduced metabolic activity supports the need for supplementation with the yeast extract in order to achieve active growth.

The increase in growth rate observed toward the end of evolution was correlated with the increase of the lactate titter observed in the broth, which indicates that lactate secretion is growth-coupled and dependent on growth rate. The secretion of lactate is coupled to the growth of the ECOM4 strain through the NADH oxidation. The necessity to oxidize the NADH forced the metabolic flux towards the D-lactate making it a major growth by-product. Utilization of this redox-coupling can lead to the

improved production capabilities of the ECOM4 strain. The lactate dehydrogenase can be substituted for a different pathway as long as the redox-coupling is preserved. In a next chapter we will present applications of the recombinant ECOM4 strain for redox-coupled production of the racemic lactate mixture and L-alanine amino acid.

5.4 MATERIALS AND METHODS

5.4.1 Strains and media

The strain described in this study was generated from the cytochrome oxidase mutant strain (ECOM3) presented before[1]. The quinol monooxygenase (*ygiN*) [30] was removed from the unevolved ECOM3; the resultant strain harbored the following mutations: *cydAB, cyoABCD, cbdAB, ygiN*, and was named ECOM4 (*Escherichia Coli* Oxidase Mutant 4). Bacterial strains were cultured at 37°C in M9 minimal liquid medium containing 4 g/L glucose and 2 g/L yeast extract, except as noted.

5.4.2 Adaptive evolution

Mutant strains were adaptively evolved in order to adapt the strain to fermentation of glucose to lactate at low pH. In short, a colony off a fresh agar plate was inoculated into 1 liter bioreactor containing 500 ml of M9 minimal media with 4 g/L glucose and 2 g/L yeast extract under oxic conditions. Cells were transferred on daily bases allowing 10 or less doubling per 24 hours. The acidity of the media was slowly adjusted from 7.0 to 4.5 over a period of 130 days. The pH was controlled by addition of 2M NaOH during cell growth to keep it within the range. The N-Methyl-

N'-nitro-N-nitrosoguanidine (NTG) was chosen as an efficient mutagen at a level of 2 μg/ml and was added on daily bases for nearly 40 days at pH 4.5 until increase in growth rate was observed. The lowest pH that was able to support the growth of the evolved ECOM4 strain was 4.5. We attempted to increase acidity to the pH 4.0 and bellow; however, this was too detrimental to the cell growth and culture was returned to the pH 4.5 for further adaptation. Evolving cultures were also supplemented with 50 μg/ml kanamycin once a week and screened daily with PCR for presence of contaminants. Samples were also frozen every 2 days throughout the evolution.

5.4.3 Phenotype assessment

To assess phenotypic characteristics of evolved and isolated strains, growth rates and byproduct secretion profiles were measured. Each strain was grown in batch culture under oxic, and anoxic conditions. Aerobic cultivation was conducted in 500 ml Erlenmeyer flasks containing 250 ml M9 medium with all appropriate additions. Growth at low pH was conducted with 2 g/L of yeast extract in the media to ensure growth of the evolved strain. Temperature was controlled by a circulating water bath, mixing and aeration was controlled with a stir bar at ~1000 rpm. Anaerobic cultivation was conducted in 250 ml Erlenmeyer flasks with 200 ml medium, sealed with rubber stoppers containing necessary inlet tubing. Anoxic conditions were achieved by continuous flushing of cultures with 95% N_2 5% CO_2 gas mixture at a flow rate of 1 ml/min. The temperature was controlled by using a circulating water bath; mixing was controlled with a stir speed of ~200 rpm. Samples were taken from

batch cultures regularly (every 30 min), filtered through a 0.2 μm filter and stored at -20°C for byproduct secretion analysis. Glucose concentration in the media was assessed using an enzymatic assay kit (R-Biopharm), while *D*-lactate secretion was measured using RI (refractive index) detection by HPLC (Waters, Milford) with a Bio-Rad Aminex HPX87-H ion exclusion column (injection volume, 10 μl) and 5mM H_2SO_4 as the mobile phase (0.5 ml/min, 45°C). The identities of metabolites and organic acids in the fermentation broth were further verified with enzymatic kits (R-Biopharm). The oxygen uptake rate of each aerobic culture was determined by measuring the rate of dissolved oxygen depletion in an enclosed respirometer chamber using a polarographic dissolved oxygen probe (YSI).

5.4.4 Fermentation

Seed cultures were prepared for large (1-liter) fermentations by suspending single colonies from fresh plates in 1ml of M9 medium and using these as inocula for 500 ml flasks containing 250 ml M9 medium with 4 g/L glucose. After 20 hours of cultivation (1000 rpm) a portion of the cells was harvested and used to inoculate a New Brunswick BioFlow 110 fermentor (0.001 OD at 600 nm) containing 700 ml M9 medium with 100 g/L glucose. Mixing was provided by a dual Rushton impeller, temperature was controlled by a heating jacket, and aeration was achieved by sparging medical grade air with flow rate of 1 ml/min. NaOH (2 M) was used to maintain the pH at 7.0. Samples were removed regularly for analysis of organic acids and cell mass.

Fermentations were stopped when base addition was no longer required to maintain pH.

5.5 ACKNOWLEDGEMENTS

The text of this chapter, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Tarasova Y., Matusov Y., Palsson, B.Ø.Redox-coupled production of organic and amino acids using *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010, *In preparation.* The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

5.6 REFERENCES

1. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. *Appl Environ Microbiol* 2008;74:7561-9.

2. Portnoy VA, Scott DA, Lewis NE, Tarasova Y, Osterman A, Palsson BO. Deletion of cytochrome oxidases and quinol monooxygenase blocks aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655 *Appl Environ Microbiol* 2010.

3. Coelho LF, Bolner de Lima CJ, Bernardo MP, Alvarez GM, Contiero J. Improvement of L(+)-lactic acid production from cassava wastewater by Lactobacillus rhamnosus B 103. *J Sci Food Agric* 2010;90:1944-50.

4. John RP, G SA, Nampoothiri KM, Pandey A. Direct lactic acid fermentation: focus on simultaneous saccharification and lactic acid production. *Biotechnol Adv* 2009;27:145-52.

5. Zhang Y, Cong W, Shi SY. Repeated fed-batch lactic acid production in a packed bed-stirred fermentor system using a pH feedback feeding method. *Bioprocess Biosyst Eng* 2010.

6. Auras R, Harte B, Selke S. An overview of polylactides as packaging materials. *Macromol Biosci* 2004;4:835-64.

7. Zhou S, Shanmugam KT, Ingram LO. Functional replacement of the *Escherichia coli* D-(-)-lactate dehydrogenase gene (ldhA) with the L-(+)-lactate dehydrogenase gene (ldhL) from *Pediococcus acidilactici*. *Appl Environ Microbiol* 2003;69:2237-2244.

8. Lowe NJ. Optimizing poly-L-lactic acid use. *J Cosmet Laser Ther* 2008;10:43-6.

9. Lee SM, Koo YM, Lin J. Production of lactic acid from paper sludge by simultaneous saccharification and fermentation. *Adv Biochem Eng Biotechnol* 2004;87:173-94.

