

Transcriptional regulation in *Escherichia coli*: A systems biology approach

Covert, Markus Willard

ProQuest Dissertations and Theses; 2003; ProQuest Dissertations & Theses A&I

pg. n/a

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Transcriptional Regulation in *Escherichia coli*:
A Systems Biology Approach

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Bioengineering, with specialization
in Bioinformatics

by

Markus Willard Covert

Committee in charge:

Professor Bernhard O. Palsson, Chair
Professor Andrew McCulloch
Professor Pavel Pevzner
Professor Milton Saier
Professor Shankar Subramaniam

2003

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



UMI Microform 3112853

Copyright 2004 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Copyright

Markus Willard Covert, 2003

All rights reserved.

The dissertation of Markus Willard Covert is approved,
and it is acceptable in quality and form for publication
on microfilm:

Paul Pen

M Saer

J McCulloch

Shanahan Subramanian

Benthal Plessen

Chair

University of California, San Diego

2003

To My Family: Past, Present and Future

*There are days we live
as if death were nowhere
in the background; from joy
to joy to joy, from wing to wing,
from blossom to blossom to
impossible blossom, to sweet impossible blossom.*
(Li-Young Lee, “From Blossoms”)

Thank you for those days, every day.

Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

—T.S. Eliot, “Choruses from the Rock”

*How often have I said to you that when you have eliminated the impossible,
whatever remains, however improbable, must be the truth?*

—Sherlock Holmes, *A Study in Scarlet*

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	viii
List of Tables	x
Acknowledgements	xi
Vita and Publications	xiii
Abstract	xvi
1 The Challenge of Systems Biology	1
A. From reductionism to systems	2
B. The need for mathematical tools	3
C. Chapter overview	4
2 Metabolic constraint-based models	8
A. Reconstruction of metabolic reaction networks	10
1. Genome annotation	10
2. Metabolic biochemistry	11
3. Cell physiology	12
B. Model construction and analysis	13
1. Model characteristics	16
2. Modeling issues	17
C. Future challenges	20
3 The Effects of Transcriptional Regulation	24
A. Metabolic constraint-based analysis	26
1. Extreme pathway analysis	26
2. Flux-balance analysis	29
B. Regulatory constraint-based analysis	29
1. Regulatory network reconstruction	29
2. Regulatory network analysis	31
3. Simulating transcriptional regulation	31
4. Regulatory constraints	32
5. Boolean representation of regulatory events	33

6.	Integration with metabolic network analysis	37
C.	Regulated constraint-based analysis of a sample network	38
1.	Regulated extreme pathway analysis	41
2.	Regulated flux-balance analysis	51
4	Modeling <i>Escherichia coli</i>	64
A.	Network reconstruction	65
B.	Simulation of <i>E. coli</i> behavior	65
1.	Calculation of steady-state regulated metabolic flux distributions	65
2.	Changing environments and time-dependent cell behavior	67
3.	Mutant strain simulation	68
C.	Model testing and validation	68
1.	Mutant growth phenotypes	70
2.	Dynamic growth simulations	76
5	Integrating high-throughput data and computational models leads to <i>E. coli</i> network elucidation	84
6	Vistas	95
A.	Looking back: lessons	96
1.	The constraint-based approach enables modular modeling	96
2.	The dynamic solution space	97
3.	Model-centric discovery	98
B.	Other mountains: challenges	99
1.	Completion of the <i>E. coli</i> model	99
2.	Accounting for fine-resolution data	100
3.	Modeling multicellular organisms	102
C.	Conclusion	104
A	Flux maps for extreme pathways in the sample network	105
B	<i>Escherichia coli</i> Central Regulatory/Metabolic Model	112
C	<i>Escherichia coli</i> Full Regulatory/Metabolic Model	124
A.	<i>Escherichia coli</i> gene list	124
1.	Complex regulation	158
B.	Metabolite list	160
C.	Detailed phenotypic analysis	166
1.	Carbon Sources	166
2.	Nitrogen Sources	168
3.	Knockout strains	171
D.	Detailed microarray analysis	173
	Bibliography	181

LIST OF FIGURES

2.1	Integrated process of microbial metabolic model construction.	9
2.2	An analogy between simulation of traffic conditions in a typical city and simulation of a microbial cell using systems analysis.	14
2.3	Venn diagram of reactions in the <i>H. pylori</i> <i>in silico</i> metabolic model.	18
2.4	Applying the Scientific Method in the post genomic era.	21
3.1	Constraint-based analysis of metabolic reaction networks.	27
3.2	Regulatory constraints change the shape of the solution space.	34
3.3	A simple regulatory circuit.	36
3.4	A schematic of the simplified core metabolic network.	38
3.5	Extreme pathway reduction by constraints, using growth of the sample system in a C1 and C2 aerobic medium as an example.	46
3.6	Aerobic growth of the sample system in a complex medium.	48
3.7	Diagrammatic representation of solution space reduction.	50
3.8	Catabolite repression in the simplified network.	55
3.9	Aerobic/Anaerobic growth calculated using the simplified network. .	56
3.10	Amino acid biosynthesis.	57
3.11	Growth on carbon and amino acid with amino acid in excess.	58
3.12	Growth on a complex medium.	60
4.1	<i>In silico</i> modeling of metabolism and transcriptional regulation using the constraints-based approach.	66
4.2	Combined regulatory/metabolic network for central metabolism in <i>E. coli</i>	69
4.3	Aerobic growth on acetate with glucose reutilization.	77
4.4	Anaerobic growth on glucose.	79
4.5	Aerobic growth on glucose and lactose.	80
5.1	Biological network characterization via a model-centric approach. .	86
5.2	Comparison of high-throughput phenotyping array data with predictions for the <i>E. coli</i> network.	88
5.3	Characterizing the regulatory network related to the aerobic-anaerobic shift.	90
5.4	Summary of the hypotheses generated using the combined experimental-computational approach.	93
6.1	Constraint-based analysis enables modular modeling.	97
6.2	Incorporating partial kinetic knowledge as a constraint.	101
6.3	Modeling multicellular organisms.	103
A.1	Flux distribution maps for the 80 extreme pathways.	106

C.1 A complete list of the new regulatory rules generated by the combined computational-experimental approach.	174
--	-----

LIST OF TABLES

3.1	A list of feasible extreme pathways for the sample system, categorized by environment.	42
3.2	Numerical values of parameters used in growth simulations of the sample metabolic/regulatory network.	53
4.1	Prediction of mutant phenotypes.	71
B.1	Core model metabolic reactions and regulatory rules.	113
B.2	Core model metabolic reactions and regulatory rules.	121
C.1	Reactions added to or changed in the iMC1010 <i>E. coli</i> model.	125
C.2	Regulatory rules for the full <i>Escherichia coli</i> model.	126
C.3	Complex regulatory rules.	159
C.4	Metabolite abbreviations for the full <i>Escherichia coli</i> model.	161
C.5	Carbon sources with inconsistent growth phenotypes.	167
C.6	Nitrogen sources with inconsistent growth phenotypes.	169
C.7	Knockout strains with inconsistent growth phenotypes.	172

ACKNOWLEDGEMENTS

It takes a village to write a dissertation. For the ideas, extended discussions, brainstorming sessions, for assistance with the writing of software, generation of data and publication of manuscripts, I owe more thanks to mentors, colleagues and friends than I can adequately express in these few paragraphs.

Bernhard Palsson tops the list of course, as an advisor, mentor and professor *par excellence*. Besides the encouragement, the advice, the opportunities, and the kindness he has regularly shared with me, he is simply an outstanding person, and an example of what I would like to be like as I develop in my career. If I were singing this dissertation, he would be the lead guitarist.

Several other mentors have been “instrumental” in helping me to bring these ideas together here. I’d like to particularly thank Shankar Subramaniam for always going the extra mile to counsel and encourage, and Milton Saier, for helpful thoughts concerning the experimental portion of this work. Other great mentors include Andrew McCulloch, Pavel Pevzner, and Willie Brown, whose infectious enthusiasm awoke me to the wonder of microbes for the first time.

Dr. Palsson’s Systems Biology Research Group (SBRG) is full of wonderful people who helped me on my way. I would like to thank particularly Christophe Schilling and Jeremy Edwards for their mentoring and friendship as I began my graduate career, Iman Famili, who has been a true friend and inspiration since my first day at UCSD, and Eric Knight, Markus Herrgard, Jennifer Reed, for the joy it has been collaborating and interacting with them. Other great friends and colleagues from the SBRG include Anuradha Raghunathan, Jason Papin, Nathan Price, Sharon 7. Wiback, Tim Allen, Susan Toyama, Steve Fong, Derren Barken, Thuy Vo, Natalie Duarte, Marc Abrams, Ryan Drogo, Andrew Joyce, Tom Fahland, and Jochen Foerster (honorary). They have all shaped my work and my life and I am grateful for it.

Other extremely helpful and friendly people are Kurt Stadsklev and Adam Fleming, who helped to code some of the more difficult calculations, Tim Durfee,

Chris Herring and Jeremy Glasner, who were very helpful with obtaining the high-throughput phenotyping data, as well as Zhongge Zhang and Guillermo Gosset who helped with knockout strain construction.

Finally, I never even would have started this work without years of encouragement and support from my friends and family. To them, and most particularly to my grandparents, Maxine and Willard Covert and Anneliese and Heinrich Schwabe, to my parents-in-law Constance and David Bennion, to my parents Ludwiga and Michael Covert, my siblings Jeff, Jen, Mathias, and Michelle, and especially to my wife Sarah, I owe everything.

The text of Chapter 2, in part or in full, is a reprint of the material as it appears in “Metabolic Modeling of Microbial Strains *In Silico*” in *Trends in the Biochemical Sciences*, the text of Chapter 3 and Appendix A, in part or in full, is a reprint of the material as it appears in “Constraints-based models: regulation of gene expression reduces the steady-state solution space” and “Regulation of gene expression in flux balance models of metabolism”, both in the *Journal of Theoretical Biology*, the text of Chapter 4 and Appendix B, in part or in full, is a reprint of the material as it appears in “Transcriptional regulation in constraint-based metabolic models of *Escherichia coli*” in *Journal of Biological Chemistry*, the text of Chapter 5 and Appendix C, in part or in full, is a reprint of the material as it appears in “Integrating high-throughput data and computational models leads to *E. coli* network elucidation”, which has been submitted for publication, and the text of Chapter 1, in part or in full, is a reprint of the material as it appears in all of the above listed articles. I was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for this chapter.

VITA

1997	B.S., Brigham Young University
2002	M.S., University of California, San Diego
2003	Doctor of Philosophy University of California, San Diego

PUBLICATIONS

“Genomic Engineering of Bacterial Metabolism”, J.S. Edwards, C.H. Schilling, M.W. Covert, S.J. Smith (Wiback), B.O. Palsson. *Encyclopedia of Microbiology*, Academic Press, New York (2000).

“Metabolic Modeling of Microbial Strains *in silico*”, M.W. Covert, C.H. Schilling, I. Famili, J.S. Edwards, I.I. Goryanin, E. Selkov, B.O. Palsson, *Trends in Biochemical Sciences*, 26(3): pp. 179–186 (2001).

“Regulation of Gene Expression in Flux Balance Models of Metabolism”, M.W. Covert, C.H. Schilling, B.O. Palsson, *Journal of Theoretical Biology*, 213(1): pp. 73–88 (2001).

“Metabolic Modeling of Microbes: the Flux-Balance Approach”, J.S. Edwards, M.W. Covert, B.O. Palsson, *Environmental Microbiology*, 4(3):133–140 (2002).

“Genome-Scale Metabolic Model of *Helicobacter pylori* 26695”, C.H. Schilling, M.W. Covert, I. Famili, G.M. Church, J.S. Edwards, B.O. Palsson, *Journal of Bacteriology*, 184(16): 4582–4593 (2002).

“Transcriptional Regulation in Constraints-Based Metabolic Models of *Escherichia coli*”, M.W. Covert, B.O. Palsson, *Journal of Biological Chemistry*, 277(31): 28058–28064 (2002).

“A More Palatable *H. pylori*: Iterative Model-Building through the Peer-Review Process”, M.W. Covert, C.H. Schilling, B.O. Palsson, *ASM News*, 68(11): 529 (2002).

“Constraints-Based Models: Regulation of Gene Expression Reduces the Steady State Solution Space”, M.W. Covert, B.O. Palsson, *Journal of Theoretical Biology*, 221(3): 309–325 (2003).

“Reconciling Gene Expression Data with Known Genome-Scale Regulatory Network Structures”, M.J. Herrgard, M.W. Covert, B.O. Palsson, *Genome Research*, in press.

“Identifying the Constraints that Govern Cell Behavior”, M.W. Covert, I. Famili, B.O. Palsson, *Biotechnology and Bioengineering*, in press.

“Reconstruction of Microbial Transcriptional Regulatory Networks”, M.J. Herrgard, M.W. Covert, B.O. Palsson, *Current Opinion in Biotechnology*, in press.

“Integrating High-Throughput Data and Computational Models Leads to *E. coli* Network Elucidation”, M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, B.O. Palsson, submitted.

ABSTRACT OF THE DISSERTATION

Transcriptional Regulation in Constraint-Based Models of Metabolism

by

Markus W. Covert

Doctor of Philosophy in Bioengineering,
with Specialization in Bioinformatics

University of California, San Diego, 2003

Professor Bernhard O. Palsson, Chair

High-throughput technologies are yielding large data sets that require network-based data analysis to reconcile heterogeneous data types, find inconsistencies, and systematically generate hypotheses. To begin this process in *Escherichia coli*, a genome-scale model of its metabolic and transcriptional regulatory networks was reconstructed. Although genome-scale constraints-based models of *Escherichia coli* metabolism have already been constructed and used to successfully interpret and predict cellular behavior under a range of conditions, such models do not account for regulation of gene transcription and thus cannot accurately predict some organism functions. I therefore expanded the constraint-based approach to modeling metabolic systems to incorporate transcriptional regulatory events in terms of time-dependent constraints on the metabolic network. The effects of regulatory constraints on metabolic behavior were studied using extreme pathway and flux-balance analysis. I demonstrate that the imposition of environmental conditions and regulatory mechanisms sharply reduces the number of active extreme pathways, and that incorporation of transcriptional regulatory events in flux-balance analysis enables interpretation, analysis and prediction of the effects of transcriptional regulation on cellular metabolism at the systemic level. The genome-scale metabolic and regulatory model accounts for 1,010 genes and was used to computationally predict growth phenotypes of 110 knockout strains under 125 growth

conditions (13,750 cases). The computations were consistent with experimental measurement in 10,828 (79%) cases, and resolution of discrepancies between prediction and observation led to identification of 18 areas where the metabolic or regulatory networks are incompletely characterized. To begin further characterization of the regulatory network, I also mRNA expression profiled wild-type and 6 knockout strains under aerobic and anaerobic conditions. Altered expression of 151 genes represented in the model were detected and 22 of these changes were due entirely to regulation that had been previously described. Model-driven analysis of the remaining cases led to the formulation of 110 new regulatory rules that represent testable hypotheses. Overall, I find that a systems biology approach that combines genome-scale experimentation and computation can systematically generate hypotheses from disparate data sources.

Chapter 1

The Challenge of Systems Biology

A principal director of the Human Genome Project, Eric Lander proved that brevity is the soul of wit last October at the 13th Ig Nobel Award Ceremony. The notoriously antic Ig Nobel awards are held each year to recognize “research that can not or should not be reproduced” [127]. Tongues firmly in cheek, the self-appointed Ig Nobel panel honored work related to the underlying preferences of pigeons for certain statues, the enlarged hippocampi of London’s taxi drivers and the physical forces which make it difficult to drag sheep.

Lander was invited to give the infamous “24/7 nanolecture,” in which one describes one’s work twice: first in twenty-four seconds and after in only seven words. His remarks are reproduced here in their entirety, beginning with the twenty-four second portion:

The genome cost US\$3 billion and gave us three billion letters: one dollar a letter. Such a deal!

And then, in seven words:

Genome. Bought the book. Hard to read.

These seven words introduce a major conceptual challenge in biology. The human genome sequence cost billions of dollars and took thirteen years between conception and completion. Sometimes called the “Book of Life”, the sequence is thought to have ushered in a new (often called “post-genomic”) era in terms of

biological discovery. But how will we understand this book, now standing tantalizingly open before us?

1.A From reductionism to systems

The complete description of the nucleotide sequence of our own DNA may be seen as the culmination of reductionism in biology. John Holland described reductionism in these words:

The idea is that you could understand the world, all of nature, by examining smaller and smaller pieces of it. When assembled, the small pieces would explain the whole [11].

This reductionist approach has enabled incredible progress in biology, to the point where many high-throughput experimental technologies have been developed in recent years that enable us to obtain full genomic sequences, perform genome-wide expression assessment and probe the protein portfolio of particular cells and organisms. These experimental technologies will only increase in speed and potential in the coming years.

These technologies, together with the massive amounts of data they produce, has led to a profound realization in biology. Once again, I turn to Eric Lander for a smart description:

We've called the human genome the blueprint, the Holy Grail, all sorts of things. It's a parts list. If I gave you the parts list for a Boeing 777 and it has 100,000 parts, I don't think you could screw it together, and you certainly wouldn't understand why it flew. [125].

In other words, with the parts list available to us, the new challenge is to go from integrate the knowledge we have about each part in such a way that we have more understanding of the whole. Using Lander's plane metaphor (and presumably with Han Solo's full approval), we must make a "bucket of bolts" fly.

This is non-trivial, because a system is more than simply the sum of its components. For example, let's say we had cut up a rubber ball into a thousand

pieces. We could study each piece in detail, recording elasticity, color, shape, melting temperature, etc. However, unless we attempted to put all of the pieces together, we would never know that the whole was intended to bounce [209]. Analogously, we need to find a way to integrate what we know about biological systems to identify interactions and emergent, or systemic, properties.

1.B The need for mathematical tools

The complex composition of a biological system requires the use of computational tools to describe its integrated function. As a result, more biologists are turning to engineers, physicists and mathematicians, who in turn are scrambling to learn biological fundamentals. Such cross-disciplinary fertilization has led to seminal studies elucidating regulation of galactose utilization in yeast [80], control of the I κ B-NF- κ B signaling module [76], dynamics of the epidermal growth factor receptor (EGFR) system [210], and mammalian cell cycle regulation [149], among others. Taken together, these developments signal an important shift in biology from conceptual modeling to computational modeling. Conceptual models describe a system in qualitative terms, whereas computational models can quantitatively simulate systemic properties to analyze, interpret and predict cell behavior. Thanks to the high-throughput generation of "omics" data [10], biologists find themselves well positioned to reconstruct fairly complicated conceptual models of metabolic, regulatory and signaling networks (in order of difficulty) [41, 42, 92, 162], culminating in the development of databases such as KEGG and MetaCyc [89, 91].

One current challenge in systems biology is to translate these conceptual models into genome-scale computational models. In the following chapters I will describe a model-driven, system-based approach that begins to answer this challenge. The particular focus of my research is in using well-designed experiments together with a mathematical framework to facilitate biological discovery. Specifically, I will describe how a genome-scale mathematical model of *Escherichia*

coli metabolism and the corresponding transcriptional regulation was constructed and used together with high-throughput data to generate hypotheses which could greatly facilitate the characterization of this organism.

1.C Chapter overview

Chapter 2 describes the large amounts of genome-scale data that are being produced and made available in databases on the World Wide Web, which are demanding the development of integrated mathematical models of cellular processes. The analysis of reconstructed metabolic networks as systems leads to the development of an *in silico* or computer representation of collections of cellular metabolic constituents, their interactions, and their integrated function as a whole. I will describe how quantitative analysis methods, such as are used in constraint-based models have been effectively used to analyze, interpret and predict the function of reconstructed genome-scale metabolic models.

Chapter 3 describes the theoretical underpinnings of this text; namely, how the effects of transcriptional regulation may be integrated with metabolic network analysis using the constraint-based approach. This is shown using flux-balance and extreme pathway analysis to analyze a simplified metabolic/regulatory network. Specifically, I will represent transcriptional regulatory events as time-dependent constraints on the capabilities of a reconstructed metabolic network to further constrain the possible function of a reconstructed metabolic network.

First, we will examine the reduction of the solution space due to regulatory constraints using extreme pathway analysis, and find that the imposition of environmental conditions and regulatory mechanisms sharply reduces the number of active extreme pathways. As a result, the method developed here provides a way to interpret how regulatory mechanisms are used to constrain network functions and produce a small range of physiologically meaningful behaviors from all allowable network functions.

Next, flux-balance analysis has been used successfully to predict time course of growth and by-product secretion, effects of mutation and knock-outs, and gene expression profiles. However, I will show that flux-balance analysis leads to incorrect predictions in situations where dominant regulatory effects influence the function of the organism. Thus, there is a need to include regulatory events within FBA to broaden its scope and predictive capabilities. Using a simplified metabolic/regulatory network, growth is simulated under various conditions to illustrate systemic effects such as catabolite repression, the aerobic/anaerobic diauxic shift and amino acid biosynthesis pathway repression. We will see that incorporation of transcriptional regulatory events in FBA enables us to interpret, analyze and predict the effects of transcriptional regulation on cellular metabolism at the systemic level.

Chapter 4 applies the theoretical framework described in Chapter 3 to *Escherichia coli*, beginning with a model of core metabolism and associated regulation. Genome-scale constraints-based models of *Escherichia coli* metabolism have been constructed and used to successfully interpret and predict cellular behavior under a range of conditions. These previous models do not account for regulation of gene transcription and thus cannot accurately predict some organism functions. I will present an *in silico* model of central *E. coli* metabolism that accounts for regulation of gene expression. This model accounts for 149 genes, the products of which include 16 regulatory proteins and 73 enzymes. These enzymes catalyze 113 reactions, 45 of which are controlled by transcriptional regulation. The combined metabolic/regulatory model can predict the ability of mutant *E. coli* strains to grow on defined media, as well as time courses of cell growth, substrate uptake, metabolic by-product secretion and qualitative gene expression under various conditions, as indicated by comparison to experimental data under a variety of environmental conditions. The *in silico* model may also be used to interpret dynamic behaviors observed in cell cultures. This combined metabolic/regulatory model is

thus an important step towards the goal of synthesizing genome-scale models that accurately represent *E. coli* behavior.

Chapter 5 demonstrates the power of an integrated experimental/computational approach in characterization of biological systems. High-throughput technologies are yielding large data sets that require network-based data analysis to reconcile heterogeneous data types, find inconsistencies, and systematically generate hypotheses. To begin this process in *Escherichia coli* we reconstructed a genome-scale model of its metabolic and transcriptional regulatory networks. The model was used to computationally predict growth phenotypes of 110 knockout strains under 125 growth conditions (13,750 cases). The computations were consistent with experimental measurement in 10,828 (79%) cases, and resolution of discrepancies between prediction and observation led to identification of 18 areas where the metabolic or regulatory networks are incompletely characterized. To begin further characterization of the regulatory network, we mRNA expression profiled wild-type and 6 knockout strains under aerobic and anaerobic conditions. Altered expression of 151 genes represented in the model were detected and 22 of these changes were due entirely to regulation that had been previously described. Model-driven analysis of the remaining cases led to the formulation of 110 new regulatory rules that represent testable hypotheses. We show that a systems biology approach that combines genome-scale experimentation and computation can systematically generate hypotheses from disparate data sources.

Chapter 6 concludes with a retrospective look at the overall implications of this work, together with a discussion of some future challenges moving forward. I will focus on expanding the constraint-based approach, describing other fundamental constraints in biology, and discussing how can we expand existing genome-scale models to include other biological processes and simulate the behavior of multi-cellular organisms.

The text of this chapter, in part or in full, is a reprint of the material as it appears in “Transcriptional regulation in constraint-based metabolic models of *Escherichia coli*” in *Journal of Biological Chemistry*, “Metabolic Modeling of Microbial Strains *In Silico*” in *Trends in the Biochemical Sciences*, “Integrating high-throughput data and computational models leads to *E. coli* network elucidation”, which has been submitted for publication, and “Constraints-based models: regulation of gene expression reduces the steady-state solution space” and “Regulation of gene expression in flux balance models of metabolism”, both in the *Journal of Theoretical Biology*. I was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for this chapter.

Chapter 2

Metabolic constraint-based models

The first constraint-based models to describe cellular properties at the genome scale were focused on metabolism. In this chapter, I will describe the process of building mathematical models of carbon and energy metabolism for microbial organisms, as shown schematically in Figure 2.1. From the annotated genome sequence and the experimentally determined biochemical and physiological characteristics for a given organism, its network of metabolic reactions can be reconstructed, as far as it is known. The reconstructed metabolic network is then analyzed using various mathematical modeling techniques. These quantitative analysis methods enable the simulation of microbial growth and behavior *in silico* and therefore have many important applications in the field of metabolic engineering, where organisms are genetically modified to enhance desirable properties. Achievements in the mathematical analysis of microbial metabolism as well as issues and challenges in the field are discussed.

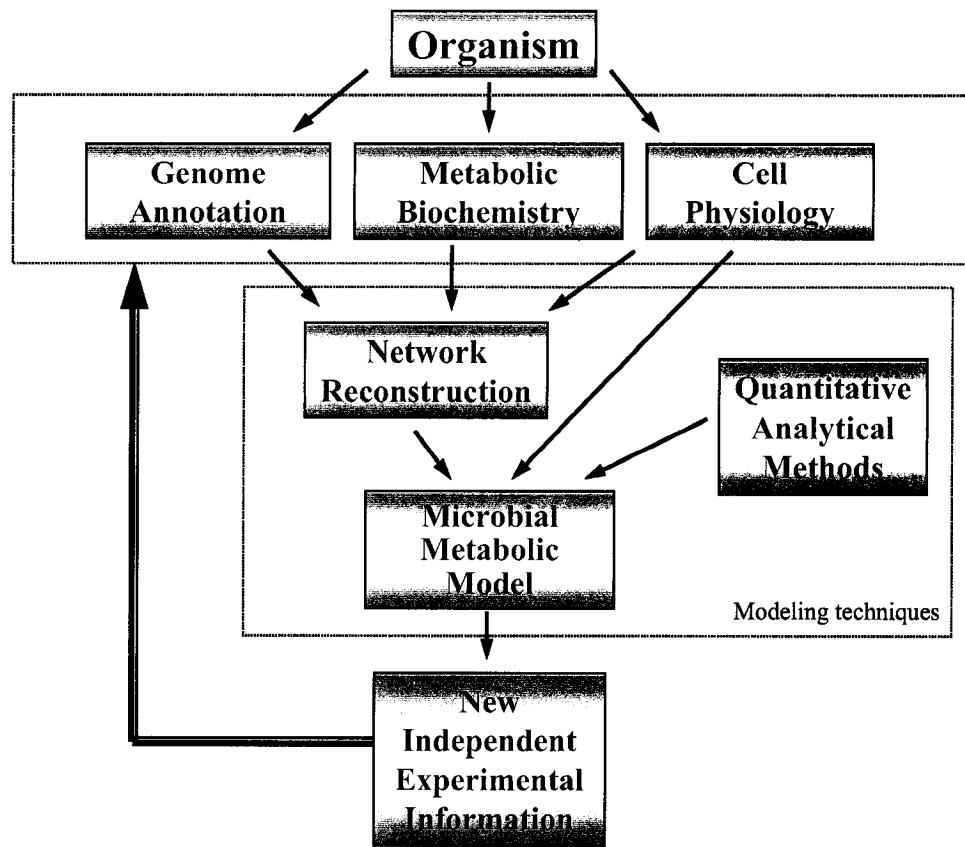


Figure 2.1: Integrated process of microbial metabolic model construction. Construction of a microbial metabolic model requires a comprehensive knowledge of the metabolism of an organism. From the annotated genome sequence and the experimentally determined biochemical and physiological characteristics of a cell, the metabolic reaction network can be reconstructed. This network is then modified in the context of other physiological constraints to produce a mathematical model, which can be used to generate quantitatively testable hypotheses *in silico*. As the model is used to direct an experimental plan, it can be important in further reexamining the biological properties of the organism.

2.A Reconstruction of metabolic reaction networks

2.A.1 Genome annotation

Reconstruction of metabolic reaction networks in an organism begins with the thorough examination of the genome. The first step in functional annotation of a genome sequence is to identify the coding regions or open reading frames (ORFs) on the sequence. Each ORF is searched initially against databases with the goal of assigning a putative function to it. Established algorithms (e.g. the BLAST and FASTA family of programs) can be used to determine the similarity between a given sequence and gene or protein sequences deposited in sequence databases [2, 142]. As the number of sequenced organisms rises, putative gene functions may also be determined by various types of gene clustering [51, 134]. A large fraction of the genes for a newly sequenced organism can usually be readily identified by these methods.

Several high-quality genomic database and metabolic network reconstruction websites that provide access to the annotated genome sequences of many organisms can be found online [91, 92]. Two types of metabolic databases are available from the Web: organism-specific and general-purpose databases. Organism-specific databases, such as EcoCyc [92], are designed to provide a user friendly interface for the inspection of the metabolic characteristics (i.e. experimental and sequence data) of a single genome. The general-purpose databases, such as the Metabolic Pathways Database (MPW) [178, 60] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [16, 131], contain sequence data for a large spectrum of organisms. Once the functional assignment of a sequenced genome is complete, a software program such as the What Is There (WIT) system [134] or KEGG can search the general-purpose database for the closest set of metabolic maps complementing the annotated genome. The metabolic maps in WIT and KEGG are then used as templates whereon the organism's metabolic network reconstruction can be represented as an organism-fitted subset of their pathway collections to be used

similarly to an organism-specific database.

The wealth of biochemistry knowledge contained in general-purpose metabolic databases enables the automated metabolic reconstruction of any sequenced organism, including those for which only a partial or “gapped” genome sequence is available [179]. These reconstructed metabolic networks, based exclusively on genomic data, can form the backbone of an *in silico* organism. However, to build a more comprehensive metabolic network, these automatically constructed metabolic networks must be evaluated in the context of experimental data, specifically the biochemical and growth characteristics of the organism.

2.A.2 Metabolic biochemistry

Genome sequencing and annotation have already outpaced the generation of biochemical and physiological data. The time required to make comprehensive experimental and literature surveys of the biochemical and physiological characteristics for each organism of interest can be excessively long. However, many of the organisms whose genomes have been sequenced completely have also been the subjects of extensive biochemical research. After locating all the known metabolic genes on an annotated genome, the additional information gained from experimental data can make the reconstructed metabolic network more complete. Continued experimental investigation of the metabolic biochemistry of an organism is important and has three main purposes:

1. To assign pertinent biochemical reactions to the enzymes found in the genome;
2. To validate and scrutinize information already found in the genome; and
3. To determine the presence of reactions or pathways not indicated by current genomic data.

The utility of a reconstructed metabolic network depends largely on its accuracy. Biochemical evidence helps to assign a function to a particular gene, and

validates the corresponding links in the reconstructed network. Also, functionality can sometimes be determined more easily by biochemical than by genomic studies. The sequencing of many organisms has shown that 20-30% of all eubacterial genes annotated so far are found to be species specific, having as yet no known homologs [161]. It follows from this observation that various organisms might have evolved widely different methods of catalyzing similar reactions or pathways. The proteins involved in these reactions would have disparate sequences despite their similar function, and would thus be undetectable by sequence comparison. The substrate specificity of many enzymes can also introduce serious errors into the metabolic reconstruction if genomes are annotated by sequence similarity alone. Combining the findings of experimentalists with the information contained in an annotated genome will reconcile these issues and lead to the most complete reconstruction of the metabolic network.

2.A.3 Cell physiology

At the current state of genomic and biochemical knowledge, a number of the metabolic genes that contribute significantly to the metabolic phenotype of an organism can not be identified. The identification of these additional genes depends on the inclusion of cell-physiological data.

Knowledge of the physiology of an organism gives indirect evidence to the presence or absence of certain metabolic reactions in a cell. For example, if experiments suggest that an organism can grow without a certain essential amino acid, but the reconstructed metabolic network is not able to produce that amino acid *in silico*, perhaps for lack of a single enzyme, then for the metabolic reaction network to have any practical meaning, the missing steps in the pathway must be included. Once a network reconstruction has been developed and evaluated in the context of available biochemical and physiological information, it can be applied to various types of mathematical analysis.

2.B Model construction and analysis

Mathematical models and their computer simulation allow us to examine the integrated function of the reconstructed metabolic network. A well defined network by itself is not enough to describe the behavior of a system quantitatively, as shown in Figure 2.2. Here, an analogy is drawn between simulating traffic conditions in a typical city and simulating the behavior of a microbial metabolic network. The first step for both situations is to generate a list of the functional components for the system. In the case of traffic simulation, this could be represented by a list of all the major roads in the city, as well as the destinations that are connected by these roads. For a cell, gene products are discovered and characterized as described earlier. The next step is to determine how these functional components are connected. This information can be integrated into a “map”, either a road map or a reconstructed metabolic map. Once a network has been described in sufficient detail, some qualitative predictions can be made. For example, a road map is used to determine whether it is possible to go from one place to another. Relative distances can be compared. Similarly, the reconstructed metabolic network can be used to study connectivity of metabolites and the like [46, 54, 85].

However, the completed road map has limitations. For example, although the possibility of driving from one destination to another can be ascertained, the actual travel time is unknown. The travel time depends on traffic conditions, which depend on the road, the time of day, the weather and a number of other contributing factors. Obtaining all the necessary data to specify each contributing factor is unfeasible. However, by estimating or approximating many of these conditions, thereby creating a reasonable model of traffic conditions, it is possible to obtain a reasonable calculation of the travel time. Such a calculation could not be made with a road map alone.

Analogously, without including other information, the reconstructed metabolic map of an organism is limited in its ability to generate quantitative predictions

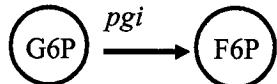
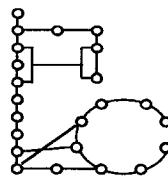
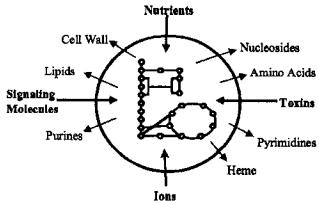
Level of Analysis	Traffic Simulation	Cellular Simulation
1. List of components		
	Isolated roads	Isolated enzymes
2. Integration and qualitative analysis		
	Road map	Metabolic map
3. Mathematical modeling and quantitative analysis		
	Traffic patterns	Flux distributions

Figure 2.2: An analogy between simulation of traffic conditions in a typical city and simulation of a microbial cell using systems analysis. For both simulations, the first step is to generate a list of all the relevant components (e.g. roads or gene products) of the system, after which the integration of these components must be determined and specified. At this point some qualitative predictions can be made about the performance of the system. Finally, mathematical modeling is used to quantitatively analyze the system as it responds to a number of environmental factors or a change in the network.

about the phenotype. The behavior of a cell depends on many factors such as temperature, substrate availability, the presence of signaling molecules, and other environmental parameters, many of which have yet to be specified completely. Properties like stoichiometry are relatively easy to establish, whereas kinetic properties of bacterial metabolism are typically much more difficult to obtain under all possible environmental conditions. Several approaches to dynamic cellular modeling have been developed [187, 71, 52, 155, 200, 120]; however, it is clear that for detailed dynamic model building to succeed on a whole-genome scale, much progress must be made in the ability to estimate kinetic parameters of enzymes from first principles [202] or by correlation to known properties of similar enzymes [177].

Although kinetic constants are hard to obtain, the network structure can be established as outlined above. As should be clear from the traffic analogy given in Figure 2.2, the study of fluxes through such a network is important and can be accomplished without having detailed kinetic information. Metabolic fluxes can be seen as a fundamental determinant of cell physiology because they show quantitatively the contributions of various pathways to overall cellular functions [187]. A common way of relating cell genotype to phenotype is therefore by analyzing the fluxes in the metabolic network.

Methods that fall into this category are pathway analysis [171], flux-balance analysis (FBA) [170, 198, 15] and metabolic flux analysis (MFA) [187]. All are based on the principle of conservation of mass for the metabolites of a given metabolic network. Pathway analysis is a method for generally defining the structure of the metabolic network as it relates to the overall metabolic capabilities [172, 176] of the organism. In contrast, FBA examines the metabolic network from a performance perspective, using linear optimization to determine optimal cellular behaviors under changing environmental and genetic conditions [46]. MFA characterizes the flux distribution in more experimental detail, estimating internal fluxes based on a combination of isotope labeling techniques [207] and mathemat-

ical analysis [24]. Pathway analysis and FBA will be described in more detail in Chapter 3.

The ability of flux-based analytical techniques to generate quantitative hypotheses has broad importance in the field of metabolic engineering, whether in the large-scale microbial generation of valuable substances or in pollutant degradation. For example, these techniques have been used effectively to model penicillin production by *Penicillium chrysogenum* [86, 73], to improve the yield of aromatic amino acids by *E. coli* [108] and lysine by *Corynebacterium glutamicum* [187] through metabolic engineering of central metabolism, and to model enhanced biological phosphorus removal for wastewater treatment [146]. Flux-based approaches also lend themselves easily to genotype-phenotype studies *in silico* and can be used to analyze enzyme deficiencies or identify drug targets as has been shown in gene deletion and metabolite connectivity studies for *E. coli* [48] and *H. influenzae* [46, 173]. Used in conjunction with an experimental study, the metabolic outcome and growth of *E. coli* using acetate and succinate as single-carbon sources has been accurately predicted [48]. Further applications of flux-analysis techniques have been reviewed [170, 187, 65].

Although the goal of developing a completely specified cellular model will require the inclusion of kinetic parameters, the development of flux analysis methods and other approaches has many applications and will continue to lead to the generation of novel and important quantitative hypotheses about microbial behavior, even in the absence of detailed kinetic information.

2.B.1 Model characteristics

A comparison of the genomic characteristics and *in silico* metabolic model characteristics for three bacterial strains is shown in Table 2 [46, 48, 174]. These *in silico* models represent between 25% and 40% of the known ORFs in their *in vivo* counterparts. Figure 2.3a shows the reaction complement of the gastric pathogen *Helicobacter pylori* 26695 in greater detail. The Venn diagram is used

to categorize the inclusion of reactions in the reconstructed network. Many of the reactions have been included with a combination of different kinds of evidence, as shown by the overlap in circles. In better-known organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, the overlap between the circles is expected to be much greater.

As shown in the figure, drawn approximately to scale, the bulk of reactions present in the network can be derived from genomic evidence (almost 73% in the case of *H. pylori*). Approximately half of the remaining reactions included in the reconstructed metabolic network were included based on observations found in the literature, whether direct biochemical evidence or indirect physiological evidence, as has been explained. The remainder of the reactions, labeled as “inferred reactions” in Figure 2.3a, have been included without experimental or genome evidence, but are inferences based on the metabolic demands of the reconstructed network. Each inferred reaction added to the reconstructed metabolic network will eventually require further experimental justification.

2.B.2 Modeling issues

Regarding the construction and analysis of microbial metabolic models, the primary issues relating to construction are that first, not all of the reactions suggested by these models are found directly in the databases or the biochemical literature; and second, not all of the metabolic genes actually present in the genotype are accounted for or even noted in the model, because their functions are as yet undiscovered (see Figure 2.3b).

For the reconstructed metabolic network (see Figure 2.3a), a “real metabolic network”, (i.e. the actual set of all the relevant reactions that occur in *H. pylori* strain 26695) exists. This network, surrounded by a dashed line, is superimposed on the network defined by our model. The lighter area is the set of all reactions that are found both in strain 26695 and in our model, the “correct” reactions. The enclosed area in white represents “false” reactions that were included in the model

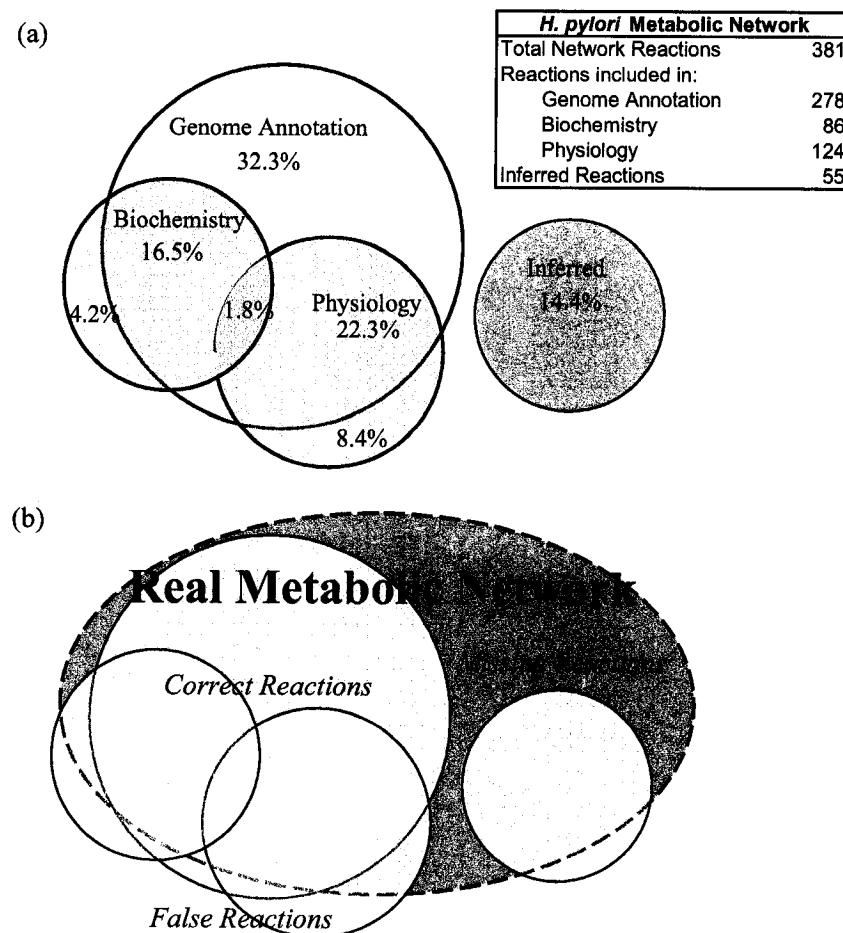


Figure 2.3: Venn diagram of reactions in the *H. pylori* *in silico* metabolic model. (a) Reactions are categorized by the type of evidence used to justify their inclusion in the reconstructed network. The number of reactions, by category, is given in the table (inset) and as a percentage of the total reactions (shown in each colored area). Many of the reactions have been included by evidence from multiple sources, as shown by overlapping circles. (b) This diagram depicts the real metabolic reaction network, illustrating the two major issues associated with metabolic reconstruction. If the actual complete metabolic network is the area enclosed by a dashed line, the reconstructed model is probably both missing some crucial reactions (dark blue area) and including some reactions which are not actually occurring in the cell (enclosed white area).

but do not actually occur in *H. pylori* strain 26695. These reactions represent mistaken assumptions used in creating the model.