10. Flieger M, Kantorova M, Prell A, Rezanka T, Votruba J. Biodegradable plastics from renewable sources. *Folia Microbiol (Praha)* 2003;48:27-44.

11. Silva AM, Kong X, Hider RC. Determination of the pKa value of the hydroxyl group in the alpha-hydroxycarboxylates citrate, malate and lactate by 13C NMR: implications for metal coordination in biological systems. *Biometals* 2009;22:771-8.

12. Galhardo RS, Hastings PJ, Rosenberg SM. Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* 2007;42:399-435.

13. Wright BE. Stress-directed adaptive mutations and evolution. *Mol Microbiol* 2004;52:643-50.

14. Hall BG. Selection, adaptation, and bacterial operons. *Genome* 1989;31:265-71.

15. Applebee MK, Herrgard MJ, Palsson BO. Impact of individual mutations on increased fitness in adaptively evolved strains of Escherichia coli. *J Bacteriol* 2008;190:5087-94.

16. Drake JW. Spontaneous mutation. *Annu Rev Genet* 1991;25:125-46.

17. Shukla VB, Zhou S, Yomano LP, Shanmugam KT, Preston JF, Ingram LO. Production of D(-)-lactate from sucrose and molasses. *Biotechnol Lett* 2004;26:689-693.

18. Zhou S, Causey TB, Hasona A, Shanmugam KT, Ingram LO. Production of optically pure D-lactic acid in mineral salts medium by metabolically engineered *Escherichia coli* W3110. *Appl Environ Microbiol* 2003;69:399-407.

19. Zhu Y, Eiteman MA, DeWitt K, Altman E. Homolactate fermentation by metabolically engineered *Escherichia coli* strains. *Appl Environ Microbiol* 2007;73:456-464.

20. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-648.

21. Chang DE, Jung HC, Rhee JS, Pan JG. Homofermentative production of D- or L-lactate in metabolically engineered Escherichia coli RR1. *Appl Environ Microbiol* 1999;65:1384-9.

22. Dien BS, Nichols NN, Bothast RJ. Recombinant Escherichia coli engineered for production of L-lactic acid from hexose and pentose sugars. *J Ind Microbiol Biotechnol* 2001;27:259-64.

23. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-8.

24. Zhou S, Causey TB, Hasona A, Shanmugam KT, Ingram LO. Production of Optically Pure D-Lactic Acid in Mineral Salts Medium by Metabolically Engineered Escherichia coli W3110. *Appl. Environ. Microbiol.* 2003;69:399-407.

25. Lee DH, Palsson BO. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. *Appl Environ Microbiol* 2010;76:4158-68.

26. Yagi T, Matsuno-Yagi A. The proton-translocating NADH-quinone oxidoreductase in the respiratory chain: the secret unlocked. *Biochemistry* 2003;42:2266-2274.

27. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC. Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* 1999;284:1961-1966.

28. Foster JW. Escherichia coli acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* 2004;2:898-907.

29. Richard HT, Foster JW. Acid resistance in Escherichia coli. *Adv Appl Microbiol* 2003;52:167-86.

30. Adams MA, Jia Z. Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. *J Biol Chem* 2005;280:8358-8363.

Table 5.1: Physiological characteristic of the evolved ECOM4LA strain at pH 7.0 and pH 4.5 following a 100 day period of adaptive evolution

	Evolved ECOM4LA	
Physiological Characteristics	pH 7.0	pH 4.5
Growth rate (1/h)	0.32 +/- 0.005	0.18 +/- 0.02
Glucose Uptake Rate (mmol/gDW/h)	26.4 +/- 0.09	12.02 +/- 2.14
Lactate Secretion Rate (mmol/gDW/h)	48.6 +/- 0.76	19.7 +/- 0.58
Acetate Secretion Rate (mmol/gDW/h)	0	0
Yield (gram/gram)	0.98 +/- 0.07	0.80 +/- 0.25
Volumetric productivity (g/L/h)	1.6 +/- 0.1	0.7 +/- 0.30

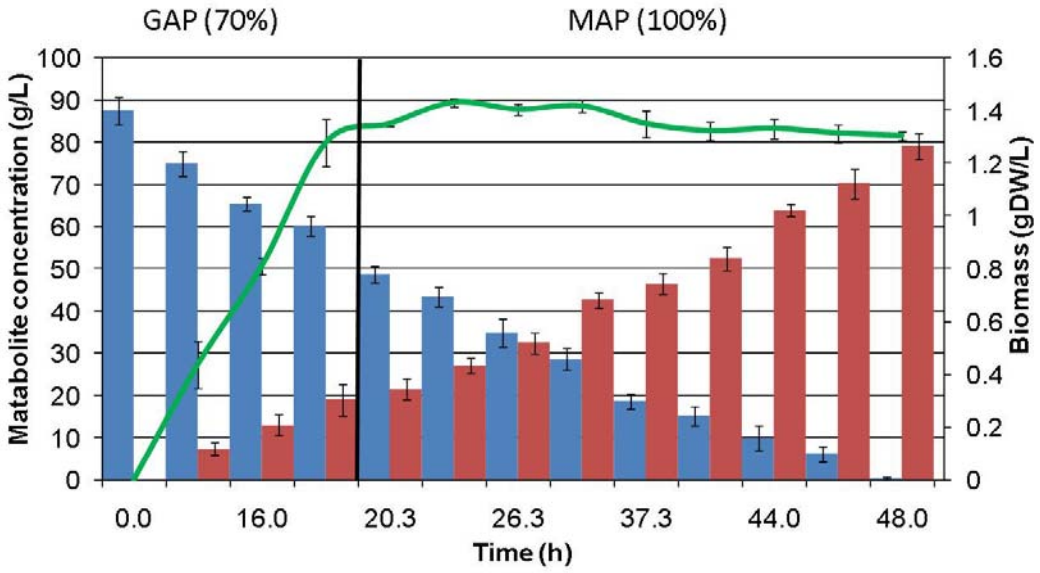


Figure 5.1: Batch fermentation of ECOM4LA strain**.** During batch fermentation the ECOM4LA strain produced D-lactic acid at 70% efficiency during the growth phase (GAP – Growth Associated Production) and at 100% efficiency during the stationary phase (MAP – Maintenance Associated Production).

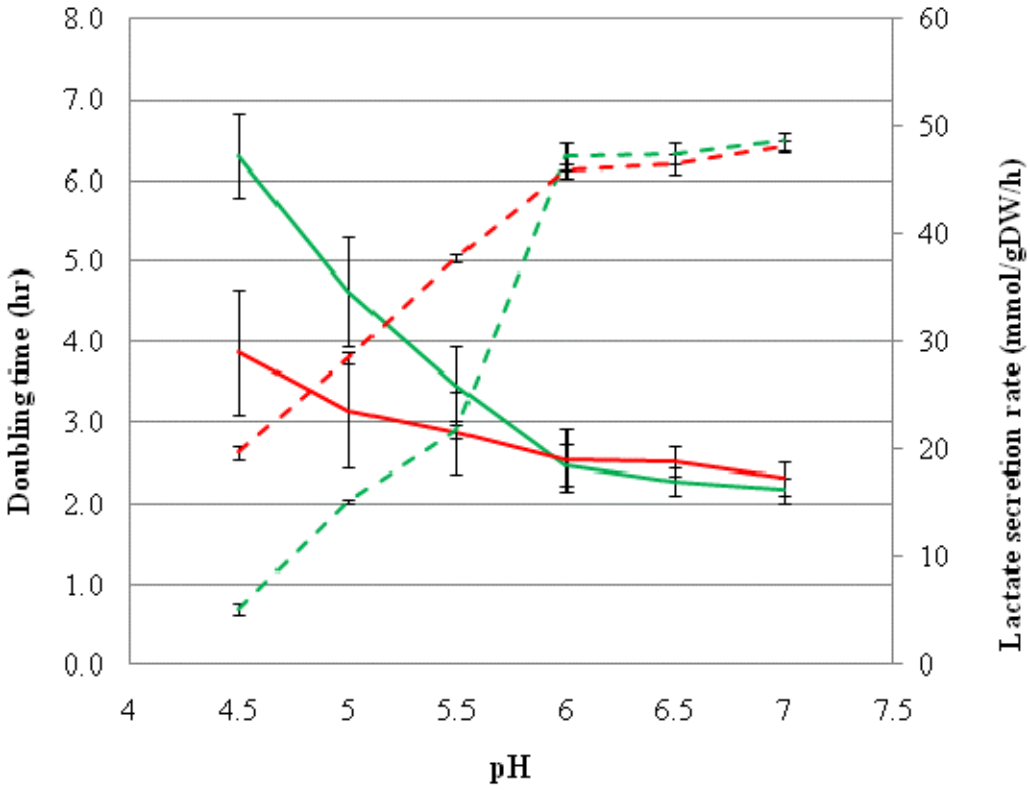


Figure 5.2: Assessment of the growth rate and lactate secretion rate of ECOM4LA at different pH before and after the adaptive evolution. Cells were cultured on M9 minimal media with glucose and yeast extract. Growth rate – solid line; lactate secretion rate – dashed line. Green – pre-adapted data; Red – after 100 days of adaptive evolution

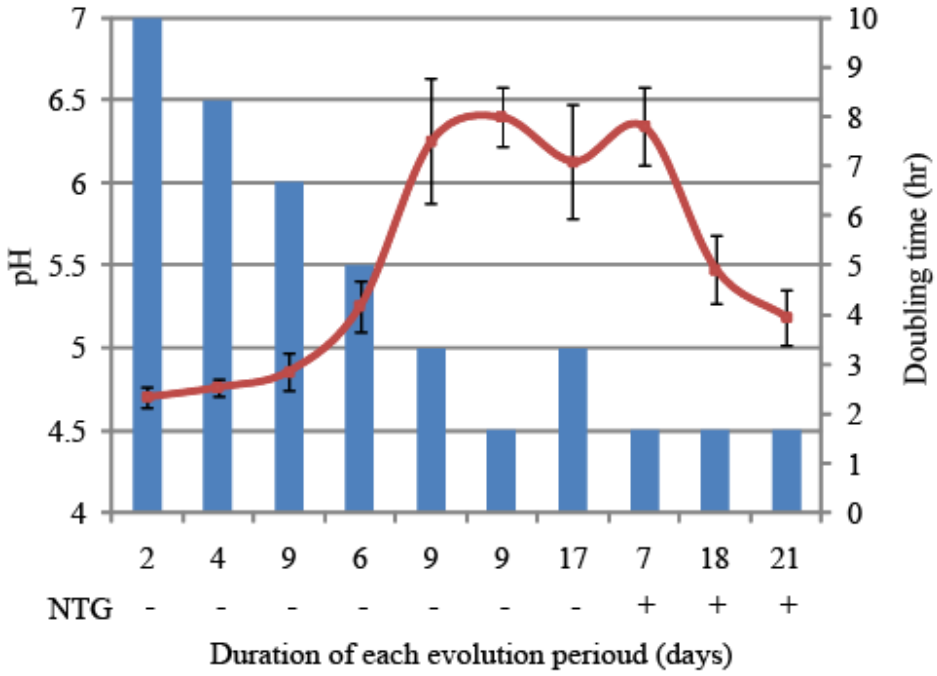


Figure 5.3: Evolutionary trajectory of the ECOM4LA strain during the 100 day adaptation to low pH environment. Blue bars indicate the pH value; red line – indicated the growth rate trajectory; the duration of each period during which the pH has not been adjusted is indicated below the x-axis as well as presence of NTG mutagen.

CHAPTER VI:

REDOX-COUPLED PRODUCTION OF ORGANIC AND AMINO ACIDS USING THE RECOMBINANT ECOM4 PLATFORM

6.1 INTRODUCTION

The depletion in fossil feedstocks, increasing oil prices, and ecological problems associated with increased CO_2 emissions result in increased interest in developing alternative sources of energy, fuels, and chemicals[1]. There is a global movement towards the replacement of non-renewable fossil resources with natural biomass. Biomass differs from crude oil in that it offers a *renewable* and *sustainable* source of carbon in the form of polymeric (cellulose, starch, lignin, hemicellulose, protein) and monomeric (carbohydrates, oils, amino acids, plant extractives) components[2].

Nowadays, practically all organic chemicals and plastics produced from crude oil and natural gas. However, it is possible to produce a wide range of commodity chemicals using microbial catalysis of natural biomass[3]. In order to be sustainable and economical, the process of bio-catalysis must utilize an organism optimized for production of different reduced by-products from variety of feedstocks. The robust

growth under various conditions and metabolic versatility make *Escherichia coli* an excellent platform microorganism for utilization in chemical production. There are a number of studied aimed at engineering *E. coli* for overproduction of commodity chemicals from the renewable resources[4-13]. Two most common groups of chemicals produced using the recombinant *E. coli* strains are organic[6,12,14-16] and amino acids[17-20].

Lactic acid is an important and versatile chemical that is commonly produced from the renewable resources. Lactic acid is used for food, pharmaceutical and polymers industries and is produced by bacterial fermentation[21]. Lactic acid has attracted a lot of attention for polymerization for polylactic acid (PLA) production. PLA is produced by co-polymerization of D- and L- lactic acid isomers in precise ratio depending on the desired properties of the resulted PLA[22]. We engineered ECOM4 strain to co-produce D- and L-lactic acid together from glucose in the mineral salt medium.

Alanine is used for pharmaceutical and veterinary applications, it is also used as a food additive due to its sweet taste[17,23]. Alanine is a central interemediate and an essential component of cellular proteins. *E. coli* produces L-alanine from pyruvate and ammonia using NADH-linked alanine dehydrogenase (*dadA*)[19].

In this chapter we will present two metabolically engineered strains capable of overproduction of racemic lactate mixture and L-alanine amino acid. Strains presented here were developed from the ECOM4 strain by introducing additional production capabilities using recombinant DNA technology. The resulted strains were capable of overproduction of both lactic acid stereo isomers with variable ratios and L-alanine amino acid with nearly maximum theoretical efficiency.