The second issue is the inverse problem: many of the proteins synthesized by the organism are not accounted for in the metabolic reconstruction. These “missing reactions” are shown by the darker area in Figure 2.3b. It is likely that some of the metabolic reactions that are catalyzed by the organism are as yet undiscovered. This implies that functionalities open to the organism are neglected by the model.

These metabolic network reconstruction issues can be resolved in part as the model is applied to various analyses. For example, the metabolic *H. pylori* model was used to reexamine the annotation of the metabolic network. All of the genes that were included in the reconstruction of *H. pylori* metabolism without direct genomic or biochemical evidence can be thought of as hypothetical. The presence of these hypothetical genes can be determined by collecting the sequences of other organisms’ copies of the hypothetical genes and using BLAST to compare them with the *H. pylori* genome sequence. The genes that are found to be significantly homologous to loci in the *H. pylori* genome sequence can then be studied experimentally to verify their proposed function based on the reconstruction and BLAST analysis.

One such gene product included in the *H. pylori* model without genomic or biochemical evidence was malate dehydrogenase. A subsequent study indicated that on locus HP0086 of the *H. pylori* genome, an open reading frame was located that showed significant similarity (36.81%) and identity (25.93%) with a malate:quinone oxidoreductase in glutamic acid bacterium *C. glutamicum* [93]. Thus, the analysis of microbial metabolic models can also have bioinformatic applications, such as functional assignment of ORFs, in addition to the more obvious experimental applications.

There are also significant issues pertaining to the analysis of microbial metabolic models. It has been noted above that flux models can successfully

predict the effects of gene knockouts and the metabolic behavior of an organism quantitatively. The specific advantage of the flux-based analyses is that such models do not require experimental information such as enzyme kinetics, regulatory mechanisms, intracellular concentrations, or enzyme activity profiles.

However, the attractive simplicity of the models also sets some inevitable limits for their predictions. Simulations of microbial behavior reflect only the topology of reconstructed systems (as contained in the reconstructed network) and the boundary conditions of the system (such as extracellular substrate concentrations). Flux-based models currently incorporate no control mechanisms of any kind, predicting a theoretical metabolic potential which assumes the constitutive expression of all genes in the metabolic reaction network. This assumption could lead to false predictions. Additionally, the flux models describe a stable stationary state of metabolism. Any projections based on their analysis toward intracellular metabolic dynamics associated with the cell cycle or cell differentiation must be made very cautiously. It is expected that these issues will be resolved as analysis methods are improved and developed to incorporate additional experimental information as listed above, resulting in models that are more complicated but perhaps more accurate in predicting the dynamic behavior of microbial organisms.

2.C Future challenges

In silico models of metabolic networks will be subjected to an ongoing iterative model building process just as complex systems in other branches of science and engineering have in the past. This process is illustrated in Figure 2.4. Here, the traditional scientific method is depicted in the context of biology in the post genomic era. Hypotheses based on the metabolic analysis of microbial strains are examined both in terms of an experimental study and using bioinformatics techniques. Experimentalists and bioinformaticists must work cooperatively to provide information to analysts, from which *in silico* representations of microbial

metabolism can be created. The analysis of these models will lead to suggestions for bioinformatic and experimental studies, which in turn will contribute to a more robust characterization of the metabolism of an organism. Once refined in this manner, the metabolic model can be used to generate a new set of hypotheses in a subsequent iteration.

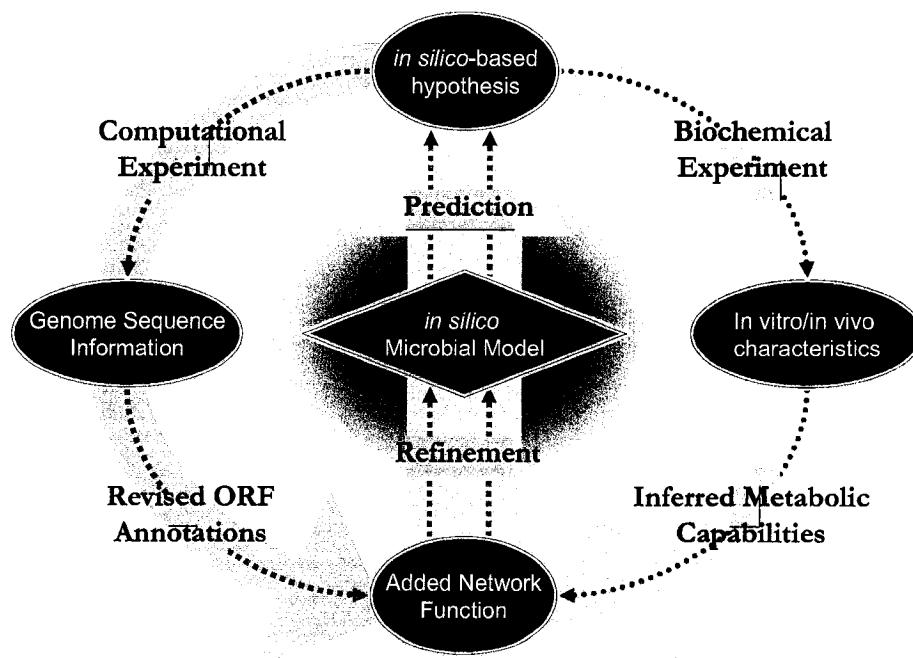


Figure 2.4: Applying the Scientific Method in the post genomic era. Once the *in silico* microbial metabolic model has been constructed, it can be used to generate testable hypotheses. These hypotheses are examined, both in traditional experimental studies (right side) and using new bioinformatics/genome sequencing techniques (left side), to discover new attributes of the metabolic network. After these discoveries have been incorporated into the *in silico* metabolic model, it can be used to generate new hypotheses in a subsequent iteration.

Concurrently, these models need to be expanded to incorporate more features of the genome than simply metabolism. This broadening in scope will occur

as a direct result of more advanced analysis methods. The further development of *in silico* microbial models that quantitatively simulate complexities such as signal transduction, control mechanisms and the dynamic behavior of microorganisms will be vitally important in the field of metabolic engineering as well as in the effort to model eukaryotic organisms, for which the genome is larger and therefore the percentage of known metabolic genes in the genome is generally far smaller. As the efficacy of analysis increases, the new *in silico* microbial models will add various functions until finally the construction of a whole cell model has been completed, an accomplishment which would greatly contribute to our understanding of the cell's essential nature. I will demonstrate this assertion in the next chapters, using transcriptional regulation as an example.

The text of this chapter, in part or in full, is a reprint of the material as it appears in “Metabolic Modeling of Microbial Strains *In Silico*” in *Trends in the Biochemical Sciences*. I was the primary author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 3

The Effects of Transcriptional Regulation

Until now, flux-based simulations have assumed that all gene products in the metabolic reaction network are available to contribute to an optimal solution, unconstrained by regulatory processes. An *in silico* model of *Escherichia coli*, for example, accounts for 660 metabolic genes thought to be in the genome [48]. However, it has been estimated that about 400 regulatory genes exist in the genome of *E. coli* [191]; of these, 178 regulatory and putative regulatory genes were found during annotation of the K-12 MG 1655 genome [12]. Currently 539 transcription units (sets of contiguously located genes with a common expression condition, promoter and terminator) are identified in RegulonDB, a database of transcriptional regulation and operon organization for *E. coli* [162]. The high level of transcriptional regulation in this and other organisms has a significant effect on cell behavior. These regulatory effects have not been accounted for in previous FBA models of *E. coli*, which leads to certain incorrect predictions of cellular-level behavior [48]. For FBA to be effectively used to predict cell behavior on a more general scale, the effects of regulatory events must be incorporated.

The ability to model transcriptional regulatory events has several important applications. As high throughput technologies such as gene chips and

microarrays have been developed to perform genome-wide expression assessment [44], there is a need to predict intracellular transcription and transcriptional regulation on a whole-genome scale based on cellular environment and intracellular conditions. Regulatory modeling is also beneficial to the field of metabolic engineering. Using metabolic control analysis, it has been demonstrated that control of biosynthetic fluxes depends on multiple enzymes [53, 120]. The engineering of cell regulation, rather than of the expression of several related genes, may be a more efficient use of an organism’s metabolism to produce a desired product. Experiments have already been described wherein a regulatory, rather than a metabolic network was manipulated to increase the flux through a particular pathway [133]. The results of such experiments may lead to significant advances in the large-scale microbial generation of valuable substances such as pharmaceuticals [23] and biocommodities [113, 4, 57] or in pollutant degradation [144].

In this chapter, we describe regulatory events and gene expression using Boolean logic equations, and show how this representation affects our metabolic network analysis, using both extreme pathway analysis and flux balance analysis to study a simplified metabolic/regulatory network. We begin with a primer on the constraint-based approach and an overview of extreme pathway analysis and dflux-balance analysis. Next, we show how regulatory events may be interpreted as further constraints on the behavior of the network.

For a given environment, we determine the corresponding regulatory constraints (e.g., repression of gene transcription) and eliminate extreme pathways that are inconsistent with the imposed regulatory constraints. This procedure reduces the solution-space and customizes it for the given environmental conditions. Furthermore, the development of a transcriptional regulatory structure may be used together with FBA to generate time profiles as well as steady state solutions of cell growth, substrate utilization and by-product secretion for organisms for which the metabolism and regulation have been characterized, either from the genome or experimentally. As we will see, the inclusion of known regulatory mechanisms ef-

fectively moves us towards the formulation of second-generation constraints-based models of complex biochemical reaction networks: models that combine metabolic flux-balance formalism and regulation of gene expression.

3.A Metabolic constraint-based analysis

The constraints-based approach to metabolic network analysis has been recently described [136] and is illustrated in Figure 3.1. The axes represent fluxes through all individual reactions in the metabolic network. Not all the points in this space are attainable because of constraints on the system, such as the interrelatedness of the fluxes, thermodynamics or maximum capacity. By imposing these constraints, one can restrict the behaviors available to the cell to an enclosed solution space (shown on the right hand side of Figure 3.1) whose edges may be thought of as extreme pathways, and in which an optimal solution (gray circle) may be determined using flux-balance analysis [65, 170]. Extreme pathway analysis and flux-balance analysis are fundamental enough to warrant a more thorough treatment here.

3.A.1 Extreme pathway analysis

As we mentioned in the last chapter, pathway analysis is a way to describe the overall metabolic capabilities of an organism. In general, mass balances are written around every metabolite concentration in a metabolic network; the resulting equation is

$$\mathbf{S}\mathbf{v} = \frac{d\mathbf{X}}{dt}, \quad (3.1)$$

where $\mathbf{S}_{m \times n}$ (m metabolites \times n fluxes) is the stoichiometric matrix for the network and $\mathbf{v}_{n \times 1}$ is a vector of the flux levels through each reaction in the system, and $\mathbf{X}_{m \times 1}$ is a vector of metabolite concentrations. Because the time constants which describe metabolic transients are fast (on the order of milli-seconds

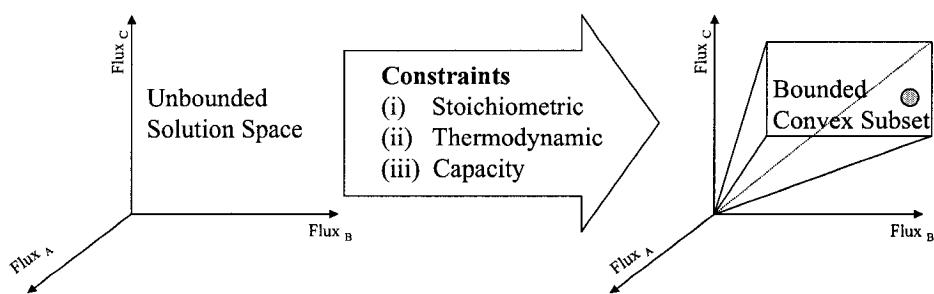


Figure 3.1: Constraint-based analysis of metabolic reaction networks. The metabolic network of an organism may be described in geometrical terms. On the left is an unbounded space containing every possible distribution of fluxes through each reaction in the network. Many of the points in the space are unattainable due to constraints on the system. By applying these constraints, such as system stoichiometry, thermodynamics (e.g. the reversibility of reactions) and maximum capacity, the set of all possible behaviors is reduced to a bounded convex subset (right) in which the solution (gray circle) must lie. These flux distributions represent phenotypes which may be exhibited by a reconstructed metabolic network.

to tens of seconds [120]) as compared to the time constants associated with cell growth, on the order of hours to days, the system may be treated by considering the behavior inside the cell to be in a quasi-steady state, and Equation 3.1 reduces to:

$$\mathbf{S}\mathbf{v} = \mathbf{0}, \quad (3.2)$$

Certain constraints on the system, such as the thermodynamics of reactions or the constraints associated with enzyme capacity (e.g. maximum metabolite uptake and/or secretion rates) may be represented as upper and lower bounds on reaction flux levels. If reaction fluxes are also constrained to positive values by decomposing reversible reactions in \mathbf{S} into forward and reverse components, a solution space may be geometrically defined for the system as a convex polyhedral cone in n -dimensional space [172]. Such a space contains every possible steady state flux distribution available to the system, subject to the given constraints.

A minimal set of generating vectors may be determined for the convex polyhedral cone using principles of convex analysis. This set is approximately analogous to a basis in linear algebra. The generating vectors span the null space of \mathbf{S} and are the edges of the cone [169]. Given that the cone represents the metabolic flux solution space at steady state, each generating vector or extreme ray corresponds to a particular pathway or active set of fluxes in the metabolic network and is termed an extreme pathway. Every possible steady state flux distribution of a metabolic network may therefore be represented as a positive combination of extreme pathways:

$$C = \left\{ v : v = \sum_{i=1}^k \alpha_i p_i, \quad \alpha_i \geq 0, \quad \forall i \right\} \quad (3.3)$$

where C is the polyhedral cone representing the metabolic network at steady state and p_i represents the extreme pathway vectors [171]. The algorithm used to generate extreme pathways has been described in detail [172, 173].

3.A.2 Flux-balance analysis

FBA can be used to quantitatively predict the time profiles of cell density as well as substrate and by-product concentrations [198]. We begin with Equation 3.2, the steady state equation which contains the mass balance constraints imposed on the system. As mentioned above, thermodynamic and capacity constraints or experimental data can be used as limits on the magnitude of each individual metabolic flux in the network:

$$a_i \leq v_i \leq b_i \quad (3.4)$$

Using Equation 3.4, a_i or b_i may be set to zero or to a finite value to constrain the direction or magnitude of a flux. Such constraints correspond to thermodynamic constraints (e.g. effective irreversibility of a given reaction due to an extremely high equilibrium constant) or capacity constraints (e.g., maximum uptake rate for a given transport protein), respectively. Experimental data on flux levels, as obtained by isotope labeling [207] or other methods, can also be used to set a_i and b_i (metabolic flux analysis, as mentioned in Chapter 2). These equality and inequality constraints define a closed solution space.

3.B Regulatory constraint-based analysis

As already stated, the incorporation of transcriptional regulation with our metabolic models could cause significant changes in behavioral predictions. Here I will describe how regulatory networks are reconstructed, analyzed, and used to simulate regulatory events. I will then show how metabolic and regulatory networks may be integrated using constraint-based analysis.

3.B.1 Regulatory network reconstruction

In a transcriptional regulatory network, transcription factors (or regulatory proteins) bind to operator regions on the chromosome and induce or repress

transcription of transcription units, such as genes and operons (for a more detailed description of microbial transcriptional regulation, see [204]). To reconstruct a regulatory network, we must identify and characterize the components (transcription factors and transcription units) and their interactions (via binding sites)[75].

For decades, biologists have set about the task of regulatory network reconstruction using a “bottom-up”, or component-based approach. The bottom-up approaches focus on a single transcription unit and attempt to characterize the effect of transcription factors on expression of that unit. Some common examples of this are construction of a plasmid which contains the promoter/operator region of the transcription unit to a gene encoding a reporter protein such as beta-galactosidase or green fluorescent protein. Once in the cellular host, activity of the reporter is measured under different environmental and genomic conditions [158, 17]. By inactivating (via mutation, for example) a given transcription factor one may determine if the presence of the factor influences the expression of the unit. Northern blots or more recently, quantitative real-time RT-PCR may also be used to measure gene expression more directly [56, 87]. Chromatin immunoprecipitation may be used to interrogate the DNA preceding the transcription unit to determine transcription factor binding sites [102], and the electrophoretic gel mobility-shift assay can be used to determine if a particular transcription factor binds to a particular piece of DNA [19]. A combination of these approaches has led to characterization of most of the regulatory network we know in *E. coli*, as collected in databases such as RegulonDB and EcoCyc [162, 92].

Full genome sequences and the development of high-throughput techniques such as gene expression arrays [45] and genome-wide location analysis [156] have opened the possibility of a top-down, or system-based, approach to regulatory network reconstruction. Although the short length of transcription factor binding sites has made sequence similarity-based approaches to network identification challenging, progress has been made in computationally identifying operons and regulatory sites [192, 126]. Similarly, computational tools have been developed

in an effort to “back-calculate” network structure from gene expression profiles [75] and genome-wide location analysis [8]. Most recently, two efforts have been made to reconcile top-down and bottom-up reconstructions [69, 74]. Such studies show some consistency between the two reconstructions but also highlight the need for better approaches to network elucidation.

3.B.2 Regulatory network analysis

Just as with metabolic networks and road maps (see Chapter 2), the structure of a regulatory network is of great interest, and some interesting topological analysis of regulatory networks has already been performed, including the determination of network connectivity [191], and network motifs [181]. For example, seven global regulators in the *E. coli* regulatory network have been shown to directly regulate expression of 51% of *E. coli* genes [117]. Most genes are known to be regulated by only one transcription factor (355 genes), but as many as seven transcription factors have been shown to regulate one gene [191, 114]. Thirty-eight of the known transcription factors in *E. coli* have other transcription factors as target genes; the resulting regulatory cascades are two to four levels deep and have important implications in terms of indirect regulatory effects [114]. The insights generated from structural and topological analyses of regulatory can therefore be an important aid in attempting to simulate such networks.

3.B.3 Simulating transcriptional regulation

Metabolic regulation and cell dynamics have been modeled using several approaches [152]. For example, Boolean logic may be used to examine the various states of a regulatory circuit [193, 94, 183], such as the regulatory networks controlling the lambda phage decision circuit [190] as well as the immune response [95]. Such mathematical descriptions enable the qualitative study of regulatory structure and lead to general analytical insights which can be usefully applied to the analysis of complex metabolic networks, but must be used in connection with

other techniques to make truly quantitative predictions. Mixed-integer linear optimization has also been used to predict optimal regulatory structures for metabolic engineering [70]. Another approach is deterministic [155, 175, 52, 71, 187], using kinetic theory to solve systems of ordinary differential equations as has been done to study *E. coli* growth on glucose and lactose [211]. Kinetic theory may also be combined with Boolean logic in hybrid models of the lambda phage [122], using fuzzy logic as has been done with *E. coli* core metabolism [104] or in conjunction with cybernetic principles [98, 67, 201]. Other approaches to the analysis of genetic regulatory circuits include the use of fractal kinetic theory [165] and stochastic modeling techniques [119, 120, 20, 121].

Detailed deterministic and stochastic models require extensive information, such as temperature, substrate availability, the presence of signaling molecules, and other environmental parameters, many of which have yet to be completely specified. The difficulty inherent in constructing detailed deterministic or stochastic models is therefore that most of the parameters required to develop them are typically very difficult to obtain under all possible environmental conditions, although some progress has been made [202]. For this reason, none of the above models have been able to be adopted on a scale large enough to represent the entire metabolism of an organism, as recently observed [7], and thus the systemic regulatory/metabolic properties of an entire organism have yet to be analyzed.

3.B.4 Regulatory constraints

A new approach is to interpret the effects of transcriptional regulation as constraints. Cells are subject to both invariant (i.e., nonadjustable) and adjustable constraints. The former are physico-chemical in origin and include stoichiometric, capacity and thermodynamic constraints. They can be used to bracket the range of possible behavior as described above. Adjustable constraints are biological in origin, and they can be used to further limit allowable behavior. These constraints will change in a condition-dependent manner.

Regulatory constraints differ from the rigid physico-chemical constraints in two important ways; they are (1) self-imposed, meaning that over time, evolution has selected the development of complex mechanisms to restrict the allowable behaviors of these organisms under various conditions; and (2) time-dependent, in that the state of the external and internal environment at a given time point determines transcriptional activity. As a result of these two features, the effects of transcriptional regulation can be treated as temporary constraints on the metabolic system. These constraints reduce the size of the solution space and change its shape from one environmental condition to another, as shown in Figure 3.2. This figure depicts a solution space and optimal solution (gray circle) for a hypothetical metabolic network. The solution space defined by non-adjustable constraints is shown in Figure 3.2(a). If the flux through a certain reaction is repressed due to transcriptional regulation, then one or more extreme vectors that define the boundaries of the solution space are removed and the volume of the space (i.e. the range of allowable cellular behaviors) is reduced. For example, when the vectors are removed as illustrated in Figure 3.2(b) and Figure 3.2(c) respectively, the solution space is restricted to a smaller space. This restricted space is analogous to a cell with fewer metabolic behavioral possibilities. Note that the optimal solution remains in the subspace shown in Figure 3.2(b) but not in the subspace shown in Figure 3.2(c). If the optimal solution is no longer in the space, the phenotype that it corresponds to cannot be expressed and a new optimal solution will be determined from the given objective, corresponding to a different behavior exhibited by the cell.

3.B.5 Boolean representation of regulatory events

The transcriptional regulatory structure can be described using Boolean logic equations. This approach involves restricting expression of a transcription unit to the value 1 if the transcription unit is transcribed and 0 if it is not. Similarly, the presence of an enzyme or regulatory protein, or the presence of certain

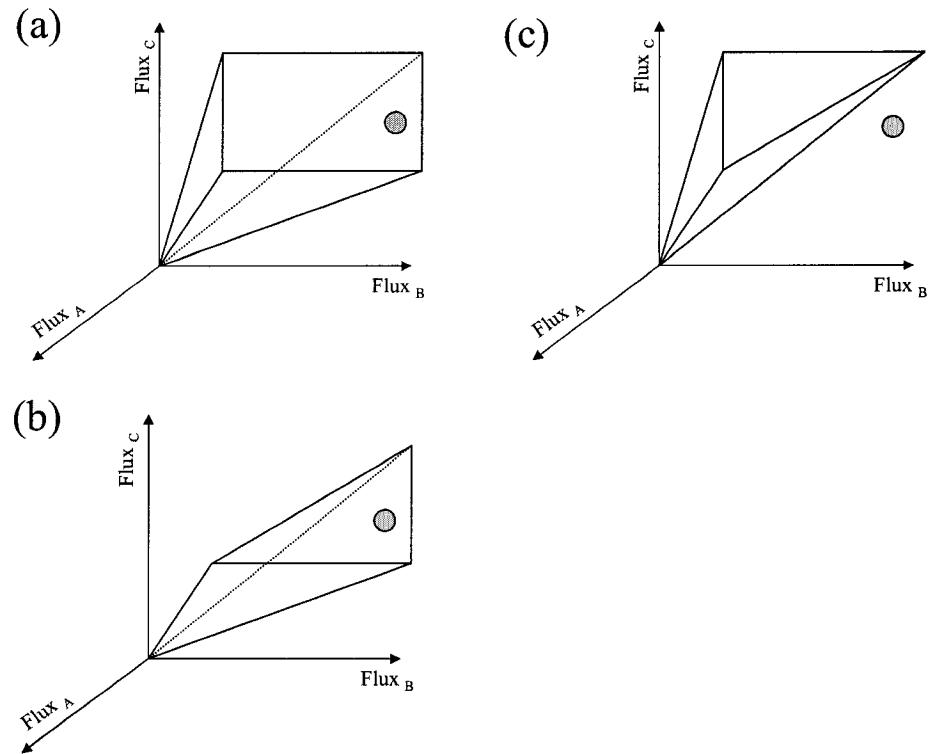


Figure 3.2: Regulatory constraints change the shape of the solution space. A hypothetical solution space and solution, defined by various non-adjustable constraints, is shown in (a). The flux through a certain reaction may be constrained by a transcriptional regulatory event, leading to the removal of one or more extreme pathway vectors from the boundaries of the solution space. These constraints further reduce the size of the solution space. After regulatory constraints have been applied, the original solution may either remain in the smaller solution space (b), or may no longer be located inside the space (c), in which case, a new solution (i.e., a new behavior) will be determined by the cell.

conditions inside or outside of the cell, may be expressed as 1 if the enzyme, protein, or condition is present and 0 if it is not.

The Boolean logic representation includes well-known modifiers such as AND, OR, and NOT, which can be used to develop equations governing the expression of transcription units. Consider a simple system, as depicted in Figure 3.3 (adapted from [193]), containing one gene G which is transcribed by a process *trans*, resulting in an enzyme E. This enzyme then catalyses the reaction *rxn* which is the conversion of substrate A to product B. The product B interacts with a binding site near G such that the transcription process *trans* is inhibited. In other words, the transcription event *trans* will occur if the gene G is present in the genome and the product B is not present to bind to the DNA. A logic equation which describes this circumstance is:

$$\textit{trans} = \text{IF (G) AND NOT (B)} \quad (3.5)$$

After a certain time for protein synthesis has lapsed, progression of the transcription/translation process *trans* will result in significant amounts of enzyme E. Similarly, after a certain protein decay time, the absence of process *trans* will result in decay and eventual depletion of E.

The requirement for the reaction *rxn* to proceed is the presence of A and of E, for which a logical equation can be written:

$$\textit{rxn} = \text{IF (A) AND (E)} \quad (3.6)$$

The presence of enzymes or regulatory proteins in a cell at a given point in time depends both on the previous transcription history of the cell and the rates of protein synthesis and decay. If sufficient time for protein synthesis has elapsed since a transcription event for a particular transcription unit occurred, we say that enzyme E is present in the cell and remains present in the cell until the time for E to decay has elapsed without the cell experiencing another transcription event for that specific transcription unit. In other words, dynamic parameters - the

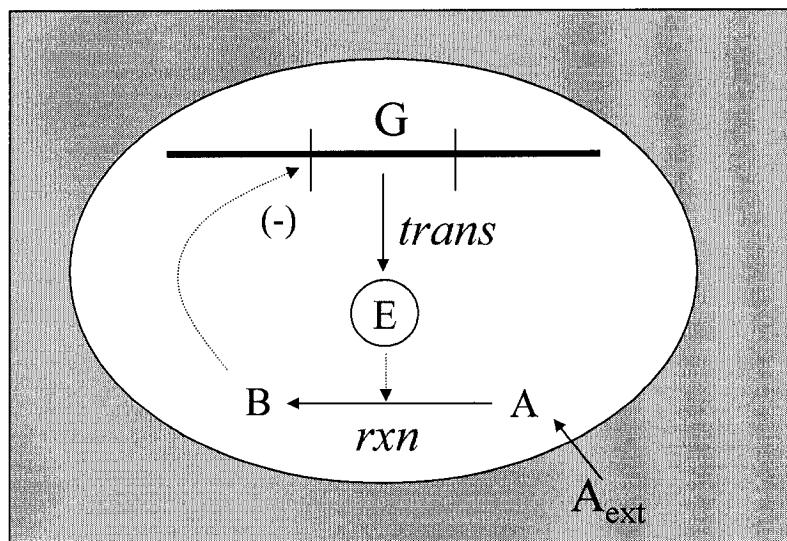


Figure 3.3: A simple regulatory circuit. Here, gene G is transcribed by a process *trans* to produce an enzyme E. This enzyme then catalyzes a reaction *rxn* which converts substrate A into product B. The product B then represses transcription of G, leading to depletion of E.

time delays of protein synthesis and degradation - are required in addition to the known causal relationships that represent regulation of gene transcription. Under steady-state conditions, the average protein synthesis and degradation times are equal.

3.B.6 Integration with metabolic network analysis

Once the presence of all regulated enzymes in the metabolic network has been determined for a given time interval (t_1 to t_2), if an enzyme has been determined “not present” for the time interval, then the flux through that enzyme is set to zero. This restriction may be thought of as adding a temporary constraint on the metabolic network,

$$v_k(t) = 0, \text{ when } t_1 \leq t \leq t_2 \quad (3.7)$$

where v_k is the flux though a reaction at the given time point t . If an enzyme is “present” during a given time interval, the corresponding flux is left unconstrained by regulation and determined instead using FBA. This approach thereby retains the quantitative aspects of FBA while incorporating qualitative regulatory information. Similarly, these regulatory restrictions may be used to determine the feasibility or infeasibility of the extreme pathways in the metabolic network as a function of time. The Boolean formalism may also be used to describe environmental constraints where a substrate is considered either present in (ON) or absent from (OFF) the external medium (OFF). Using these restrictions, an extreme pathway may be determined infeasible (1) if the external nutrient is absent from the external medium; or (2) when expression of the gene responsible for producing a metabolic flux has been repressed. Conversely, a pathway is feasible for a certain condition if it is consistent with all the applicable environmental and regulatory constraints.

Thus, known regulation of gene expression in cells can be represented by Boolean logic and incorporated into constraint-based models. We can determine

the effect of regulatory events on such models by applying this approach to analyze a simple example that represents the skeleton of core metabolism.

3.C Regulated constraint-based analysis of a sample network

A simplified metabolic network, together with the corresponding reaction stoichiometry and regulatory rules, is represented in Figure 3.4. The network contains some 20 reactions, 7 of which are regulated by 4 regulatory proteins. For the purposes of this example, the following instances of transcriptional regulation were examined:

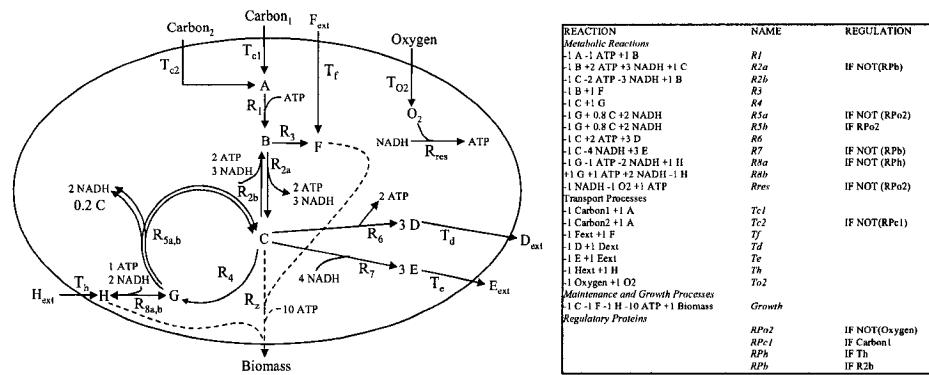


Figure 3.4: A schematic of the simplified core metabolic network, together with a table containing the stoichiometry of the 20 metabolic reactions, 7 of which are regulated by 4 regulatory proteins. This network is a highly simplified representation of core metabolic processes, including a glycolytic pathway with primary substrates Carbon₁ and Carbon₂, as well as a pentose phosphate pathway and a TCA cycle, through which “amino acid” H enters the system. Fermentation pathways as well as amino acid biosynthesis are also represented. The regulation modeled in this study includes simplified versions of catabolite repression (e.g., preferential uptake of C1 over C2), aerobic/anaerobic regulation, amino acid (H) biosynthesis regulation and carbon storage regulation and is also listed. The *Growth* reaction is indicated by a dashed line.

1. Preferential carbon source uptake/catabolite repression [160]. For this example, *Carbon1* is arbitrarily defined as the preferred carbon source. For our purposes, we say that the presence of *Carbon1* in the extracellular medium activates a regulatory protein which inhibits the transcription of *Tc2*, the gene which encodes a protein for transport of *Carbon2* into the cell. Framed in terms of Thomas' Boolean formalism, the resulting equations for this system are:

$$RPc1 = IF(Carbon1) \quad (3.8)$$

$$tTc2 = IF NOT(RPc1) \quad (3.9)$$

where *RPc1* is the regulatory protein which senses extracellular *Carbon1*, *tTc2* is the occurrence of a transcription event (which will eventually result in transport protein *Tc2* and the relaxation of one regulatory constraint, *Tc2 = 0*, on the solution space).

2. Anaerobic growth. The transcription of many enzymes is regulated according to whether or not oxygen is available to the cell [112]. In this case, the presence of *Oxygen* will inactivate regulatory protein *RPO2*, which inhibits transcription of the genes for *Rres* and *R5a* but induces transcription of the gene for *R5b*. Note that *R5a* and *R5b* are isozymes. The logic equations follow this form:

$$RPO2 = IF NOT(Oxygen) \quad (3.10)$$

$$tRres = IF NOT(RPO2) \quad (3.11)$$

$$tR5a = IF NOT(RPO2) \quad (3.12)$$

$$tR5b = IF (RPO2) \quad (3.13)$$

3. Amino acid biosynthesis pathway repression [141]. The transcription of amino acid biosynthesis genes is often induced by a low intracellular concentration of the amino acid. Intracellular concentrations as yet can not be obtained using FBA; instead we use fluxes to approximate the regulation. Metabolite H represents the “amino acid” in this example, and can be made by the cell via flux $R8a$ or transported from the extracellular media through flux Th . For the regulatory structure, flux Th will be used to activate RPh which will repress transcription of $R8a$. In other words, we assume that if flux Th is active (due to the presence of extracellular H), the concentration of amino acid is relatively high and therefore that transcription of H biosynthesis genes will not be induced.

$$RPh = IF (Th > 0) \quad (3.14)$$

$$tR8a = IF NOT (RPh) \quad (3.15)$$

4. This case is the transcriptional regulation to maintain concentration levels of important metabolites [160]. The activation or repression of these genes depends on the level of B in the cell. Again, rather than attempting to determine an internal concentration, we may use a flux rather than concentration to turn ‘off’ an enzyme. We choose $R2b$ as the determining factor; it will activate $R Pb$ which in turn will inactivate $tR2a$ and $tR7$.

$$RPb = IF (R2b > 0) \quad (3.16)$$

$$tR2a = IF NOT (RPb) \quad (3.17)$$

$$tR7 = IF\ NOT\ (RPb) \quad (3.18)$$

We will now analyze this network, first using extreme pathway analysis to determine how the metabolic potential of the cell is reduced by the application of regulatory constraints.

3.C.1 Regulated extreme pathway analysis

A total of 80 extreme pathways were calculated for the simplified metabolic system shown in Figure 3.4. They are all shown graphically in the appendix. The numbering of the extreme pathways is arbitrary and corresponds to the order in which the algorithm generates them. Given the five inputs to the metabolic network and representing these inputs using Boolean logic, considering each as ON if present or OFF if absent, there are a total of $2^5 = 32$ possible environments which may be recognized by the cell. These environments are listed in Table 3.1. For each environment, the transcription of several of the enzymes in the network may be restricted due to regulation. The constraints imposed on the metabolic system by both the substrates available to (i.e. the external environment) and the enzymes expressed in the cell (i.e. the internal environment) reduce the number of extreme pathways accessible to the cell at a given time.

Table 3.1: A list of all the possible environments which can be recognized by the system shown in Fig. 1, of which six environments do not enable the cell to produce biomass. For each environment, there is a set of enzymes which are repressed under the given environmental conditions. The extreme pathways which remain feasible even under the combination of environmental and regulatory constraints are listed. For a schematic of each of the pathways, see the Appendix. C1 = Carbon1, C2 = Carbon2, O2 = Oxygen. Two of the environments are shown in more detail in Figures 3 and 4 and are labeled correspondingly.

		Environments				Repressed Enzymes			Pathways			Pathway List	
		C1	C2	F	H	O2	R5b	R8a	Tc2	26	P2, P4, P5, P6, P8, P9, P10, P12, P29, P30	P31, P32, P33, P34, P35, P36, P37, P38, P45, P46	P47, P48, P49, P50, P51, P52
C1	C2	F	H				R5a	R8a	Rres	Tc2	10	P39, P40, P41, P42, P43, P44, P49, P50, P51, P52	P29, P30, P33, P34, P45, P46, P49, P50
C1	C2	F		O2			R5b	R8a	Rres	Tc2	8	P41, P42, P49, P50	P2, P5, P6, P9, P10, P30, P31, P34, P35, P37
C1	C2	F		H	O2		R5a	R8a	Rres	Tc2	4	P46, P47, P50, P51	P39, P42, P43, P50, P51
C1	C2			H			R5a	R8a	Rres	Tc2	5	P30, P34, P46, P50	P30, P34, P46, P50
C1	C2			O2			R5b	R8a	Rres	Tc2	4	P42, P50	P46, P47, P50, P51
C1	C2			H	O2		R5a	R8a	Rres	Tc2	2	P39, P42, P43, P50, P51	P31, P32, P33, P34, P35, P36, P37, P38, P45, P46
C1		F	H				R5b	R8a	Rres	Tc2	26	P2, P4, P5, P6, P8, P9, P10, P12, P29, P30	P47, P48, P49, P50, P51, P52
C1		F	H				R5a	R8a	Rres	Tc2	10	P39, P40, P41, P42, P43, P44, P49, P50, P51, P52	P29, P30, P33, P34, P45, P46, P49, P50
C1		F		O2			R5b	R8a	Rres	Tc2	8	P41, P42, P49, P50	P2, P5, P6, P9, P10, P30, P31, P34, P35, P37
C1		F		H	O2		R5a	R8a	Rres	Tc2	4	P46, P47, P50, P51	P39, P42, P43, P50, P51
C1		H					R5a	R8a	Rres	Tc2	5		

Table 3.1, con't.

Environments			Repressed Enzymes			Pathways		Pathway List	
C1	O2	R5b	Tc2	Tc2	4	P30, P34, P46, P50			
C1	F	H	R5a	R8a	26	P3, P4, P5, P7, P8, P9, P11, P12, P57, P58, P59			
C2	F	H	O2	R5b	2	P42, P50			
C2	F	H	O2	R5a	P60, P61, P62, P63, P64, P65, P66, P73, P74, P75	P3, P4, P5, P7, P8, P9, P11, P12, P57, P58, P59			
C2	F	H	O2	R5b	P76, P77, P78, P79, P80	P60, P61, P62, P63, P64, P65, P66, P73, P74, P75			
C2	F	H	O2	R5a	R8a	10	P67, P68, P69, P70, P71, P72, P78, P79, P80		
C2	F	H	O2	R5b	R8a	8	P57, P58, P61, P62, P73, P74, P77, P78		
C2	H	O2	R5b	R8a	R8a	4	P69, P70, P77, P78		
C2	H	O2	R5a	R8a	R8a	14	P3, P5, P7, P9, P11, P58, P39, P62, P63, P65		
C2	H	O2	R5b	R8a	R8a	5	P74, P75, P78, P79		
C2	H	O2	R5a	R5b	R8a	5	P67, P70, P71, P78, P79		
C2	F	H	O2	R5a	R8a	4	P58, P62, P74, P78		
C2	F	H	O2	R5b	R8a	2	P70, P78		
F	F	H	O2	R5a	R8a	5	P4, P5, P8, P9, P12		
F	F	H	O2	R5b	R8a	0	P4, P5, P8, P9, P12		
F	H	O2	R2a	R5b	R7	0	P5, P9		
F	H	O2	R2a	R5a	R7	0	P5, P9		
F	H	O2	R5b	R8a	R8a	0	P5, P9		
						0	P5, P9		

Several interesting observations may be made from Table 3.1: first, 21 extreme pathways (P1, P13-P28 and P53-56, enclosed by a box in the appendix), although stoichiometrically feasible, are always impossible due to regulatory constraints. Pathways P13-28 and P53-56 are infeasible due to the fact that R_{res} is only expressed aerobically while R_{5b} is only expressed anaerobically. Therefore, any pathway that includes a flux through both R_{res} and R_{5b} is eliminated. Similarly, pathways P1 and P13 are eliminated because a flux through R_{2b} activates a regulatory protein that represses transcription of $R7$. Therefore, R_{2b} and $R7$ can not both be expressed together. Note that P13 is infeasible in either case.

Another interesting observation from Table 3.1 is that several environments have identical or near-identical sets of available extreme pathways. For example, the environment containing *Carbon1* (C1), *Carbon2* (C2), *F*, *H* and *Oxygen* (O2) has an identical extreme pathway list to that for the environment containing C1, F, H and O2. The reason is that $Tc2$, the transport flux for C2, is repressed in the presence of C1. Furthermore, the extreme pathway list for the environment containing C2, F, H and O2 is similar to the pathway lists for the previously mentioned environments, different only in that the pathways which utilize the $Tc1$ flux in the former pathway list are replaced by pathways which utilize the $Tc2$ flux in the latter.

Finally, Table 3.1 shows that the highest number of extreme pathways available to the cell is 26; the lowest is 2, corresponding to a reduction in the number of available extreme pathways from between 67.5-97.5%. A relatively simple dual-substrate environment and the most complex environment were examined in more detail to more closely investigate the effect of regulation on available pathways.

Example 1: Growth on C1, C2 and O2

The metabolic network was given C1, C2 and O2 as inputs and allowed Biomass, Dext and Eext as outputs. These conditions reduce the number of ex-

treme pathways available to the cell (Figure 3.5). Initially, all 80 pathways are considered, and represented schematically in Figure 3.5a. Twenty-one of the extreme pathways are always restricted by regulation, as discussed earlier; the boxes representing these pathways are darkened in gray. By considering only the pathways with appropriate inputs and outputs based on the cell environment, 49 more pathways are eliminated (shaded in light gray). Of the ten remaining pathways, six are inconsistent with the given regulation (C1 catabolite repression of the C2 transport protein Tc2 or regulation due to the aerobic environment) (shown in black) and the flux maps for the remaining four extreme pathways are shown in Figure 3.5b.