6.2 RESULTS

6.2.1 Production of high purity racemic lactate mixture by recombinant ECOM4LA strain

The ECOM4 strain was further improved to include additional production capabilities. The ECOM4 strain (*cydABcyoABCDcbdABygiN*) was genetically modified to produce a racemic mixture of D- and L- lactic acid. The new strain was named ECOM4DL (ECOM4: D- and L- lactic acid producer). Originally the ECOM4 strain produced dextrorotatory isomer (D-) of lactic acid as a result of carbohydrate fermentation under both oxic and anoxic growth conditions. The production efficiency was nearly 100% and productivity was between 1.4-1.7 g/L/h at moderate cell densities. In order to engineer the ECOM4 strain to produce racemic lactate mixture, suitable for PLA production we introduced the L-lactate dehydrogenase (EC: 1.1.1.27) from the *Lactococcus Lactis* bacteria. The recombinant strain (Figure 6.1) was able to metabolize glucose to D- and L-lactate simultaneously with comparable yields. Gene expression analysis, performed in chapter 4, showed that the native lactate dehydrogenase (*ldhA*) was upregulated over 5 fold in ECOM4 strain[24,25]. Comparable expression of L-lactate dehydrogenase gene was achieved by using an IPTG inducible promoter. The inducible promoter was used to vary the dosage of *ldh* gene, and therefore alter the ratio of D- and L- isomers in the final mixture (Figure 6.2). Production rates for D- and L- isomers with the 10 μM IPTG induction varied between 17 – 21 mmol/gDW/h for both stereo isomers for a total lactate secretion rate

of 34 – 42 mmol/gDW/h. The growth rate of the ECOM4DL strain with 10 μM IPTG induction was on the order to 0.25 +/- 0.84 h^{-1}.

6.2.2 Production of high purity L-alanine using recombinant ECOM4 platform

The ECOM4 strain was further modified to produce L-alanine from glucose. To this end, we introduced an additional copy of the alanine dehydrogenase (*dadA*) instead of the lactate dehydrogenase (*ldhA*) (Figure 6.3). The new strain was named ECOM4A (ECOM4: L-alanine producer). The ECOM4A strain produced L-alanine from glucose with the following yield and productivity: 70 +/- 2.54 % and 27.3 +/- 0.46 mmol/gDW/h respectively. The alanine dehydrogenase was expressed using the similar expression system: pGEX-6p-1 plasmid with the IPTG induction. Low levels of induction (20 μM IPTG) was sufficient to produce enough of alanine dehydrogenase to sustain the elevated glycolytic flux from glucose to pyruvate. We observed trace amounts of pyruvate present in the fermentation broth in mid-log growth phase; however, no pyruvate was detected at the end of fermentation. The growth of the ECOM4A strain during alanine fermentation with 20 μM IPTG induction was 0.18 +/- 0.08 h^{-1}.

6.3 DISCUSSION

As a result of the perturbation to the electron transport chain resulting in the inability to transfer electrons from the NADH to oxygen the lactic acid secretion

became growth coupled. The secretion of the lactic acid is triggered by the need to oxidize the global redox carried (NADH) and supply NAD^+ for other metabolic functions. The removal of lactate dehydrogenase (*ldhA*) ultimately results in a weak phenotype incapable of growing without supplementation, illustrating the essentiality of the redox-coupled lactate secretion. The need for NADH recycling can be used as key feature to couple the "desired" metabolic flux to the growth of the ECOM strain. In other words the redox-coupling can be employed to produce not only the D-lactic acid but other commodity chemicals as long as the new pathway depends on NADH oxidation. Particularly, we demonstrated that the lactate dehydrogenase can be substituted by an entirely different pathway as long as the ability to oxidize NADH is preserved. We presented L-alanine production by the recombinant ECOM4A strain (ECOM4: L-alanine producer) capable of overproduction of an amino acid L-alanine at the industrially relevant yields with volumetric productivity of: 0.8 g/L/h.

The idea of coupling production to growth or other metabolic function is not a new one. It has been shown that growth-coupling is extremely useful for the overproduction of natural and latent metabolites using *E. coli* or yeast[14,26-28]. Some of the computational methods that allow to design producing strains using growth-coupling approaches and algorithms such as OptKnock [29,30] were presented in Chapter 2. Redox-coupling is a new concept; however, it has a similar goal: coupling production to the key metabolic function ensuring that the carbon flux is always supplied to the desired product. The need to oxidize the NADH forces the carbon flux through the lactate dehydrogenase reaction resulting in formation of D-lactic acid. We used this coupling to introduce other functionalities into the ECOM4 strain in order to

overproduce other metabolites. We successfully demonstrate that ECOM4 can be engineered to produce racemic mixture of lactic acid with predictable content (ECOM4DL) and to produce L-alanine (ECOM4A) at nearly maximum theoretical efficiency. Due to the absence of by-products the racemic lactate mixture and L-alanine have high purity.

6.4 MATERIALS AND METHODS

6.4.1 Strains and media

The strain described in this study was generated from the cytochrome oxidase mutant strain (ECOM3) presented before (10). The quinol monooxygenase (*ygiN*) (1) was removed from the unevolved ECOM3; the resultant strain harbored the following mutations: *cydAB, cyoABCD, cbdAB, ygiN*, and was named ECOM4 (*E*scherichia *C*oli Oxidase Mutant 4). Genetically, modified strain containing the copy of the lactate dehydrogenase from the *L. lactis* was named ECOM4DL (ECOM4: D- and L- lactic acid producer). Genetically modified strain containing an extra copy of the alanine dehydrogenase from *E. coli* was named ECOM4A (ECOM4: L-alanine producer). Bacterial strains were cultured at 37°C in M9 minimal liquid medium containing 4 g/l glucose, except as noted.

6.4.2 Cloning

The lactate dehydrogenase (*ldh*) and alanine dehydrogenase (*dadA*) were cloned into the pGEX-6p-1 expression system using the standard cloning protocol. In short, flanking oligonucleotide primers (20 base pairs) were setup using the genomic sequence and ordered from the IDT (San Diego). Restriction sites for BamHI and NotI restriction enzymes were included in primers outside of the gene regions. Gene sequences were amplified using a standard PCR reaction from the genome of *L. lactic* and *E. coli* respectively for *ldh* and *dadA*[31]. Resulted DNA fragments were digested with BamHI and NotI restriction enzymes (New England Biolabs, MA) using a standard restriction digest protocol at 37°C overnight. The resulted digested gene sequences were ligated into the pGEX-6p-1 plasmid (GE Healthcare) using standard ligation reation[32,33]. Ligated plasmids were transformed into *E.coli* BL21 for expression and confirmation. Plasmids were also purified and digested with BamHI and NotI to verify the correct insert size. Once plasmid integrity was confirmed, they were transformed into the ECOM4 strain. The resulted strains were selected on solid LB medium containing dual antibiotic resistance (kan/amp). The native lactate dehydrogenase (*ldhA*) was removed from the ECOM4 strain containing an additional copy of the *dadA* gene, using the technique of homologous recombination[34] (see Chapter 3 for the detailed protocol).

6.4.3 Phenotype assessment

To assess phenotypic characteristics of evolved and isolated strains, growth rates and byproduct secretion profiles were measured. Each strain was grown in batch

culture under oxic, and anoxic conditions. Aerobic cultivation was conducted in 500 ml Erlenmeyer flasks containing 250 ml M9 medium. Temperature was controlled by a circulating water bath, mixing and aeration was controlled with a stir bar at ~1000 rpm. Anaerobic cultivation was conducted in 250 ml Erlenmeyer flasks with 200 ml medium, sealed with rubber stoppers containing necessary inlet tubing. Anoxic conditions were achieved by continuous flushing of cultures with 95% N_2 5% CO_2 gas mixture at a flow rate of 1 ml/min. The temperature was controlled by using a circulating water bath; mixing was controlled with a stir speed of ~200 rpm. Samples were taken from batch cultures regularly (every 30 min), filtered through a 0.2 μm filter and stored at -20°C for byproduct secretion analysis. Glucose concentration in the media was assessed using an enzymatic assay kit (R-Biopharm), while *D*-lactate secretion was measured using RI (refractive index) detection by HPLC (Waters, Milford) with a Bio-Rad Aminex HPX87-H ion exclusion column (injection volume, 10 μl) and 5mM H_2SO_4 as the mobile phase (0.5 ml/min, 45°C). The identities of metabolites and organic acids in the fermentation broth were further verified with enzymatic kits (R-Biopharm). The oxygen uptake rate of each aerobic culture was determined by measuring the rate of dissolved oxygen depletion in an enclosed respirometer chamber using a polarographic dissolved oxygen probe (YSI).

6.4.4 Fermentation

Seed cultures were prepared for large (1-liter) fermentations by suspending single colonies from fresh plates in 1ml of M9 medium and using these as inocula for

500 ml flasks containing 250 ml M9 medium with 4 g/l glucose. After 20 hours of cultivation (1000 rpm) a portion of the cells was harvested and used to inoculate a New Brunswick BioFlow 110 fermentor (0.001 OD at 600 nm) containing 700 ml M9 medium with 100 g/L glucose. Mixing was provided by a dual Rushton impeller, temperature was controlled by a heating jacket, and aeration was achieved by sparging medical grade air with flow rate of 1 ml/min. NaOH (2 M) was used to maintain the pH at 7.0. Samples were removed regularly for analysis of organic acids and cell mass. Fermentations were stopped when base addition was no longer required to maintain pH.

6.5 ACKNOWLEDGEMENTS

The text of this chapter, in part or in full, is a reprint of material as it appears in Portnoy, V.A., Tarasova Y., Matusov Y., Palsson, B.Ø.Redox-coupled production of organic and amino acids using *Escherichia coli* K-12 MG1655, *Appl Environ Microbiol*. 2010, *In preparation.* The dissertation author was the primary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

6.7 REFERENCES

1. Scott E, Peter F, Sanders J. Biomass in the manufacture of industrial products--the use of proteins and amino acids. *Appl Microbiol Biotechnol* 2007;75:751-62.

2. Bozell JJ. Renewable feedstocks for the production of chemicals. Golden: National Renewable Energy laboratory:204- 209.

3. Patel MK, Dornburg V, Hermann BG, Shen L, van Overbeek L. Sustainable production of bulk chemicals by application of "white biotechnology". *Sheng Wu Gong Cheng Xue Bao* 2008;24:2022-6.

4. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* 2006;281:8024-8033.

5. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007;3:149.

6. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol* 2005;71:7880-7887.

7. Altaras NE, Cameron DC. Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli. *Appl Environ Microbiol* 1999;65:1180-5.

8. Antoniewicz MR, Kraynie DF, Laffend LA, Gonzalez-Lergier J, Kelleher JK, Stephanopoulos G. Metabolic flux analysis in a nonstationary system: Fed-batch fermentation of a high yielding strain of E. coli producing 1,3-propanediol. *Metab Eng* 2007.

9. Aristidou AA, San KY, Bennett GN. Metabolic engineering of *Escherichia coli* to enhance recombinant protein production through acetate reduction. *Biotechnology Progress* 1995;11:475-478.

10. Chang DE, Jung HC, Rhee JS, Pan JG. Homofermentative production of D- or L-lactate in metabolically engineered Escherichia coli RR1. *Appl Environ Microbiol* 1999;65:1384-9.

11. Dharmadi Y, Murarka A, Gonzalez R. Anaerobic fermentation of glycerol by Escherichia coli: a new platform for metabolic engineering. *Biotechnol Bioeng* 2006;94:821-9.

12. Donnelly MI, Millard CS, Clark DP, Chen MJ, Rathke JW. A novel fermentation pathway in an *Escherichia coli* mutant producing succinic acid, acetic acid, and ethanol. *Appl Biochem Biotechnol* 1998;70-72:187-98.

13. Farmer WR, Liao JC. Precursor balancing for metabolic engineering of lycopene production in Escherichia coli. *Biotechnol Prog* 2001;17:57-61.

14. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643-8.

15. Zhu Y, Eiteman MA, DeWitt K, Altman E. Homolactate fermentation by metabolically engineered *Escherichia coli* strains. *Appl Environ Microbiol* 2007;73:456-464.

16. Lee SY, Hong SH, Moon SY. In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered Escherichia coli as an example. *Genome Inform* 2002;13:214-23.

17. Lee M, Smith GM, Eiteman MA, Altman E. Aerobic production of alanine by Escherichia coli aceF ldhA mutants expressing the Bacillus sphaericus alaD gene. *Appl Microbiol Biotechnol* 2004;65:56-60.

18. Raunio RP, Jenkins WT. D-alanine oxidase form Escherichia coli: localization and induction by L-alanine. *J Bacteriol* 1973;115:560-6.

19. Zhang X, Jantama K, Moore JC, Shanmugam KT, Ingram LO. Production of L -alanine by metabolically engineered Escherichia coli. *Appl Microbiol Biotechnol* 2007;77:355-66.

20. Smith GM, Lee SA, Reilly KC, Eiteman MA, Altman E. Fed-batch two-phase production of alanine by a metabolically engineered Escherichia coli. *Biotechnol Lett* 2006;28:1695-700.

21. Okano K, Tanaka T, Ogino C, Fukuda H, Kondo A. Biotechnological production of enantiomeric pure lactic acid from renewable resources: recent achievements, perspectives, and limits. *Appl Microbiol Biotechnol*;85:413-23.

22. Flieger M, Kantorova M, Prell A, Rezanka T, Votruba J. Biodegradable plastics from renewable sources. *Folia Microbiol (Praha)* 2003;48:27-44.

23. Hols P, Kleerebezem M, Schanck AN, Ferain T, Hugenholtz J, Delcour J, de Vos WM. Conversion of Lactococcus lactis from homolactic to homoalanine fermentation through metabolic engineering. *Nat Biotechnol* 1999;17:588-92.

24. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. *Appl Environ Microbiol* 2008;74:7561-9.

25. Portnoy VA, Scott DA, Lewis NE, Tarasova Y, Osterman A, Palsson BO. Deletion of cytochrome oxidases and quinol monooxygenase blocks aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655 *Appl Environ Microbiol* 2010.

26. Burgard AP, Van Dien SJ. Methods and Organisms for the Growth-Coupled Production of Succinate. U.S.A., 2006.

27. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *In preparation* 2008.

28. Lee DH, Palsson BO. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. *Appl Environ Microbiol* 2010;76:4158-68.

29. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;84:647-57.