The resulting solution space is projected in three dimensions (C1 uptake rate, Oxygen uptake rate and growth rate), as shown in Figure 3.5c, with the four feasible extreme pathways labeled. All the volume defined by these edges is accessible to the cell. The corresponding range of growth and uptake rates can be attained by the cell under these conditions.

We can also represent the solution space in two dimensions as projections which are sometimes called phenotypic phase planes (PhPPs)[47, 172]. Once the axes of the PhPP are set (generally uptake rates of two nutrients such as oxygen and a carbon source), flux maps which are optimal with respect to growth may be calculated for each point in the plane. The lines in a PhPP demarcate a change in the flux map; these lines are determined using the shadow prices (the dual solution in linear programming). PhPP analysis has recently been used to demonstrate the optimal growth of *E. coli* on succinate and acetate minimal media [49]. The 2-dimensional PhPP for growth on C1 and O2 is shown in Figure 3.5d. This PhPP has two feasible phases between the lines shown, which represent the four extreme pathways available to the system. For this case, the two-dimensional projection of the extreme pathways lies on the region boundaries of the PhPP; pathways 46 and 50 are both fermentative and therefore overlap in the PhPP (Oxygen Uptake Rate = 0). Pathway 30 is the line of optimality [49, 50] as none of the carbon is lost

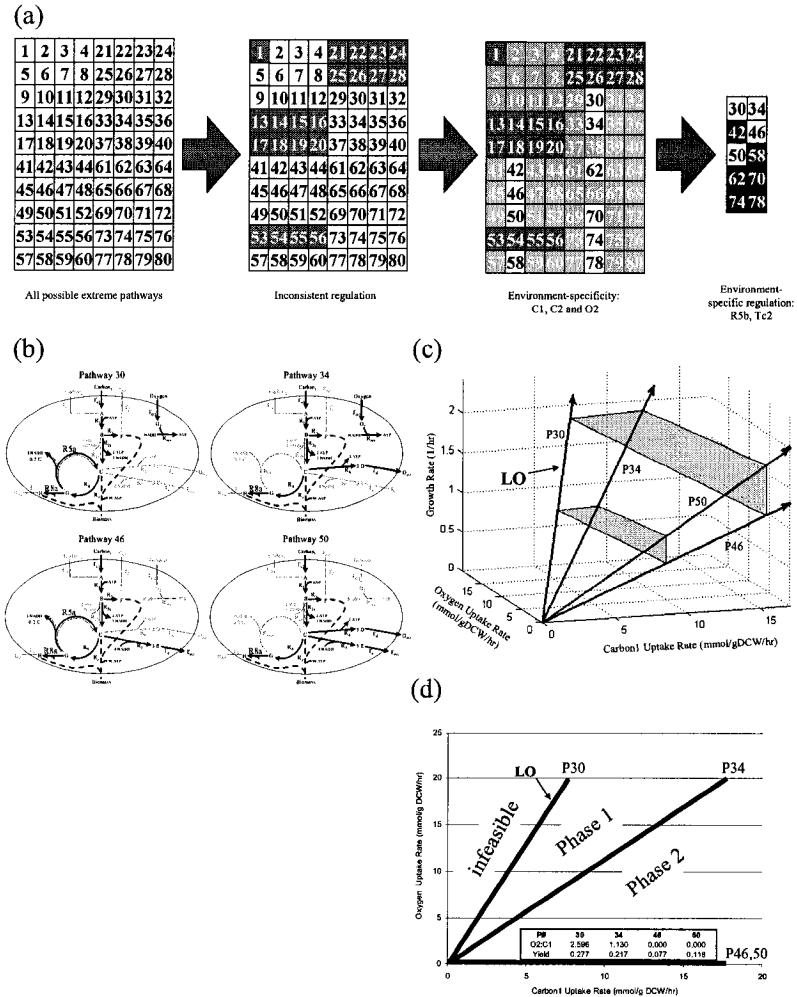


Figure 3.5: Extreme pathway reduction by constraints, using growth of the sample system in a C1 and C2 aerobic medium as an example. (a) The 80 total extreme pathways calculated for the system are represented by a grid (see Appendix). The number of pathways is reduced by 21 when pathways are removed which are always inconsistent with the regulatory rules (dark gray), then by 49 due to the specific environmental constraints (light gray), then by 6 as the regulation corresponding to the specific environment is considered (black). The four remaining pathways are shown schematically in (b), where the thick dark arrows represent active fluxes. (c) The solution space of the system, projected on a three-dimensional space, with the pathways and the line of optimality (the pathway with the greatest growth yield) noted. (d) A two-dimensional projection of the space, superimposed on a two-dimensional phenotypic phase plane for the C1 and Oxygen uptake. The region at left lies outside of the space and is therefore infeasible.

in secretion of by-products; pathway 34 includes secretion of Dext and therefore gives a lower biomass yield (inset) than Pathway 30.

Example 2: Growth on C1, C2, F, H and O2

The allowable extreme pathways for growth on a medium containing C1, C2, F, and H in an aerobic environment were determined. For this case, the environment offers no restrictions - all possible inputs are available - and therefore the restriction of the solution space by elimination of extreme pathways is entirely due to regulatory effects. From Table 3.1 it is shown that R2a, R5b, R7, R8a and Tc2 are constrained to zero by the regulatory rules. Consequently, the 33 corresponding extreme pathways were removed from the solution space, resulting in a list of 26 available pathways which may be used by the cell under these conditions (Figure 3.6b). These remaining pathways, normalized by the total input of C1, C2, F and H for comparison purposes, are shown in Figure 3.6a.

The 26 allowable pathways shown in Figure 3.6a may be grouped by biomass yield. In the top two sets all pathways are optimal or very near-optimal in terms of biomass yield, with no by-product secretion. The middle pathway sets involve secretion of either D or E while biomass is generated, and the bottom set of pathways represents purely fermentative use of the network. Again, it is seen that even with a higher number of allowable extreme pathways, the actual degree of variation in possible network behavior is surprisingly small once regulatory constraints are taken into account. The multiplicity of extreme pathways with near optimal biomass yield gives the metabolic network robustness characteristics as the cell has many alternatives with nearly the same outcome. The reduced solution space is projected in three and two dimensions, as shown in Figure 3.6c and Figure 3.6d, respectively. The dimensions are the same as in Figure 3.5, with the exception of the C1 uptake rate axis, which has been replaced by a normalized axis of all possible routes for substrate uptake ($C1 + C2 + F + H$). The three dimensional projection is bounded by the solid black vectors, many of which overlap in the

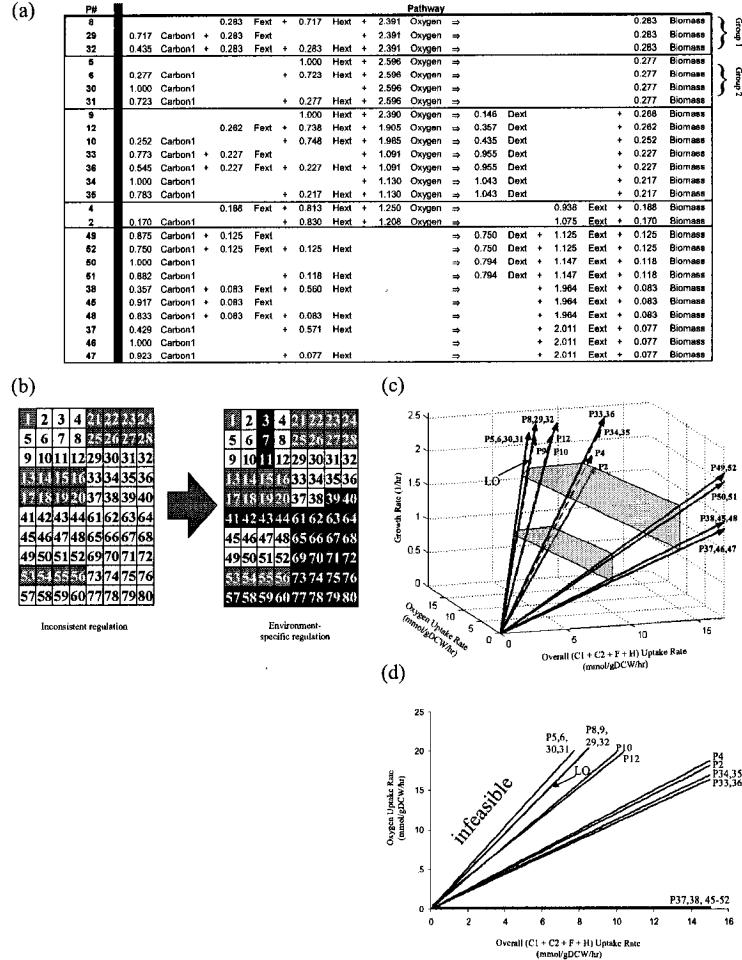


Figure 3.6: Aerobic growth of the sample system in a complex medium. (a) A table summarizing the 26 extreme pathways which are feasible for this environment. The four boxes indicate significant pathway groupings based on by-product secretion and yield. Group 1 and Group2 are the sets of pathways with the highest growth yield. (b) Extreme pathway reduction in the complex medium, similar in format to Fig. 3a. Note that no environmental constraints are imposed upon the space in this example, and reduction of the solution space is therefore only due to regulatory constraints. Here the pathways have been normalized by the combined uptake of C1, C2, F and H in order to be represented together in the plots shown in (c) and (d), which are also similar in format to their counterparts in Fig. 3c and Fig. 3d, respectively. In (c), four extreme pathways lie inside the 3-dimensional projection (dashed lines).

projection (note that these extreme pathways do not actually overlap; however, in a projection of the high-dimensional solution space onto a lower dimension, they share certain characteristics such as the relationship between growth, substrate uptake rate and oxygen uptake rate). Five pathways (P4, P9, P10, P34, P35, shown with dashed black lines) lie inside the three-dimensional projection. The structure of the solution space is more complex as compared to the simple growth condition analyzed in Example 1.

The solution spaces are compared on the same axis to illustrate the concept of solution space reduction further (Figure 3.7a). The solution space discussed in Example 1 is a subset of the space defined by the complex medium. Figures 3.7b and 3.7c show cross sections of the space where combined ($C_1 + C_2 + F + H$) uptake and Oxygen uptake are set at a constant rate of 5 mmol/gDCW/hr, respectively. The line of optimality, P30 in Example 1, is shifted to any of three pathways, P8, P29 and P32, in Example 2. These three pathways obtain similar growth yields but exhibit different behaviors in terms of substrate uptake. With slightly smaller growth yields, extreme pathways P5, P6, P30 and P31 also bound the space and are in close proximity to the optimal pathways. The cross section where the Oxygen uptake rate is constant (bottom) is also unbounded, as shown by the dotted lines.

Regulatory constraints reduce the number of feasible extreme pathways

The most important finding of this study is that the imposition of regulatory constraints significantly reduces the size of the solution space. For the skeleton core metabolic network used, the number of extreme pathways was reduced from 80 to as few as 2, in some cases, as a result of the imposition of relatively simple environmental and regulatory constraints. In a simulated rich medium, the skeleton network was unconstrained with respect to the environment and yet 67.5% of the pathways were eliminated by regulation. This large reduction in the solution space seems to indicate that despite the complex interaction of many genes to produce

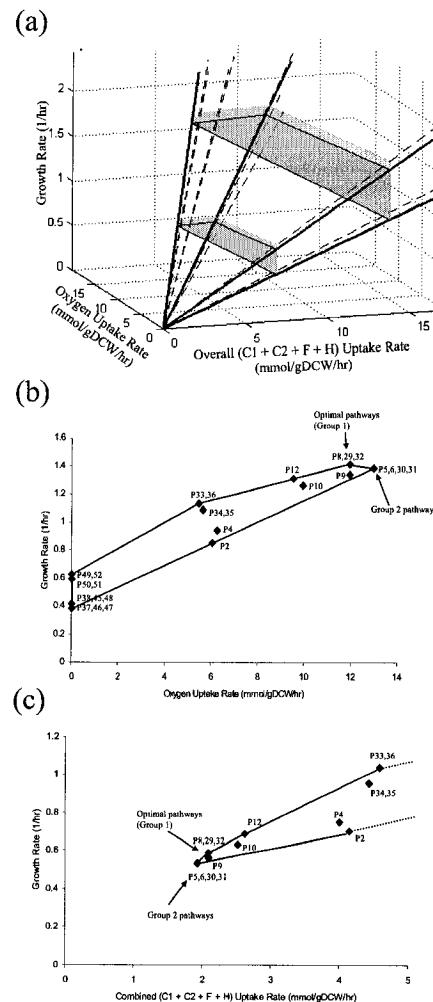


Figure 3.7: Diagrammatic representation of solution space reduction. (a) Schematic of the two solution spaces superimposed on one another. The solution space of Example 1 is a subspace of the space of the complex medium. (b) Cross section of the solution space on the Oxygen Uptake Rate - Growth Rate plane where the combined ($C_1 + C_2 + F + H$) Uptake is set at 5 mmol/gDCW/hr. (c) Cross section of the solution space on the Combined ($C_1 + C_2 + F + H$) Uptake Rate - Growth Rate plane where the Oxygen Uptake Rate is set at 5 mmol/gDCW/hr.

an integrated cellular function, simple behaviors can relatively easily be selected by the cell.

Another noteworthy observation that can be made from these results regards the extreme pathways that remain after regulatory constraints are applied. In both examples given here, many fermentative pathways were left available to the network despite the presence of oxygen. Although these extreme pathways may be unused, the ability to implement them without the delays associated with transcription and translation would give the cell the ability to rapidly adapt to oxygen deprivation.

Finally, the close proximity and even overlap of optimal and near-optimal extreme pathways to one another in the 3-dimensional and 2-dimensional projections of the solution space suggest that the system has numerous means by which to obtain its growth objectives. Extreme pathway analysis of metabolic networks at the genome scale has indicated a high degree of pathway redundancy [137]. Although regulatory constraints greatly reduce such redundancy as has been shown, it seems that a certain amount of flexibility is beneficial to the cell. Such flexibility may be useful to an organism, for example, in colonizing diverse and changing environments.

3.C.2 Regulated flux-balance analysis

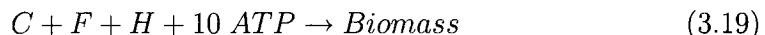
We now look at quantitative simulations of cell growth using FBA, after a short treatment of dynamic simulations.

Simulating time courses of growth

The quasi-steady state assumption on the metabolic network can be used to generate dynamic profiles of cell growth. The experimental time is divided into small time steps, Δt . Beginning at $t = 0$ where the initial conditions of the experiment are specified, the metabolic model may be used to predict the optimal flux distribution for the metabolic network. From the transport fluxes, the extracel-

lular concentrations may be calculated in a time-dependent fashion, as previously described [199]. These concentrations are then used as the initial conditions for the next time step. This type of dynamic modeling was shown to correlate well with the growth of *E. coli* on glucose minimal media under aerobic and anaerobic conditions, predicting quantitatively the uptake of glucose and growth rate as well as by-product secretion [199].

The time constants characterizing transcriptional regulation are generally on the order of a few minutes or slower [215, 157, 120], which are slower than those time constants associated with metabolism. Therefore, the FBA generation of time profiles for dynamic cellular behavior may be integrated with a set of transcriptional regulatory rules which are represented by Boolean logic equations. The status of transcription is found from the given conditions at the particular time interval. Specifically, transcription may be altered by the presence or surplus of an intracellular metabolite, an extracellular metabolite, regulatory proteins, signaling molecule, or any combination of these or other factors. The logical equation governing transcription of each transcriptional unit is used to determine whether transcription occurs or does not occur. At a given time point, a commercially available linear programming package (LINDO, Lindo Systems, Chicago) is used to identify an optimal metabolic flux distribution within the solution space. The optimal metabolic flux distribution is identified here as the flux distribution which maximized the Growth flux, a flux which represents growth of an organism by removing necessary precursors of growth from the system:



The hypothesis that microbial cells behave in such a way that their growth is optimized has been verified experimentally under certain conditions [49]. Using the resulting flux distribution and the conditions of the system in a previous time step, the conditions of the next time step were calculated on a commercially available spreadsheet package (EXCEL, Microsoft Corporation, Redmond), following

a procedure previously described [199] to obtain biomass as well as extracellular substrate and byproduct concentrations. Numerical values for the parameters used in the simulation, such as the protein transcription and decay time as well as maximum uptake rates for all possible substrates, are shown in Table 3.2.

Table 3.2: Numerical values of parameters used in growth simulations of the sample metabolic/regulatory network.

Parameter	Value
<i>Maximum uptakes (mmol/gDCW/hr)</i>	
Carbon1	10.5
Carbon2	10.5
D	12
E	12
F	5
H	5
O ₂	15
<i>Protein Synthesis/Decay Delay (hrs)</i>	0.25

Using the sample metabolic network shown in Figure 3.4, modeling the dynamic growth of a cell with the incorporation of temporary regulatory constraints as shown in the inset table can be illustrated with several insightful examples. The results of five simulations, chosen to illustrate each regulatory element separately and in a complex medium, are described below.

Example 1: Diauxie on two carbon sources

The first example concerns the growth of the cell on two carbon sources, C1 and C2. For this example, the initial concentrations of both carbon sources in the media were set to 10mM and the simulation was run with oxygen in excess. The simulation was run until both carbon sources had been completely exhausted. The results are shown in Figure 3.8. The upper half of Figure 3.8 is a time plot which shows the concentrations of C1, C2, byproduct D, and the biomass

X. The typical diauxic growth curve found in instances of catabolite repression is easily seen. The time plot is divided into three regions with dotted lines. The flux maps generated using FBA are qualitatively identical in each region, and the bars beneath the flux maps indicate whether the regulated genes are being transcribed (dark gray) or not (white). In region A (flux map shown on bottom left of Figure 3.8), the cell grows using C1 as the preferred carbon source; in region C (bottom right), C2 is used. Region B (not shown) is a short period where the cell does not grow while the transport protein Tc2 is being upregulated and synthesized. Without the addition of regulatory constraints, the system would grow on both carbon sources together to maximize production of the biomass conditions and fail to predict a diauxic shift.

Example 2 - Anaerobic/aerobic diauxie

The second example of transcriptional regulatory modeling is the diauxic shift associated with a sudden removal of available oxygen to the culture (Figure 3.9). Again, the dynamic profile may be divided into three areas. First, in region A, the culture grows with C2 as the primary carbon source. At two hours, the oxygen supply to the culture is removed and the culture grows anaerobically (regions B & C). It can readily be seen from Figure 3.9 that regions B and C have similar flux distributions and that the regulatory constraints, unlike in Figure 3.8 where they actually restrict the optimal solution, act redundantly (the regulation is not necessary to the solution). Therefore, the regulatory structure allows unnecessary proteins to decay in this case, without changing the optimal solution.

Example 3 - Growth on carbon and amino acid with carbon in excess

In this case, shown in Figure 3.10, the simplified amino acid H is present in the medium with C2, where it is used both to satisfy the H biomass requirement and as a supplementary source of C (Region A). After the extracellular supply of H is depleted there is a region of no growth (Region B) while the cell upregulates its

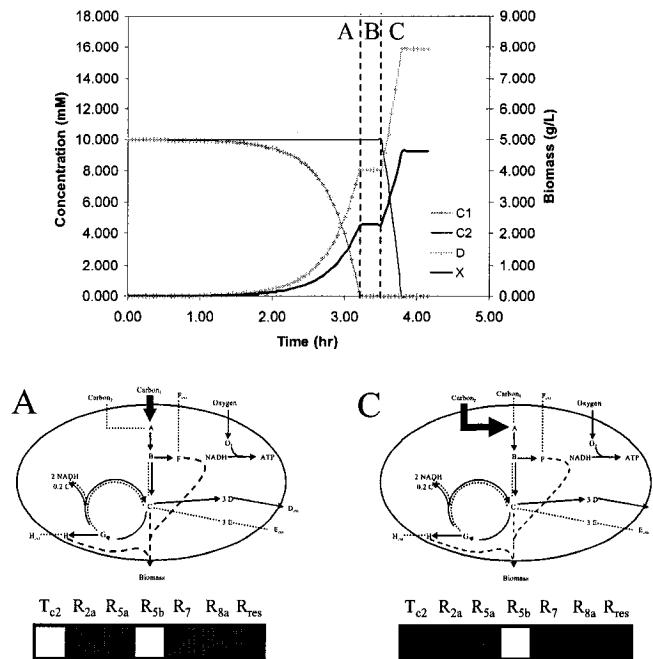


Figure 3.8: Catabolite repression in the simplified network. A time profile of calculated growth and metabolism is shown (top), divided into three regions with dotted lines. For regions A and C, the network maps are shown (bottom) with the inactive fluxes denoted by thin dotted lines and the active fluxes shown as solid, or in the case of the biomass flux, dashed black lines. Certain fluxes are emphasized with a thick arrow to indicate the change in flux distributions due to regulation. Underneath each network map is an in silico array which shows whether a particular reaction is activated (dark gray) or inactivated (white). Labels correspond to metabolites shown in Table 1, with three additions: C1 = Carbon1, C2 = Carbon2, X = Biomass.

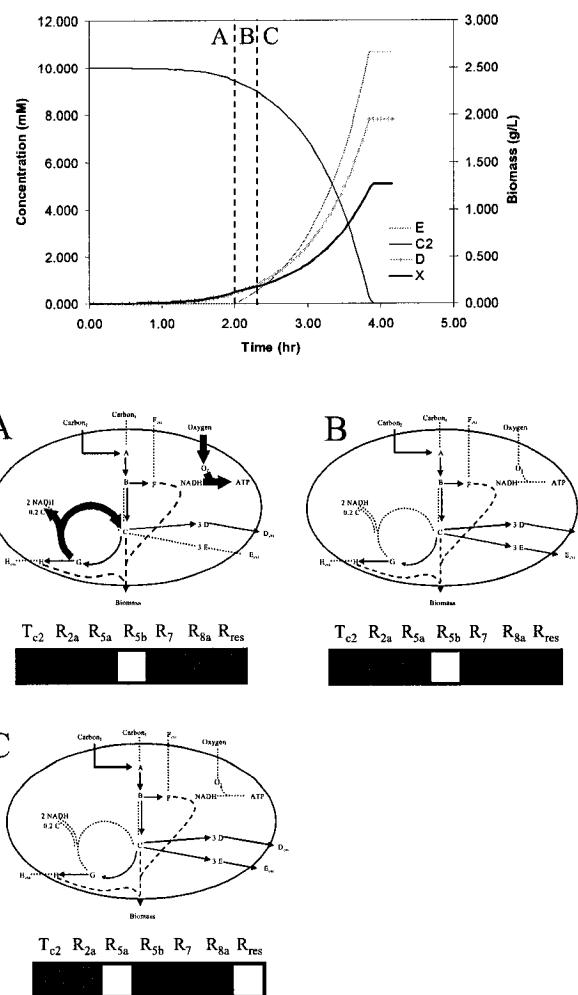


Figure 3.9: Aerobic/Anaerobic growth calculated using the simplified network. Similar in format to Figure 3.8. Again, the dynamic profile may be divided into three areas, with the metabolic network maps and in silico arrays shown below.

H biosynthesis machinery, R8a. This example is similar to Example 1 in that there is a pause where no growth occurs while the regulatory structure is expanding the solution space to allow the cell to synthesize H.

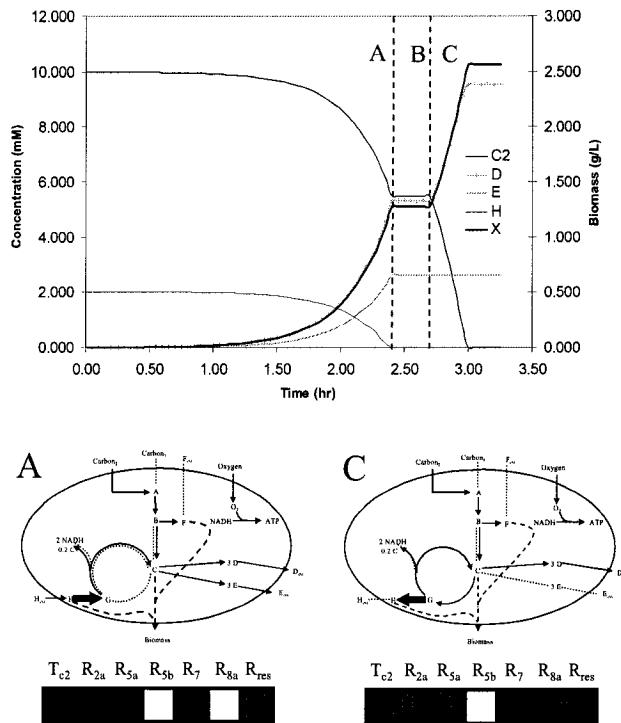


Figure 3.10: Amino acid biosynthesis. Similar in format to Figures 3.8 and 3.9.

Example 4 - Growth on carbon and amino acid with amino acid in excess

Figure 3.11 shows a graph of simulated growth on C2 and H where H is in excess. This case is designed to demonstrate the regulation of RPb; initially, the C2 and H are taken up together as in Example 3 (Region A). However, as the C2 is depleted, RPb is activated and the transcription of R2a and R7 ceases (Region C). As in Example 2, the flux distribution changes as the process moves to Region

B and remains constant as regulatory events constrain the cell to stop producing certain unnecessary proteins.

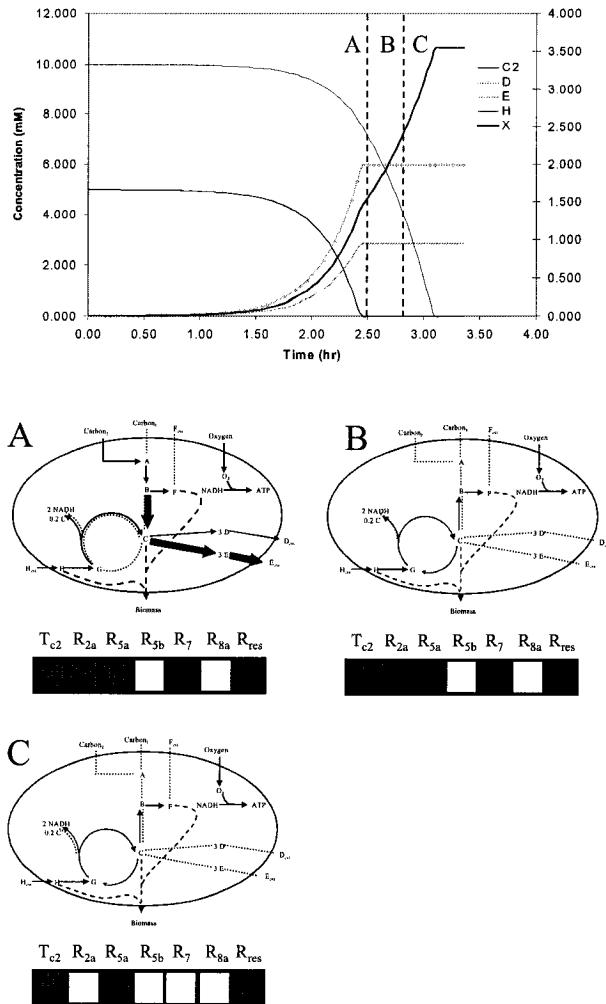


Figure 3.11: Growth on carbon and amino acid with amino acid in excess. Similar in format to Figures 3.8- 3.10.

Example 5 - Complex medium

Figure 3.12 depicts growth of the cell on a complex medium, with initial substrate concentrations strategically chosen so that the status of all regulatory proteins will be changed over the course of the experiment. C1, C2, F and H are all initially present in the medium; at one hour, oxygen is also allowed into the culture. The figure illustrates the interplay of several regulatory actions to control the growth of the cell on multiple substrates and under changing conditions. The plot is divided into 15 regions, with Regions 12 and 15 as transitory periods with no growth while certain regulatory changes are occurring. The genes which are being regulated are thick black arrows, with the exceptions of R5a and R5b which are light and dark gray arrows, respectively. In seven of these (Regions 1, 2, 4, 5, 6, 8, 10) the internal flux distribution is qualitatively similar, with minor changes as C1 or F is depleted. Another more interesting change occurs in the transition from Region 2 to Region 3. The simplified TCA cycle uses R5b in Regions 1 and 2 and begins using R5a (an isozyme which is expressed under aerobic conditions) in Region 3, after oxygen has been allowed to enter the system. Because R5a and R5b are equivalent stoichiometrically, FBA alone does not favor one reaction over the other and would fail to predict which of the two isozymes are active under given conditions. The growth of the cell on the complex medium also exhibits unusual flux distributions, which represent unstable transitory states while the organism's regulation is changing. For example, Region 9 shows a situation where C2 is used only to fulfill the F biomass requirement while H is used as a C, H and energy source, due to the repression of R2a.

Regulatory constraints lead to more accurate simulations

From these examples, we can see that incorporating transcriptional regulatory structure into FBA enables us to more accurately predict dynamic flux profiles of microbial growth. This procedure has major advantages over FBA in the following areas:

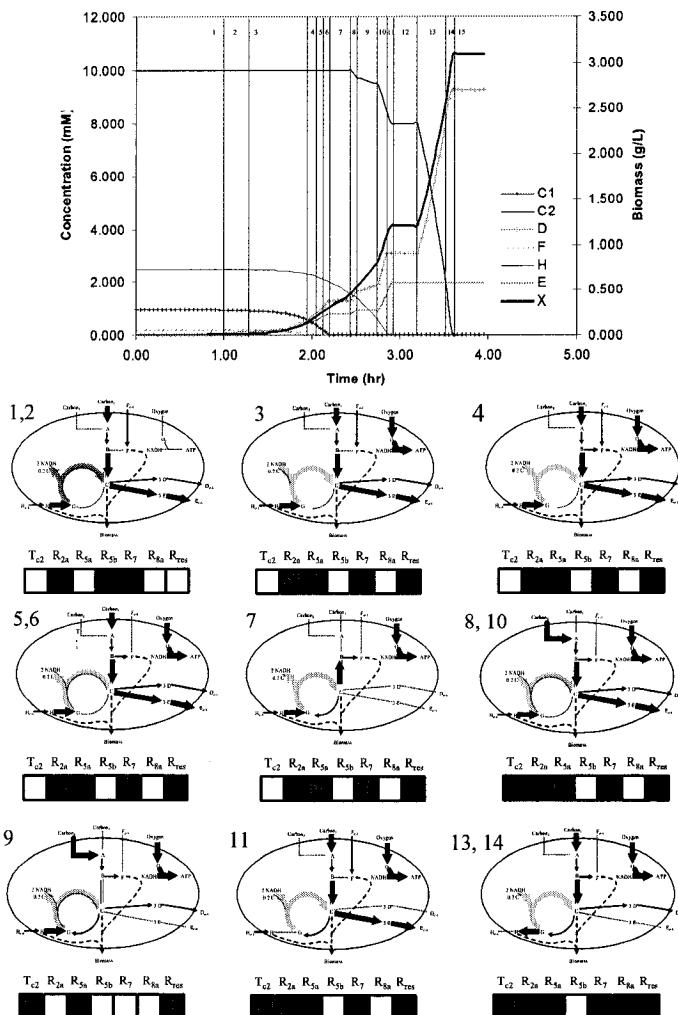


Figure 3.12: Growth on a complex medium. Similar in format to Figures 3.8- 3.11, except in the case of R5a and R5b, where the activity of these isozymes has been highlighted with a light gray arrow in the case of R5a and with a dark gray arrow in the case of R5b. Additionally, similar flux distributions have been grouped together as shown.

1. Quantitative dynamic simulation of substrate uptake, cell growth and by-product secretion;
2. Qualitative simulation of gene transcription events and the presence of proteins in the cell; and
3. Investigation of the systemic effects of imposing temporary regulatory constraints on the solution space.

The quantitative predictions made by the combined regulatory/metabolic model are completely unpredictable using FBA alone under many conditions. Example 1 illustrates this point. The diauxic growth curve shown in Figure 3.8 is a completely different result than would be obtained by FBA alone, which would incorrectly predict the maximal possible uptake of both C1 and C2. It is interesting that the addition of one simple constraint to the solution space ($Tc2 = 0$) results in such a dramatic change in the predicted phenotype. Similarly, from Example 5 it is clear that FBA alone would fail to predict which of two isozymes plays the more active role in catalysis under given conditions. Again, it is the addition of one simple constraint which results in the more correct prediction.

The combined regulatory/metabolic model is also capable of making qualitative predictions about the up- and down-regulation of enzyme production. This capability is shown in Examples 2 and 4, where FBA and the combined regulatory/metabolic approach predict similar flux distributions, but the combined approach also predicts the down-regulation of certain enzymes which are no longer required to obtain the optimal solution. FBA is unable to consistently make such predictions due to the fact that an enzyme may be present in a cell and still have a zero flux.

Finally, this approach allows an investigation of the systemic effects of transcriptional regulation. The individual operations required to model the combined metabolic/regulatory network as described are simple. However, when combined to represent biological networks of even modest complexity, they reproduce

quite complex behavioral patterns. The sample network examined here, although two orders of magnitude smaller than the metabolic networks of commonly studied bacteria, exhibits surprisingly complex behavior, as shown by the unusual intermediate flux distributions during growth on the complex medium. For a more complicated network, the multiple constraints applied to the system can cause the solution space to change dramatically in response to a changing environment.

Besides simply determining whether or not regulatory constraints are implemented, the environment also has an important influence on the regulatory constraints themselves. Unlike the non-adjustable physico-chemical constraints, regulatory constraints are biological in nature and can change with evolution. The demand theory of gene regulation indicates that the evolution of gene regulation may be governed by rules, predicting a correlation between demand for gene expression and the mode of control exhibited [163, 165]. This correlation has been supported experimentally [164]. The constraints-based approach to modeling regulatory events, by providing a framework for analyzing, interpreting and predicting the systemic effects of transcriptional regulation, may therefore also contribute to our understanding of the rules which govern evolution and the subsequent effects of evolutionary forces on an organism's solution space.

The method is presented here in its most simple form. However, the use of Boolean logic to represent genetic regulatory networks qualitatively has grown in sophistication, including such features as multilevel logic variables and asynchronous updating of protein synthesis [194]. These features may be incorporated into this approach at some time in the future. Another important development in this approach will be its application to a real network. Databases have already been developed and made available online which detail the known regulation of *E. coli* [91, 162]; such information, made available for this and other organisms will enable the construction of genome-scale metabolic/regulatory microbial models. The ability of the approach described here to generate quantitative hypotheses which may be experimentally tested will lead to an ongoing iterative model-building process,

resulting in advanced models and augmented scientific knowledge.

In summary, the application of transcriptional regulatory constraints to metabolic networks results in a large reduction of behaviors available to the network under a given environment. The present study leads to the formulation of a second-generation of constraints-based models that can be used to interpret how regulation is used to keep a restricted portion of the total solution space accessible, and thereby, by the process of elimination, force a particular set of phenotypic behaviors to be expressed, an approach which is potentially more versatile and may be used to simulate a wider range of experimental conditions, as we will see more clearly with the real-world example of *E. coli*.

The text of this chapter, in part or in full, is a reprint of the material as it appears in “Constraints-based models: regulation of gene expression reduces the steady-state solution space” and “Regulation of gene expression in flux balance models of metabolism”, both in the *Journal of Theoretical Biology*. I was the primary author and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 4

Modeling *Escherichia coli*

The annotated genome sequence of *Escherichia coli* and other organisms [143] has made it possible to develop *in silico* models of cellular behavior at a genome-scale.

The constraint-based approach to studying *E. coli* metabolism has more than a decade's worth of history and has been used to predict evolutionary trajectories, wild type and knockout phenotypes, and the like [153]. Earlier constraint-based models of *E. coli* ignored the constraints associated with gene expression and treated all genes and reactions as available to the system at any given time. However, it has been observed that only about 50% of the *E. coli* genome is expressed under typical growth conditions [97]. This reduction in the number of expressed genes corresponds to fewer available metabolic reactions and a far smaller metabolic potential [32, 30]. We recently developed a model of central *E. coli* regulation and metabolism, and demonstrated that the imposition of time-dependent regulatory constraints substantially reduces the metabolic solution space and has a significant effect on the scope and accuracy of model predictions, enabling dynamic growth simulations together with predicted shifts in gene expression [31].

In this chapter, I will report on a regulatory network that was reconstructed for central metabolism in *E. coli*. This regulatory network was then incorporated into an existing *E. coli* central metabolic model and used to sim-

ulate *E. coli* behavior under a variety of environmental conditions and genome perturbations.

4.A Network reconstruction

The metabolic network was reconstructed by identifying a set of biochemical reactions in central *E. coli* metabolism, taken from the annotated genome sequence [12, 180] as well as biochemical and physiological literature [130, 162]. The regulatory network was derived from the literature data (cited in the supplemental material) and represented as a set of regulatory rules following established procedures [30]. These rules were based on external (e.g., the presence or absence of extracellular molecules) and/or internal (e.g., the activity or inactivity of an enzymatic flux in a given environment) conditions of the system. To accommodate the generally qualitative nature of existing transcriptional regulatory data, regulatory constraints were described using a Boolean formalism [193], where gene products are either available (ON) or unavailable (OFF) to the cell. The regulatory rules can be computed using commercially available software packages, including EXCEL (Microsoft Corporation, Redmond).

4.B Simulation of *E. coli* behavior

4.B.1 Calculation of steady-state regulated metabolic flux distributions

Once the metabolic and regulatory networks were defined, FBA was used to determine an optimal metabolic flux distribution for the given conditions. FBA has been described and reviewed in detail in Chapter 3[47, 15] and is based on the location of optimal flux distributions within the solution space defined by a metabolic network and governing constraints, as described above. For the purposes of these simulations, capacity constraints included maximum uptake rates of oxygen as well as substrates such as glucose, acetate and lactose, as determined

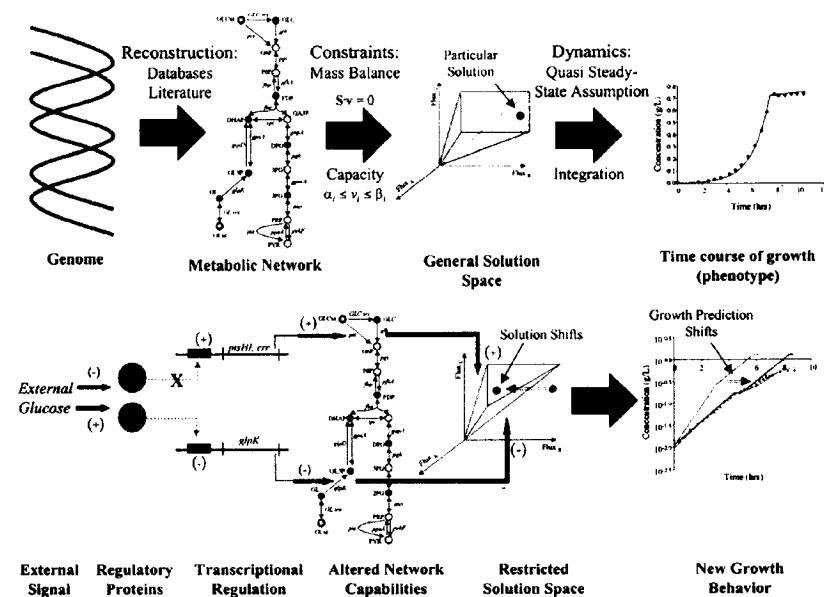


Figure 4.1: *In silico* modeling of metabolism and transcriptional regulation using the constraints-based approach, as reviewed in the last chapter. (A) The constraints-based approach to metabolic modeling. The metabolic network is defined and described geometrically as a closed solution space, wherein flux-balance analysis can be used to identify particular optimal solutions and simulate dynamic behavior. (B) Transcriptional regulation reduces the steady-state solution space. As stimuli are sensed by various regulatory proteins in the cell, transcription of various genes are repressed or activated, leading to the availability or unavailability of the respective reactions or transport processes in the metabolic network, and possibly to the removal of available extreme pathway basis vectors from the steady-state solution space. The result is a new predicted behavior if regulatory effects have a dominant influence on metabolism under the given conditions.

from growth experiments found in the literature [99, 199] (see Figures 4.3- re-JBC5 for numerical values). It has been demonstrated for *E. coli* under various conditions that cell growth may be used as an objective function [49, 98]; the production of growth precursors in certain ratios was used here as an approximation [129]. LINDO was used to calculate the optimal flux distributions (Lindo Systems, Chicago).

The incorporation of regulatory constraints in the flux-balance framework has been described in detail [30]. Regulatory constraints were applied to the network under steady-state conditions by evaluating whether genes were expressed or repressed for the given conditions (based on the aforementioned regulatory rules), and constraining the fluxes of repressed gene products to zero. If a gene product was expressed, the actual flux value was determined by FBA. Thus, although the regulatory network is based on qualitative rules, the flux distribution determined by FBA is a quantitative result.

4.B.2 Changing environments and time-dependent cell behavior

The time constants which describe metabolic transients are fast (on the order of milli-seconds to tens of seconds) as compared to the time constants associated with transcriptional regulation (generally on the order of a few minutes or slower) or cell growth (on the order of hours to days) [30]. Therefore, dynamic simulations may be performed by considering the behavior inside the cell to be in a quasi-steady state during short time intervals relative to the environment [199]. Beginning at time zero, the simulation was run in small time steps (3 seconds for this study). For each time step, the regulatory rules were evaluated based on current environmental and internal conditions of the cell to determine the up- or down-regulation of every regulated gene. If a gene was determined to be up-regulated at a given time, the corresponding reactions in the metabolic system were considered unconstrained (by gene regulation), after a time delay to allow for protein synthesis. Conversely, the down-regulation of a gene resulted in the corre-

sponding reactions being constrained to zero after a time delay to allow for protein degradation. The resulting sets of regulated genes which were present or absent for a particular time interval were used to generate the qualitative *in silico* expression arrays, which indicate whether or not a certain gene is being transcribed for a given time interval. Similarly, an *in silico* protein activity array was generated to show the activity of a certain regulatory protein at a given time. The time delay associated with gene transcription and protein synthesis is a required parameter and was estimated for each dynamic simulation (shown in Figures 4.3- 4.5).