30. Pharkya P, Burgard AP, Maranas CD. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* 2003;84:887-99.

31. Lee DH, Lee YJ, Ryu YW, Seo JH. Molecular cloning and biochemical characterization of a novel erythrose reductase from *Candida magnoliae* JH110. *Microb Cell Fact* 2010;9:43.

32. Harper S, Speicher DW. Expression and purification of GST fusion proteins. *Curr Protoc Protein Sci* 2008;Chapter 6:Unit 6 6.

33. Wei J, Liu XJ, Li GD, Wang Y, Zhang ZC, Wang YQ, Lu ZM. Expression, Purification and Preliminary Clinical Use of Recombinant HBsAg GST-PreS1(21--47 aa) Fusion Proteins. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 2001;33:379-385.

34. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000;97:6640-6645.

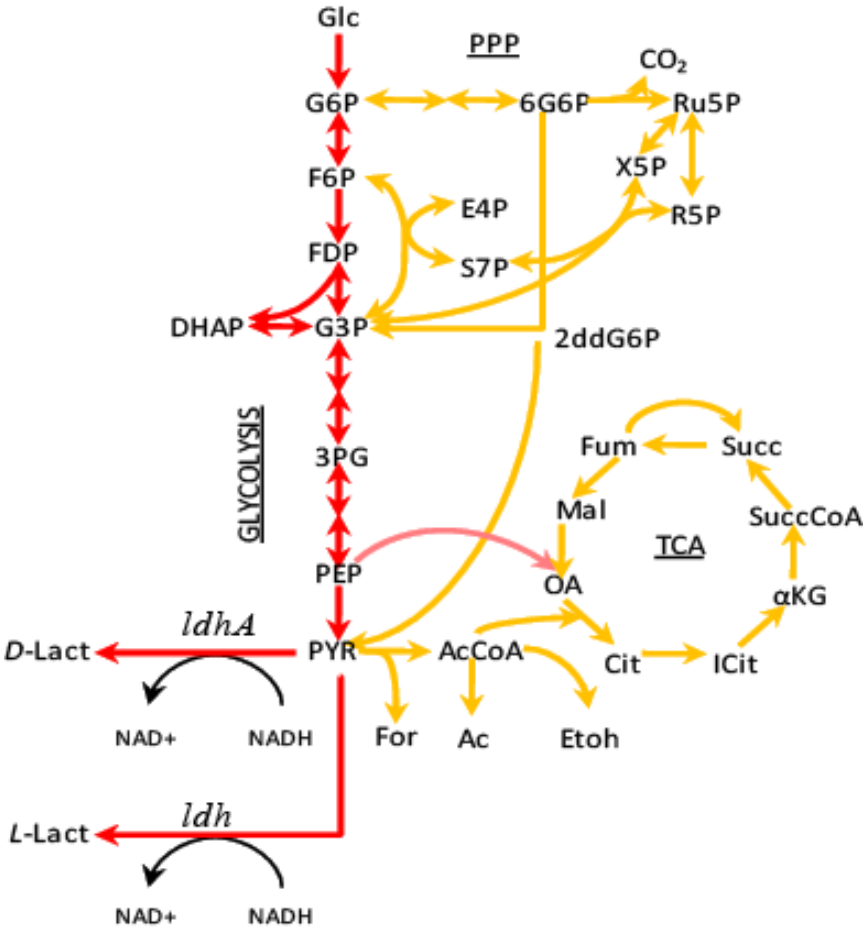


Figure 6.1: Metabolic flux distribution in the recombinant ECOM4 strain capable of overproducing the racemic mixture of lactic acid. Fluxes are assumed based on the ^{13}C measurements conducted in Chapter 4. *Ldh* – lactate dehydrogenase (*L. lactis*).

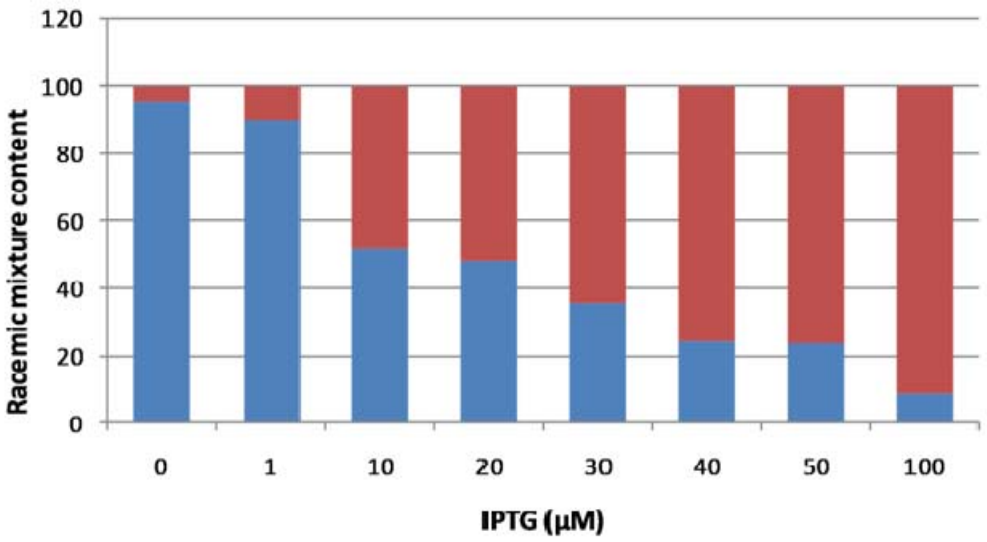


Figure 6.2: Composition of racemic mixture depends on the amount of IPTG added. Percentage of (*D*-) and (*L*-) lactate produced by ECOM4DL mutant with *ldh* (*L*-lactate dehydrogenase) with respect to amount of IPTG added. Bars show total amount of lactate produced, red part – *L*-lactate; blue part – *D*-lactate.

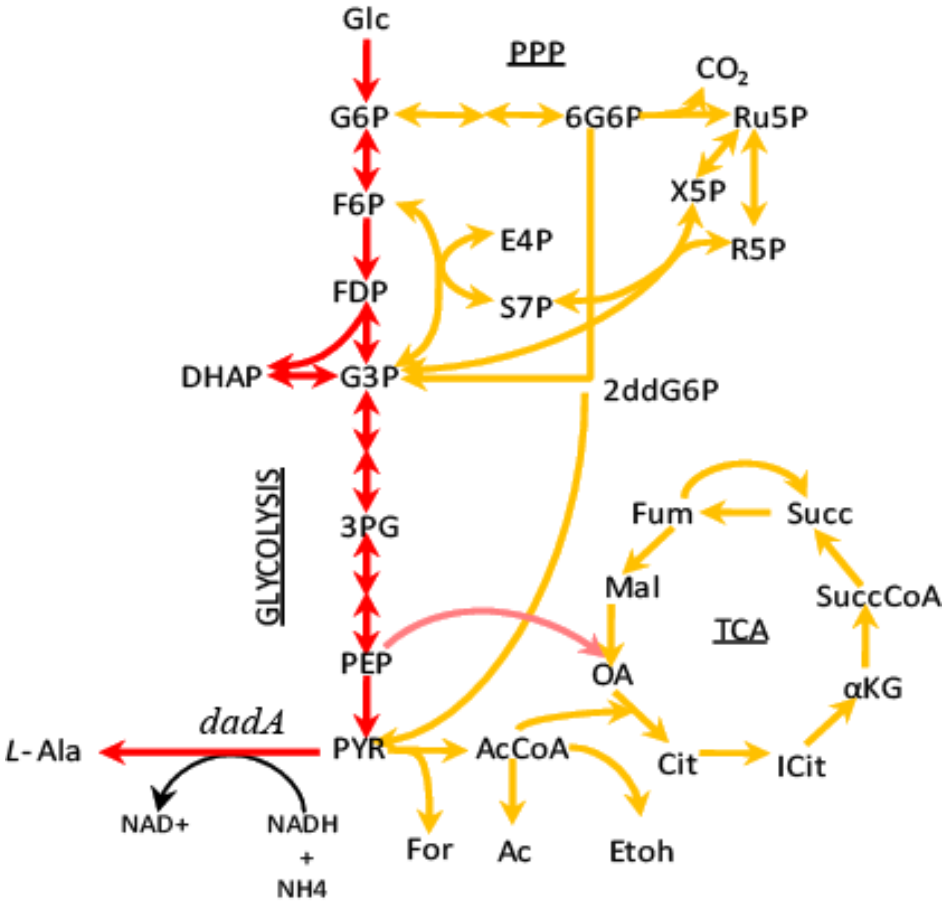


Figure 6.3: Metabolic flux distribution in the recombinant ECOM4 strain capable of overproducing L-Alanine. Fluxes are assumed based on the ^{13}C measurements conducted in Chapter 4. *dadA* – Alanine dehydrogenase (*E. coli*).

CHAPTER VII:

CONCLUDING REMARKS AND FUTURE DIRECTIONS

7.1 INTRODUCTION

The work described in this dissertation combines aspects of computational biology, adaptive evolution, molecular biology and metabolic engineering, all fitting under the umbrella of systems biology. While the term systems biology has been used often interchangeably with bioinformatics and computational biology based on high-throughput phenotype measurements, my personal views are different. To me, systems biology encapsulates the work that is aimed at understanding of systemic behavior of biological systems using high-throughput measurements and interactions between individual components. Thus, obtaining high-throughput measurements for every individual component (gene, protein, transcript, metabolite, etc.) of the biological system is extremely important, but these datasets would only be fundamental building blocks of systems biology. A systems-based understanding of these datasets is essential to a complete understanding of the biological systems, and the computational metabolic models are the only means available to truly understand the interactions between these individual components within the biological system.

The work presented in these dissertation relies heavily on the genome-scale metabolic model *i*JR906 [1] and *i*AF1260 [2], and it is a constant improvement of this metabolic model that allowed me to indentify the mechanism of oxygen utilization in the ECOM3 strain [3] (Chapter 3). Furthermore, the high-throughput data acquired for the gene expression analysis of the oxygen –independent platform strain presented in Chapter 4 was mapped onto the metabolic reconstruction of *E. coli* (*i*AF1260) in order to identify the perturbed pathways, which underlie observed changes in physiology. This clearly illustrates the usefulness of the metabolic reconstructions as a tool to understand and analyze high-throughput data sets and draw valid and relevant conclusions.

Herein, this chapter is devoted to summarize the major results in the following focus areas: i) development of oxygen-independent platform strain; ii) understanding the physiology of the developed platform strain; iii) redox-coupled production of organic and amino acids using the platform strain. Also future directions will be discussed.

7.2 SUMMARY OF RESULTS

In an effort to develop an oxygen independent strain that could be used as a platform strain for implementation of metabolic engineering designs, we developed a strain of *Escherichia coli* capable of efficient fermentation even under fully aerobic conditions. *E. coli* cytochrome oxidase mutant (ECOM3) strain was generated through inactivation of three known cytochrome oxidases by complete removal of coding sequence from the genome [4]. The resulting strain was adapted to growth on minimal

media without amino acid supplementation and end-point strains were extensively physiologically characterized. Lactic acid was observed as a major aerobic byproduct of three independently evolved ECOM3 strains consistent with expectations based on known physiology of lactic acid bacteria. However, under anaerobic conditions, two of the evolved strains showed mixed acid fermentation. Even though we expected that removal of terminal oxidases would fully eliminate oxygen uptake, our experimental results indicated residual oxygen uptake in all three end-point strains. Based on *in silico* modeling and gene expression analysis we attributed the residual oxygen uptake to elevated expression of *ygiN* gene encoding the quinole monooxygenase that is known to use oxygen as an electron acceptor in order to allow recycling of the quinone pool [5]. We identified the mechanisms accounting for and the observed residual oxygen uptake using a combination of genome-scale metabolic model of *Escherichia coli* and gene expression analysis of specific pathways. The removal of the *ygiN* gene nearly completely eliminated the oxygen uptake in ECOM3 cell line: 0.03 +/- 0.04 mmol O_2/gDW/h. This experimental evidence demonstrates that observed oxygen uptake can be attributed to the activity of YgiN. The resulted strain was named ECOM4 (*Escherichia coli* Cytochrome Oxidase Mutant 4). The ECOM4 strain have been shown to be amenable to genetic manipulation and thus can be used as a platform strain for further metabolic engineering that redirect lactate flux into other desirable byproducts.

As a result of the deletion of major components of the electron transport chain we present an *Escherichia coli* mutant strain capable of exhibiting very similar phenotypes when grown under oxic and anoxic conditions. The resulted strain

harbored following mutations: *cydABcyoABCDcbdABygiN*. We conducted whole genome transcriptomics analysis coupled with ^{13}C tracing experiments to understand metabolic changes resulted from deletions and adaptation.

The ECOM4 strain exhibited a similar physiological behavior under oxic and anoxic growth conditions and produced D-lactic acid as a major growth associated by-product. The D-lactate production is strongly coupled to NADH oxidation under both oxic and anoxic growth conditions. No other by-products were observed in a final media; however, we noted formation of succinate during the exponential growth. The oxygen uptake rate was diminished nearly 60 times comparing to WT *E. coli*. It is difficult to attribute the remaining oxygen uptake rate, observed in ECOM4 (0.25 +/- 0.12 mmol/gDW/h), to any significant metabolic function as very subtle differences were observed between oxic and anoxic physiology. We showed that the action mode of one of the global transcription regulators (ArcA) has been altered as a result of major metabolic adjustment, which affected the gene expression in a non-intuitive way. In particular, the change in a composition of the quinone pool has lead to activation of ArcA and subsequential activation of fermentative metabolism during the aerobic and anaerobic growth of ECOM4 cell line.

The phenotype of all ECOM strains has one unique feature: the D-lactic acid production. As a result of the perturbation to the electron transport chain resulting in the inability to transfer electrons from the NADH to oxygen the lactic acid secretion became growth coupled. The secretion of the lactic acid is triggered by the need to oxidize the global redox carried (NADH) and supply NAD^+ for other metabolic functions. The removal of lactate dehydrogenase (*ldhA*) ultimately results in a weak

phenotype incapable of growing without supplementation, illustrating the essentiality of the redox-coupled lactate secretion. The need for NADH recycling can be used as key feature to couple the “desired” metabolic flux to the growth of the ECOM strain. In other words the redox-coupling can be employed to produce not only the D-lactic acid but other commodity chemicals as long as the new pathway has the NADH oxidation step.