Once the regulatory constraints for a particular time step were applied to the system, the optimal flux distribution was calculated as described above. The optimal flux distribution was then used to calculate cell growth, substrate uptake and metabolic by-product secretion using numerical integration as described earlier [199]. The time courses of cell growth, substrate uptake and by-product secretion were generated using a spreadsheet software package (EXCEL, Microsoft Corporation, Redmond).

4.B.3 Mutant strain simulation

To simulate deletion of a metabolic gene from the network, the flux through the corresponding gene product was set to zero for all conditions. Regulatory gene deletion was simulated by setting the expression of the gene to OFF.

4.C Model testing and validation

The regulatory network associated with central metabolism for *E. coli* was reconstructed, and combined with a previously developed central metabolic model [172, 48] (Figure 4.2). The analysis of the combined metabolic/regulatory network using FBA may be called regulatory flux-balance analysis (rFBA). The rFBA network accounts for 149 genes, the products of which include 16 regulatory proteins and 73 enzymes, which catalyze 113 reactions. The synthesis of 43 of the

enzymes in this model is controlled by transcriptional regulation; as a result, the availability of 45 of the reactions to the system is controlled by a logic statement. The details of the rFBA network may be found in the supplemental material.

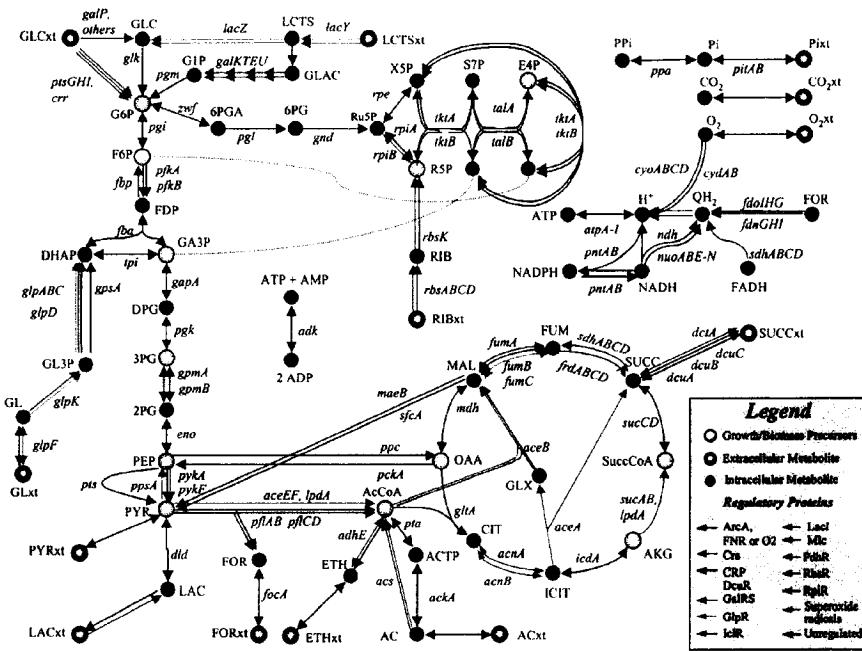


Figure 4.2: Combined regulatory/metabolic network for central metabolism in *E. coli*. All of the metabolic genes considered are shown. The genes which are regulated are indicated by the color code shown in the legend. Genes or reactions regulated by multiple regulatory proteins or molecules are shown with multiple arrows. The complete details of the network may be found in Appendix B.

The metabolic/regulatory network was used in a retrospective analysis of experimental data to determine the ability of the rFBA model to make accurate phenotypic predictions. A mutant study as well as the dynamic simulation of growth under three environmental conditions (aerobic growth on glucose with acetate reutilization, glucose fermentation, and a mixed aerobic glucose-lactose batch culture) is described here.

4.C.1 Mutant growth phenotypes

The rFBA model was used to ascertain the ability of mutant strains of *E. coli* to grow on defined media. For this study, 116 cases were examined and are shown in Table 4.1. In each case, the experimental data were compared with predictions of both the rFBA model and the FBA (i.e. purely metabolic) model [48]. The FBA model alone was able to correctly predict growth characteristics in 97, or 83.6% of the given cases. Incorrect predictions were made for 16 of the 116 cases, and in three of the cases, predictions were not possible for the FBA model alone because *rpiR* is a regulatory gene and therefore only included in the rFBA model.

Table 4.1: Table of comparisons between model predictions and experimental findings for mutant phenotypes. Results are scored as + or - meaning growth or no growth determined from *in vivo*/FBA/rFBA data. An 'N' indicates that data were not available for these conditions. Cases where rFBA makes a correct prediction either unpredicted or incorrectly predicted by FBA alone are denoted by a shaded box. In 106 of 116 cases the rFBA prediction matches the experimentally observed behavior. glc, glucose, gl, glycerol, suc, succinate, ac, acetate, rib, ribose, (-O₂), anaerobic conditions.

	<i>in vivo</i> /FBA/rFBA	glc	gl	suc	ac	rib	glc (-O ₂)	Duals	Ref
<i>aceA</i>	+//+/+			+//+/-	-/-/-		+///+		[33]
<i>aceB</i>				-/-/-					[34]
<i>aceEF</i>	-/+/-			-/+/-	+//+/+		+///+		[103]
<i>ackA</i>					+//+/+				[100]
<i>ackA + pta + acs</i>	+//+/+	+//+/+	+//+/+	+//+/+	-/-/-		+//+/+		[100]
<i>acnA</i> ¹	+//+/+	+//+/+	+//+/+	+//+/+	+//+/+		+//+/+		[34, 66]
<i>acnB</i>	+//+/+	+//+/+	+//+/+	+//+/+	-/+/-		+//+/+		[66]
<i>acnA + acnB</i>	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-		-/-/-		[66]
<i>acs</i>	+//+/+	+//+/+	+//+/+	+//+/+	+//+/+		-/+/-		[100]
<i>adh</i> ²	+//+/+	+//+/+	+//+/+	+//+/+	+//+/+		-/+/-		[35]
<i>cyd</i>									[18]
<i>cyo</i>	+//+/+								[18]
<i>eno</i>	-/-/-	-/-/-	-/-/-				(gl + succ)	+///+	[81]
<i>fbaA</i> ³	-/+/-	-/+/-	-/+/-						[59]
<i>fbp</i>	+//+/+	+//+/+	+//+/+						[58]
<i>frdA</i>	+//+/+	+//+/+	+//+/+						[33]
<i>fumA</i>					-/+/-		+//+/+		[34]

Table 4.1, continued.

<i>in vivo</i> /FBA/rFBA	glc	gl	suc	ac	rib	glc (-O ₂)	Duals	Ref
<i>gap</i>	-/-/-	-/-/-	-/-/-				(gl + succ)	[81]
<i>glk</i>	+/-/+						+/-/+	[59]
<i>glk</i> + <i>pfkA</i>	+/-/+							[59]
<i>glk</i> + <i>pts</i>	+/-/+							[59]
<i>gltA</i>	-/-/-							[66]
<i>gnd</i>	-/-/-							[59]
<i>icd</i> (<i>idh</i>)	+/-/+							[66]
<i>mdh</i>	-/-/-							[28]
<i>ndh</i>	+/-/+	+/-/+	+/-/+					[195]
<i>nuo</i>	+/-/+	+/-/+	+/-/+					[195]
<i>pfl</i>	+/-/+	+/-/+	+/-/+					[118]
<i>pgi</i> ^a + <i>gnd</i>	+/-/-	+/-/-	+/-/-					[59]
<i>pgi</i> + <i>zwf</i>	-/-/-	-/-/-	-/-/-					[59]
<i>pgk</i>	-/-/-	-/-/-	-/-/-					[81]
<i>pgl</i>	+/-/+							[59]
<i>ppc</i>	-/+/-	-/+/-	+/-/+				(gl + succ)	[59, 28]
<i>pta</i>							+/-/+	[100]
<i>pts</i>	+/-/+						(glc + succ)	[59]

Table 4.1, continued.

<i>in vivo</i> /FBA/rFBA	glc	gl	suc	ac	rib	glc (-O ₂)	Duals	Ref
<i>pykA</i>	+/-/+							[59]
<i>pykA + pykF</i>	+/-/+							[59]
<i>pykF</i>	+/-/+							[59]
<i>rpiA</i>	-/+/-				+/-/+		(glc + rib)	[185]
<i>rpiA + rpiB⁵</i>	-/-/-				-/+/+		(glc + rib)	[185]
<i>rpiB</i>	+/-/+				+/-/+		(glc + rib)	[185]
<i>rpiR + rpiA</i>	+/-/(na)/+				+/-/(na)/+		(glc + rib)	[185]
<i>sdhABCD</i>	+/-/+				-/-/-	+/-/+	(glc + succ)	[33]
<i>sucAB-tpd⁶</i>	-/-/+		-/+/+	-/-/-	-/+/+	+/-/+	(glc + succ)	[33, 103]
<i>tpd⁷</i>	-/-/+	-/-/-	-/-/-	-/-/-		+/-/+	(glc + gl)	[81, 3]
<i>zwf</i>		+/-/+				+/-/+		[59]

Table 4.1, continued.

-
- ¹The acnA gene allows the model to grow; the expression of this isozyme is thought to be adapted to stress responses which have yet to be incorporated into the regulatory network described here.
- ²The simulations use the pyridine nucleotide transhydrogenase to grow.
- ³The loss of fba is thought to inhibit stable RNA synthesis, which would have an effect on growth unpredictable by the models.
- ⁴The pgi mutation is lethal in the FBA simulations due to the inability to synthesize glucose
- ⁶-phosphate, a precursor of biomass components which may not be essential for cell growth.
- ⁵It is likely that ribose inhibits growth by an accumulation of ribose 5-phosphate in rpi double mutants.
- ⁶The absence of α -ketoglutarate dehydrogenase results in an inability to produce succinyl CoA, which is used as a precursor metabolite in trace amounts for the synthesis of heme and hemelike compounds, but is not included as an essential biomass component in the model.
- ⁷Dihydroxyacetone phosphate can accumulate in tpi mutants, resulting in formation of the bacteriocidal compound methylglyoxal.

The rFBA model made correct predictions about growth characteristics in 106 or 91.4% of the 116 cases, an improvement of nine correct predictions over the unregulated metabolic model. The mutants whose growth capabilities were correctly predicted by the rFBA, but not the FBA model alone were *aceEF*, *fumA*, *ppc*, *rpiA*, and *rpiR*. The remaining incorrect predictions are detailed in Table 4.1 and are usually due to accumulation of toxic substances, an effect not yet predictable by this approach.

The rFBA model was used to examine the above nine correct predictions in more detail. According to the predictions of the rFBA model, pyruvate dehydrogenase, encoded by the *aceEF-lpdA* operon, is a lethal mutation in *E. coli* for growth on minimal glucose and minimal succinate media under aerobic conditions due to the aerobic down-regulation of its fermentative counterpart, pyruvate formate-lyase. Similarly, fumarase A (*fumA*) is the only fumarase which is generally transcribed under aerobic conditions. Phosphoenolpyruvate carboxylase (*ppc*) was correctly predicted to be a lethal mutation due to the down-regulation of the glyoxylate shunt.

The ribose phosphate isomerase A (*rpiA*) and the ribose repressor protein RpiR illustrate how regulatory gene mutant phenotypes may be simulated using rFBA. Two isomerases exist in *E. coli* for the interconversion of ribulose 5-phosphate and ribose 5-phosphate, encoded for by the *rpiA* and *rpiB* genes. While the expression of *rpiA* is thought to be constitutive, expression of *rpiB* occurs in the absence of RpiR, which is inactivated by ribose [185, 59]. As a result, *rpiA* mutants are ribose auxotrophs while *rpiB* mutants exhibit a null phenotype. The further mutation of *rpiR* in *rpiA* mutants disables repression of *rpiB* and restores the ability to grow in the absence of ribose, as correctly predicted by the rFBA model.

4.C.2 Dynamic growth simulations

The rFBA model was used to simulate growth of *E. coli* quantitatively over the course of growth experiments. The resulting time courses of growth, substrate uptake, and by-product secretion were then compared with experimental data.

E. coli has been observed to secrete acetate when grown aerobically on glucose in batch cultures; when glucose is depleted from the environment, the acetate is then reutilized as a substrate. Using the rFBA and FBA models, an aerobic batch culture of *E. coli* on glucose minimal medium was simulated; the calculations are shown together with experimental data [199] (Figure 4.3). The major difference between the rFBA and FBA simulations is in the delayed reaction of the system to depletion of glucose in the growth medium. The stand-alone metabolic network does not account for the delays associated with protein synthesis.

An *in silico* expression array and a regulatory protein activity array were also generated (Figure 4.3C). The *in silico* array predicted the up-regulation of four gene products, *aceA*, *aceB*, *acs*, and *ppsA*, as well as the down-regulation of three gene products, *adhE*, *ptsGHI-crr*, and *pykF*. DNA microarray technology has been used to detect differential transcription profiles on a collection of 111 genes in *E. coli* [132]. The difference in gene expression for aerobic growth on acetate versus growth on glucose as reported in [132] is included in Figure 4.3C. The calculated expression of the eight genes included in the rFBA model for which expression data were published were in qualitative agreement with the predictions of the rFBA model. The ability of the rFBA model to reutilize acetate depends on the up-regulation of the glyoxylate shunt genes, *aceA* and *aceB*, which explains the high magnitude of transcription difference (20-fold) reported in [132].

Importantly, the rFBA model suggests an interpretation for the regulation of two genes which were found to be regulated but by unknown causes, *ppsA* and *adhE* [132]. The rFBA model indicates that a second regulatory shift is induced by catabolite activator protein Cra, which responds to falling intracellular

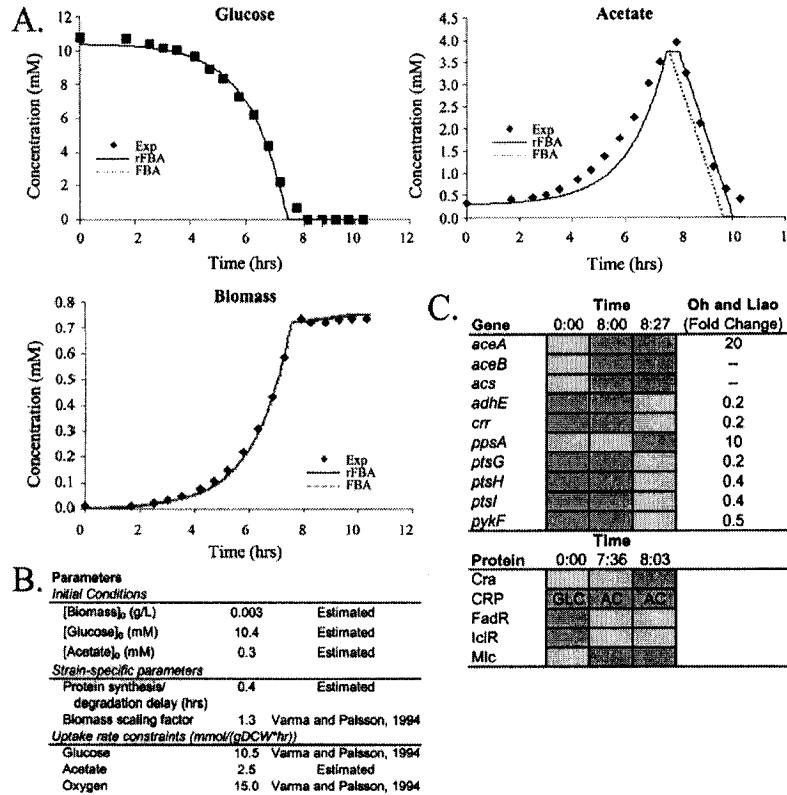


Figure 4.3: Aerobic growth on acetate with glucose reutilization. (A) Three time plots showing experimental data obtained by Varma and Palsson [199], and the corresponding simulations performed using FBA and rFBA. In the acetate plot, the regulatory/metabolic model predictions differ from that of the regulatory plot alone, as shown. (B) A table containing the parameters required to generate the time plots. (C) In silico arrays showing the up- or down-regulation of selected genes, or activity of regulatory proteins, in the regulatory network (dark gray - gene transcription / protein activity, light gray - transcriptional repression / protein inactivity). Data from Oh and Liao [132] showing the experimentally determined transcriptional fold changes for certain genes (acetate: glucose) are shown where applicable. Note that the regulation of CRP is represented by a complex set of Boolean statements (see Appendix B). As a result, CRP activity is represented here as GLC or AC to denote when glucose or acetate is accepted by the system, respectively.

concentrations of fructose 6-phosphate and fructose 1,6-bisphosphate once glucose is depleted from the medium. This second regulatory shift is responsible for up-regulation of *ppsA*, and down-regulation of *adhE*, *crr*, *ptsG*, *ptsHI*, and *pykF*, according to the rFBA model.

The next case studied using the rFBA model was anaerobic growth on glucose [199] (Figure 4.4). Under these conditions, the FBA model makes similar predictions as the rFBA model, with one exception: the rFBA model is able to make predictions about the use of a particular isozyme. For example, both models require fumarase activity as part of the optimal flux distribution; however, only the rFBA model is able to specify the *fumB* gene product which is expressed anaerobically [197].

The final case examined for this study involved aerobic growth of *E. coli* on glucose and lactose. This case has been studied in detail by many groups [99, 211, 1]; for this study, mixed batch culture data as well as the predictions of a detailed kinetic model recently derived by Kremling, et al., were used for comparison [99]. The rFBA model predictions are in good agreement with the data, comparable with the predictions made by the Kremling model, and far better than the predictions of the stand-alone FBA model (Figure 4.5). The failure of the stand-alone FBA model to accurately predict the results of this experiment is due to the concurrent uptake of glucose and lactose, resulting in much more rapid depletion of the substrates and a higher growth rate. Interestingly, because of the larger flux of carbon source uptake, the FBA model predicts that *E. coli* growth should be oxygen-, rather than carbon-limited in this case. Accordingly, the secretion of acetate and formate is predicted. The rFBA model predicts that no secretion will occur under these conditions.

The *in silico* arrays for the experiment (Figure 4.5C) show one shift in gene expression, occurring just under five hours. The up-regulation of the lactose uptake and degradation machinery, together with key enzymes in galactose metabolism, enables the system to use lactose as a carbon source once the glucose

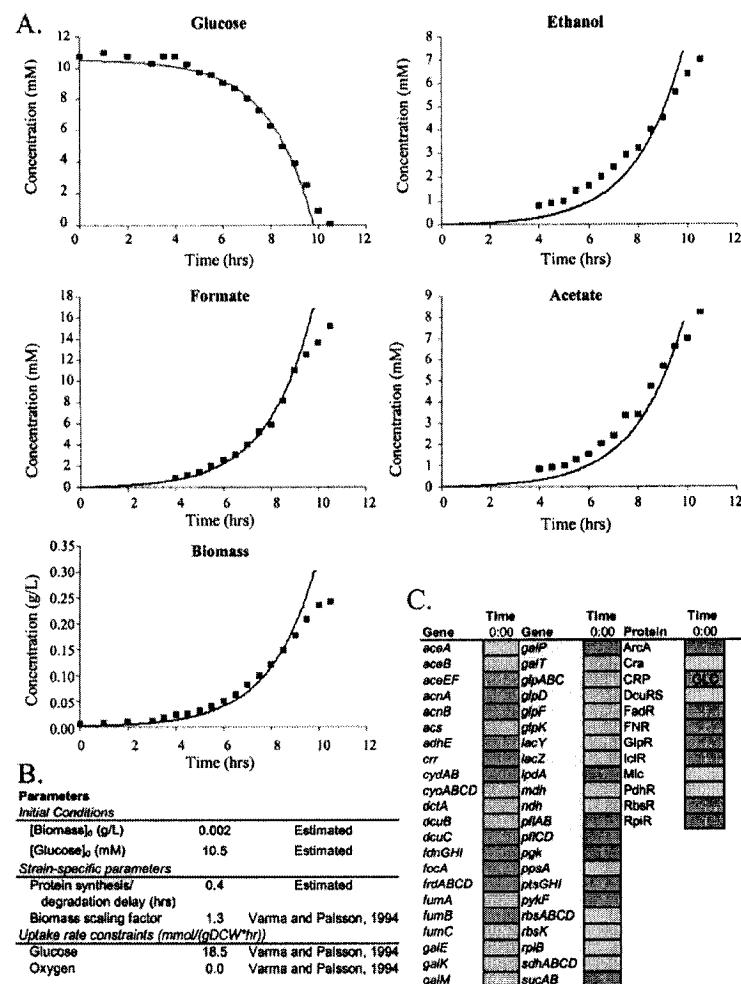


Figure 4.4: Anaerobic growth on glucose. (A) Five time plots similar to those shown in Figure 4.3, again using experimental data obtained by Varma and Palsson [199]. (B) A table containing the parameters required to generate the time plots. (C) In silico arrays for the simulation (similar in format to Figure 4.3).

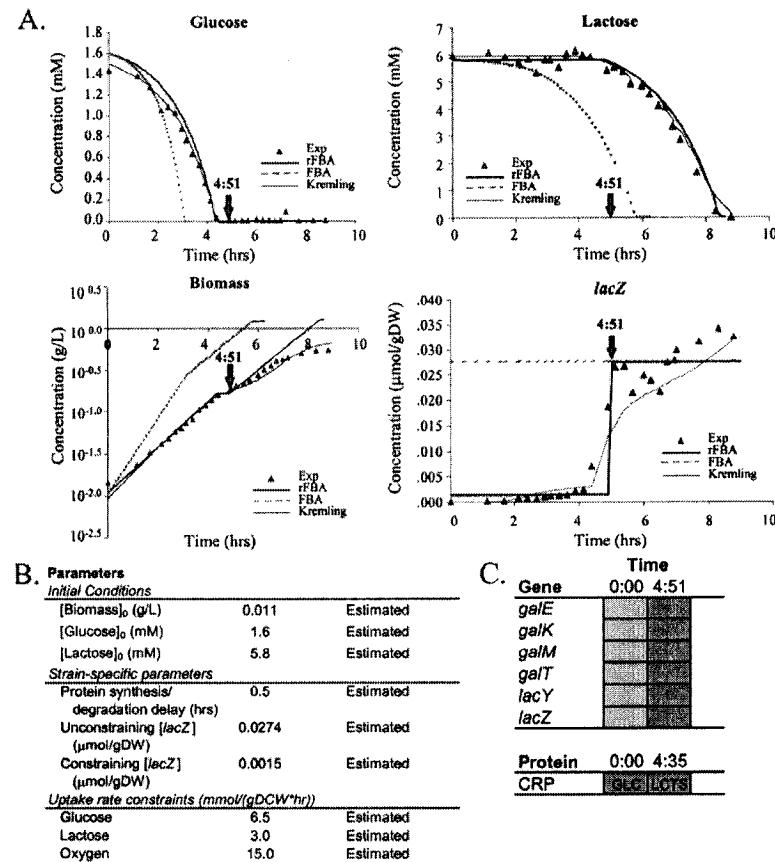


Figure 4.5: Aerobic growth on glucose and lactose. (A) Four time plots showing experimental data and simulation results obtained by Kremling et al. (LacZ refers to the cell concentration of mRNA encoding the *lacZ* gene product) [99], together with the corresponding simulations performed using FBA and rFBA. (B) A table containing the numerical values of the parameters required to generate the time plots for the FBA models. (C) *In silico* arrays for the simulation, similar in format to Figure 4.3, where CRP activity is represented as GLC (glucose) or LCTS (lactose) to denote when lactose is accepted by the system.

in the medium has been depleted.

In summary, several approaches have been developed to mathematically model metabolic regulation and cell dynamics, as recently reviewed [182, 120]. The majority of these approaches rely on extensive kinetic and other environmental information which is difficult to obtain. A combined model of *E. coli* central metabolism and regulation was reconstructed and analyzed using the constraint-based approach [136]. The regulated model was used to simulate the viability of mutant strains, cellular growth and by-product secretion under various environments, and internal regulatory shifts under changing environmental conditions. The results of this study indicate that:

1. The addition of regulatory constraints significantly improves the predictive capacity of flux-balance models where regulatory effects have a dominant role in flux-balance models of metabolism;
2. Regulatory constraints are also useful in the interpretation of dynamic behaviors observed in cell cultures; and
3. Large-scale metabolic models constructed using the constraint-based framework can capture the essential behavioral features and systemic characteristics of an organism with relatively few parameters.

The imposition of regulatory constraints on the solution space formed by the reconstructed network for an organism's metabolism results in a smaller solution space that more accurately represents its behavior. This improved accuracy is readily seen in the results of the mutant study, where regulatory constraints were responsible for the correction of six false predictions made by the stand-alone FBA model. Furthermore, regulatory constraints enable the prediction of phenotype for regulatory gene mutations, as demonstrated by the three *rpiR* mutant predictions made using rFBA. The time courses generated for *E. coli* growth on glucose with acetate reutilization and glucose-lactose diauxie are examples where

regulatory constraints have a substantial impact on the simulation results, causing the simulation to better reflect the actual phenotype of the cell.

The addition of regulatory constraints can be used to interpret simulation results of cellular growth and by-product secretion. The glucose/acetate simulation clearly suggests that upregulation of the glyoxylate shunt enables the reutilization of acetate, and that a second regulatory shift is responsible for regulation of genes such as *ppsA* and *adhE*, both of which were found to be regulated for unknown reasons in a recent microarray study of these conditions [132]. More intuitively, the simulation of glucose-lactose diauxic growth indicates that upregulation of the gal and lac operons is vital to the diauxic shift observed.

By comparing the rFBA simulations with those produced by the stand-alone FBA model, it may also be possible to infer causes of regulatory evolution. For example, in the case of glucose fermentation, the relatively small effect of regulation on the observed phenotype could suggest that this organism has evolved a system which can respond instantaneously to sudden oxygen deprivation. For the case of glucose-lactose diauxic growth, the stand-alone FBA model shows that combined uptake of lactose and glucose could cause the system to be oxygen-, rather than carbon-limited for biomass production, resulting in the secretion of acetate and formate and reducing the growth yield. This finding, combined with evidence that *E. coli* evolves to optimize its growth yield during growth on single-carbon source media [78, 49] and that catabolite repression does not occur under starvation conditions, where the cell is carbon, rather than oxygen-limited [105], suggests the hypothesis that regulation of substrate uptake may have evolved as a means of maintaining optimal growth yields on single substrates. Such a hypothesis obviously remains to be examined in detail, yet is illustrative of the usefulness of such a model in generating hypotheses which address such broad and fundamental topics as regulatory network strategy.

Finally, the ability of the combined model to accurately capture the essential behavioral features and systemic characteristics of central metabolism and

regulation of *E. coli* with relatively few parameters illustrates the utility of the constraint-based framework for modeling microbial systems. This capability to interpret and predict complex biological functions in the absence of detailed parameters is particularly evident in the glucose-lactose example discussed above.

In summary, a central regulatory network for *E. coli* has been generated and the corresponding regulatory constraints have been applied to central metabolism, resulting in improved interpretation and prediction of the behavior of this organism under specified conditions. As this combined metabolic/regulatory network is expanded to reflect the entire state of knowledge about *E. coli* transcriptional regulation, it is expected that the resulting constraints will further constrict the solution space and contribute to a fuller understanding of the systemic properties of this organism. With efforts underway to reconstruct regulatory networks using microarray data [116], it may be expected that rFBA models may soon be generated for other less-characterized organisms. Interestingly, such *in silico* model development does permit the generation of hypotheses about the systems biology of metabolism and regulation which may be difficult to formulate without an *in silico* model.

We have now expanded this central regulatory model to the genome-scale, reconstructing a transcriptional regulatory network for interaction with the existing genome-scale metabolic network previously reconstructed in our lab [154]. The specifics of this model and its ability to generate testable hypotheses based on large-scale high-throughput data sets, is the subject of the next chapter.

The text of this chapter, in part or in full, is a reprint of the material as it appears in “Transcriptional regulation in constraint-based metabolic models of *Escherichia coli*” in *Journal of Biological Chemistry*. I was the primary author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 5

Integrating high-throughput data and computational models leads to *E. coli* network elucidation

High-throughput technologies are yielding large data sets that require network-based data analysis to reconcile heterogeneous data types, find inconsistencies, and systematically generate hypotheses. To begin this process in *Escherichia coli* we reconstructed a genome-scale model of its metabolic and transcriptional regulatory networks. The model was used to computationally predict growth phenotypes of 110 knockout strains under 125 growth conditions (13,750 cases). The computations were consistent with experimental measurement in 10,828 (79%) cases, and resolution of discrepancies between prediction and observation led to identification of 18 areas where the metabolic or regulatory networks are incompletely characterized. To begin further characterization the regulatory network, we mRNA expression profiled wild-type and 6 knockout strains under aerobic and anaerobic conditions. Altered expression of 151 genes represented in the model were detected and 22 of these changes were due entirely to regulation that had been previously described. Model-driven analysis of the remaining cases led to the formulation of 110 new regulatory rules that represent testable hypotheses. We

show that a systems biology approach that combines genome-scale experimentation and computation can systematically generate hypotheses from disparate data sources.

The flood of high-throughput biological data has led to the expectation that genome-scale *in silico* models can be used as a basis for iterative model building [29, 79, 96]. Each iteration involves prediction of experimental outcomes based on an existing model, obtaining the corresponding experimental data, reconciling the predicted outcomes and the computed ones, and using discrepancies to update the *in silico* model by making predictions consistent with the available data (Figure 5.1). This process is implemented here to expand our knowledge about the genome-scale transcriptional regulatory network in *E. coli*.

Over a 13 year period [153], metabolic models of *E. coli* have been developed, culminating in a metabolic reconstruction that accounts for 904 metabolic genes [154]. These models have been used to simulate various data types and whole-cell functions [49, 147]. A small scale transcriptional regulatory network comprised of 16 transcription factors has been used to develop the methods to compute the effects of transcriptional regulation of metabolic network function yielding a limited range of validated phenotypic predictions [31]. Here we have synthesized a genome-scale transcriptional regulatory network that accounts for 104 regulatory proteins, which together with other stimuli regulate the expression of 572 of the 906 metabolic genes in the updated genome-scale metabolic network. The contents of this *in silico* strain (*E. coli* iMC1010) are detailed in the supplemental material.

The *E. coli* iMC1010 strain was, first validated against 13,750 growth phenotypes contained in the ASAP data base [62], and second, was used to select transcription factors for prospective gene knock-out studies. Comparison with the growth phenotypes led to identification of several substrates and knockout strains whose growth behavior did not match predictions (Figure 5.2); further investigation of these conditions and strains led to identification of 5 environmental conditions where dominant regulatory interactions are active but uncharacterized,

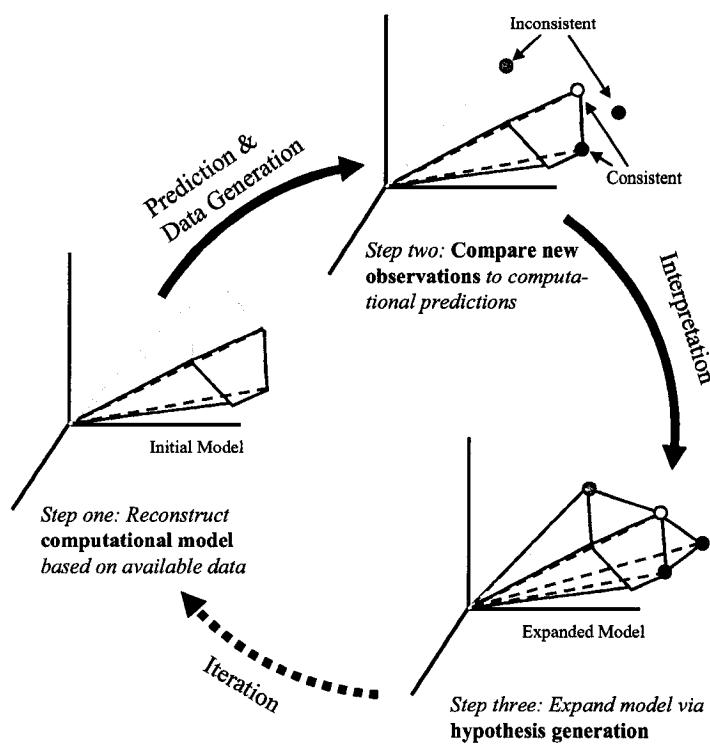


Figure 5.1: Biological network characterization via a model-centric approach. The overall capability of a metabolic network is represented as a solution space (grey and blue), of which only certain of the enzymes are expressed (blue). With such a model we can predict the outcome of experiments *in silico* to compare with actual experimental outcomes, which may be classified as consistent or inconsistent with respect to each other (colors described in more detail in Figure 5.2). If model predictions are consistent with experimental observations, the model is adequately characterized. If not, the model identifies a knowledge gap and may be used to generate testable hypotheses to reconcile predictions and observations. In the process, testable hypotheses are generated, building our understanding of the biological system.

and 5 environmental conditions as well as 8 knockout strains which highlight uncharacterized enzymes or non-canonical pathways which are predicted to exist in the organism (Figure 5.2; see also Appendix C).

In order to select a particular subnetwork for further experimentation, results from an earlier study, which evaluated the consistency between existing gene expression data sets and the known transcriptional regulatory network of *E. coli*, were used [74]. In this study the network was found to have highly consistent sub-networks, or modules (such as the flagellar biosynthesis system) and other modules that were partially consistent with the available data. The response to oxygen deprivation was an example of a partially consistent module and for this reason it was selected for an iterative model driven process of network elucidation. Six knock-out strains involving key transcriptional regulators in the oxygen response ($\Delta arcA^-$, $\Delta appY^-$, Δfnr^- , $\Delta oxyR^-$, $\Delta soxS^-$, and double knockout $\Delta arcA^- \Delta fnr^-$) were constructed [39] and these strains as well as the wild-type strain were mRNA expression profiled in aerobic and anaerobic conditions. The data was analyzed in the context of iMC1010 to generate new causal links in this network that are stated as testable hypotheses.

Beginning with the wild-type strain, we found 437 genes that experienced a significant shift in transcription (t-test, multiple testing corrected to give false discovery rate FDR < 5%). We then used iMC1010 to predict gene expression as well as growth and by-product secretion rates, finding that 75 genes in the model were predicted to change expression significantly. Comparisons between computation and experimental data are shown in Figure 5.3a and Figure 5.3c. Given that no experimental data was obtained for 28 of the predicted shifts, only about one half of the predicted shifts were actually observed in the data. Additionally, 151 of the 437 genes which shifted were included in the model, and 23 of these (15.2%) were identified correctly. The predicted:experimental shift comparisons fell into the following five categories: correctly predicted shifts (represented in Figure 5.3 as P:E), correctly predicted stable expression (0:0), predicted shifts

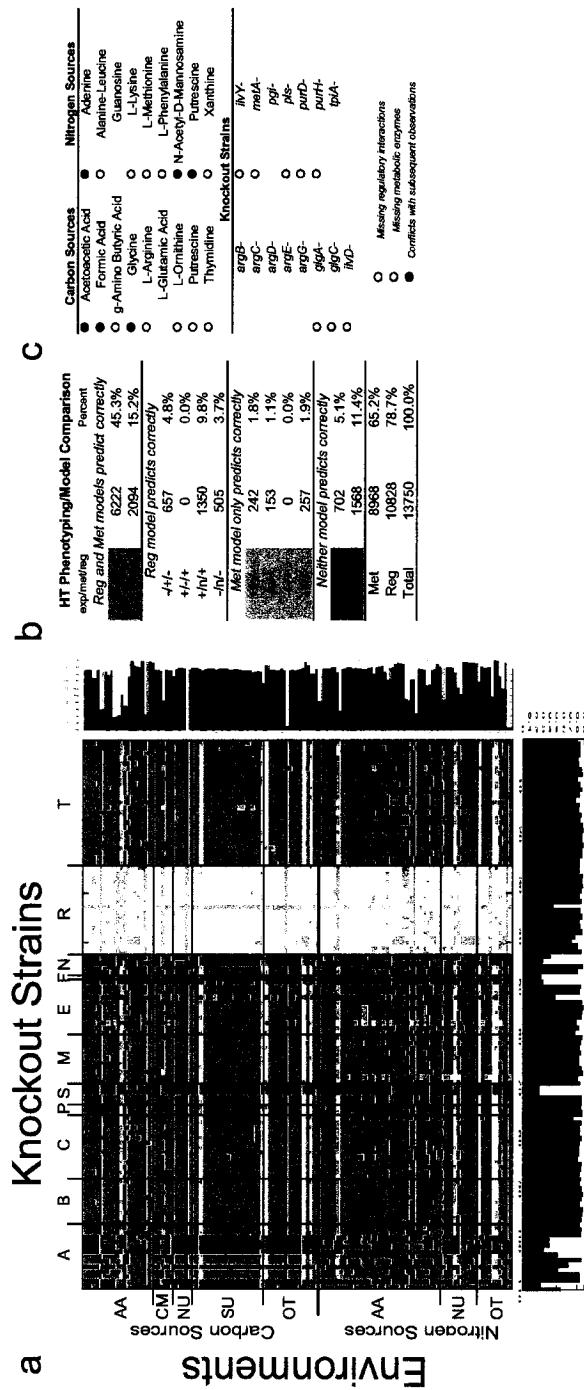


Figure 5.2: Comparison of high-throughput phenotyping array data with predictions for the *E. coli* network, both considering regulatory constraints (reg) and ignoring such constraints as a control (met). The results of the comparison are divided into four subgroups, represented by four colors as indicated on the chart. Also shown is a table which contains all environments or knockout strains for which the measured growth phenotypes were consistent with model predictions for < 60% of the cases. Several of these substrates or knockout strains point to as-yet uncharacterized metabolic or regulatory capabilities in this organism, as indicated.

which were opposite to measured shifts (P:-E), falsely predicted shifts which were measured as stable (P:0), and significant shifts which were predicted to be stable (0:E). Because the model represents our current knowledge about the *E. coli* transcriptional regulatory network, these results indicate substantial gaps in our knowledge of regulation of the aerobic/anaerobic shift.

We compared the gene expression data for the wild type and each knockout strain separately using two-way analysis of variance (ANOVA) to determine whether the shift was significantly altered in the knockout strain compared to the wild type shift. A large portion of the shifts observed for the wild-type strain were not significantly disturbed in any of the strains (195/437 overall, 63/151 in the model, FDR < 5%), suggesting that none of the five transcription factors deleted regulate their expression. The remainder of the genes exhibited abolished shifts in one or more of the knockout strains (Figure 5.3b).

Identification of various transcription factors that contribute to shifts in expression enables us to rewrite, relax or remove various regulatory rules in the model. For example, a gene exhibiting a significant positive shift in its expression under anaerobic conditions which is abolished in the $\Delta arcA^-$ and $\Delta arcA^-fnr^-$ knockout strains would be assigned a new regulatory rule: if ArcA is active, then the gene is expressed. We used ANOVA to generate similar rules to resolve the discrepancies between iMC1010 and the experimentally determined wild-type gene shifts. For many of the cases, a rule already existed and had to be reconciled with our new data to accommodate previously determined transcription factor dependencies. In cases where none of the knockouts had an altered expression shift, we simply based the regulatory rule on the presence of oxygen rather than a transcription factor. In cases where a shift was predicted but not observed, we removed the oxygen dependency from the existing rule. Even in cases where the predicted shifts agreed with the observed expression in the wild type, our perturbation analysis indicated that the underlying factors in the model needed to be changed. Such changes occurred with 10 genes.

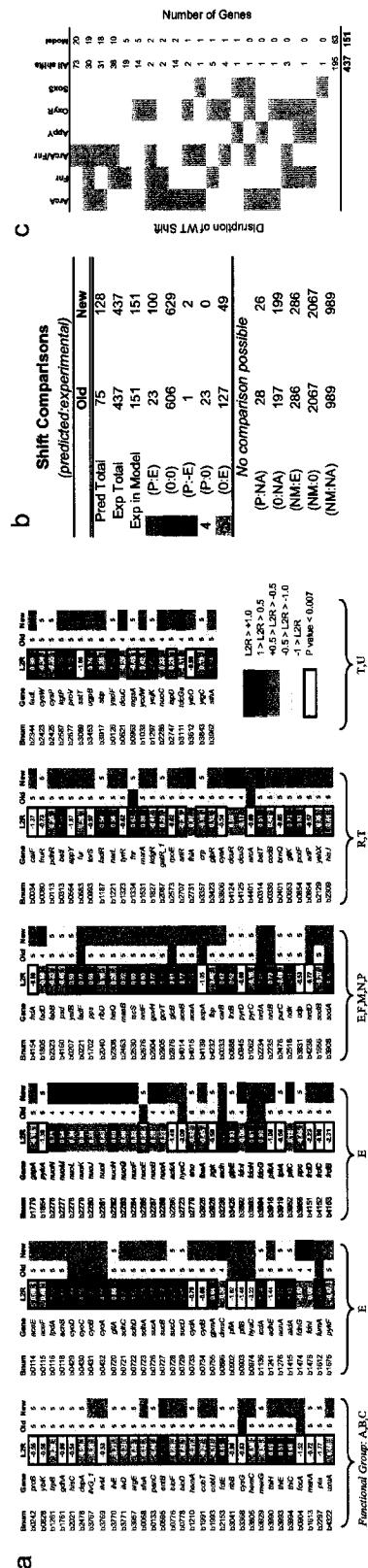


Figure 5.3: Characterizing the regulatory network related to the aerobic-anaerobic shift. One hundred and fifty-one of the genes in the model exhibited a significant shift from anaerobic to aerobic conditions in the wild type strain (divided by functional group, with the same abbreviations as in Figure 2). The B numbers, gene names and the log₂ ratio of gene expression are shown for all the genes with either predicted or observed shifts. Log₂ ratios enclosed by a box indicate a statistically significant ($P < 0.007$, FDR = 5%) shift. The predictions of the original model (Old) and updated model (New) are shown, together with a table tallying the results of comparing the predicted and observed shifts for both models.

(b) Using systematic perturbation analysis to determine the transcription factors responsible for the shift. Each combination of contributing factors which abolished the shift, as determined by analysis of variance, is shown together with the number of genes affected.

The addition of these regulatory rules led to three main observations. First, some of the results of the knockout perturbation analysis were complex enough to defy definition of a rule. For example, the interplay of Fnr and ArcA can lead to complex behaviors where the $\Delta arcA^-$ or the Δfnr^- strains, but not the $\Delta arcA^-fnr^-$ strain, leads to abolition of the expression shift observed in wild type. Such complex interplay between transcription factors can lead to specialized shifts, as has been observed in the *cydAB* response to anaerobic, microaerobic and aerobic conditions [25, 27]. Second, in revising transcription rules for transcription factors, we found that while in many cases, such as *arcA*, expression of a regulatory protein correlates positively with its activity, in several cases, including *fnr*, *betI*, and *fur*, among others, the transcription of a regulatory gene is reduced when in fact, the protein is activated. For example, under anaerobic conditions, when Fnr is known to be active, its expression is significantly reduced. Such behavior suggests that identification of regulatory networks, and particularly transcription factors, will not be accomplished by the determination of co-regulated gene sets alone. Third, many of these gene expression shifts involve complex interactions and indirect effects. Transcription factors may be affected, for example, by the presence of fermentation by-products, build-up of internal metabolites and the like. Such effects would be extremely difficult to identify or account for without a computational model.

The updated model (iMC1010+) was used to recalculate all of the predictions for both the aerobic/anaerobic shift data and the high-throughput phenotyping arrays. For the aerobic and anaerobic predictions, and comparison with the experimental data was substantially higher, as predicted (Figure 5.3a,b). Specifically, 99 of the 151 shifts were correctly computed, and the number of false positive (P:0) predictions was reduced to zero. In resolving many of the unpredicted shift (0:E) cases we found that implementation of the given rule resulted in the inability of the wild type or knockout *in silico* strain to grow under aerobic or anaerobic conditions. Because the growth rates of wild type and knockout strains under aerobic

($0.68 \pm 0.04/\text{hr}$) and anaerobic ($0.43 \pm 0.07/\text{hr}$) clearly indicated normal growth under these conditions, we relaxed the rule to allow for a correct phenotype. Surprisingly, data/prediction comparison for the high-throughput phenotyping data revealed very little difference from Figure 5.2.

This study has demonstrated several significant results (Figure 5.4). First, it has shown that high-throughput data from disparate sources can be assembled into genome-scale models that in turn can be used to prospectively design experiments whose results can be used to generate specific hypothesis. Second, the number of hypotheses so generated is large, calling for the development of high-throughput methods for testing such hypothesis. Third, the results for the oxygen shift in *E. coli* suggest that we may only know about one-fourth of the transcriptional regulatory interactions in this commonly studied process, and further suggest that model-driven interrogation of genome-scale transcriptional regulatory networks will accelerate their elucidation greatly. Taken together, the results show that the promise of the so-called systems biology approach in cell and molecular biology is beginning to be realized.

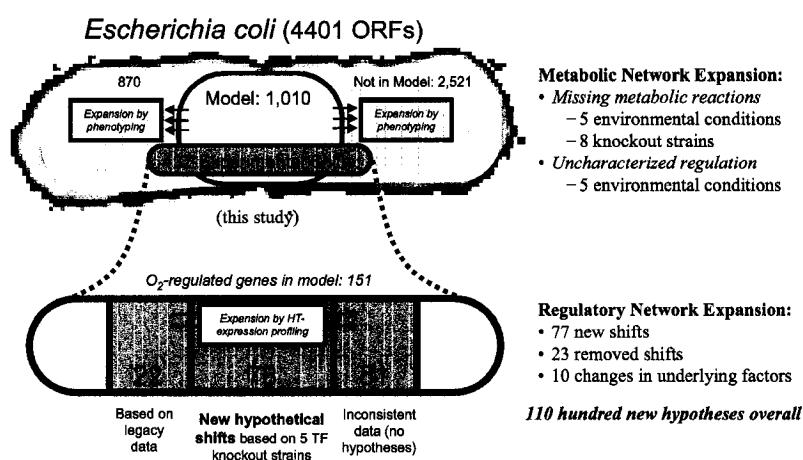


Figure 5.4: Summary of the hypotheses generated in this study (included in more detail in Appendix C). The metabolic and regulatory networks may be expanded using high-throughput phenotyping data and gene expression data to update, validate and generate hypotheses about organism function. *E. coli* image courtesy of www.denniskunkel.com.

The text of this chapter, in part or in full, is a reprint of the material as it appears in “Integrating high-throughput data and computational models leads to *E. coli* network elucidation” which has been submitted for publication. I was the primary author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 6

Vistas

Even on top of Everest, I was still looking at other mountains and thinking of how one might climb them. – Sir Edmund Hilary[188].

We have now reached the summit of this work. At the top of any peak, the first reward is to enjoy the view. First, we look back at where we have been. From the top, one gets a sense of the journey that doesn't easily come from the circuitous path. We see the precarious ledge, the almost-too-wide chasm, and the slake-covered peak in their proper perspective. Similarly, we look back on the fresh-water spring, the unexpected meadow, and the living forest, remembering the pleasure in their discovery. I will spend the first half of this chapter looking back, remembering the difficult leaps as well as the unexpected discoveries.

Looking back can only take us so far, however. Just as Sir Hilary, one of the greatest adventurers of all time, if we would discover new horizons, we must look ahead, identify the next seemingly insurmountable peaks and consider how these may also be conquered. I will conclude this chapter and this work with some thoughts on which future challenges await, and where to start.

6.A Looking back: lessons

6.A.1 The constraint-based approach enables modular modeling

The first, and pivotal lesson of this work is that regulatory events can be incorporated with metabolic network behavior using the constraint-based framework, and that they have strong effects on metabolic behavior. This was not intuitive, for several reasons. Most importantly, for simple conditions, the optimal growth behavior of the regulated and unregulated networks is sometimes the same, or very similar. Growth on glucose, glycerol or succinate minimal media are some examples [49]. For this reason it seemed as though the simple introduction of an objective function could obviate the need for regulatory information. In fact, some of my earlier efforts involved trying to infer regulatory structure itself from a calculated optimal flux distribution. It is now clear that gene expression and metabolic flux are not necessarily linearly related.

Another hurdle involved the choice of a formalism to represent regulation. My eventual choice of a Boolean representation was somewhat controversial because it seemed “too simple” to capture complex regulatory interactions. Although other descriptions may describe the process in more detail [99], logic rules are simply the only representation which could accommodate the limits of available regulatory data and computation to allow for genome-scale simulations [14]. We will explore this point in more detail in the challenges section of this chapter.

In retrospect, we can now see how overcoming these challenges led to development of the first large-scale modular model. Modularization is often discussed as a way of bringing different modeling methods together, representing each in the most appropriate way [152]. Modular or “hybrid” models have been seen as the only way to computationally capture the distinct processes involved in an organism [5]. This work demonstrated that the constraint-based framework allows quite naturally for modularization, simply by representing regulatory events as constraints on metabolic processes (see Figure 6.1). Such modularization will be

key in reconciling the different processes which are part of biological systems.

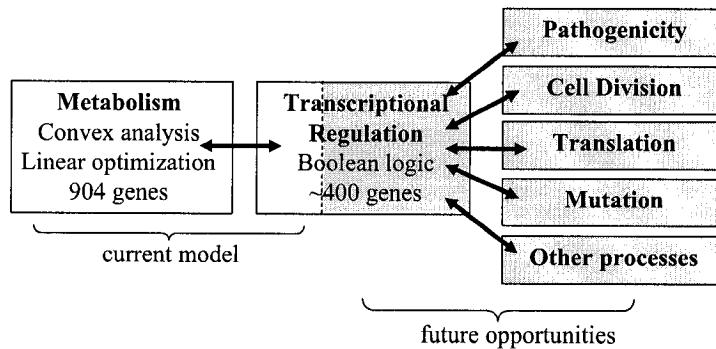


Figure 6.1: Constraint-based analysis enables modular modeling. The model described here integrates metabolism and regulatory structure seamlessly, simply by interpreting regulatory events as imposing constraints on the metabolic system. By expanding the model to include all transcription factors and regulated genes, we may be able to include other processes as shown on the right. Such processes also impose constraints which may potentially be incorporated in our model.

6.A.2 The dynamic solution space

The second lesson is that the solution space of a metabolic network is in fact dynamic. Although it is bounded by its components, it is constantly shifting over time as gene expression is repressed and induced. There is a core part of the network which remains static. We could probably guess that a part of this network is always upregulated or constitutively expressed (this part of the network is related to the environment-independent “minimal gene set” [77], and has not yet been expressly defined). A non-obvious finding of this study is the inverse: a part of this network is in fact never expressed due to inconsistent regulation.

This finding has several important applications. The first relates to recent studies of metabolic network topology. Recent studies of cellular metabolic networks [85, 205] have characterized such networks as scale-free or “small world”.

These studies have only been based on genomic constraints (i.e., any gene located in the genome was considered a part of the network). Given that the entire metabolic network of a cell is never completely expressed at a given time, one is left to wonder whether the topology of the metabolic network remains "small world" over time. This question could be answered through further transcriptomic and proteomic studies.

Another important by-product of this finding is the enhanced power of the constraint-based framework in simulating dynamics. Some dynamic studies using flux-balance analysis have already been attempted [199, 115], but were not broadly applicable because regulatory events were not accounted for [199, 109]. We saw in Chapter 3 how incorporation of relatively few constraints could lead to very complex behaviors over time. This new framework enables us to analyze, interpret and predict these behaviors.

6.A.3 Model-centric discovery

The most exciting development came in Chapter 5, with the utilization of the model together with high-throughput data to better characterize the *E. coli* regulatory and metabolic networks. It is my conviction that a model-centric approach such as the one described in Chapter 5 will be the only way to completely characterize any biological system. I further believe that high-throughput data generation will lack meaning in the absence of a predictive model to evaluate, validate, and interpret our results in the context of what is already known. Interpretation of large-scale data sets can not be accomplished by any human mind.

Taken together, this points to the necessity of a dialogue between experimental biologists, who are generally more aware of biological constraints, and computational biologists, who are equipped to describe constraints mathematically. As such dialogues multiply and grow, I believe that genome-scale constraint-based models will develop to include all major cellular processes and enable prediction and interpretation of large-scale datasets, challenges which we discuss as we turn

to look down the opposite slope of the mountain.

6.B Other mountains: challenges

I believe the major challenge we now face is in generating whole-system models for many systems, especially the model organisms. This is a veritable Himalayan range of challenges and opportunities and ends with “Mount Everest” itself: *Homo sapiens*. Such models which relate the genotype to phenotype could be critical in helping us to understand the causes of complex diseases. Some of these accomplishments are decades away, while others are already in progress. I will discuss some of the preliminary steps towards these goals here: broadening the genome-scale *E. coli* model to account for the entire genome, rendering the model in ever-increasing specificity as information becomes available, and expanding the constraint-based approach to simulate behavior of multi-cellular organisms.

6.B.1 Completion of the *E. coli* model

Our first challenge is to build the *E. coli* model to account for all the genes in the K-12 genome. What I have described here is primarily focused on metabolism and the corresponding regulation in *E. coli*. The logical next step in this work is to build the *E. coli* model described here to account for all of the transcription factors, diverse stimuli, and regulated genes in the system.

Such a model would have two major impacts: first, it would enable further and more encompassing studies such as described in Chapter 5, where we target sections of the regulatory network for better characterization and systematically perturb the system to identify regulatory interactions. It should be emphasized here that the perturbations need to be environment-focused as well as gene knockout-focused, in part because systems and metabolic flux distributions are often robust to a genetic perturbation, but far less so to environmental variation (Uwe Sauer, personal communication). Such a strategy will have a far-reaching

impact in completely and systematically characterizing *E. coli*.

The second major impact of such an expanded model is the opportunity to account for more biological processes as part of the constraint-based framework. By including the up- and down-regulation of genes responsible for certain processes in our model (see Figure 6.1), we can also begin to incorporate the constraints imposed by such processes. What are the constraints imposed on the cell, for example, by cell division, and how can we incorporate these effects in a meaningful way? Certainly the expression of genes is associated with constraints such as the energy required to synthesize proteins, energy and charge balance issues associated with highly charged macromolecules, and simply the volume of the cell, which can only contain a certain total amount of protein. Modeling transcriptional regulation opens the door to address such questions because it has widespread effects which are now clearly definable in terms of regulatory rules.

The result of this effort would be a broader, more complete model of *E. coli* which could be used to address a variety of questions. For the first time, the interactions of subsystems normally assumed to operate in relative isolation, could be examined at the genome-scale. Like the complex feedback between metabolism and regulation we saw in Chapter 5, I expect that such system-level studies will generate exciting new insights about the *E. coli* system which may be relevant to many other organisms.

6.B.2 Accounting for fine-resolution data

The first challenge focused on breadth, while our next challenge is concerned with depth. I have mentioned several times here that we chose linear optimization and Boolean logic for our metabolic and regulatory representations because more detailed information was not available at the genome-scale. However, for the case of *E. coli*, a relatively large amount of detail has in fact been determined in terms of characterized enzymes, transcription factor-operator site binding efficiencies, and the like. It would be unfortunate not to be able to account

for such data as it becomes available. How can we incorporate this knowledge into our models?

The accommodation of new, more detailed knowledge can also be addressed using the constraint-based approach. Such an attempt should be initiated at the small scale, perhaps by revisiting the small network introduced in Chapter 3, as shown in Figure 6.2. As described earlier, Carbon1 in the small network has a regulatory effect on transcription of transport protein Tc2, preventing uptake of Carbon2. Instead of using a Boolean description of this process, if we know some details about the mechanism of this regulatory interaction, as depicted on the right side of Figure 6.2, we should be able to simulate the dynamic behavior of this response over time. The inputs and outputs are identical in both representations: extracellular concentration of Carbon1 and flux of Carbon2 traveling into the cell via the transport protein. However, the simulations will be quite different.

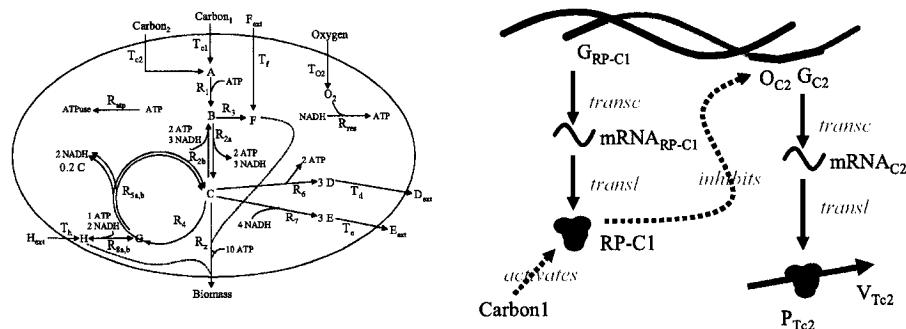


Figure 6.2: Incorporating partial kinetic knowledge as a constraint. Here the simplified network from Chapter 2 is shown at the left, and a more detailed mechanism of the catabolite repression of Carbon1 on the transport of Carbon2 is shown at on the right.

A major gain from adding kinetic information, even if limited, will be in the prediction of metabolite or protein concentrations. The steady-state assumption required for flux-balance analysis makes it difficult to determine con-

centrations, although approaches are being devised to bracket the ranges of such concentrations. How much inclusion of such information will affect prediction capability remains to be seen. Overall, one might expect that for simple cases like the LacI transcription factor, it would seem to be less important from a predictive sense than for a case like regulation via the Crp or Cra proteins, whose activities depend on the concentration of internal cAMP and fructose 1,6-bisphosphate, respectively, and where therefore kinetic representations might be much more accurate. By adding detailed process descriptions to our broad genome-scale model, we gain the opportunity to see how such processes affect the whole organism and vice versa.

6.B.3 Modeling multicellular organisms

The completion of a broad and somewhat detailed model of *E. coli* will pave the way for reconstruction of many other microorganisms, hopefully assisting discovery and advances in infectious disease, metabolic engineering and bioremediation as well as basic science. The next major hurdle will be to describe the behavior of multicellular organisms. Such organisms have cells which are differentiated to perform various functions, expressing genes particular to their objectives. Multiple protein complements implies that more than one solution space will be involved in multicellular simulations.

A simple example of this may be found among the cyanobacteria. The cyanobacterium *Anabaena* sp. PCC 7120 grows two types of cells: vegetative cells and terminally-differentiated heterocysts which fix atmospheric nitrogen. With the annotated genome sequence recently available, *Anabaena* is a strong candidate for developing the first constraints-based model of a multi-cellular organism (Figure 6.3). To reconstruct the metabolic and where possible the regulatory networks of vegetative and heterocyst *Anabaena* cells, and model the interaction between the cells [88], would allow us to model and investigate the effects on cell behavior of interaction between two constrained solution spaces with different objectives.

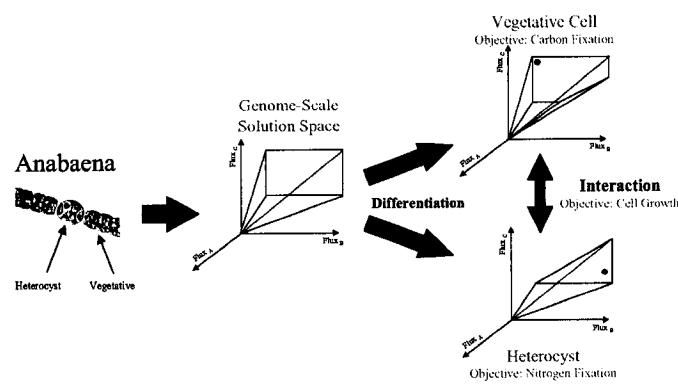


Figure 6.3: Using the constraints-based approach to model multicellular organism *Anabaena* PCC 7120. The known metabolic and transcriptional regulatory networks are reconstructed as has been described, resulting in an overall solution space for the organism. This solution space is not truly achievable to the cell due to gene expression differences in the vegetative cell and the terminally differentiated heterocyst. To describe the cell more accurately, a solution space must be defined for each cell. The interaction between the cells delivers fixed nitrogen to the vegetative cell, enabling it to grow.

This in turn would enable us to address important questions about the nature of differentiation. How much do the metabolic solution spaces of vegetative and terminally differentiated cells overlap? What advantage does the organism have by utilizing two solution spaces instead of one? It is possible that certain microorganism populations grow and evolve in such a way as to have the minimal set of reactions (i.e., edges of a solution space) that still enables them to have the best possibilities for growth in their environment? The reasons for such a “minimal solution space” hypothesis may include the energy required for transcription and translation, or perhaps more importantly, the adverse osmotic effects of holding highly charged protein in the cell. What is the overall objective of multicellular organism, and can we infer computationally the objective for the component cells?

6.C Conclusion

During the year of this writing we celebrated the 50th anniversary of two major human achievements: Edmund Hilary and Tenzing Norgay’s ascent of Mount Everest, and Jim Watson and Francis Crick’s landmark paper identifying the structure of DNA. While Hilary’s success signaled the end of a journey, Watson and Crick’s marked the beginning of one, the culmination of which arguably occurred with the sequence of the human genome. However, even from that summit there are further questions and uncharted territory as far as the eye can see, with one challenge leading naturally to next “as Alps on Alps arise” [145]. It is my firm belief that these peaks can only be scaled when experimentalists tightly integrate their work with that of computational modelers, both groups tied together as tightly as Hilary and Norgay were, working together to reach our common end – unraveling the complexity of life.

Appendix A

Flux maps for extreme pathways in the sample network

The flux distribution maps for all 80 of the extreme pathways calculated for the sample network in Chapter 3. The active fluxes are shown with thick dark lines, inactive fluxes are denoted by thin light dotted lines and the biomass flux is shown with a dashed line. Note that reactions R_{5a} and R_{5b} are isozymes. The 21 extreme pathways which are infeasible under every environmental condition are enclosed in boxes.

Figure A.1: Flux distribution maps for the 80 extreme pathways.

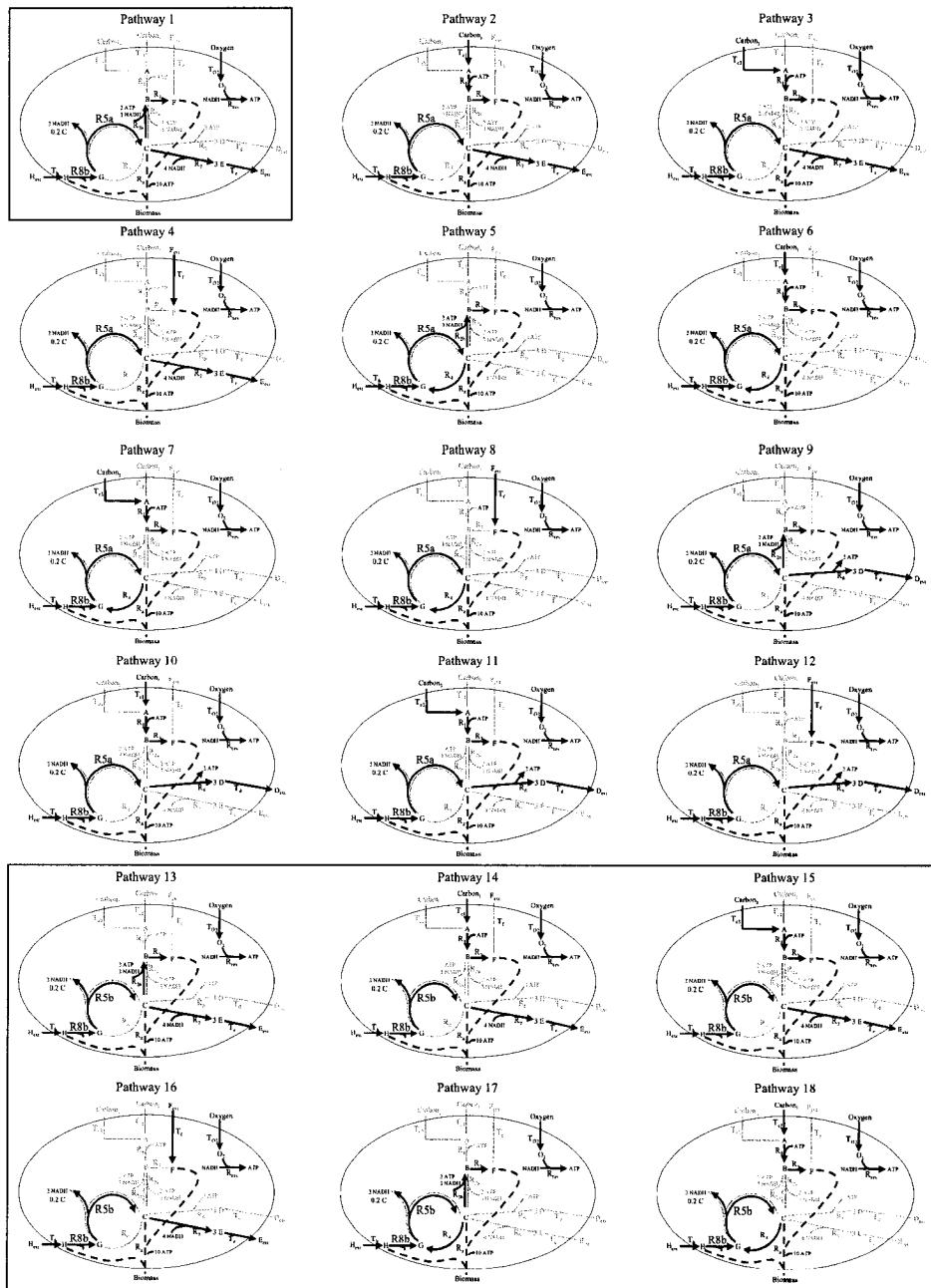


Figure A.1, continued.

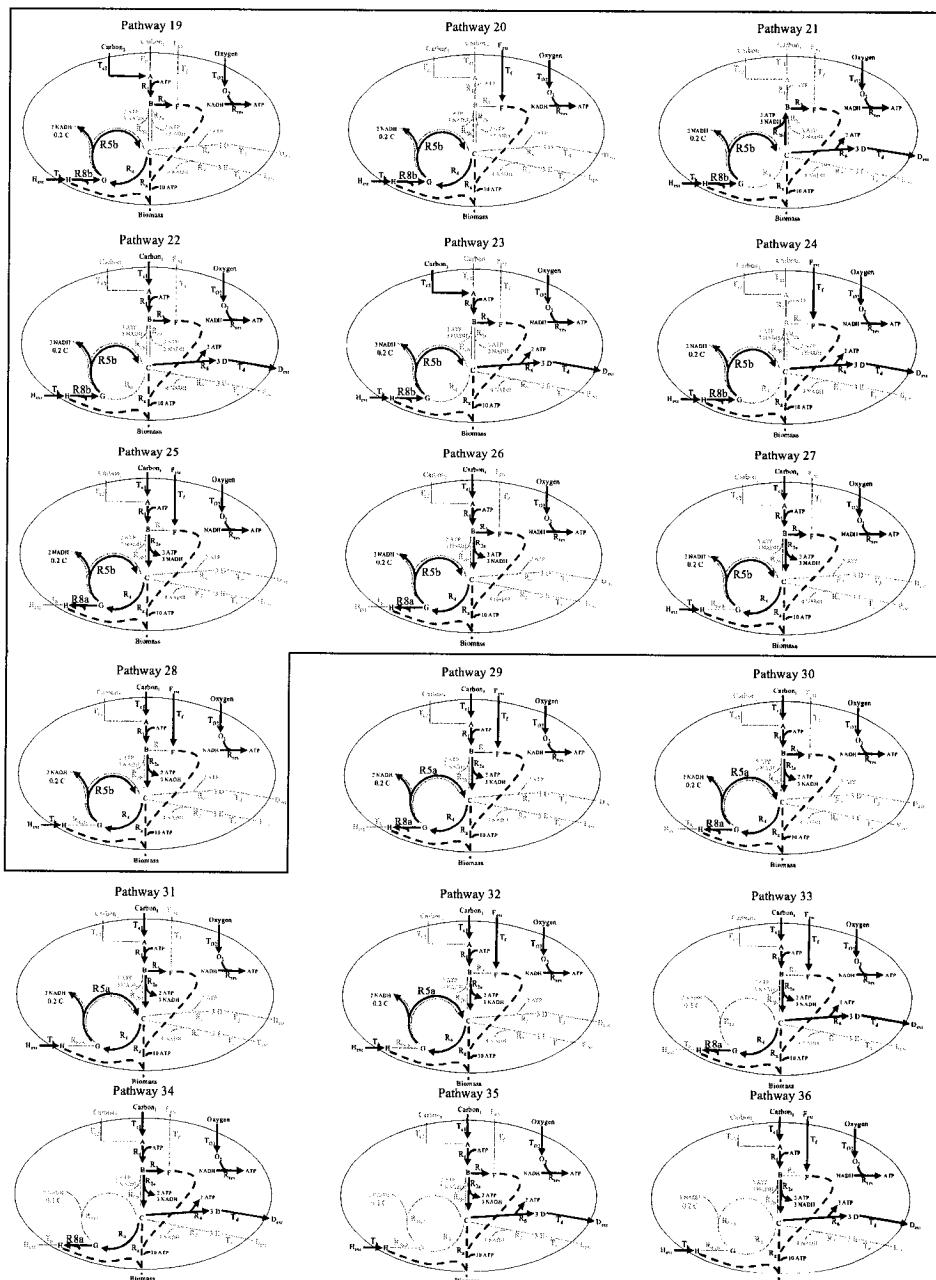


Figure A.1, continued.

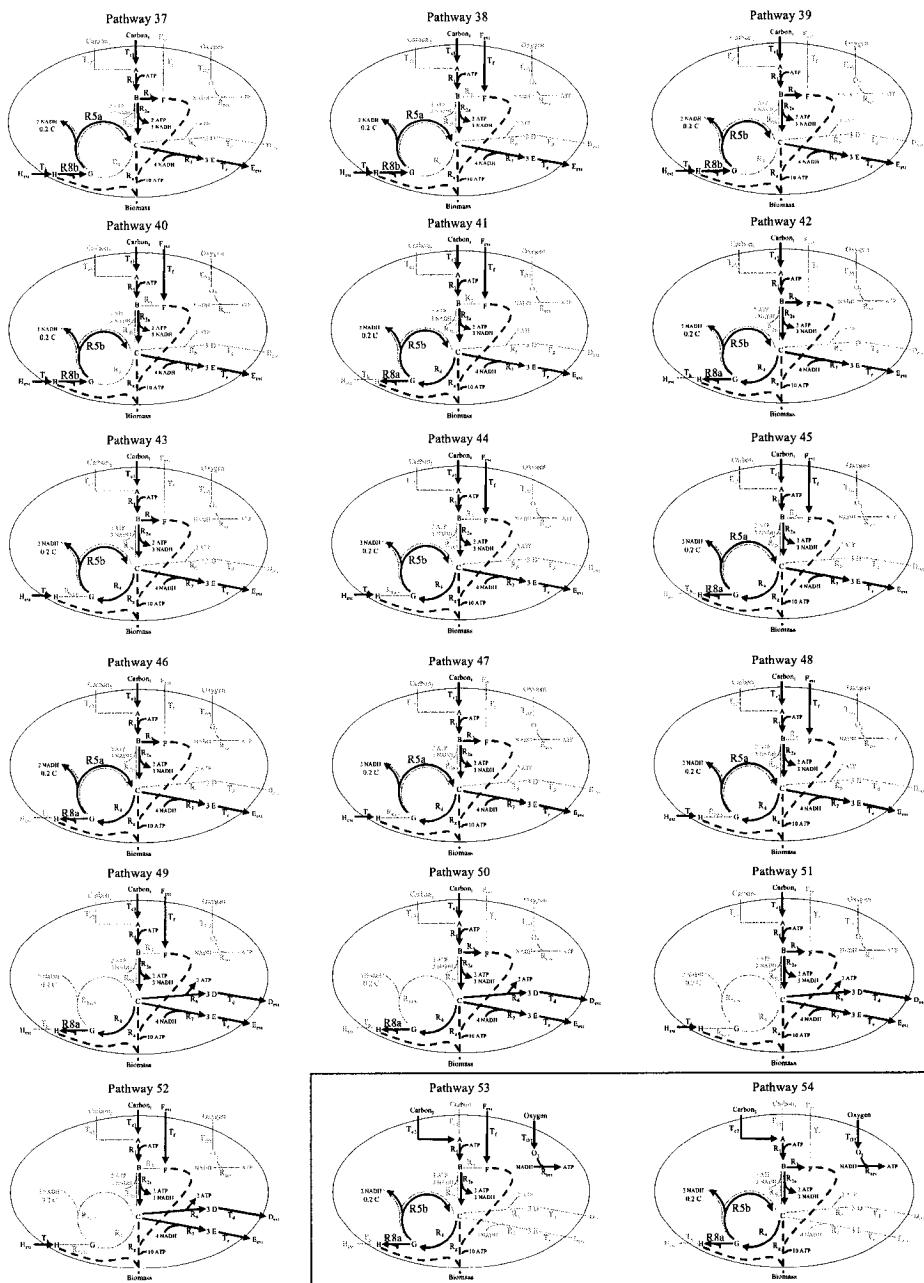


Figure A.1, continued.

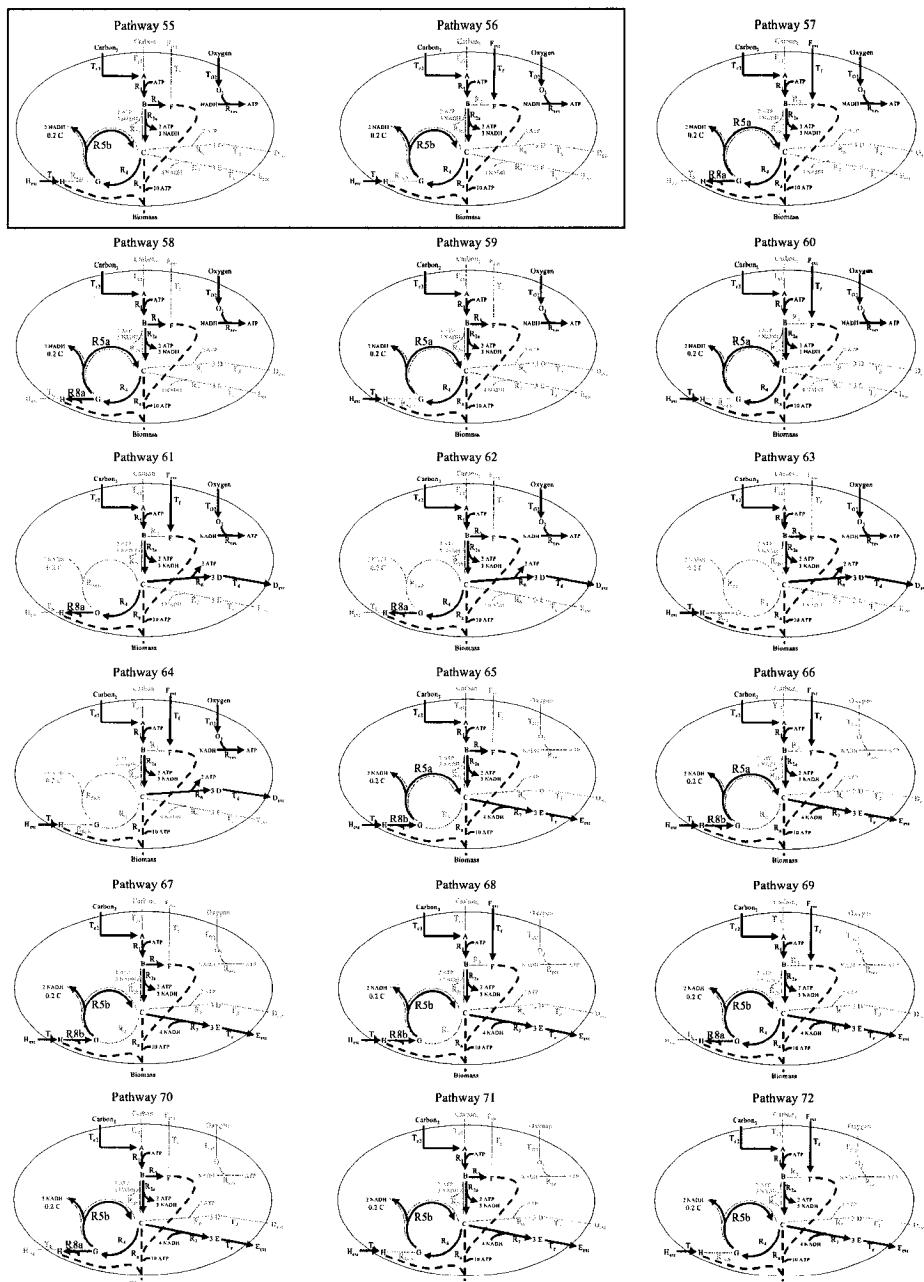
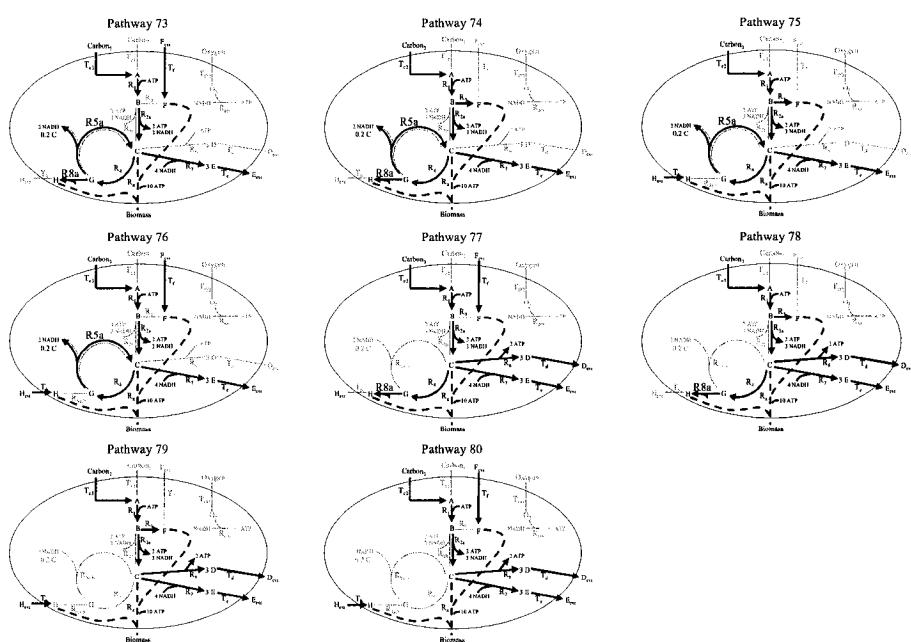


Figure A.1, continued.



The material in this appendix, in part or in full, is a reprint of the material as it appears in “Constraints-based models: regulation of gene expression reduces the steady-state solution space”, in the *Journal of Theoretical Biology*. I was the primary author and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

Appendix B

Escherichia coli Central Regulatory/Metabolic Model

This appendix corresponds to the work described in Chapter 4. There are two tables here:

1. A list of the metabolic reactions and regulatory rules in the *Escherichia coli* central regulatory/metabolic model, and
2. A table with the metabolite abbreviations.

Table B.1: A list of all the metabolic reactions and regulatory rules in the *Escherichia coli* core model.