In particular we demonstrate that the ECOM4LA platform strain can be genetically improved to produce racemic mixture of lactic acid isomers. This was archived through the addition of a second copy of lactate dehydrogenase specific L-lactic acid. The second copy of L-lactate dehydrogenase was cloned from *L.lactis* and expressed in the pGEX-6-1 plasmid under IPTG control. The utilization of the inducible IPTG dependant promoter allowed varying the ratio of L- and D- lactic acid isomers. The ability to control the ratio of the stereo isomers provides a tuning mechanism for production of polylactic acid with different physical properties suitable for different applications [6]. Moreover, the utilization of the homo-fermentative platform strain leads to production of ultra pure D- and L-lactic acid, which significantly simplifies the downstream processing and product purification.

The ECOM4 strain was further modified to produce high purity L-alanine at yields close to theoretical maximum. An extra copy of the alanine dehydrogenase was introduced into ECOM4 strain using a similar expression system as presented for the racemic lactate mixture production. The resulted strain produced L-alanine with nearly 70% yield and specific production rate of 27.3 +/- 0.46 mmol/gDW/h and volumetric productivity of 0.8 g/L/h.

7.3 FUTURE DIRECTIONS

Practically all organic chemicals and plastics are nowadays produced from the crude oil and natural gas [7]. However, it is possible to produce a wide bulk of chemicals from the renewable resources though the use of biotechnology and metabolic engineering. A conversion technology for biomass lags behind that for petrochemicals leading to a divergence on the two industries. However, a number of companies are beginning to realize the potential of microbial production for bio-catalysis and switching its manufacturing strategies.

In the chapter 6, I presented an application of ECOM4 platform strain for the redox-coupled overproduction of metabolites, in particular L-lactate acid and L-alanine. However, one can imagine that alternative pathways can be introduced into ECOM4 strain to produce other commodity chemicals or bio-fuels through the redox-coupling. For example, it is possible to engineer the ECOM4 strain to produce 2,3-butanediol (2,3-BDO), that can be easily converted into 1.3-butainedien [8] or dehydrated into methyl-ethyl keton [9]. The 2,3-BDO production using the ECOM4 platform strain is possible since the overall stoichiometry of the reaction is presented below and allows for redox-coupling through the NADH oxidation.

$$2\ pyruvate + NADH \longrightarrow 2\ CO_2 + \ 2{,}3 - butanediol + NAD^+$$

Additionally to the 2,3-BDO production, there are a number of pathways that can be coupled to pyruvate by NADH oxidation and therefor implemented in the ECOM4 platform strain. The overall scheme of the utilization of ECOM4 strain as a redox-coupled production platform is presented in Figure 7.1.

7.4 REFERENCES

1. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biology* 2003;4:R54.1-R54.12.

2. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3.

3. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. *Appl Environ Microbiol* 2008;74:7561-9.

4. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000;97:6640-6645.

5. Adams MA, Jia Z. Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. *J Biol Chem* 2005;280:8358-8363.

6. Flieger M, Kantorova M, Prell A, Rezanka T, Votruba J. Biodegradable plastics from renewable sources. *Folia Microbiol (Praha)* 2003;48:27-44.

7. Patel MK, Dornburg V, Hermann BG, Shen L, van Overbeek L. [Sustainable production of bulk chemicals by application of "white biotechnology"]. *Sheng Wu Gong Cheng Xue Bao* 2008;24:2022-6.

8. Yan Y, Lee CC, Liao JC. Enantioselective synthesis of pure (R,R)-2,3-butanediol in Escherichia coli with stereospecific secondary alcohol dehydrogenases. *Org Biomol Chem* 2009;7:3914-7.

9. Voloch M, Jansen NB, Ladisch MR, Tsao GT, R. N, W. RV. 2,3 - Butanediol:933-947.

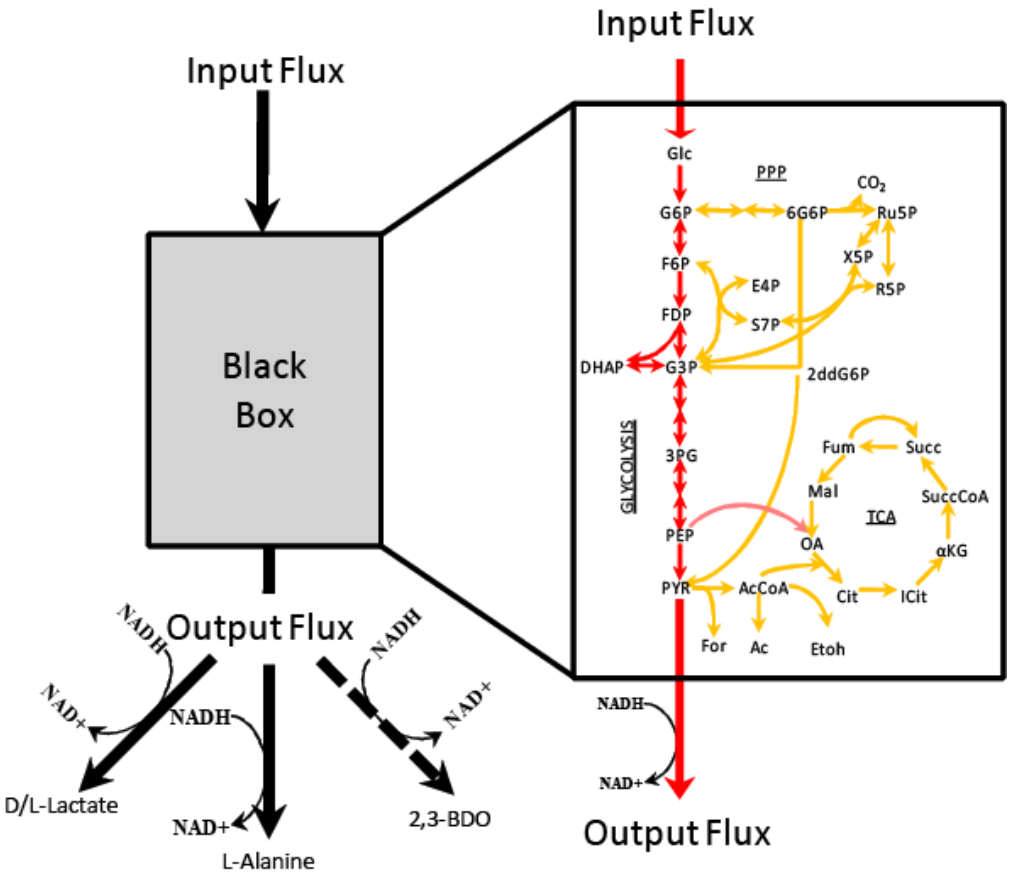


Figure 7.1: Used of ECOM4 strain as a redox-coupled production platform. The output flux is coupled to the NADH oxidation as a key feature required for the survival of this strain. The native product of fermentation is D-lactate can be substituted for other NADH coupled reactions such as L-lactate dehydrogenase; L-Alanine dehydrogenase, 2,3-BDO pathway (dashed line – multistep pathway), etc.