Protein	Gene	Reaction	Regulatory logic	References
Isocitrate lyase	<i>aceA</i>	$\text{ICIT} \rightarrow \text{GLX} + \text{SUCC}$	IF not (IclR)	[34, 68, 26]
Malate synthase A	<i>aceB</i>	$\text{ACCOA} + \text{GLX} \rightarrow \text{COA} + \text{MAL}$	IF not (ArcA or IclR)	[34, 68, 112, 26]
Pyruvate dehydrogenase	<i>aceEF, lpdA</i>	$\text{PYR} + \text{COA} + \text{NAD} \rightarrow \text{NADH}$	IF (not PdhR)	[34, 151]
Acetate kinase A		$\text{ACTP} + \text{ADP} \leftrightarrow \text{ATP} + \text{AC}$		
Aconitase A	<i>ackA</i>	$\text{CIT} \leftrightarrow \text{ICIT}$	IF (GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt)	[36]
Aconitase B	<i>acnB</i>	$\text{CIT} \leftrightarrow \text{ICIT}$	IF (GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt)	[36]
Acetyl-CoA synthetase	<i>acs</i>	$\text{ATP} + \text{AC} + \text{COA} \rightarrow \text{AMP} + \text{PPI} + \text{ACCOA}$	IF not (GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt and (not IclR))	[101]
Acetaldehyde dehydrogenase	<i>adhE</i>	$\text{ACCOA} + 2 \text{ NADH} \leftrightarrow \text{ETH} + 2 \text{ NAD} + \text{COA}$	IF not (O2xt) or not (O2xt and Cra)	[123, 124]
Adenylylate kinase	<i>adk</i>	$\text{ATP} + \text{AMP} \leftrightarrow 2 \text{ ADP}$		
F0F1-ATPase	<i>atpABCFDEFGHI</i>	$\text{ATP} \leftrightarrow \text{ADP} + \text{Pi} + 4 \text{ HEXT}$	IF (not FNR) or ArcA	[196]
Cytochrome oxidase bd	<i>cydAB</i>	$\text{QH}_2 + .5 \text{ O}_2 \rightarrow \text{Q} + 2 \text{ HEXT}$		
Cytochrome oxidase bo3	<i>cyoABCDF</i>	$\text{QH}_2 + .5 \text{ O}_2 \rightarrow \text{Q} + 2.5 \text{ HEXT}$	IF not (ArcA or FNR)	[196]
D-Lactate dehydrogenase 1	<i>dld</i>	$\text{PYR} + \text{NADH} \leftrightarrow \text{NAD} + \text{LAC}$		
D-Lactate dehydrogenase	<i>dld</i>	$\text{LAC} + \text{Q} \rightarrow \text{PYR} + \text{QH}_2$		

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
Enolase	<i>eno</i>	2PG \leftrightarrow PEP		
Fructose-1,6-bisphosphate aldolase	<i>fba</i>	FDP \leftrightarrow T3P1 + T3P2		
Fructose-1,6-bisphosphatase	<i>fp</i>	FDP \rightarrow F6P + PI	IF FNR	[106, 38]
Formate dehydrogenase-N	<i>fdnGH</i>	FOR + Q \rightarrow QH2 + CO2 +2 HEXT	IF FNR or DeuR	[64, 196]
Formate dehydrogenase-O	<i>fdtHG</i>	FOR + Q \rightarrow QH2 + CO2 +2 HEXT	IF not (ArcA or FNR)	[139]
Fumarate reductase	<i>frdABCD</i>	FUM + FADH \rightarrow SUCC + FAD	IF FNR or FNR	[197]
Fumarase A	<i>fumA</i>	FUM \leftrightarrow MAL	IF FNR	[139]
Fumarase B	<i>fumB</i>	FUM \leftrightarrow MAL	IF (superoxide radicals)	[139]
Fumarase C	<i>fumC</i>	FUM \leftrightarrow MAL	IF not (GLCxt) and not (GalR or GalS)	[1]
UDP-glucose 4-epimerase	<i>gale</i>	UDPGAL \leftrightarrow UDPG	IF not (GLCxt) and not (GalR or GalS)	
Galactokinase	<i>galK</i>	GLAC + ATP \leftrightarrow GAL1P + ADP	IF not (GLCxt) and not (GalR or GalS)	[1]
Aldose 1-epimerase (mutorotase)	<i>galM</i>	bDGLAC \leftrightarrow GLAC	IF not (GLCxt) and not (GalR or GalS)	[1]
Aldose 1-epimerase (mutorotase)	<i>galM</i>	bDGLC \leftrightarrow GLC	IF not (GLCxt) and not (GalR or GalS)	[1]
Galactose-1-phosphate uridylyltransferase	<i>galT</i>	GAL1P + UTP \leftrightarrow PPI + UDPGAL	IF not (GLCxt) and not (GalR or GalS)	[1]
UDP-glucose-1-phosphate uridylyltransferase	<i>galU</i>	G1P + UTP \leftrightarrow UDPG + PPI		
Glyceraldehyde-3-phosphate dehydrogenase-A Complex	<i>gapA</i>	T3P1 + PI + NAD \leftrightarrow NADH + 13PDG		[22]
Glucokinase	<i>glk</i>	GLC + ATP \rightarrow G6P + ADP		

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
Glycerol-3-phosphate dehydrogenase (anaerobic)	<i>gphABC</i>	$\text{GL3P} + \text{Q} \rightarrow \text{T3P2} + \text{QH2}$	IF not (GLCxt or LCTSxt or RIBxt) and FNR and not GpR	[135, 110]
Glycerol-3-phosphate dehydrogenase (aerobic)	<i>gphD</i>	$\text{GL3P} + \text{Q} \rightarrow \text{T3P2} + \text{QH2}$	IF not (GLCxt or LCTSxt or RIBxt) and not (ArcA or GpR)	[84]
Glycerol kinase	<i>gpkK</i>	$\text{GL} + \text{ATP} \rightarrow \text{GL3P} + \text{ADP}$	IF not (GLCxt or LCTSxt or RIBxt) and not GpR	[135, 110]
Citrate synthase	<i>gltA</i>	$\text{ACCOA} + \text{OA} \rightarrow \text{COA} + \text{CIT}$		[206]
6-Phosphogluconate dehydrogenase (decarboxylating)	<i>gnd</i>	$\text{D6PGC} + \text{NADP} \rightarrow \text{NADPH}$		[138]
Phosphoglycerate mutase 1	<i>pgmA</i>	$+ \text{CO2} + \text{RL5P}$		
Phosphoglycerate mutase 2	<i>pgmB</i>	$3\text{PG} \leftrightarrow 2\text{PG}$		
Glycerol-3-phosphate-dehydrogenase -[NAD(P)+]	<i>gpsA</i>	$3\text{PG} \leftrightarrow 2\text{PG}$		
		$\text{GL3P} + \text{NADP} \leftrightarrow \text{T3P2}$		
		$+ \text{NADPH}$		
		$\text{ICIT} + \text{NADP} \leftrightarrow \text{CO2}$		
		$+ \text{NADPH} + \text{AKG}$		
		$\text{LCTS} \rightarrow \text{GLC} + \text{bDGGLAC}$		
		$\text{MAL} + \text{NADP} \rightarrow \text{CO2}$		
		$+ \text{NADPH} + \text{PYR}$		
Beta-galactosidase (LACTase)	<i>lacZ</i>		IF not (GLCxt) and not (<i>LacI</i>)	[1]
Malic enzyme (NADP)	<i>maeB</i>			
Malate dehydrogenase	<i>mdh</i>	$\text{MAL} + \text{NAD} \leftrightarrow \text{NADH} + \text{OA}$	IF not ArcA	
NADH dehydrogenase II	<i>ndh</i>	$\text{NADH} + \text{Q} \rightarrow \text{NAD} + \text{QH2}$	IF not FNR	
NADH dehydrogenase I	<i>nuoABEF</i>	$\text{NADH} + \text{Q} \rightarrow \text{NAD} + \text{QH2}$		
		+3.5 HEKT		
Phosphoenolpyruvate carboxykinase	<i>pckA</i>	$\text{OA} + \text{ATP} \rightarrow \text{PEP} + \text{CO2} + \text{ADP}$		

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
Phosphofructokinase	<i>pfkA</i>	F6P + ATP → FDP + ADP		
Phosphofructokinase B	<i>pfkB</i>	F6P + ATP → FDP + ADP		
Pyruvate formate lyase 1	<i>pflAB</i>	PYR + COA → ACCOA + FOR	IF (ArcA or FNR)	[112, 168]
Pyruvate formate lyase 2	<i>pflCD</i>	PYR + COA → ACCOA + FOR	IF (ArcA or FNR)	[112, 168]
Phosphoglucose isomerase	<i>pgi</i>	G6P ↔ F6P		
Phosphoglycerate kinase	<i>pgk</i>	13PDG + ADP ↔ 3PG + ATP	IF (GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt)	[22]
6-Phosphogluconolactonase	<i>pgl</i>	D6PGL → D6PGC		
Phosphoglucomutase	<i>pgm</i>	G1P ↔ G6P		
Pyridine nucleotide transhydrogenase	<i>pntAB</i>	NADPH + NAD → NADP + NADH		
Pyridine nucleotide transhydrogenase	<i>pntAB</i>	NADP + NADH + 2 HEXT → NADPH + NAD		
Inorganic pyrophosphatase	<i>ppa</i>	PPI → 2 PI		
Phosphoenolpyruvate carboxylase	<i>ppc</i>	PEP + CO2 → OA + PI		
Phosphoenolpyruvate synthase	<i>ppsA</i>	PYR + ATP → PEP + AMP + PI	IF (Cra)	[128]
Phosphotransacetylase	<i>pta</i>	ACCOA + PI ↔ ACTP + COA		
Pyruvate Kinase II	<i>pykA</i>	PEP + ADP → PYR + ATP		
Pyruvate Kinase I	<i>pykF</i>	PEP + ADP → PYR + ATP	IF (not Cra)	[13]
Ribokinase	<i>rbsK</i>	RIB + ATP → R5P + ADP	IF not (GLCxt or LCTSxt) and not (RbsR)	[184, 90] [111]
Ribulose phosphate 3-epimerase	<i>rpe</i>	RL5P ↔ X5P		
Ribose-5-phosphate isomerase A	<i>rpiA</i>	RL5P ↔ R5P		
Ribose-5-phosphate isomerase B	<i>rpiB</i>	RL5P ↔ R5P	IF not RpiR	[186]

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
Succinate dehydrogenase	<i>sdhABCD</i>	SUCC + FAD \rightarrow FADH + FUM	IF not (ArcA or FNR)	[140]
Succinate dehydrogenase complex	<i>sdhABCD</i>	FADH + Q \leftrightarrow FAD + QH2	IF not (ArcA or FNR)	[140]
Malic enzyme (NAD)	<i>sfcA</i>	MAL + NAD \rightarrow CO2 + NADH + PYR		
2-Ketoglutarate dehydrogenase	<i>sucAB, lpdA</i>	AKG + NAD + COA \rightarrow CO2 + NADH + SUCCOA	IF (not PdhR)	[34, 151]
Succinyl-CoA synthetase	<i>sucCD</i>	SUCCOA + ADP + PI \leftrightarrow ATP + COA + SUCC		[140, 37]
Transaldolase A	<i>talA</i>	T3P1 + S7P \leftrightarrow E4P + F6P		
Transaldolase B	<i>talB</i>	T3P1 + S7P \leftrightarrow E4P + F6P		
Transketolase I	<i>tktA</i>	R5P + X5P \leftrightarrow T3P1 + S7P		
Transketolase I	<i>tktA</i>	X5P + E4P \leftrightarrow F6P + T3P1		
Transketolase II	<i>tktB</i>	R5P + X5P \leftrightarrow T3P1 + S7P		
Transketolase II	<i>tktB</i>	X5P + E4P \leftrightarrow F6P + T3P1		
Triosephosphate Isomerase	<i>tpiA</i>	T3P1 \leftrightarrow T3P2		
Glucose 6-phosphate-1-dehydrogenase	<i>zwf</i>	G6P + NADP \leftrightarrow D6PGL + NADPH		
B. Transport Fluxes				
Acetate transport		ACxt + HEXT \leftrightarrow AC		
Carbon dioxide transport		CO2xt \leftrightarrow CO2		
Ethanol transport		ETHxt + HEXT \leftrightarrow ETH		
Formate transport	<i>focA</i>	FORxt \leftrightarrow FOR	IF (ArcA or FNR)	[112, 168]
Glucose transport	<i>ptsGHI, crr</i>	GLCxt + PEP \rightarrow G6P + PYR	IF (GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt) and (not (Mlc) or not (Cra))	[189, 159] [82, 43]

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
Glucose transport (low affinity)	<i>galP</i> , etc.	GLC _{xxt} + HEXT → GLC	IF (GLC _{xxt} or LCTS _{xxt} or RIB _{xxt} or GLxt or LAC _{xxt} or PYR _{xxt} or SUCC _{xxt} or ETH _{xxt} or AC _{xxt} or FOR _{xxt})	[55, 61, 72]
Glycerol transporter	<i>glpF</i>	GLxt ↔ GL	IF not (GLC _{xxt} or LCTS _{xxt} or RIB _{xxt}) and not GlpR	[206]
Lactate uptake		LAC _{xxt} + HEXT → LAC	IF not (GLC _{xxt} or LCTS _{xxt} or RIB _{xxt} or GLxt)	[135]
Lactate secretion		LAC → LAC _{xxt} + HEXT	IF not (GLC _{xxt}) and not (lacI)	[1]
Lactose permease	<i>lacY</i>	LCTS _{xxt} + HEXT ↔ LCTS _{xxt}		
Oxygen transport		O ₂ xt ↔ O ₂		
Phosphate transport	<i>pitAB</i>	PIxt + HEXT ↔ PI		
Pyruvate transport		PYR _{xxt} + HEXT ↔ PYR		
Ribose transport	<i>rbsABCD</i>	RIB _{xxt} + ATP → RIB + ADP + PI	IF not (GLC _{xxt} or LCTS _{xxt}) and not RbsR	[184, 90, 111]
Succinate transport	<i>dctA</i>	SUCC _{xxt} + HEXT ↔ SUCC	IF not (GLC _{xxt} or LCTS _{xxt} or RIB _{xxt} or GLxt or LAC _{xxt} or PYR _{xxt}) and not ArcA and DcuR	[40]
Succinate transport	<i>dcuA</i>	SUCC _{xxt} + HEXT ↔ SUCC		[63]
Succinate transport	<i>dcuB</i>	SUCC _{xxt} + HEXT ↔ SUCC		[63]
Succinate efflux	<i>dcuC</i>	SUCC → SUCC _{xxt} + HEXT	IF (FNR or ArcA)	[214]

Table B.1, continued.

Protein	Gene	Reaction	Regulatory logic	References
C. Maintenance/Biomass				
ATP non-growth associated maintenance flux		ATP → ADP + PI		
Biomass production flux		41.25 ATP + 3.54 NAD + 18.22 NADPH + 0.2 G6P + 0.07 F6P + 0.89 R5P + 0.36 E4P + 0.12 T3P1 + 1.49 3PG + 0.51 PEP + 2.83 PYR + 3.74 ACCOA + 1.78 OA + 1.07 AKG → 3.74 COA + 41.25 ADP + 41.25 PI + 3.54 NADH + 18.22 NADP + 1 Biomass		
D. Exchange Fluxes				
Acetate exchange	ACxt	→		
Carbon dioxide exchange	CO2xt	→		
Ethanol exchange	ETHxt	→		
Formate exchange	FORxt	→		
Glucose exchange	GLCxt	→		
Glycerol exchange	GLxt	→		
Biomass exchange (growth)	Biomass	+ 13 ATP → 13 ADP + 13 PI		
Lactate exchange	LACxt	→		
Lactose exchange	LCTSxt	→		
Oxygen exchange	O2xt	→		
Phosphate exchange	PIxt	→		
Pyruvate exchange	PYRxxt	→		
Ribose exchange	RIBxt	→		
Succinate exchange	SUCCxt	→		

Protein	Gene	Reaction	Regulatory logic	References
E. Regulatory Proteins				
Aerobic/Anaerobic response regulator	<i>arcA</i>		active IF not (O2xt)	
Catabolite activator protein	<i>cra</i> (<i>fruR</i>)		active IF not (surplus FDP or F6P)*	
Catabolite repressor protein	<i>crp</i>		action is complex and highlighted in italics above	
Dicarboxylate response regulator	<i>dcuR</i>		active IF DcuS	
Dicarboxylate response sensor	<i>dcaS</i>		active IF (SUCCxt)	
Fatty acid/Acetate response regulator	<i>fadR</i>		active IF (GLCxt) or not (ACxt)	
Aerobic/Anaerobic response regulator	<i>fnr</i>		active IF not (O2xt)	
Galactose operon repressor	<i>galR</i>		active IF not (GLAC)	
Galactose operon repressor	<i>galS</i>		active IF not (GLAC)	
Glycerol response regulator	<i>glpR</i>		active IF not (GLxt)	
Fatty acid/Acetate response regulator	<i>iclR</i>		active IF FadR	
Lactose operon repressor	<i>lacI</i>		active IF not (LCTxt)	
Glucose response regulator	<i>mlc</i>		active IF not (GLCxt)	
Pyruvate response regulator	<i>pdhR</i>		active IF not (surplus PYR)**	
Ribose response regulator	<i>rbsR</i>		active IF not (RIBxt)	
Ribose response regulator	<i>rpiR</i>		active IF not (RIBxt)	

Table B.2: A list of all the metabolic reactions and regulatory rules in the *Escherichia coli* core model.

Abbreviation	Metabolite
13PDG	1,3-bis-Phosphoglycerate
2PG	2-Phosphoglycerate
3PG	3-Phosphoglycerate
AC	Acetate
ACCOA	Acetyl-CoA
ACTP	Acetyl-phosphate
ACxt	External acetate
ADP	Adenosine diphosphate
AKG	α -Ketoglutarate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
bDGLAC	b-D-Galactose
bDGLC	b-D-Glucose
Biomass	Cell biomass
CIT	Citrate
CO2	Carbon dioxide
CO2xt	External carbon dioxide
COA	Coenzyme A
D6PGC	D-6-Phosphate-gluconate
D6PGL	D-6-Phosphate-glucono-delta-lactone
E4P	Erythrose 4-phosphate
ETH	Ethanol
ETHxt	External ethanol
F6P	Fructose 6-phosphate
FAD	Flavin adenine dinucleotide
FADH	Flavin adenine dinucleotide reduced
FDP	Fructose 1,6-diphosphate
FOR	Formate
FORxt	External Formate
FUM	Fumarate
G1P	Glucose 1-phosphate
G6P	Glucose 6-phosphate
GAL1P	Galactose 1-Phosphate
GL	Glycerol
GL3P	Glycerol 3-phosphate
GLAC	Galactose
GLC	α -D-Glucose
GLCxt	External glucose
GLX	Glyoxylate
GLxt	External glycerol
HEXT	External H ⁺

Table B.2, continued.

Abbreviation	Metabolite
ICIT	Isocitrate
LAC	D-Lactate
LACxt	External lactate
LCTS	Lactose
LCTSxt	External Lactose
MAL	Malate
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide reduced
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Dihydronicotinamide adenine dinucleotide phosphate reduced
O2	Oxygen
O2xt	External Oxygen
OA	Oxaloacetate
PEP	Phosphoenolpyruvate
PG	Phosphatidyl glycerol
PI	Phosphate (inorganic)
PIxt	External phosphate
PPI	Pyrophosphate
PYR	Pyruvate
PYRxt	External pyruvate
Q	Ubiquinone
QH2	Ubiquinol
R5P	Ribose 5-phosphate
RIB	Ribose
RIBxt	External ribose
RL5P	Ribulose 5-phosphate
S7P	sedo-Heptulose
SUCC	Succinate
SUCCOA	Succinate CoA
SUCCxt	External succinate
T3P1	Glyceraldehyde 3-phosphate
T3P2	Dihydroxyacetone phosphate
UDPG	UDP Glucose
UDPGAL	UDP Galactose
UTP	Uridine triphosphate
X5P	Xylulose-5-phosphate

The text of this appendix, in part or in full, is a reprint of the material as it appears in "Transcriptional regulation in constraint-based metabolic models of *Escherichia coli*" in *Journal of Biological Chemistry*. I was the primary author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Appendix C

Escherichia coli Full Regulatory/Metabolic Model

This appendix corresponds to the work described in Chapter 5. It consists of the following sections:

1. A list of the genes in the *Escherichia coli* full regulatory/metabolic model, together with the corresponding regulatory rule.
2. A table with the metabolite abbreviations.
3. A more in-depth treatment of the genes and environments for which the predicted and observed phenotypic behavior was found to be inconsistent in Figure 5.3.
4. A list of the new rules determined from the analysis shown in Figure 5.4.

C.A *Escherichia coli* gene list

The regulatory model is based on a pre-existing model of metabolism in *Escherichia coli*[154]. There are some differences between the metabolic network used in this study and that which was presented earlier; the changes are summarized in Table C.1.

Table C.1: Reactions added to or changed in the iMC1010 *E. coli* model.

Rxn Abbr	Gene	Reaction	Comment
GALU	galU	[c]g1p + h + utp ↔ ppi + udpg	Removed
ORNTA	ygiG	[c] : akg + orn → glu-L + glu5sa	Removed
PTRCA	ygiG	[c] : akg + ptrc → 4abutn + glu-L	New GPR Assoc.
ABUTD	aldH	[c]4abutn + h2o + nad → 4abut + (2) h + nadh	New Isozyme
TRPAS1	none	[c]cys-L + h2o → h2s + nh4 + pyr	New Isozyme
CYSabc	none	atp[c] + cys-L[e] + h2o[c] → adp[c] + cys-L[c] + h[c] + pi[c]	New Isozyme
CYSabc	none	atp[c] + cys-L[e] + h2o[c] → adp[c] + cys-L[c] + h[c] + pi[c]	New Isozyme
INDOLEt	acrEF	h[c] + indole[c] → h[e] + indole[e]	New Reaction
H2SO	none	[c] : h2s + (2) o2 → (2) h + so4	New Reaction
H2St	none	h2s[c] → h2s[e]	New Reaction
EX_h2s(e)	none	[e]h2s ↔	New Reaction

Each gene is listed by B number and gene name, and the regulatory rules for expression or activity (in the case of transcription factors) are included with references listed by their PubMed Ids (which can be entered into PubMed to obtain the full references). Chapters from the book “*Escherichia coli and Salmonella: cellular and molecular biology*” edited by F.C. Neidhardt, are indicated first by an NH (e.g. chapter 22 is listed as NH 22). We also have additional comments on certain aspects of the regulatory rules as listed after the table.

Table C.2: Regulatory rules for the full *Escherichia coli* model. Every gene in the model is listed here by B number, together with gene abbreviation, regulatory rule, and references. To keep the bibliography of this text manageable, the PubMed ID number for many references is listed with the preceding abbreviation PMID. Similarly, a chapter from *Escherichia coli and Salmonella: cellular and molecular biology* (F. C. Neidhardt, editor, Washington, DC., ASM Press, 1996) may be listed with “NH” preceding.

bNum	Gene	Rule	Reference
b0002	<i>thrA</i>	(NOT (THRxt > 0 OR ILExt > 0))	NH 32
b0003	<i>thrB</i>	(NOT (THRxt > 0 OR ILExt > 0))	NH 32
b0004	<i>thrC</i>	(NOT (THRxt > 0 OR ILExt > 0))	NH 32
b0007	<i>yaaJ</i>		
b0008	<i>talB</i>		
b0019	<i>nhaA</i>	((NhaR) OR (RpoS))	PMID: 11133959
b0020	<i>nhaR</i>	(NAxt > 0)	PMID: 11133959
b0025	<i>ribF</i>		
b0029	<i>lytB</i>		
b0031	<i>dapB</i>	(NOT LYSxt > 0)	NH 32
b0032	<i>carA</i>	(NOT ArgR)	NH 25; PMID: 9457878
b0033	<i>carB</i>	(NOT ArgR)	NH 25; PMID: 9457878
b0034	<i>caif</i>	(Fnr AND Crp AND NOT NarL)	PMID: 10564497, 8631699
b0036	<i>caid</i>	(Crp AND CaiF)	PMID: 10564497
b0038	<i>caib</i>	(Crp AND CaiF)	PMID: 10564497
b0040	<i>cait</i>	(Crp AND CaiF)	PMID: 10564497
b0048	<i>folA</i>		
b0049	<i>apaH</i>		
b0052	<i>pdxA</i>	(RpoE)	PMID: 11844765
b0061	<i>araD</i>	(AraC OR (AraC AND Crp))	NH 20
b0062	<i>araA</i>	(AraC OR (AraC AND Crp))	NH 20
b0063	<i>araB</i>	(AraC OR (AraC AND Crp))	NH 20
b0064	<i>araC</i>	(ARABxt > 0)	NH 20
b0066	<i>sfuC</i>		
b0067	<i>sfuB</i>		
b0068	<i>sfuA</i>		
b0071	<i>leuD</i>	(NOT(LEUxt > 0) OR Lrp)	
b0072	<i>leuC</i>	(NOT(LEUxt > 0) OR Lrp)	
b0073	<i>leuB</i>	(NOT(LEUxt > 0) OR Lrp)	NH 27
b0074	<i>leuA</i>	(NOT(LEUxt > 0) OR Lrp)	NH 27
b0077	<i>ilvI</i>	(Lrp AND NOT (LEUxt > 0))	NH 27; PMID: 12218014

Table C.2, continued.

bNum	Gene	Rule	Reference
b0078	<i>ilvH</i>	(Lrp AND NOT (LEUxt > 0))	NH 27; PMID: 12218014
b0080	<i>fruR</i>	(NOT ("Surplus FDP"))	PMID: 8550429
b0085	<i>murE</i>		
b0086	<i>murF</i>		
b0087	<i>mraY</i>		
b0088	<i>murD</i>		
b0090	<i>murG</i>		
b0091	<i>murC</i>		
b0092	<i>ddlB</i>		
b0096	<i>lpxC</i>		
b0099	<i>mutT</i>		
b0104	<i>guaC</i>	(NOT ((GLNxt > 0) OR (GNxt > 0)))	NH 34; PMID: 2999079
b0109	<i>nadC</i>		
b0112	<i>aroP</i>	(NOT (TyrR AND ((PHExt>0) OR (TYRxt>0) OR (TRPxt>0))))	NH 22, 28; PMID: 9209035, 9765583
b0113	<i>pdhR</i>	(NOT "Surplus PYR")	PMID: 7783622
b0114	<i>aceE</i>	((NOT(PdhR)) OR (Fis))	NH 16; PMID: 7783622
b0115	<i>aceF</i>	((NOT(PdhR)) OR (Fis))	NH 16; PMID: 7783622
b0116	<i>lpdA</i>	(ON)	NH 16, NH 23; PMID: 7783622, 9209026, 9720032, 12101307
b0118	<i>acnB</i>	(ON)	PMID: 9421904
b0120	<i>speD</i>		NH 25
b0121	<i>speE</i>		NH 25
b0124	<i>gcd</i>	(NOT Crp)	PMID: 11810262
b0125	<i>hpt</i>	(Crp)	NH 34; PMID: 11810262
b0126	<i>yadF</i>		
b0131	<i>panD</i>		
b0133	<i>panC</i>		
b0134	<i>panB</i>		
b0142	<i>folK</i>		
b0154	<i>hemL</i>		
b0158	<i>yadT</i>		
b0159	<i>mtn</i>		
b0160	<i>dgt</i>		PMID: 2157212
b0162	<i>sdaR</i>	((GLRTxt>0) OR (GALRTxt>0))	PMID: 10762278
b0166	<i>dapD</i>		NH 32
b0167	<i>glnD</i>	(Lrp)	NH 23
b0171	<i>pyrH</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b0173	<i>dxr</i>		
b0174	<i>uppS</i>		
b0175	<i>cdsA</i>		
b0179	<i>lpxD</i>		
b0180	<i>fabZ</i>	((NOT((Stringent>0) OR (Stringent<0)))) OR CpxR OR RpoE OR FadR2)	NH 37; PMID:11859088, 11566998, 864995
b0181	<i>lpxA</i>		
b0182	<i>lpxB</i>		
b0185	<i>accA</i>		NH 37
b0186	<i>ldcC</i>		PMID: 9339543
b0197	<i>yaeC</i>	(NOT MetJ)	PMID: 12218041
b0198	<i>yaeE</i>	(NOT MetJ)	PMID: 12218041
b0199	<i>abc</i>	(NOT MetJ)	PMID: 12218041
b0200	<i>yaeD</i>		
b0207	<i>yafB</i>		
b0212	<i>gloB</i>		
b0221	<i>fadF</i>	(NOT (FadR2) OR NOT (ArcA))	NH 21
b0222	<i>gmhA</i>		
b0238	<i>gpt</i>		NH 34
b0242	<i>proB</i>		NH 25,26
b0243	<i>proA</i>		NH 25,26
b0273	<i>argF</i>	(NOT ArgR)	NH 25
b0312	<i>betB</i>	(NOT (ArcA OR BetI))	PMID: 8626294
b0313	<i>betI</i>	(CHOLxt > 0)	PMID: 8626294
b0314	<i>betT</i>	(NOT (BetI))	PMID: 8626294
b0323	<i>yahI</i>		
b0331	<i>prpB</i>	(PPNTxt>0)	PMID: 12473114
b0333	<i>prpC</i>	(PPNTxt>0)	PMID: 12473114
b0334	<i>prpD</i>	(PPNTxt>0)	PMID: 12473114
b0335	<i>prpE</i>	(PPNTxt>0)	PMID: 12473114
b0336	<i>codB</i>	(NOT (PurR) OR (NRL_hi))	NH 35; PMID 7500333
b0337	<i>codA</i>	(NOT (PurR) OR NRL_hi)	PMID: 2673119
b0338	<i>cynR</i>	(CNOxt>0)	PMID: 7961413, 8253686
b0339	<i>cynT</i>		PMID: 7961413, 8253686
b0340	<i>cynS</i>	(CynR)	PMID: 8083164, 7961413, 8253686
b0341	<i>cynX</i>	(CNOxt>0)	PMID: 2670891
b0343	<i>lacY</i>	("CRP noGLC" AND NOT(LacI))	[1]
b0344	<i>lacZ</i>	("CRP noGLC" AND NOT(LacI))	[1]

Table C.2, continued.

bNum	Gene	Rule	Reference
b0345	<i>lacI</i>	(NOT (LCTSxt>0))	PMID: 9104037
b0346	<i>mhpR</i>	(HPPPNTxt>0)	PMID: 9098055
b0347	<i>mhpA</i>	(MhpR)	PMID: 9098055
b0348	<i>mhpB</i>	(MhpR)	PMID: 9098055
b0349	<i>mhpC</i>	(MhpR)	PMID: 9098055
b0350	<i>mhpD</i>	(MhpR)	PMID: 9098055
b0351	<i>mhpF</i>	(MhpR)	PMID: 9098055
b0352	<i>mhpE</i>	(MhpR)	PMID: 9098055
b0353	<i>mhpT</i>		
b0356	<i>adhC</i>		
b0365	<i>tauA</i>	(Cbl AND CysB)	
b0366	<i>tauB</i>	(Cbl AND CysB)	
b0367	<i>tauC</i>	(Cbl AND CysB)	
b0368	<i>tauD</i>	(Cbl AND CysB)	PMID: 11479697, 9401024
b0369	<i>hemB</i>		
b0381	<i>ddlA</i>		
b0386	<i>proC</i>		NH 25,26
b0388	<i>aroL</i>	(NOT((TyrR AND (TYRxt>0)) OR (TyrR AND (TYRxt>0) AND TrpR)))	NH 28
b0399	<i>phoB</i>	(PhoR)	NH 87, PMID: 11489853
b0400	<i>phoR</i>	(PIxt<0.004E-6 M)	NH 87, PMID: 11489853
b0401	<i>brnQ</i>		
b0403	<i>malZ</i>	(MalT)	PMID:11931562, 11867639, 9529892
b0414	<i>ribD</i>		
b0415	<i>ribH</i>		
b0417	<i>thiL</i>		
b0418	<i>pgpA</i>		
b0420	<i>dxs</i>		
b0421	<i>ispA</i>		
b0423	<i>thiI</i>		
b0425	<i>panE</i>		
b0428	<i>cyoE</i>	(NOT (ArcA OR Fnr))	
b0429	<i>cyoD</i>	(NOT (ArcA OR Fnr))	PMID: 8576043
b0430	<i>cyoC</i>	(NOT (ArcA OR Fnr))	PMID: 8576043
b0431	<i>cyoB</i>	(NOT (ArcA OR Fnr))	PMID: 8576043
b0432	<i>cyoA</i>	(NOT (ArcA OR Fnr))	PMID: 8576043
b0451	<i>amtB</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b0469	<i>apt</i>		
b0474	<i>adk</i>		NH 34
b0475	<i>hemH</i>		
b0477	<i>gsk</i>		
b0480	<i>ushA</i>		
b0485	<i>ybaS</i>		NH 22
b0504	<i>ybbS</i>	(NOT(O2xt>0) AND NOT AllR AND NOT (NH3xt>0))	PMID: 12460564
b0505	<i>allA</i>	(NOT AllR)	PMID: 12460564
b0506	<i>allR</i>	(OFF)	PMID: 12460564
b0507	<i>gcl</i>	(NOT AllR)	PMID: 12460564
b0508	<i>hyi</i>		PMID: 10561547, 8440684
b0509	<i>glxR</i>	(NOT AllR)	PMID: 12460564
b0511	<i>allP</i>		
b0512	<i>allB</i>	(NOT AllR)	PMID: 12460564
b0514	<i>glxK</i>	(NOT AllR)	PMID: 12460564
b0516	<i>allC</i>	(AllS)	PMID: 12460564
b0521	<i>arcC</i>		
b0522	<i>purK</i>	(NOT (PurR))	NH 34
b0523	<i>purE</i>	(NOT (PurR))	NH 34
b0529	<i>folD</i>		
b0564	<i>appY</i>	(NOT CitB)	PMID 11889485, 9701802
b0576	<i>pheP</i>		
b0583	<i>entD</i>	(NOT (Fur))	
b0586	<i>entF</i>		
b0593	<i>entC</i>	(NOT (Fur))	NH 39; PMID: 8655506
b0594	<i>entE</i>	(NOT (Fur))	
b0595	<i>entB</i>	(NOT (Fur))	
b0596	<i>entA</i>	(NOT (Fur))	
b0612	<i>citT</i>	(CitB AND (NOT (O2xt > 0)))	PMID: 11889485
b0615	<i>citF</i>	(CitB)	PMID: 11889485, 9701802
b0616	<i>citE</i>	(CitB)	PMID: 11889485, 9701802
b0617	<i>citD</i>	(CitB)	PMID: 11889485, 9701802
b0619	<i>dpiB</i>	(CITxt>0)	PMID: 11889485
b0620	<i>dpiA</i>	(CitA)	PMID: 11889485
b0621	<i>dcuC</i>	(Fnr OR ArcA)	PMID: 8955408
b0638	<i>phpB</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b0639	<i>nadD</i>		
b0652	<i>gltL</i>	(NOT (GLCxt > 0))	NH 22
b0653	<i>gltK</i>	(NOT (GLCxt > 0))	NH 22
b0654	<i>gltJ</i>	(NOT (GLCxt > 0))	NH 22
b0655	<i>gltI</i>	(NOT (GLCxt > 0))	NH 22
b0662	<i>ubiF</i>		
b0674	<i>asnB</i>		NH 24
b0676	<i>nagC</i>	(NOT((NAGxt > 0) OR NAGA>0))	NH 75
b0677	<i>nagA</i>	(NOT (NagC))	NH 20; PMID 1766379, 11139621
b0678	<i>nagB</i>	(NOT (NagC) OR (GLCAxt>0))	NH 20; PMID 1766379, 11139621
b0679	<i>nagE</i>	(NOT (NagC))	PMID: 1766379, 11139621
b0683	<i>fur</i>	((FE2xt > 0) AND (OxyR OR SoxS))	PMID: 10419964
b0688	<i>pgm</i>		
b0692	<i>potE</i>		
b0693	<i>speF</i>		NH 22,25
b0694	<i>kdpE</i>	(KdpD)	PMID: 12115059
b0695	<i>kdpD</i>	(NOT (Kxt > 1))	PMID: 11248697
b0696	<i>kdpC</i>	(KdpE)	PMID: 8437514, 11248697
b0697	<i>kdpB</i>	(KdpE)	PMID: 8437514, 11248697
b0698	<i>kdpA</i>	(KdpE)	PMID: 8437514, 11248697
b0720	<i>gltA</i>		PMID: 8051021
b0721	<i>sdhC</i>	(NOT((ArcA) OR (Fnr)) OR (Crp) OR (Fis))	PMID: 9209026, 9720032
b0722	<i>sdhD</i>	(NOT((ArcA) OR (Fnr)) OR (Crp) OR (Fis))	PMID: 9209026, 9720032
b0723	<i>sdhA</i>	(NOT((ArcA) OR (Fnr)) OR (Crp) OR (Fis))	PMID: 9209026, 9720032
b0724	<i>sdhB</i>	(NOT((ArcA) OR (Fnr)) OR (Crp) OR (Fis))	PMID: 9209026, 9720032
b0726	<i>sucA</i>		PMID: 9209026, 9720032
b0727	<i>sucB</i>		PMID: 9209026, 9720032
b0728	<i>sucC</i>		NH 16; PMID: 9209026, 9720032, 7783622, 8057842
b0729	<i>sucD</i>		NH 16; PMID: 9209026, 9720032, 7783622, 8057842
b0733	<i>cydA</i>	((NOT Fnr) OR (ArcA))	PMID: 8576043

Table C.2, continued.

bNum	Gene	Rule	Reference
b0734	<i>cydB</i>	((NOT FnR) OR (ArcA))	PMID: 8576043
b0750	<i>nadA</i>		NH 48
b0751	<i>pnuC</i>		
b0754	<i>aroG</i>	(NOT(((PHExt>0) OR (TRPxxt>0)) AND TyrR))	NH 28
b0755	<i>gpmA</i>		
b0757	<i>galK</i>	(NOT(GLCxt > 0) AND (NOT(GalR OR GalS)) OR NOT (Rob))	[1]
b0758	<i>galT</i>	(NOT(GLCxt > 0) AND (NOT(GalR OR GalS)) OR NOT (Rob))	[1]
b0759	<i>galE</i>	(NOT(GLCxt > 0) AND (NOT(GalR OR GalS)) OR NOT (Rob))	[1]
b0774	<i>bioA</i>	(NOT (BirA))	NH 45; PMID: 12368242
b0775	<i>bioB</i>	(NOT (BirA))	NH 45; PMID: 12368242
b0776	<i>bioF</i>	(NOT (BirA))	NH 45; PMID: 12368242
b0778	<i>bioD</i>	(NOT (BirA))	NH 45; PMID: 12368242
b0809	<i>glnQ</i>		NH 22,23,24
b0810	<i>glnP</i>		NH 22,23,24
b0811	<i>glnH</i>		NH 22,23,24
b0825	<i>fsa</i>		
b0828	<i>ybiK</i>		PMID: 12007658
b0840	<i>deoR</i>	(NOT((DEOB1R>0) OR (DEOB1R<0)))	NH 20
b0854	<i>potF</i>		NH 25
b0855	<i>potG</i>		NH 25
b0856	<i>potH</i>		NH 25
b0857	<i>potI</i>		NH 25
b0860	<i>artJ</i>		NH 25
b0861	<i>artM</i>		NH 25
b0862	<i>artQ</i>		NH 25
b0864	<i>artP</i>		NH 25
b0870	<i>ltaA</i>		
b0871	<i>poxB</i>	((NOT (Growth > 0)) AND (RpoS))	NH 93
b0888	<i>trxR</i>		PMID: 10788450
b0889	<i>lrp</i>	(NOT LEUxt > 0)	NH 94
b0894	<i>dmsA</i>	(FnR AND NOT NarL)	PMID: 12079504
b0895	<i>dmsB</i>	(FnR AND NOT NarL)	PMID: 12079504
b0896	<i>dmsC</i>	(FnR AND NOT NarL)	PMID: 12079504
b0902	<i>pflA</i>	(ArcA OR FnR AND (Crp OR NOT(NarL)))	NH 95; PMID: 7934836

Table C.2, continued.

bNum	Gene	Rule	Reference
b0903	<i>pflB</i>	(ArcA OR Fnr AND (Crp OR NOT(NarL)))	NH 95; PMID: 7934836
b0904	<i>focA</i>	(ArcA OR Fnr AND (Crp OR NOT (NarL)))	NH 95; PMID: 7934836
b0907	<i>serC</i>	(Lrp OR (NOT (Crp)))	NH 30; PMID: 9171388
b0908	<i>aroA</i>		NH 28
b0910	<i>cmk</i>		
b0915	<i>lpzK</i>		
b0918	<i>kdsB</i>		PMID: 7543480
b0928	<i>aspC</i>		NH 22,24,28
b0931	<i>pncB</i>	(NOT (NadR))	NH 48
b0945	<i>pyrD</i>	((NOT (CYTSxt > 0)) OR (GNxt > 0) OR NOT PurR)	NH 35
b0954	<i>fabA</i>	((NOT((Stringent>0) OR (Stringent<0))) OR CpxR OR RpoE OR FadR2)	NH 37; PMID: 11859088, 11566998, 864995
b0963	<i>mgsA</i>		
b0972	<i>hyaA</i>	((ArcA OR Fnr) AND (AppY))	NH 17; PMID 10537212
b0973	<i>hyaB</i>	((ArcA OR Fnr) AND (AppY))	NH 17; PMID 10537212
b0974	<i>hyaC</i>	((ArcA OR Fnr) AND (AppY))	NH 17; PMID 10537212
b0993	<i>torS</i>	(TMAOxt>0)	PMID: 9135110, 11004177
b0995	<i>torR</i>	(TorS)	PMID: 9135110, 11004177
b0996	<i>torC</i>	(TorR OR NOT (NarL))	PMID: 9135110, 11004177
b0997	<i>torA</i>	(TorR OR NOT (NarL))	PMID: 9135110, 11004177
b1002	<i>agp</i>		NH 87
b1006	<i>ycdG</i>		
b1014	<i>putA</i>	((PROxt > 0) OR Crp OR Nac)	NH 22
b1015	<i>putP</i>	((PROxt > 0) OR Crp OR Nac)	NH 22; PMID: 2464125
b1033	<i>ycdW</i>		PMID: 11237876
b1054	<i>lpzL</i>		
b1062	<i>pyrC</i>	((NOT (CYTSxt > 0)) OR (GNxt > 0) OR NOT PurR)	NH 35
b1091	<i>fabH</i>	((NOT((Stringent>0) OR (Stringent<0))))	PMID: 8649995
b1092	<i>fabD</i>	((NOT((Stringent>0) OR (Stringent<0))))	PMID: 8649995
b1093	<i>fabG</i>	((NOT((Stringent>0) OR (Stringent<0)))) OR CpxR OR RpoE OR FadR2)	NH 37; PMID: 11859088, 11566998, 864995
b1095	<i>fabF</i>	((NOT((Stringent>0) OR (Stringent<0)))) OR CpxR OR RpoE OR FadR2)	
b1096	<i>pabC</i>		
b1098	<i>tmk</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b1101	<i>ptsG</i>	(NOT(Mlc) OR NOT(Cra))	PMID: 10469172, 8106445, 7518773, 1324322, 11931562, 11867639, 9529892, 9148912
b1109	<i>ndh</i>	(NOT (Fnr))	
b1123	<i>potD</i>		NH 25
b1124	<i>potC</i>		NH 25
b1125	<i>potB</i>		NH 25
b1126	<i>potA</i>		NH 25
b1131	<i>purB</i>	(NOT (PurR))	NH 34
b1136	<i>icdA</i>		PMID: 9209047, 9922253
b1186	<i>nhaB</i>		PMID: 11779554
b1187	<i>fadR</i>	(GLCxt > 0 OR NOT (ACxt > 0))	PMID: 8755903
b1189	<i>dadA</i>	(ALAxt > 0 AND NOT Crp)	NH 22
b1190	<i>dadX</i>	((ALAxxt > 0) OR (DALAxxt > 0)) AND Crp)	NH 22,24
b1197	<i>treA</i>	(RpoS)	PMID: 9148912, 8892826
b1198	<i>dhaH</i>		PMID: 11021910
b1199	<i>dhaK2</i>		PMID: 11021910
b1200	<i>dhaK1</i>		PMID: 11021910
b1207	<i>prsA</i>	(NOT PurR)	PMID: 8388874
b1208	<i>ispE</i>		
b1210	<i>hemA</i>		PMID: 8997718
b1215	<i>kdsA</i>		PMID: 7543480
b1216	<i>chaA</i>		PMID: 11779554, 9518629
b1221	<i>narL</i>	((NO3xt>0) OR (NO2xt>0))	NH 17
b1223	<i>narK</i>	(Fnr OR NarL)	PMID: 1474901
b1224	<i>narG</i>	(Fnr AND NarL)	NH 17; PMID: 8736541, 10464201
b1225	<i>narH</i>	(Fnr AND NarL)	NH 17; PMID: 8736541, 10464201
b1226	<i>narJ</i>	(Fnr AND NarL)	NH 17; PMID: 8736541, 10464201
b1227	<i>narI</i>	(Fnr AND NarL)	NH 17; PMID: 8736541, 10464201
b1232	<i>purU</i>		
b1236	<i>galU</i>		
b1238	<i>tdk</i>		
b1241	<i>adhE</i>	(NOT (O2xt > 0) OR (NOT	PMID: 10601216, 9371462

Table C.2, continued.

bNum	Gene	Rule	Reference
		((O2xt > 0) AND (Cra))) OR (Fis) OR NOT (NarL) OR (RpoS))	
b1249	<i>cls</i>		
b1260	<i>trpA</i>	(NOT TrpR)	NH 28
b1261	<i>trpB</i>	(NOT TrpR)	NH 28
b1262	<i>trpC</i>	(NOT TrpR)	NH 28
b1263	<i>trpD</i>	(NOT TrpR)	NH 28
b1264	<i>trpE</i>	(NOT TrpR)	NH 28
b1270	<i>btuR</i>		
b1275	<i>cysB</i>	(NOT (CYStxt > 0))	NH 31
b1276	<i>acnA</i>	(SoxS)	PMID: 9421904
b1277	<i>ribA</i>		PMID: 8709966
b1278	<i>pgpB</i>		
b1281	<i>pyrF</i>		NH 35
b1288	<i>fabI</i>	((NOT((Stringent>0) OR (Stringent<0))) OR CpxR OR RpoE OR FadR2)	NH 37; PMID:11859088, 11566998, 864995
b1297	<i>ycjK</i>		
b1300	<i>aldH</i>		
b1302	<i>goaG</i>		
b1323	<i>tyrR</i>	((TRPxxt>0) OR (TYRxxt>0) OR (PHExt>0))	NH 28
b1334	<i>fnr</i>	(NOT (O2xt>0))	PMID: 2964639, NH 95
b1363	<i>trkG</i>		NH 72
b1380	<i>ldhA</i>		PMID: 11535784
b1384	<i>feaR</i>	(Crp)	PMID: 8631685
b1385	<i>feaB</i>	(FeaR)	PMID: 8631685
b1386	<i>tynA</i>	(MaoB)	PMID: 8631685
b1398	<i>paaK</i>		PMID: 9748275
b1415	<i>aldA</i>		PMID: 9202484
b1416	<i>gapC_2</i>		
b1417	<i>gapC_-1</i>		
b1440	<i>ydcS</i>		
b1441	<i>ydcT</i>		
b1442	<i>ydcU</i>		
b1443	<i>ydcV</i>		
b1469	<i>narU</i>		PMID: 7747940
b1474	<i>fdnG</i>	(Fnr OR NarL)	PMID: 1629153, 8736541
b1475	<i>fdnH</i>	(Fnr OR NarL)	PMID: 1629153, 8736541
b1476	<i>fdnI</i>	(Fnr OR NarL)	PMID: 1629153, 8736541

Table C.2, continued.

bNum	Gene	Rule	Reference
b1479	<i>sfcA</i>		
b1492	<i>xasA</i>		
b1493	<i>gadB</i>	((NOT (Growth > 0)) OR (pH < 4))	NH 22; PMID: 11976288
b1519	<i>tam</i>	(NOT (Growth>0))	PMID: 10224113
b1521	<i>uxaB</i>	(NOT ExuR)	NH 20
b1524	<i>yncH</i>	((NOT (GLCxt > 0) OR ((NH3xt > 0) AND NOT Crp)))	NH 22
b1531	<i>marA</i>	(Salicylate > 0)	PMID: 8522515
b1584	<i>speG</i>		PMID: 10986239
b1594	<i>mlc</i>	(NOT (GLCxt>0))	PMID: 10469172
b1602	<i>pntB</i>		
b1603	<i>pntA</i>		
b1605	<i>arcD</i>		
b1611	<i>fumC</i>	(MarA OR Rob OR SoxS AND (NOT(ArcA)))	PMID: 7592392
b1612	<i>fumA</i>	(NOT(ArcA OR Fnr))	PMID: 7592392
b1613	<i>manA</i>		NH 20
b1620	<i>malI</i>	(NOT (MLTxt > 0))	PMID: 2670898
b1621	<i>malX</i>	((MalT AND Crp) OR MalT)	PMID: 1856179
b1622	<i>malY</i>	(NOT (MalI))	PMID:11931562, 11867639, 9529892
b1623	<i>add</i>	((ADxt > 0) OR (HYXNxt > 0))	NH 34
b1636	<i>pdxY</i>		
b1638	<i>pdxH</i>		NH 32
b1646	<i>sodC</i>	(NOT(Growth>0) AND NOT Fnr)	PMID: 8626323, 10216871
b1651	<i>gloA</i>		
b1656	<i>sodB</i>		PMID: 11782507
b1658	<i>purR</i>	((HYXNxt > 0) OR (GNxt > 0))	NH 34
b1662	<i>ribE</i>		
b1676	<i>pykF</i>	(NOT(Cra))	PMID: 8550429
b1692	<i>ydiB</i>		
b1693	<i>aroD</i>		NH 28
b1702	<i>pps</i>	(Cra)	PMID: 9512708
b1704	<i>aroH</i>	(NOT TrpR)	NH 28
b1709	<i>btuD</i>		
b1711	<i>btuC</i>		
b1723	<i>pfkB</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b1732	<i>katE</i>	(NOT(Growth>0))	PMID: 12589799
b1740	<i>nadE</i>		
b1744	<i>astE</i>	((NOT(Growth>0) AND RpoS) OR (NRI_hi AND RpoN))	PMID: 12003934
b1745	<i>astB</i>	((NOT(Growth>0) AND RpoS) OR (NRI_hi AND RpoN))	PMID: 12003934
b1746	<i>astD</i>	((NOT(Growth>0) AND RpoS) OR (NRI_hi AND RpoN))	PMID: 12003934
b1747	<i>astA</i>	((NOT(Growth>0) AND RpoS) OR (NRI_hi AND RpoN))	PMID: 12003934
b1748	<i>astC</i>	((NOT(Growth>0) AND RpoS) OR (NRI_hi AND RpoN))	PMID: 12003934
b1761	<i>gdhA</i>	(NOT ((Nac) OR (GLUxt > 0)))	NH 22,24; PMID: 9785451
b1764	<i>selD</i>		PMID: 1650339
b1767	<i>ansA</i>		NH 22
b1768	<i>pncA</i>		
b1773	<i>b1773</i>		
b1779	<i>gapA</i>		PMID: 9851989
b1801	<i>yeaV</i>		
b1805	<i>fadD</i>	(NOT (FadR2) OR NOT (ArcA))	NH 21
b1812	<i>pabB</i>		
b1814	<i>sdaA</i>	((GLYxt > 0 OR LEUxt > 0 OR NOT (O2xt > 0)) AND ((NOT Lrp) OR (Lrp AND LEUxt > 0)))	NH 22
b1817	<i>manX</i>	((“CRP noLAC”) OR (NOT (Mlc)))	NH 20; PMID: 9484892, 11934616
b1818	<i>manY</i>	((“CRP noLAC”) OR (NOT (Mlc)))	NH 20; PMID: 9484892, 11934616
b1819	<i>manZ</i>	((“CRP noLAC”) OR (NOT (Mlc)))	NH 20; PMID: 9484892, 11934616
b1827	<i>kdgR</i>	(NOT(KDGxt>0) AND NOT (UXUA>0) AND NOT(UXAA>0))	NH 20
b1849	<i>purT</i>	(NOT (PurR))	NH 34
b1850	<i>eda</i>	(ON)	PMID: 1624451,8655507
b1851	<i>edd</i>	(NOT (GntR))	PMID: 1624451,8655507
b1852	<i>zwf</i>		
b1854	<i>pykA</i>		
b1855	<i>msbB</i>		
b1865	<i>ntpA</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b1872	<i>torZ</i>		PMID: 11004177
b1873	<i>torY</i>		PMID: 11004177
b1896	<i>otsA</i>	(RpoS)	PMID: 9148912, 12105274
b1897	<i>otsB</i>	(RpoS)	PMID: 9148912, 12105274
b1898	<i>araH_2</i>	(AraC OR (AraC AND Crp))	NH 20
b1899	<i>araH_1</i>	(AraC OR (AraC AND Crp))	NH 20
b1900	<i>araG</i>	(AraC OR (AraC AND Crp))	NH 20
b1901	<i>araF</i>	(AraC OR (AraC AND Crp))	NH 20
b1907	<i>tyrP</i>	(NOT(TyrR AND (TYRxt>0)))	NH 22,28
b1912	<i>pgsA</i>		
b1982	<i>amn</i>		NH 34
b1987	<i>cbl</i>	(NOT ((SLFxt > 0) OR (CYStx > 0)) AND CysB)	PMID: 10506196
b1988	<i>nac</i>	(NRJ_low AND RpoN)	NH 23
b1991	<i>cobT</i>	(CBIxt>0)	PMID: 7592411
b1992	<i>cobS</i>	(CBIxt>0)	PMID: 7592411
b1993	<i>cobU</i>	(CBIxt>0)	PMID: 7592411
b2019	<i>hisG</i>		NH 29
b2020	<i>hisD</i>		NH 29
b2021	<i>hisC</i>		NH 29
b2022	<i>hisB</i>		NH 29
b2023	<i>hisH</i>		NH 29
b2024	<i>hisA</i>		NH 29
b2025	<i>hisF</i>		NH 29
b2026	<i>hisI</i>		NH 29
b2028	<i>ugd</i>		
b2029	<i>gnd</i>		
b2036	<i>glf</i>		
b2038	<i>rfbC</i>		
b2039	<i>rfbA</i>		
b2040	<i>rfbD</i>		
b2041	<i>rfbB</i>		
b2042	<i>galF</i>		
b2045	<i>wcaK</i>		
b2048	<i>cpsG</i>		
b2049	<i>manC</i>		
b2052	<i>fcl</i>		
b2053	<i>gmd</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b2065	<i>dcd</i>		
b2066	<i>udk</i>	(NOT ((THYxt > 0) OR (CYTSxt > 0) OR (URAxt > 0)))	NH 35
b2087	<i>gatR_1</i>	(NOT (GLTLxt > 0))	PMID: 7772602
b2090	<i>gatR_2</i>	(NOT (GLTLxt > 0))	PMID: 7772602
b2091	<i>gatD</i>	(NOT (GatR))	N20; PMID: 8955298
b2092	<i>gatC</i>	(NOT (GatR))	N20; PMID: 8955298
b2093	<i>gatB</i>	(NOT (GatR))	N20; PMID: 8955298
b2094	<i>gatA</i>	(NOT (GatR))	N20; PMID: 8955298
b2095	<i>gatZ</i>	(NOT (GatR))	N20; PMID: 8955298
b2096	<i>gatY</i>	(NOT (GatR))	N20; PMID: 8955298
b2097	<i>fbaB</i>	((PYRxt>0) OR (LACxt>0) AND NOT(GLCxt>0))	PMID: 9531482
b2103	<i>thiD</i>		
b2104	<i>thiM</i>		
b2128	<i>yehW</i>		
b2129	<i>yehX</i>		
b2130	<i>yehY</i>		
b2131	<i>yehZ</i>		
b2132	<i>bgI</i> X		PMID: 8757730
b2133	<i>ddl</i>		
b2143	<i>cdd</i>	(Crp AND NOT (CytR))	PMID: 2575702
b2148	<i>mglC</i>	(Crp AND NOT (GalS))	PMID: 12101127
b2149	<i>mglA</i>	(Crp AND NOT (GalS))	PMID: 12101127
b2150	<i>mglB</i>	(Crp AND NOT (GalS))	PMID: 12101127
b2151	<i>galS</i>	(NOT (LCTSxt>0) OR NOT (GLACxt>0))	PMID: 8982002
b2153	<i>folE</i>		
b2156	<i>lysP</i>		
b2167	<i>fruA</i>	(NOT (Cra))	PMID: 7852310
b2168	<i>fruK</i>	(NOT (Cra))	PMID: 7852310
b2169	<i>fruB</i>	(NOT (Cra))	PMID: 7852310
b2210	<i>mqa</i>	(NOT ArcA)	PMID: 11092847
b2219	<i>atoS</i>	(ACTOACxt < 0)	NH 21
b2220	<i>atoC</i>	(AtoS)	NH 21
b2221	<i>atoD</i>	(AtoC)	NH 21
b2222	<i>atoA</i>	(AtoC)	NH 21
b2223	<i>atoE</i>	(AtoC)	

Table C.2, continued.

bNum	Gene	Rule	Reference
b2224	<i>atoB</i>	(AtoC)	NH 21
b2232	<i>ubiG</i>	((O2xt > 0) AND Crp)	NH 39; PMID: 2830238
b2234	<i>nrdA</i>	(NOT (ArcA))	NH 34; PMID: 9680219, 8954104
b2235	<i>nrdB</i>	(NOT (ArcA))	NH 34; PMID: 9680219, 8954104
b2239	<i>glpQ</i>	((NOT GlpR OR Fnr) AND Crp)	PMID: 9179845, 1521763
b2240	<i>glpT</i>	(NOT (GlpR) AND Crp)	PMID: 9179845
b2241	<i>glpA</i>	("CRP noRIB" AND (Fnr OR ArcA) AND (NOT(GlpR)))	NH 20; PMID:2403539; [135]
b2242	<i>glpB</i>	("CRP noRIB" AND (Fnr OR ArcA) AND (NOT(GlpR)))	NH 20; PMID:2403539; [135]
b2243	<i>glpC</i>	("CRP noRIB" AND (Fnr OR ArcA) AND (NOT(GlpR)))	NH 20; PMID:2403539; [135]
b2260	<i>menE</i>		
b2261	<i>menC</i>		
b2262	<i>menB</i>		
b2264	<i>menD</i>		
b2265	<i>menF</i>		
b2276	<i>nuoN</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2277	<i>nuoM</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2278	<i>nuoL</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2279	<i>nuoK</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2280	<i>nuoJ</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2281	<i>nuoI</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2282	<i>nuoH</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2283	<i>nuoG</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2284	<i>nuoF</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2285	<i>nuoE</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2286	<i>nuoC</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2287	<i>nuoB</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2288	<i>nuoA</i>	(NOT (ArcA OR Fnr) OR NarL)	PMID: 10628873
b2296	<i>ackA</i>		
b2297	<i>pta</i>		
b2306	<i>hisP</i>	(NOT (LYSxt > 0))	NH 22
b2307	<i>hisM</i>	(NOT (LYSxt > 0))	NH 22
b2308	<i>hisQ</i>	(NOT (LYSxt > 0))	NH 22

Table C.2, continued.

bNum	Gene	Rule	Reference
b2309	<i>hisJ</i>	(NOT (LYSxt > 0))	NH 22
b2310	<i>argT</i>	(NOT (LYSxt > 0))	NH 22
b2311	<i>ubiX</i>	(NOT (PurR))	
b2312	<i>purF</i>	(NOT (PurR))	NH 34
b2315	<i>folC</i>		
b2316	<i>accD</i>		NH 37
b2320	<i>pdxB</i>		PMID: 11844765
b2323	<i>fabB</i>	((NOT((Stringent>0) OR (Stringent<0))) OR CpxR OR RpoE OR FadR2)	PMID: 11859088, 11566998
b2329	<i>aroC</i>		NH 28
b2344	<i>fadL</i>	((NOT (Crp OR FadR OR OmpR)))	NH 21;
b2364	<i>dsdC</i>	(DSERxt > 0)	PMID: 7592420
b2366	<i>dsdA</i>	((Crp AND DsdC) OR DsdC)	NH 22
b2378	<i>lpnP</i>		PMID: 10092655
b2388	<i>glk</i>		
b2393	<i>nupC</i>	(Crp OR NOT (CytR))	NH 35
b2400	<i>gltX</i>		
b2405	<i>xapR</i>	(XTSNxt > 0)	PMID: 7559336
b2406	<i>xapB</i>	(XapR)	PMID: 7559336
b2407	<i>xapA</i>	(XapR)	PMID: 7559336
b2411	<i>lig</i>		
b2413	<i>cysZ</i>	(CysB)	NH 31
b2414	<i>cysK</i>	(CysB)	NH 31
b2415	<i>ptsH</i>	(ON)	PMID: 10469172, 8106445, 7518773, 1324322, 11931562, 11867639, 9529892, 9148912
b2416	<i>ptsI</i>	(ON)	PMID: 10469172, 8106445, 7518773, 1324322, 11931562, 11867639, 9529892, 9148912
b2417	<i>crr</i>	(ON)	PMID: 10469172, 8106445, 7518773, 1324322, 11931562, 11867639, 9529892, 9148912
b2418	<i>pdxK</i>		
b2421	<i>cysM</i>	(CysB)	NH 31
b2422	<i>cysA</i>	(CysB)	NH 31
b2423	<i>cysW</i>	(CysB)	NH 31
b2424	<i>cysU</i>	(CysB)	NH 31
b2425	<i>cysP</i>	(CysB)	NH 31

Table C.2, continued.

bNum	Gene	Rule	Reference
b2429	<i>yfeV</i>		
b2436	<i>hemF</i>		PMID: 8990283
b2440	<i>eutC</i>		
b2441	<i>eutB</i>		
b2458	<i>eutD</i>		
b2463	<i>maeB</i>		
b2464	<i>talA</i>		
b2465	<i>tktB</i>		
b2472	<i>dapE</i>		NH 32
b2476	<i>purC</i>	(NOT (PurR))	NH 34
b2478	<i>dapA</i>		NH 32
b2479	<i>gcvR</i>	(NOT (GLYxt > 0))	PMID: 12101307
b2492	<i>focB</i>	(ArcA OR Fnr AND (Crp OR NOT (NarL)))	PMID: 12426353
b2497	<i>uraA</i>		NH 35
b2498	<i>upp</i>		NH 35
b2499	<i>purM</i>	(NOT (PurR))	NH 34
b2500	<i>purN</i>	(NOT (PurR))	NH 34
b2507	<i>guaA</i>	(NOT (PurR AND Crp))	NH 34; PMID: 10856643
b2508	<i>guaB</i>	(NOT (PurR AND Crp))	NH 34; PMID: 10856643
b2515	<i>gcpE</i>		
b2518	<i>ndk</i>		NH 34
b2530	<i>iscS</i>		
b2533	<i>suhB</i>		PMID: 8831954
b2536	<i>hcaT</i>		PMID: 9603882
b2537	<i>hcaR</i>	(PPPNTxt>0)	PMID: 9603882
b2538	<i>hcaE</i>	(HcaR AND (NOT((LBMedia>0) OR (LBMedia<0))))	PMID: 9603882
b2539	<i>hcaF</i>	(HcaR AND (NOT((LBMedia>0) OR (LBMedia<0))))	PMID: 9603882
b2540	<i>hcaC</i>	(HcaR AND (NOT((LBMedia>0) OR (LBMedia<0))))	PMID: 9603882
b2541	<i>hcaB</i>	(HcaR AND (NOT((LBMedia>0) OR (LBMedia<0))))	PMID: 9603882
b2542	<i>hcaD</i>	(HcaR AND (NOT((LBMedia>0) OR (LBMedia<0))))	PMID: 9603882
b2551	<i>glyA</i>	(NOT (GLYxt > 0) OR MetR OR NOT (PurR))	NH 22,30; PMID 8900067

Table C.2, continued.

bNum	Gene	Rule	Reference
b2553	<i>glnB</i>	(Lrp)	NH 23
b2557	<i>purL</i>	(NOT (PurR))	NH 34
b2563	<i>acpS</i>		
b2564	<i>pdxJ</i>	(RpoE)	PMID: 11844765
b2573	<i>rpoE</i>	("heat shock")	PMID: 7751307
b2574	<i>nadB</i>	(NOT (NadR))	NH 48
b2585	<i>pssA</i>		
b2587	<i>kgtP</i>		PMID: 1556144
b2599	<i>pheA</i>	(NOT (PHExt>0))	NH 28
b2600	<i>tyrA</i>	(NOT((PHExt>10) OR (TYRxt>0)) AND TyrR))	NH 28
b2601	<i>aroF</i>	(NOT((PHExt>10) OR (TYRxt>0)) AND TyrR))	NH 28
b2615	<i>yfjB</i>		
b2661	<i>gabD</i>		NH 22,23
b2662	<i>gabT</i>		NH 22,23; PMID: 12446648
b2663	<i>gabP</i>		PMID: 9829938
b2675	<i>nrdE</i>		PMID: 11278973
b2676	<i>nrdF</i>		PMID: 11278973
b2677	<i>proV</i>		NH 25, 26
b2678	<i>proW</i>		NH 25, 26
b2679	<i>proX</i>		NH 25, 26
b2687	<i>luxS</i>		
b2688	<i>gshA</i>		
b2690	<i>yqaB</i>		
b2702	<i>srlA</i>	((NOT GutR) AND "CRP noGL")	NH 20
b2703	<i>srlE</i>	((NOT GutR) AND "CRP noGL")	NH 20
b2704	<i>srlB</i>	((NOT GutR) AND "CRP noGL")	NH 20
b2705	<i>srlD</i>	(GutM AND (NOT GutR) AND "CRP noGL")	NH 20
b2706	<i>gutM</i>	(ON)	PMID: 3062173
b2707	<i>srlR</i>	(NOT (GLTxt > 0))	NH 20
b2719	<i>hycG</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b2720	<i>hycF</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b2721	<i>hycE</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18

Table C.2, continued.

bNum	Gene	Rule	Reference
b2722	<i>hydC</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b2723	<i>hydC</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b2724	<i>hydB</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b2731	<i>fhlA</i>	((NOT (O2xt > 0)) AND (NOT (NO3xt > 0)) AND (NOT (NO2xt > 0)) AND (NOT (TMAOxt > 0)) AND (NOT (DMSOxt > 0)) AND (FORxt>0))	NH 18
b2738	<i>ygbL</i>		
b2741	<i>rpoS</i>	(NOT (Growth > 0))	NH 93
b2746	<i>ispF</i>		
b2747	<i>ispD</i>		
b2750	<i>cysC</i>	(CysB)	NH 31
b2751	<i>cysN</i>	(CysB)	NH 31
b2752	<i>cysD</i>	(CysB)	NH 31
b2762	<i>cysH</i>	(CysB)	NH 31
b2763	<i>cysI</i>	(CysB)	NH 31
b2764	<i>cysJ</i>	(CysB)	NH 31
b2779	<i>eno</i>		
b2780	<i>pyrG</i>		NH 35
b2781	<i>mazG</i>		
b2787	<i>gudD</i>	(SdaR)	PMID: 10762278
b2788	<i>ygeY</i>		
b2789	<i>gudP</i>		
b2796	<i>sdaC</i>	(Crp OR (NOT (Lrp) AND (LEUxt > 0) AND Crp))	NH 22
b2797	<i>sdaB</i>	(ON)	NH 22
b2799	<i>fucO</i>	(((((FucR) OR (RMNxt > 0)) AND (NOT (O2xt > 0))) AND Crp) OR (((FucR) OR (RMNxt > 0)) AND (NOT (O2xt > 0))))	PMID: 3325779
b2800	<i>fucA</i>	((FucR AND Crp) OR FucR)	PMID: 3325779
b2801	<i>fucP</i>	((FucR AND Crp) OR FucR)	PMID: 3325779
b2802	<i>fucI</i>	((FucR AND Crp) OR FucR)	PMID: 3325779
b2803	<i>fucK</i>	((FucR AND Crp) OR FucR)	PMID: 3325779

Table C.2, continued.

bNum	Gene	Rule	Reference
b2805	<i>fucR</i>	(FUCxt > 0)	PMID: 3325779
b2808	<i>gcvA</i>	(NOT GcvR)	PMID: 12101307
b2818	<i>argA</i>	(NOT ArgR)	NH 25
b2827	<i>thyA</i>		
b2836	<i>aas</i>		
b2837	<i>galR</i>	(NOT (LCTSxt>0) OR NOT (GLACxt>0))	PMID: 8982002
b2838	<i>lysA</i>	(LysR AND NOT LYSxt > 0)	NH 32
b2839	<i>lysR</i>	(NOT (LYSxt > 0))	NH 32
b2841	<i>araE</i>	(Crp)	NH 20
b2874	<i>yqeA</i>		
b2883	<i>ygfP</i>		
b2889	<i>idi</i>		
b2901	<i>bgmA</i>		
b2903	<i>gcvP</i>	((Fis AND NOT PdhR) AND ((NOT(GcvR) AND GcvA) OR Lrp OR NOT PurR))	NH 23; PMID: 12101307
b2904	<i>gcvH</i>	((Fis AND NOT PdhR) AND ((NOT(GcvR) AND GcvA) OR Lrp OR NOT PurR))	NH 23; PMID: 12101307
b2905	<i>gcvT</i>	((Fis AND NOT PdhR) AND ((NOT(GcvR) AND GcvA) OR Lrp OR NOT PurR))	NH 23; PMID: 12101307
b2907	<i>ubiH</i>		
b2913	<i>serA</i>		NH 30
b2914	<i>rpiA</i>		
b2917	<i>sbm</i>		
b2919	<i>ygfG</i>		
b2920	<i>ygfH</i>		
b2925	<i>fbaA</i>		
b2926	<i>pgk</i>	(ON)	PMID: 9851989
b2927	<i>epd</i>	(Crp)	
b2935	<i>tktA</i>		
b2937	<i>speB</i>		NH 22
b2938	<i>speA</i>	(NOT (PurR))	NH 25; PMID: 8388874
b2942	<i>metK</i>		NH 33
b2943	<i>galP</i>	((NOT (GalR)) OR (GalS) AND (Crp OR NOT (Crp)))	PMID: 8703508, 8982002, 1970645, 12101127
b2947	<i>gshB</i>		
b2957	<i>ansB</i>	(Fnr AND Crp)	NH 22
b2964	<i>nupG</i>	(Crp OR NOT (CytR) OR NOT (DeoR))	PMID: 8596434, 2115441
b2965	<i>speC</i>	(NOT (Crp))	NH 22; PMID 3021588
b2975	<i>glcA</i>	(NOT ArcA AND GlcC)	PMID: 8606183, 9880556

Table C.2, continued.

bNum	Gene	Rule	Reference
b2976	<i>glcB</i>	(NOT (ArcA) AND (GlcC))	PMID: 8606183, 9880556
b2978	<i>glcF</i>		PMID: 9880556
b2979	<i>glcD</i>		PMID: 9880556
b2980	<i>glcC</i>	((ACxt>0) OR (GLCLTxt>0))	PMID 9880556
b2987	<i>pitB</i>	(NOT (PhoB))	PMID: 11489853
b2988	<i>gsp</i>		
b2994	<i>hybC</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	PMID 10537212
b2997	<i>hybO</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	PMID 10537212
b3008	<i>metC</i>	(NOT MetJ)	NH 33
b3012	<i>yqhE</i>		
b3018	<i>plsC</i>		
b3041	<i>ribB</i>		
b3052	<i>rfaE</i>		NH 69
b3058	<i>folB</i>		
b3061	<i>ttdA</i>	(NOT(O2xt>0) AND (TARTxt>0))	PMID: 8371115
b3062	<i>ttdB</i>	(NOT(O2xt>0) AND (TARTxt>0))	PMID: 8371115
b3063	<i>yggE</i>		
b3073	<i>yggG</i>		
b3089	<i>sstT</i>		PMID: 12097162
b3091	<i>uxaA</i>	(NOT ExuR)	NH 20
b3092	<i>uxaC</i>	(NOT ExuR)	NH 20
b3093	<i>exuT</i>	(NOT ExuR)	NH 20
b3094	<i>exuR</i>	(NOT(UXAC1R>0) AND NOT(UXAC2R>0) AND NOT(UXUBR>0) AND NOT(UXABR>0) AND NOT(UXAC1R<0) AND NOT(UXAC2R<0) AND NOT(UXUBR<0) AND NOT(UXABR<0))	NH 20
b3111	<i>tdcGa</i>	(Crp OR NOT(O2xt>0))	PMID: 11251844
b3112	<i>tdcGb</i>	(Crp OR NOT(O2xt>0))	PMID: 11251844
b3114	<i>tdcE</i>	(Crp OR Fnr OR LysR OR TdcA OR NOT TdcR)	PMID: 11251844
b3115	<i>tdcD</i>	(Crp OR Fnr OR LysR OR TdcA OR NOT TdcR)	PMID: 11251844

Table C.2, continued.

bNum	Gene	Rule	Reference
b3116	<i>tdcC</i>	(Crp OR Fnr OR LysR OR TdcA OR NOT TdcR)	PMID: 7928991, 8413189
b3117	<i>tdcB</i>	(Crp OR Fnr OR LysR OR TdcA OR NOT TdcR)	NH 22
b3118	<i>tdcA</i>	((THRxt > 0) AND (SERxt > 0) AND (VALxt > 0) AND (ILExt > 0) AND NOT (O2xt > 0))	PMID: 7928991, 2573820, 8413189
b3119	<i>tdcR</i>	((THRxt > 0) AND (SERxt > 0) AND (VALxt > 0) AND (ILExt > 0) AND NOT (O2xt > 0))	PMID: 7928991, 2573820, 8413189
b3124	<i>garK</i>	(SdaR)	PMID: 10762278
b3125	<i>garR</i>	(SdaR)	PMID: 10762278
b3126	<i>garL</i>	(SdaR)	PMID: 10762278
b3127	<i>garP</i>		
b3128	<i>garD</i>	(SdaR)	PMID: 10762278
b3132	<i>agaZ</i>		
b3137	<i>agaY</i>		
b3161	<i>mtr</i>	(NOT TrpR OR (TyrR AND ((PHExt>0) OR (TYRxt>0))))	NH 28
b3172	<i>argG</i>		NH 25
b3176	<i>mrsA</i>		
b3177	<i>folP</i>		
b3187	<i>ispB</i>		
b3189	<i>murA</i>		
b3202	<i>rpoN</i>	(ON)	NH 24
b3212	<i>gltB</i>	((Lrp AND NOT (LEUxt > 0)) OR NOT(NRL_hi AND ((GLUxt > 0) OR (ARGxt > 0) OR (ASPxt > 0) OR (HISxt > 0) OR (PROxt > 0))))	NH 22,23,24
b3213	<i>gltD</i>	((Lrp AND NOT (LEUxt > 0)) OR NOT(NRL_hi AND ((GLUxt > 0) OR (ARGxt > 0) OR (ASPxt > 0) OR (HISxt > 0) OR (PROxt > 0))))	NH 22,23,24
b3222	<i>nanK</i>		PMID: 9864311
b3223	<i>nanE</i>		PMID: 9864311
b3224	<i>nanT</i>		
b3225	<i>nanA</i>		
b3236	<i>mdh</i>	(NOT(ArcA))	

Table C.2, continued.

bNum	Gene	Rule	Reference
b3237	<i>argR</i>	(ARGxt > 0)	NH 25, PMID: 3116542, 12003934
b3255	<i>accB</i>		NH 37
b3256	<i>accC</i>		NH 37
b3258	<i>panF</i>		PMID: 8226664
b3261	<i>fis</i>	(Growth > 0)	NH 90
b3265	<i>acrE</i>		
b3266	<i>acrF</i>		
b3281	<i>aroE</i>		NH 28
b3357	<i>crp</i>	("CRP noGLC")	PMID: 5337847
b3359	<i>argD</i>	(NOT ArgR)	NH 25
b3360	<i>pabA</i>		PMID: 2050628
b3365	<i>nirB</i>	(Fnr AND NarL)	PMID: 11004182
b3366	<i>nirD</i>	(Fnr AND NarL)	PMID: 11004182
b3367	<i>nirC</i>	(Fnr AND NarL)	PMID: 11004182
b3368	<i>cysG</i>	(Fnr OR NarL)	
b3380	<i>yhfW</i>		
b3385	<i>gph</i>		PMID: 10572959
b3386	<i>rpe</i>		
b3389	<i>aroB</i>		NH 28
b3390	<i>aroK</i>		NH 28
b3403	<i>pckA</i>		
b3405	<i>ompR</i>	("high osmolarity")	PMID: 7932717
b3409	<i>feoB</i>		NH 71
b3415	<i>gntT</i>	(NOT (GntR) AND "CRP GLCN")	PMID: 9537375
b3416	<i>malQ</i>	(MalT)	PMID: 11931562, 11867639, 9529892
b3417	<i>malP</i>	(MalT)	PMID: 11931562, 11867639, 9529892
b3418	<i>malT</i>	((MLTxt > 0) OR (MLTTRxt > 0) OR (MLTTTxt > 0) OR (MLTHXxt > 0) OR (MLTPTxt > 0))	PMID: 10973069
b3423	<i>glpR</i>	(NOT (GLxt>0))	PMID: 1372899 , 9524241
b3425	<i>glpE</i>	(Crp)	PMID: 1846566, 9524241
b3426	<i>glpD</i>	("CRP noMAL" AND (NOT (ArcA OR GlpR)))	NH 20; PMID: 2403539, 8955388; [135]
b3428	<i>glgP</i>	(Crp)	NH 67;

Table C.2, continued.

bNum	Gene	Rule	Reference
b3429	<i>glgA</i>		NH 67; PMID: 8576033
b3430	<i>glgC</i>		PMID: 12067347
b3433	<i>asd</i>		NH 32
b3437	<i>gntK</i>	(NOT (GntR) AND "CRP GLCN")	PMID: 8655507
b3438	<i>gntR</i>	(NOT (GLCNxt > 0))	PMID 9537375
b3450	<i>ugpC</i>	(Crp OR PhoB)	PMID: 1987150, 1745236
b3451	<i>ugpE</i>	(Crp OR PhoB)	PMID: 1987150, 1745236
b3452	<i>ugpA</i>	(Crp OR PhoB)	PMID: 1987150, 1745236
b3453	<i>ugpB</i>	(Crp OR PhoB)	PMID: 1987150, 1745236
b3454	<i>livF</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3455	<i>livG</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3456	<i>livM</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3457	<i>livH</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3458	<i>livK</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3460	<i>livJ</i>	(NOT(LEUxt > 0) OR Lrp)	NH 22; PMID: 1729203
b3493	<i>pitA</i>		PMID: 11489853
b3500	<i>gor</i>	(OxyR OR RpoS)	PMID: 8593953
b3517	<i>gadA</i>	((NOT (Growth > 0) AND NOT Crp) OR (pH < 4))	NH 22; PMID: 11976288
b3519	<i>treF</i>	(RpoS)	PMID: 9148912, 8892826
b3526	<i>kdgK</i>	(NOT KdgR)	NH 20
b3528	<i>dctA</i>	((("CRP noMAN") AND NOT(ArcA) AND (DcuR)) AND RpoN)	PMID: 10482502
b3551	<i>bisC</i>		
b3553	<i>yiaE</i>		PMID: 11237876, 9811658
b3564	<i>xylB</i>	((XylR AND Crp) OR XylR)	PMID: 9371449
b3565	<i>xylA</i>	((XylR AND Crp) OR XylR)	PMID: 9371449
b3566	<i>xylF</i>	((XylR AND Crp) OR XylR)	PMID: 9371449
b3567	<i>xylG</i>	((XylR AND Crp) OR XylR)	PMID: 9371449
b3568	<i>xylH</i>	((XylR AND Crp) OR XylR)	PMID: 9371449
b3569	<i>xylR</i>	(XYLxt > 0)	PMID: 9371449
b3572	<i>avtA</i>	(NOT (ALAxt > 0 OR LEUxt > 0))	NH 24,27; PMID: 6373721
b3574	<i>yiaJ</i>	(NOT (FUCxt>0))	PMID: 10913096
b3575	<i>yiaK</i>	((NOT YiaJ) AND Crp)	PMID: 10913096
b3579	<i>yiaO</i>	(Crp OR YiaJ)	
b3581	<i>sgbH</i>	(NOT YiaJ)	PMID: 10913096
b3583	<i>sgbE</i>	(NOT YiaJ)	PMID: 10913096

Table C.2, continued.

bNum	Gene	Rule	Reference
b3588	<i>aldB</i>	(RpoS AND (Crp))	PMID: 7768815
b3599	<i>mtlA</i>	(NOT (MtlR))	NH 20
b3600	<i>mtlD</i>	(NOT MtlR)	NH 20
b3601	<i>mtlR</i>	(NOT (MNTxt > 0))	PMID: 8300537
b3603	<i>lldP</i>	(NOT (ArcA))	PMID: 8892825
b3605	<i>lldD</i>	(LLACxt \downarrow 0 AND O2xt \downarrow 0)	PMID: 8407843
b3607	<i>cysE</i>	(CysB)	NH 31
b3608	<i>gpsA</i>		
b3612	<i>yibO</i>		
b3616	<i>tdh</i>	(NOT (Lrp) AND (LEUxt > 0))	NH 22
b3617	<i>tbl</i>	(NOT (Lrp) AND (LEUxt > 0))	NH 22
b3619	<i>rfaD</i>		NH 69
b3620	<i>rfaF</i>		NH 69
b3621	<i>rfaC</i>		NH 69
b3622	<i>rfaL</i>		NH 69
b3626	<i>rfaJ</i>		NH 69
b3627	<i>rfaI</i>		NH 69
b3631	<i>rfaG</i>		NH 69
b3633	<i>kdtA</i>		
b3634	<i>coaD</i>		
b3640	<i>dut</i>		
b3642	<i>pyrE</i>	(NOT (URAxxt > 0 OR GNxt > 0))	NH 35
b3648	<i>gmk</i>		NH 34
b3653	<i>gltS</i>	(ASPxt > 0)	NH 22
b3654	<i>yicE</i>		
b3665	<i>yicP</i>		
b3666	<i>uhpT</i>	(Crp OR UhpA)	PMID: 11702079
b3668	<i>uhpB</i>	(G6Pxt > 0)	PMID: 11702079
b3669	<i>uhpA</i>	(UhpB)	PMID: 7596290
b3670	<i>ilvN</i>	(NOT(LEUxt > 0 OR VALxt > 0) AND Crp)	NH 27
b3671	<i>ilvB</i>	(NOT(LEUxt > 0 OR VALxt > 0) AND Crp)	NH 27
b3691	<i>dgoT</i>	(GALNTxt>0)	NH 20
b3692	<i>dgoA</i>	(GALNTxt>0)	NH 20
b3693	<i>dgoK</i>	(GALNTxt>0)	NH 20
b3708	<i>tnaA</i>	(Crp AND (TRPxxt>0 OR CYStxt > 0))	NH 22

Table C.2, continued.

bNum	Gene	Rule	Reference
b3709	<i>tnaB</i>	(Crp AND (TRPx _t >0))	NH 22, 28
b3725	<i>pstB</i>	(PhoB)	PMID: 2651888, 3054125
b3726	<i>pstA</i>	(PhoB)	PMID: 2651888, 3054125
b3727	<i>pstC</i>	(PhoB)	PMID: 2651888, 3054125
b3728	<i>pstS</i>	(PhoB)	PMID: 2651888, 3054125
b3729	<i>glmS</i>		PMID: 11139621
b3730	<i>glmU</i>	(NagC)	PMID: 11139621
b3731	<i>atpC</i>		
b3732	<i>atpD</i>		
b3733	<i>atpG</i>		
b3734	<i>atpA</i>		
b3735	<i>atpH</i>		
b3736	<i>atpF</i>		
b3737	<i>atpE</i>		
b3738	<i>atpB</i>		
b3739	<i>atpI</i>		
b3743	<i>asnC</i>	(NOT (ASNxt > 0) AND NRI_hi)	NH 24
b3744	<i>asnA</i>	(NOT (ASNxt > 0) AND AsnC)	NH 24
b3748	<i>rbsD</i>	("CRP noXYL" AND NOT(RbsR))	PMID: 9666469, 9673030, 6327616
b3749	<i>rbsA</i>	("CRP noXYL" AND NOT(RbsR))	PMID: 9666469, 9673030, 6327616
b3750	<i>rbsC</i>	("CRP noXYL" AND NOT(RbsR))	PMID: 9666469, 9673030, 6327616
b3751	<i>rbsB</i>	("CRP noXYL" AND NOT(RbsR))	PMID: 9666469, 9673030, 6327616
b3752	<i>rbsK</i>	("CRP noXYL" AND NOT(RbsR))	PMID: 9666469, 9673030, 6327616
b3753	<i>rbsR</i>	(NOT (RIBxt>0))	PMID: 9666469, 9673030, 6327616
b3767	<i>ilvG_1</i>	(NOT(LEUxt > 0 OR ILExt > 0 OR VALxt > 0) AND Lrp)	NH 27
b3768	<i>ilvG_2</i>	(NOT(LEUxt > 0 OR ILExt > 0 OR VALxt > 0) AND Lrp)	NH 27
b3769	<i>ilvM</i>	(NOT(LEUxt > 0 OR ILExt > 0 OR VALxt > 0) AND Lrp)	NH 27
b3770	<i>ilvE</i>		NH 27
b3771	<i>ilvD</i>		NH 27
b3772	<i>ilvA</i>		NH 27

Table C.2, continued.

bNum	Gene	Rule	Reference
b3773	<i>ilvY</i>	(NOT VALxt > 0)	NH 27, PMID: 10588699
b3774	<i>ilvC</i>	(IlvY)	NH27; PMID: 10588699
b3784	<i>wecA</i>		
b3786	<i>wecB</i>		
b3787	<i>wecC</i>		
b3788	<i>rffG</i>		
b3789	<i>rffH</i>		
b3790	<i>wecD</i>		
b3791	<i>wecE</i>		
b3793	<i>wecF</i>		
b3794	<i>wecG</i>		
b3803	<i>hemX</i>		
b3804	<i>hemD</i>		PMID: 8997718
b3805	<i>hemC</i>		PMID: 8997718
b3806	<i>cyaA</i>	(NOT Crp)	
b3809	<i>dapF</i>		NH 32
b3821	<i>pldA</i>		
b3825	<i>pldB</i>		
b3828	<i>metR</i>	(NOT (METxt > 0))	NH 33
b3829	<i>metE</i>	((NOT MetJ) AND MetR)	NH 33
b3831	<i>udp</i>	(NOT (CytR) OR Crp)	NH 35
b3833	<i>ubiE</i>		
b3835	<i>ubiB</i>		
b3843	<i>yigC</i>		
b3845	<i>fadA</i>	(NOT (FadR2) OR NOT (ArcA))	NH 21
b3846	<i>fadB</i>	(NOT (FadR2) OR NOT (ArcA))	NH 21
b3849	<i>trkH</i>		NH 72
b3850	<i>hemG</i>		
b3868	<i>glnG</i>	(NOT (NH3xt > 2))	NH 23, NH 24
b3869	<i>glnL</i>	(ON)	NH 24
b3870	<i>glnA</i>	(Crp AND RpoN)	NH 24; PMID: 12218022
b3892	<i>fdoI</i>	((O2xt > 0) OR ((NOT (O2xt > 0) AND (NO3xt > 0))))	PMID: 8522521
b3893	<i>fdoH</i>	((O2xt > 0) OR ((NOT (O2xt > 0) AND (NO3xt > 0))))	PMID: 8522521

Table C.2, continued.

bNum	Gene	Rule	Reference
b3894	<i>fdoG</i>	((O2xt > 0) OR ((NOT (O2xt > 0) AND (NO3xt > 0))))	PMID: 8522521
b3902	<i>rhaD</i>	(RhaS OR (RhaS AND Crp))	PMID: 10852886
b3903	<i>rhaA</i>	(RhaS OR (RhaS AND Crp))	PMID: 10852886
b3904	<i>rhaB</i>	(RhaS OR (RhaS AND Crp))	PMID: 10852886
b3905	<i>rhaS</i>	(RhaR)	PMID: 10852886
b3906	<i>rhaR</i>	(RMNxt > 0)	PMID: 10852886
b3907	<i>rhaT</i>	(RhaS OR (RhaS AND Crp))	PMID: 8757746
b3908	<i>sodA</i>	(NOT (ArcA OR Fur) OR (MarA OR Rob OR SoxS))	PMID: 8412671
b3909	<i>kdgT</i>	(NOT KdgR)	NH 20
b3912	<i>cpxR</i>	(Stress > 0)	PMID: 10671468
b3916	<i>pfkA</i>		
b3917	<i>sbp</i>	(CysB)	NH 31
b3918	<i>cdh</i>		
b3919	<i>tpiA</i>		
b3926	<i>glpK</i>	("CRP noMAL" AND (NOT(GlpR)))	NH 20; PMID:1372899; [135]
b3927	<i>glpF</i>		PMID: 1372899
b3929	<i>menG</i>		
b3930	<i>menA</i>		
b3934	<i>cytR</i>	(CYTDxt > 0)	PMID: 8626289
b3938	<i>metJ</i>	(METxt > 0)	PMID: 12218041
b3939	<i>metB</i>	(NOT MetJ)	NH 32,33
b3940	<i>metL</i>	(NOT MetJ)	NH 32,33
b3941	<i>metF</i>		NH 36
b3942	<i>katG</i>	((Growth>0) AND OxyR AND RpoS)	PMID: 12589799
b3945	<i>gldA</i>		PMID: 8132480
b3946	<i>talC</i>		
b3951	<i>pflD</i>	(ArcA OR Fnr)	NH 95; PMID: 7934836
b3952	<i>pflC</i>	(ArcA OR Fnr)	NH 95; PMID: 7934836
b3956	<i>ppc</i>		
b3957	<i>argE</i>	(NOT ArgR)	NH 25
b3958	<i>argC</i>	(NOT ArgR)	NH 25
b3959	<i>argB</i>	(NOT ArgR)	NH 25
b3960	<i>argH</i>	(NOT ArgR)	NH 25
b3961	<i>oxyR</i>	(H2O2xt > 0)	PMID: 12589799
b3962	<i>sthA</i>		
b3966	<i>btuB</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b3967	<i>muri</i>		
b3972	<i>murB</i>		
b3973	<i>birA</i>	(BTxt > 0)	PMID 12368242, NH 46
b3974	<i>coaA</i>		
b3990	<i>thiH</i>		
b3991	<i>thiG</i>		
b3992	<i>thiF</i>		
b3993	<i>thiE</i>		
b3994	<i>thiC</i>		
b3997	<i>hemE</i>		
b4005	<i>purD</i>	(NOT (PurR))	NH 34
b4006	<i>purH</i>		NH 34
b4013	<i>metA</i>	(NOT (MetJ) OR MetR)	NH 33
b4014	<i>aceB</i>	(NOT (IclR) AND (NOT (ArcA) OR NOT (Cra)))	NH 16, 95; PMID: 8755903, 2001680
b4015	<i>aceA</i>	(NOT (IclR) AND (NOT (ArcA) OR NOT (Cra)))	NH 16, 95; PMID: 8755903, 2001680
b4018	<i>iclR</i>	(FadR)	PMID: 2001680, 8755903
b4019	<i>metH</i>	(MetR)	NH 33
b4024	<i>lysC</i>	(NOT LYSxt > 0)	NH 32
b4025	<i>pgi</i>		
b4031	<i>xylE</i>	(XylR)	PMID: 9371449
b4032	<i>malG</i>	((MalT AND Crp) OR MalT)	PMID:11931562, 11867639, 9529892
b4033	<i>malF</i>	((MalT AND Crp) OR MalT)	PMID:11931562, 11867639, 9529892
b4034	<i>malE</i>	((MalT AND Crp) OR MalT)	PMID:11931562, 11867639, 9529892
b4035	<i>malK</i>	((MalT AND Crp) OR MalT)	PMID:11931562, 11867639, 9529892
b4036	<i>lamB</i>	((MalT AND Crp) OR MalT)	PMID:11931562, 11867639, 9529892
b4039	<i>ubiC</i>	((NOT FnR) AND Crp)	PMID: 9315722
b4040	<i>ubiA</i>	((NOT FnR) AND Crp)	PMID: 9315722, 7765507
b4041	<i>plsB</i>		
b4042	<i>dgrA</i>		
b4053	<i>alr</i>		NH 22,24

Table C.2, continued.

bNum	Gene	Rule	Reference
b4054	<i>tyrB</i>	(NOT(((PHExt>0) OR (TYRxt>0)) AND TyrR))	NH 24:PMID:12207706
b4062	<i>soxS</i>	(SoxR)	NH 95
b4063	<i>soxR</i>	((H2O2xt > 0) OR ("Oxidative Stress" > 0))	NH 95
b4069	<i>acs</i>	(RpoS OR Fnr OR ((NOT IclR) AND ("CRP noSUCc")))	PMID: 10894724
b4077	<i>gltP</i>		NH 22
b4079	<i>fdhF</i>	(FhlA AND RpoN AND (NOT (O2xt > 0)))	NH 18
b4089	<i>rpiR</i>	(NOT (RIBxt>0))	PMID: 8576032
b4090	<i>rpiB</i>	(NOT(RpiR))	PMID: 8572885, 8576032
b4111	<i>proP</i>	(NOT (Crp) AND Fis AND RpoS)	NH 22; PMID: 9079929
b4116	<i>adiY</i>	((pH < 7) AND (NOT (O2xt > 0)) AND NOT ("Rich Medium" > 0))	PMID: 8704970
b4117	<i>adiA</i>	(AdiY)	NH 22; PMID 8704970
b4118	<i>melR</i>	((MELIxt > 0) OR (MELIxt > 0 AND Crp))	PID: 10747919, 10760178
b4119	<i>melA</i>	((MeLR) OR (MeLR AND Crp))	PMID: 10747919, 10760178
b4120	<i>melB</i>	((MeLR) OR (MeLR AND Crp))	PMID: 10747919, 10760178
b4122	<i>fumB</i>	((Fnr) OR NOT (Crp) OR (DcuR) OR NOT (NarL))	PMID: 9418241
b4123	<i>dcuB</i>	((("CRP noMAN") AND (Fnr) AND (DcuR)) AND NOT (NarL))	PMID: 9852003, 9973351
b4124	<i>dcuR</i>	(DcuS)	PMID: 9973351
b4125	<i>dcuS</i>	((SUCCxt > 0) OR (ASPxt > 0) OR (FUMxt > 0) OR (MALxt > 0))	PMID: 9973351
b4131	<i>cadA</i>	(ArcA OR CadC)	PMID: 9075621, 7830562
b4132	<i>cadB</i>	(ArcA OR CadC)	PMID: 7830562, 9075621
b4133	<i>cadC</i>	(LYSxt > 0)	NH 33, PMID: 7830562
b4138	<i>dcuA</i>		PMID: 9852003
b4139	<i>aspA</i>	((Crp AND NOT (Fnr)) OR Fnr)	NH 22
b4151	<i>frdD</i>	(Fnr OR DcuR OR NOT (NarL))	PMID: 9973351, 8576043
b4152	<i>frdC</i>	(Fnr OR DcuR OR NOT (NarL))	PMID: 9973351, 8576043
b4153	<i>frdB</i>	(Fnr OR DcuR OR NOT (NarL))	PMID: 9973351, 8576043
b4154	<i>frdA</i>	(Fnr OR DcuR OR NOT (NarL))	PMID: 9973351, 8576043
b4160	<i>psd</i>		

Table C.2, continued.

bNum	Gene	Rule	Reference
b4177	<i>purA</i>	(NOT (PurR) OR RpoE)	NH 34
b4196	<i>sgaH</i>		
b4197	<i>sgaU</i>		
b4198	<i>sgaE</i>		
b4208	<i>cycA</i>		
b4226	<i>ppa</i>		
b4227	<i>ytfQ</i>		
b4228	<i>ytfR</i>		
b4229	<i>ytfS</i>		
b4230	<i>ytfT</i>		
b4231	<i>yjfF</i>		
b4232	<i>fbp</i>		
b4238	<i>nrdD</i>	(Fnr)	PMID: 8954104
b4239	<i>treC</i>	((NOT TreR) AND Crp) OR (NOT TreR))	PMID: 9148912
b4240	<i>treB</i>	((NOT TreR) AND Crp) OR (NOT TreR))	PMID: 9148912
b4241	<i>treR</i>	(NOT (TRExt > 0))	PMID: 9148912
b4244	<i>pyrI</i>	(NOT (URAx _t > 0 OR GN _x t > 0))	NH 35
b4245	<i>pyrB</i>	(NOT (URAx _t > 0 OR GN _x t > 0))	NH 35
b4254	<i>argI</i>	(NOT ArgR)	NH 25
b4264	<i>idnR</i>	((IDN _x t > 0) OR (5KG _x t > 0))	PMID: 9658018
b4265	<i>idnT</i>	(IdnR)	PMID: 9658018
b4266	<i>idnO</i>	(IdnR)	PMID: 9658018
b4267	<i>idnD</i>	(IdnR)	PMID: 9658018
b4268	<i>idnK</i>		PMID: 9658018
b4301	<i>sgcE</i>		
b4321	<i>gntP</i>	(Crp AND NOT (GLCN _x t > 0))	PMID: 8550444
b4322	<i>uxuA</i>	(NOT ExuR AND NOT UxuR)	NH 20; PMID: 3083215
b4323	<i>uxuB</i>	(NOT ExuR AND NOT UxuR)	NH 20; PMID: 3083215
b4324	<i>uxuR</i>	(NOT(UXUBR > 0)) AND NOT(UXAC1R > 0) AND NOT(UXUBR < 0) AND NOT(UXAC1R < 0))	NH 20
b4381	<i>deoC</i>	((NOT DeoR) OR ((NOT DeoR) AND (Crp) AND (NOT CytR)))	NH 20
b4382	<i>deoA</i>	(NOT (DeoR OR CytR) AND Crp)	NH 35

Table C.2, continued.

bNum	Gene	Rule	Reference
b4383	<i>deoB</i>	((NOT DeoR) OR ((NOT DeoR) AND (Crp) AND (NOT CytR)) OR ((INSxt > 0) OR (GNxt > 0)))	NH 20
b4384	<i>deoD</i>	((NOT DeoR) OR ((NOT DeoR) AND (NOT CytR)) OR (INSxt > 0) OR (GNxt > 0))	NH 34
b4388	<i>serB</i>		NH 30
b4390	<i>nadR</i>	("high NAD")	PMID: 10464228
b4393	<i>trpR</i>	(TRPxxt>0)	NH 28
b4395	<i>gpmB</i>		
b4396	<i>rob</i>	(dipyridyl > 0)	PMID: 11844771
b4401	<i>arcA</i>	(NOT (O2xt > 0))	PMID: 2964639, NH 95
b4407	<i>thiS</i>		

C.A.1 Complex regulation

Complex regulation made description of simple rules difficult in some cases, listed in Table C.3. For example, Crp has complex regulation based on the level of cAMP in the cell. To describe this using Boolean logic, we divided the responses into categories based on the data of PMID: 5337847 and assuming that repression by a “higher” level substrate was complete until the substrate was exhausted. Additionally, because most Crp testing involved only glucose, we have a more general Crp statement which depends on glucose only. Furthermore, The method we describe does not currently allow for the calculation of internal metabolite concentrations. For most cases where internal metabolite concentrations are involved in activation/repression, we simply use concentration of related external metabolites as an approximation (e.g., for induction of the *lac* operon, we consider external lactose, rather than internal allolactose, the inducer). However, in the case of important central metabolites, we use the values of connected fluxes to approximate concentration qualitatively. Transcription factor FadR seems to respond to two different stimuli and regulates different sets of genes accordingly. A second rule was written for FadR activity to accomodate this action. The nitrogen response has a fast (low-level) and a slow (high-level) response, which we describe using two rules. Finally, the following stimuli were recorded in the literature and therefore included in the model but are not yet defined or accounted for strictly: dipyridyl, heat shock, high NAD external concentration, high osmolarity, LB media/rich media, oxidative stress, salicylate, stress, and the stringent response.

Table C.3: A separate table of complex regulatory rules in the model

CRP GLCN	(glcn(e) > 0)
CRP noARAB	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0)))
CRP noGL	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (mal-L(e) > 0) OR (glyc(e) > 0))	
CRP noGLC	(NOT((glcn(e) > 0) OR (glc-D(e) > 0)))
CRP noGlcN	(NOT((glcn(e) > 0)))
CRP noGLT	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (mal-L(e) > 0) OR (glyc(e) > 0) OR (sbt-D(e) > 0))	
CRP noLAC	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (sbt-D(e) > 0)))
OR (rib-D(e) > 0) OR (mal-L(e) > 0) OR (glyc(e) > 0) OR (sbt-D(e) > 0) OR (lac-D(e) > 0))	
CRP noMAL	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (mal-L(e) > 0))	
CRP noMAN	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (mal-L(e) > 0) OR (glyc(e) > 0) OR (sbt-D(e) > 0) OR (lac-D(e) > 0) OR (man(e) > 0)))	
CRP noREB	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (suc(e) > 0))	
CRP noSUCC	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0))
OR (rib-D(e) > 0) OR (mal-L(e) > 0) OR (glyc(e) > 0) OR (sbt-D(e) > 0) OR (lac-D(e) > 0) OR (man(e) > 0))	
CRP noXYL	(NOT((glcn(e) > 0) OR (glc-D(e) > 0) OR (arab-L(e) > 0) OR (xyL-D(e) > 0)))
Surplus FDP	((NOT (FBP > 0) AND NOT (TKT2 > 0 OR TALA > 0 OR PGI > 0)) OR fru(e) > 0)
Surplus PYR	((NOT ((ME2 > 0 OR ME1 > 0) AND NOT (GLCpts > 0 OR PYK > 0 OR PFK > 0 OR LDHD < 0 OR LDHD2 > 0 OR SUCCt2.2 > 0 OR SUCCt2.3 > 0)))
FadR2	((NOT (titaca(e)>0 OR hdea(e)>0 OR ocdea(e)>0))
NRI_hi	(NRI_low AND Rp0N)
NRI_low	(GlnG AND GlnB AND GlnD)

C.B Metabolite list

The metabolite list matches that reported in iJR904[154], with the exception of six additional extracellular metabolites: 5dglcn(e), btn(e), cbi(e), h2o2(e), ppa(e), and thym(e). These eight metabolites act as stimuli for the regulatory network.

Table C.4: Metabolite abbreviations for the full *Escherichia coli* model.

Abbreviation	Metabolite
12PPDxt	(S)-Propane-1,2-diol
CADVxt	1,5-Diaminopentane
MDAPxt	meso-2,6-Diaminoheptanedioate
KDGxt	2-Dehydro-3-deoxy-D-gluconate
HCIxt	3-hydroxycinnamic acid
HPPPNTxt	3-(3-hydroxy-phenyl)propionate
GABAxt	4-Aminobutanoate
ACxt	Acetate
ACTOACxt	Acetoacetate
ACALxt	Acetaldehyde
NAGxt	N-Acetyl-D-glucosamine
NAMANxt	N-Acetyl-D-mannosamine
SLAxt	N-Acetylneuraminate
ADxt	Adenine
ADNxt	Adenosine
AKGxt	2-Oxoglutarate
DALAxt	D-Alanine
ALAxt	L-Alanine
ALTNxt	Allantoin
AMPxt	AMP
ARABxt	L-Arabinose
ARGxt	L-Arginine
ASNxt	L-Asparagine
ASPxt	L-Aspartate
BUTRATxt	Butyrate
VITB12xt	Cob
CHOLxt	Choline

Table C.4, continued.

Abbreviation	Metabolite
CITxt	Citrate
CO2xt	CO2
CRNxt	L-Carnitine
CYTSxt	Cytosine
CNOxt	Cyanate
CYSxt	L-Cysteine
CYTDxt	Cytidine
DAxt	Deoxyadenosine
DCxt	Deoxycytidine
DGxt	Deoxyguanosine
DHAxt	Dihydroxyacetone
DINxt	Deoxyinosine
DMSxt	Dimethyl sulfide
DMSOxt	Dimethyl sulfoxide
DUxt	Deoxyuridine
ETHxt	Ethanol
FE2xt	Fe2+
FORxt	Formate
FRUxt	D-Fructose
FUC1Pxt	L-Fucose 1-phosphate
FUCxt	L-Fucose
FUMxt	Fumarate
G6Pxt	D-Glucose 6-phosphate
GLACxt	D-Galactose
GALRTxt	D-Galactarate
GALNTxt	D-Galactonate
GLTLxt	Galactitol
GALRNTxt	D-Galacturonate
GLCAxt	D-Glucosamine
BBTxt	gamma-butyrobetaine
GLCxt	D-Glucose
GLCNxt	D-Gluconate
GLRTxt	D-Glucarate
GLCRNTxt	D-Glucuronate
GLNxt	L-Glutamine
GLUxt	L-Glutamate
GLYxt	Glycine

Table C.4, continued.

Abbreviation	Metabolite
GLALxt	D-Glyceraldehyde
BETxt	Glycine betaine
GLxt	Glycerol
GL3Pxt	Glycerol 3-phosphate
GLCLTxt	Glycolate
GSNxt	Guanosine
GNxt	Guanine
HEXT	H ⁺
H2Oxt	H ₂ O
C160xt	Hexadecanoate
HISxt	L-Histidine
HYXNxt	Hypoxanthine
IDNxt	L-Idonate
IEXT	L-Isoleucine
INDOLExt	Indole
INSxt	Inosine
Kxt	K ⁺
LACxt	D-Lactate
LLACxt	L-Lactate
LCTSxt	Lactose
LEUxt	L-Leucine
LYSxt	L-Lysine
MALxt	L-Malate
MLTxt	Maltose
MLTHXxt	Maltohexaose
MLTPTxt	Maltopentaose
MLTTRxt	Maltotriose
MLTTTxt	Maltotetraose
MANxt	D-Mannose
MAN6Pxt	D-Mannose 6-phosphate
MELIxt	Melibiose
METDxt	D-Methionine
METxt	L-Methionine
MNTxt	D-Mannitol
NAxt	Sodium
NACxt	Nicotinate

Table C.4, continued.

Abbreviation	Metabolite
NADxt	Nicotinamide adenine dinucleotide
NH3xt	ammonium
NMNxt	NMN
NO2xt	Nitrite
NO3xt	Nitrate
O2xt	O ₂
C180xt	octadecanoate
ORNxt	Ornithine
PHExt	L-Phenylalanine
PIxt	Phosphate
PNTOxt	
PPPNTxt	Phenylpropanoate
PROxt	L-Proline
PTRCxt	Putrescine
PYRxt	Pyruvate
RIBxt	D-Ribose
RMNxt	L-Rhamnose
GLTxt	D-Sorbitol
DSERxt	D-Serine
SERxt	L-Serine
SLFxt	Sulfate
SPMDxt	Spermidine
SUCCxt	Succinate
SUCxt	Sucrose
TARTxt	L-tartrate
TAURNxt	Taurine
THMNxt	Thiamin
THRxt	L-Threonine
DTxt	Thymidine
TMAxt	Trimethylamine
TMAOxt	Trimethylamine N-oxide
TRExt	Trehalose
TRPxt	L-Tryptophan
SSO3xt	Thiosulfate
C140xt	tetradecanoate
TYRxt	L-Tyrosine
URAxt	Uracil

Table C.4, continued.

Abbreviation	Metabolite
UREAxt	Urea
URIxt	Uridine
VALxt	L-Valine
XANxt	Xanthine
XTSNxt	Xanthosine
XYLxt	D-Xylose
MANxt	D-Mannose
FRUxt	D-Fructose
C180xt	octadecanoate
C160xt	Hexadecanoate
C140xt	tetradecanoate

C.C Detailed phenotypic analysis

The following sections detail possible reasons for the discrepancies and model predictions for carbon and nitrogen sources as well as for knockout strains. Some of the high discrepancy growth conditions were retested on a Bioscreen C (Helsinki, Finland) with five replicates; Bioscreen measures growth rates by monitoring OD. M-9 minimal media with 0.2% carbon source was used to test K-12 MG1655 growth on different carbon sources; W-salts media (10.5 g of K₂HPO₄, 4.5 g of KH₂PO₄, and 0.241 ml of 1 M MgSO₄ per liter) supplemented with 0.2% succinate and 0.2% nitrogen source was used to test wildtype growth on different nitrogen sources. Two controls were used: M-9 minimal media with no carbon source and 0.2% succinate W-salts media with no nitrogen source. Cells were precultured overnight in 0.2% succinate M-9 minimal media and transferred into the different media conditions Bioscreen was run over three days, and the relative growth rates (growth rate divided by the appropriate control growth rate) are reported in the tables below. MEME and MAST , sequence alignment and comparison tools, were used as reported previously[154] to identify putative genes for some of the enzymes that could resolve model and experiment discrepancies. Data is listed as (Experimental observation/FBA prediction/rFBA prediction).

C.C.1 Carbon Sources

Table C.5 lists the carbon sources which have a fraction of agreement (FA) between model prediction and experimental observation which is ≤ 0.60 . The “Biolog results” column reports for the 110 knockouts, how many grow and do not grow on the media specified, and the “Relative growth rate” is with respect to the control. “NA” indicates that this source was not tested.

Table C.5: Carbon sources with inconsistent growth phenotypes.

Carbon Source	FA	Biolog Results (G/NG)	BioScreen Data (WT)
Acetoacetic Acid	0.42	43 / 67	2.7
Formic Acid	0.05	104 / 6	0.8
Glycine	0.51	54 / 56	1
Thymidine	0	110 / 0	NA
g-Amino Butyric Acid	0.21	13 / 97	0.9
L-Arginine	0.35	37 / 73	NA
L-Ornithine	0.21	15 / 95	1
Putrescine	0.41	43 / 67	0.9
L-Glutamic Acid	0.3	22 / 88	NA

Formic Acid (+/-/-), Glycine (+/+/-) and Acetoacetic Acid (-/+/+)

The metabolic and regulatory models incorrectly predict growth phenotypes as measured on the Biolog plates with formate and acetoacetate as carbon sources, while only the regulatory model disagrees with experimental observations with glycine. According to the Biolog plates, *E. coli* grows with formate as a carbon source and does not grow with acetoacetate as a carbon source (growth on the latter carbon source has been observed, PMID: 3025185). Mixed Biolog results are observed for growth on glycine. Wildtype K-12 was retested for growth on all three carbon sources using the Bioscreen; in all cases the results are in agreement with the regulatory model predictions and disagree with the Biolog results.

Thymidine (+/-/-)

Both the regulated and unregulated models predict that thymidine can not be used as the sole carbon or nitrogen source. Thymidine can be converted to thymine by thymidine phosphorylase, this enzyme is already in the metabolic network. Older experimental studies have shown that thymine can be degraded by some strains of *E. coli* 4,5 and it has been proposed that *E. coli* B contains the reductive pathway involved in uracil and thymine degradation 5 (EC numbers 1.3.1.2 or 1.3.1.1, 3.5.2.2, 3.5.1.6). Sequence comparisons using MEME and MAST indicate that 1.3.1.2 might be encoded by b2106 and 3.5.2.2 might be encoded by

b2873 or b0512. Identification of this pathway in *E. coli* K-12 MG1655 would explain the observed Biolog data. Incorporating the associated metabolic genes and knowledge on how they are regulated would increase the predictive ability of the model.

L-Glutamic Acid (-/+/+)

The inability to grow on glutamate as the sole carbon source is believed to be due to a low transport capacity (NH 20). If measured a maximum rate for the uptake of glutamate can be used to further constrain the solutions predicted by the models.

g-Amino Butyric Acid, L-arginine, Ornithine and Putrescine (-/+/+)

Both models predict growth on g-amino butyrate (GABA), arginine, ornithine and putrescine as a sole carbon source. This is in disagreement with the Biolog and Bioscreen data, which indicate that these substrates are not suitable carbon sources. The *gab* pathway, needed for the degradation of GABA and putrescine, is reported to be expressed at a low constitutive level that is not sufficient to support growth on GABA (NH22) (strain W3110 is able to utilize GABA as a carbon source PMID: 12446648). In addition to the *gab* pathway arginine and ornithine, can also be degraded by enzymes in the *ast* pathway, but this latter pathway is only expressed under nitrogen limitation. The *gabDPTC* operon is induced under nitrogen limitation allowing these compounds to be used as a nitrogen sources PMID: 12446648. Constraining the maximum allowable fluxes through the *gab* pathway or including regulation of these genes in the model would explain the lack of growth and increase the predictive capabilities of the models.

C.C.2 Nitrogen Sources

Table C.6 is similar in format to Table C.5, with the exception that these substrates were tested as nitrogen sources with succinate serving as the carbon

source.

Table C.6: Nitrogen sources with inconsistent growth phenotypes.

Nitrogen Source	FA	Biolog Results (G/NG)	BioScreen Data (WT)
Adenine	0.42	44 / 66	NA
N-Acetyl-D-Mannosamine	0.48	52 / 58	1.8
Putrescine	0.6	64 / 46	3.5
L-Lysine	0.44	62 / 48	2.4
L-Methionine	0.42	64 / 46	2.1
L-Phenylalanine	0.24	84 / 26	1.5
Xanthine	0.04	106 / 4	NA
Guanosine	0.51	64 / 46	NA
Alanine-Leucine	0.28	79 / 31	NA

Adenine,N-Acetyl-D-Mannosamine and Putrescine (-/+/+)

These three nitrogen sources do not support growth according to the Biolog data, but are predicted to support growth by the regulated and unregulated models. It has been shown previously that *E. coli* can use adenine as a sole nitrogen source (PMID: 10986234), suggesting that the Biolog results might be inaccurate. N-acetyl-D-mannosamine and putrescine were also tested as nitrogen sources using the Bioscreen- growth rates were significantly higher than the control indicating that the Biolog results are incorrectly measuring a lack of growth.

L-Lysine, L-Methionine, L-Phenylalanine and Xanthine (+/-/-)

Both the Biolog data and Bioscreen data indicate that lysine, methionine, phenylalanine, and xanthine can be used as an alternate nitrogen sources. Neither the regulated or unregulated model predicts growth with these substrates as nitrogen sources, indicating that the metabolic enzymes, which allow incorporation of nitrogen from these substrates, are missing from the metabolic network. For the case of lysine, we could not find any data on how nitrogen is removed from lysine. Proposed pathways for methionine, phenylalanine, and xanthine utilization are summarized below. Methionine aminotransferase activity

has been observed in *E. coli* B, where methionine and α-ketoglutarate are converted to 2-oxo-4-methylthiobutyric acid and glutamate (PMID: 3541827), 2-oxo-4-methylthiobutyric acid is then converted into ethylene (PMID: 2693600). The pathway and associated genes have not been found in K-12 and so have not yet been included in the models. Including the phenylpyruvate decarboxylase reaction, which converts phenylpyruvate to phenylacetate (EC 4.1.1.43), as well as the complete phenylacetate degradation pathway (which has not yet been fully characterized) would enable the model to use phenylalanine as a nitrogen source. A xanthine dehydrogenase activity has been assigned to the *xdhA* gene product, where xanthine would be converted to uric acid and then presumably to allantoin (PMID: 10986234). Allantoin can not be used as a nitrogen source under aerobic conditions, so how nitrogen is removed from the base remains unclear biochemically.

Alanine-Leucine (+/+/-)

Leucine represses the synthesis of biosynthetic enzymes for isoleucine and valine, which is why the model predicts that *E. coli* won't grow with leucine or alanine+leucine as the sole nitrogen source. Experimentally growth with just leucine as the nitrogen source does not permit growth, but growth with both alanine and leucine allows for growth. A lower concentration of leucine might allow for growth with alanine if the repression of the isoleucine and valine biosynthetic enzymes is relaxed.

Guanosine (-/+/+)

Biolog data indicates growth with guanosine in 64 knockouts and no growth with 46 knockouts. Performing more replicates of the Biolog data and possibly testing the knockout strains on the Bioscreen would provide more information as to whether the model or the Biolog data is more accurate.

C.C.3 Knockout strains

All of the major failure modes between model predictions of knockouts and Biolog data are the case where the regulated and unregulated models predict the knockout to be lethal but the experimental data seems to suggest that they are not lethal. Most of these discrepancies involve knockouts which prevent the production of a biomass component.

glgA⁻, glgC⁻ (+/-)

These two genes are involved in the synthesis of glycogen. Three different hypothesis can be made from the model and data discrepancies: (1) glycogen is not an essential biomass component, (2) glycogen phosphorylase is reversible, or (3) there is a new redundant pathway for glycogen synthesis. If incorporated into the model any of these possibilities could resolve the model and data disagreements.

argB⁻, argC⁻, argD⁻, argE⁻, argG⁻ (+/-)

The following genes involved in arginine biosynthesis: *argB*, *argC*, *argD*, *argE*, and *argG*, are all lethal deletions according to the model but not in the Biolog data. For *argB*, *argC*, *argD*, and *argE* the growth phenotype can be explained by making a few reactions reversible in the model (ABUTD, PTRCTA, and ORNDC); the backwards reactions allow for a new route converting glutamate into ornithine (and then arginine). No information could be found regarding the reversibility of these enzymes. For *argG* there must be another isozyme.

purD⁻, purH⁻, metA⁻ (+/-)

The genes, *purD* and *purH* are responsible for the enzymes needed in the early and late steps of purine biosynthesis. One of the early reactions of methionine biosynthesis is carried out by the *metA* gene product. *E. coli* will obviously need to still make purines and methionine, so isozymes or alternate synthesis routes must

Table C.7: Knockout strains with inconsistent growth phenotypes.

Knockout strain	Affected Enzymes	FA	Biolog Results
<i>argB</i> (b3959) -	acetylglutamate kinase	0.39	76/34
<i>argC</i> (b3958) -	N-acetyl-g-glutamyl-phosphate reductase	0.31	86/24
<i>argD</i> (b3359) -	acylornithine transaminase	0.27	91/19
<i>argE</i> (b3957) -	acylornithine deacetylase	0.52	64/46
<i>argG</i> (b3172) -	argininosuccinate synthase	0.57	54/56
<i>glgA</i> (b3429) -	glycogen synthase	0.25	94/16
<i>glgC</i> (b3430) -	glucose-1-phosphate adenylyltransferase	0.25	94/16
<i>ilvD</i> (b3771) -	dihydroxy-acid hydratase	0.52	60/50
<i>ilvY</i> (b3773) -	transcriptional activator for isoleucine and valine synthesis	0.45	69/41
<i>metA</i> (b4013) -	homoserine O-succinyltransferase	0.39	77/33
<i>pgi</i> (b4025) -	glucose-6-phosphate isomerase	0.3	95/15
<i>pls</i> (b4041) -	glycerolphosphate acyltransferase	0.28	90/20
<i>purD</i> (b4005) -	phosphoribosylglycinamide synthase	0.3	92/18
<i>purH</i> (b4006) -	phosphoribosylaminoimidazolecarboxamide formyltransferase and IMP cyclohydrolase	0.41	74/36
<i>tpiA</i> (b3919) -	triose-phosphate isomerase	0.5	91/19

be available.

pgi⁻, tpiA⁻ (+/-/-)

Both *pgi* and *tpiA* are predicted to be lethal under most conditions by the model because with these knockouts there is no way of making glucose-6-phosphate from carbon sources that do not directly feed into upper glycolysis. Like with the *arg* knockouts, making some of the reactions in the model reversible (6-phosphogluconolactonase and either the entner doudoroff pathway or phosphogluconate dehydrogenase) would change the model predictions.

ilvD⁻, ilvY⁻ (+/-/-, +/n/-)

Both the *ilvD* and *ilvY* are incorrectly predicted by the model to be lethal because both are needed to make the biomass components leucine, valine, and isoleucine. *ilvD* encodes an enzyme in the metabolic pathways and *ilvY* is an transcriptional activator for *ilvC* (PMID: 9556617) encoding another essential enzyme in the pathway. There must be another way of making these amino acids, either alternate isozymes exist for *ilvD* and *ilvC* or in the case of *ilvY*, the level of IlvC is still high enough to permit growth.

C.D Detailed microarray analysis

This section contains the information from Figure 5.3 in more detail, together with the old and new model rules and notes.

	Gene	L2R	Old	New	New Rule	Comments
Bnum	<i>proB</i>	-0.55	5	5	(NOT (O2d > 0))	Essential for growth on Arginine
b0242	<i>ybiK</i>	-0.56	5	5	(ON)	Essential for WT growth
b0828	<i>trpB</i>	0.28	5	5	(NOT (TrpR))	Essential for WT growth
b1261	<i>gdhA</i>	-0.96	5	5	(NOT (Nac)) OR (GLU:t1 > 0))	Essential for WT growth
b1761	<i>hisC</i>	-0.54	5	5	(ON)	Essential for WT growth
b2021	<i>dapA</i>	-0.24	5	5	none	Very small shift: ON to 0 (ArcA and Fnr) or OxyR
b2478	<i>lvG_1</i>	-0.36	5	5	(NOT (LEU:t1 > 0 OR (ILE:t1 > 0 OR VAL:t1 > 0)) AND Lrp)) AND NOT (O2d:R))	
b3767	<i>lvM</i>	-0.53	5	5	(NOT (LEU:t1 > 0 OR (ILE:t1 > 0 OR VAL:t1 > 0)) AND Lrp)) AND (Fnr:))	
b3770	<i>lvE</i>	-0.27	5	5	(ON)	Essential for WT growth
b3771	<i>lvD</i>	-0.33	5	5	(ON)	Essential for WT growth
b3957	<i>argE</i>	0.39	5	5	(NOT ArgR)	Essential for WT growth
b0068	<i>sfuA</i>	0.33	5	5	(O2d > 0)	No knockout exhibited abolished shift
b0133	<i>panC</i>	-0.48	5	5	(ON)	Essential for WT growth
b0595	<i>entB</i>	1.20	5	5	(NOT (Fur))	Fur transcription is directly opposite to activity
b0776	<i>bioF</i>	0.48	5	5	(NOT (BioA)) AND (O2d > 0))	No knockout exhibited abolished shift
b0778	<i>bioD</i>	0.43	5	5	(NOT (BioA)) AND (O2d > 0))	No knockout exhibited abolished shift
b1210	<i>hemA</i>	-0.31	5	5	(ON)	Essential for WT growth
b1991	<i>cobT</i>	-0.27	5	5	(CB:t1 > 0) OR (Fnr:))	
b1993	<i>cobJ</i>	-0.17	5	5	(CB:t1 > 0) OR (Fnr:))	
b2153	<i>folE</i>	0.85	5	5	(ON)	Essential for WT growth
b3041	<i>rbB</i>	-0.96	5	5	(Fnr OR NarL)	Essential for WT growth
b3368	<i>cysG</i>	-0.63	5	5	(NOT (O2d > 0))	Correct
b3805	<i>hemC</i>	-0.19	5	5	(Fnr OR NarL)	No knockout exhibited abolished shift
b3929	<i>menG</i>	-0.50	5	5	none	Complex rule? 0 (O2d > 0) and not ArcA.
b3990	<i>thiH</i>	0.69	5	5	(O2d > 0)	No knockout exhibited abolished shift
b3993	<i>thiE</i>	0.51	5	5	(NOT (Fnr OR ArcA))	
b3994	<i>thiC</i>	0.69	5	5	(NOT (Fnr OR ArcA))	
b3994	<i>focA</i>	-1.52	5	5	(ArcA OR Fnr AND (Cip OR NOT (NarL)))	Correct
b1613	<i>manA</i>	-0.72	5	5	(ON)	Essential for WT growth
b2297	<i>pfa</i>	-1.77	5	5	(ON)	Essential for WT growth
b4322	<i>uxUA</i>	-0.40	5	5	(NOT UXUR AND NOT UXuR) OR (O2d > 0))	No knockout exhibited abolished shift

Figure C.1: A complete list of the new regulatory rules generated by the combined computational-experimental approach.

		L2R	Old	New	New Rule	Comments
Bnum	Gene					
b0114	<i>aceE</i>	0.48	5	5	((NOT(Fm)) OR (Fs)) AND ((NOT(AcA) AND Fm))	-
b0115	<i>aceF</i>	0.48	5	5	((NOT(Fm)) OR (Fs)) AND ((NOT(AcA) AND Fm))	-
b0116	<i>lpdA</i>	1.32	5	5	(NOT (AcA AND Fm))	-
b0118	<i>acnB</i>	2.63	5	5	(NOT (AcA))	Correct
b0429	<i>cyoD</i>	3.89			(NOT (AcA) OR Fm)	Correct
b0430	<i>cyoC</i>	4.15			(NOT (AcA) OR Fm)	Correct
b0431	<i>cyoB</i>	3.73			(NOT (AcA) OR Fm)	Correct
b0432	<i>cyoA</i>	4.16			(NOT (AcA) OR Fm)	Correct
b0720	<i>glfA</i>	0.86	5	5	(ON)	Essential for WT growth
b0721	<i>sdhC</i>	4.70	5	5	((NOT((AcA) AND (Fm)) AND ((Cr) OR (Fs)))	AND, OR change
b0722	<i>sdhD</i>	4.63	5	5	((NOT((AcA) AND (Fm)) AND ((Cr) OR (Fs)))	AND, OR change
b0723	<i>sdhA</i>	3.01	5	5	((NOT((AcA) AND (Fm)) AND ((Cr) OR (Fs)))	AND, OR change
b0726	<i>sucA</i>	2.17	5	5	(NOT AcA)	-
b0727	<i>sucB</i>	2.07	5	5	(NOT AcA)	-
b0728	<i>succC</i>	2.27	5	5	(ON)	Essential for WT growth
b0729	<i>sucD</i>	2.81	5	5	(ON)	Essential for WT growth
b0733	<i>cyaA</i>	-0.79	5	5	(NOT (O2d > 0))	No knockout exhibited abolished shift
b0734	<i>cydB</i>	-0.66	5	5	(NOT (O2d > 0))	No knockout exhibited abolished shift
b0755	<i>gpmA</i>	0.84	5	5	(NOT (AcA AND Fm))	-
b0896	<i>dmscC</i>	-0.34	4	2	(NOT Nat.)	No shift
b0902	<i>pflA</i>	-1.02			(AcA OR Fm AND (Cp OR NOT(Nat.)))	Correct
b0903	<i>pflB</i>	-1.48			(AcA OR Fm AND (Cp OR NOT(Nat.)))	Correct
b0974	<i>hyvC</i>	-3.22			(AcA OR Fm) AND (App'Y)	Correct
b1136	<i>icdA</i>	1.34	5	5	(ON)	Essential for WT growth
b1241	<i>adnE</i>	-1.44	5	5	(NOT (O2d > 0) AND (NOT ((O2d > 0) AND (Cm))) AND ((Fs) OR NOT (Nat.) OR (FcsS)))	AND, OR change
b1276	<i>acnA</i>	1.30	5	5	(ON)	Essential: 10/1000 AcnA
b1415	<i>addA</i>	1.43	5	5	(ON)	Essential for WT growth
b1474	<i>fngG</i>	-0.07	4	2	(Nat.)	No shift
b1476	<i>fdnl</i>	0.00	4	2	(Nat.)	No shift
b1612	<i>fumA</i>	2.65			(NOT(AcA) OR NOT (O2d > 0))	Correct
b1676	<i>pykF</i>	-0.47	5	5	(NOT(AcA) OR NOT (O2d > 0))	No knockout exhibited abolished shift

Figure C.1, continued.

Bnum	Gene	L2R	Old		New Rule	Comments
			5	5		
b1779	<i>gapA</i>	-0.18	5	5	(NOT (O2d > 0))	No knockout exhibited abolished shift
b1854	<i>pykA</i>	-1.33	5	5	none	Complex rule? 0/10000 Fnr and not ArcA
b2276	<i>nuoN</i>	0.71	4	2	(ON)	Essential for growth on acetate
b2277	<i>nuoM</i>	0.54	4	2	(ON)	Essential for growth on acetate
b2278	<i>nuoL</i>	0.34	4	2	(ON)	Essential for growth on acetate
b2279	<i>nuoK</i>	0.34	4	2	(ON)	Essential for growth on acetate
b2280	<i>nuoJ</i>	0.34	4	2	(ON)	Essential for growth on acetate
b2281	<i>nuoI</i>	0.43	4	2	(ON)	Essential for growth on acetate
b2282	<i>nuoH</i>	0.60	4	2	(ON)	Essential for growth on acetate
b2283	<i>nuoG</i>	0.46	4	2	(ON)	Essential for growth on acetate
b2284	<i>nuoF</i>	0.32	4	2	(ON)	Essential for growth on acetate
b2285	<i>nuoE</i>	0.38	5	5	(ON)	Essential for growth on acetate
b2287	<i>nuoB</i>	0.28	4	2	(ON)	Essential for growth on acetate
b2288	<i>nuoA</i>	0.23	4	2	(ON)	Essential for growth on acetate
b2296	<i>ackA</i>	-1.49	5	5	(Fnr AND ArcA)	-
b2723	<i>hyuC</i>	-3.08	5	5	(Fnr AND RpoN AND (NOT (O2d > 0)))	Make this an OR?
b2779	<i>eno</i>	-0.42	5	5	(ON)	Essential for WT growth
b2925	<i>fbaA</i>	-0.45	5	5	(NOT (O2d > 0))	No knockout exhibited abolished shift
b2926	<i>pgk</i>	-0.59	5	5	(ON)	Essential for WT growth
b3236	<i>mdh</i>	2.23	5	5	(NOT(ArcA))	Correct
b3425	<i>glpE</i>	0.83	5	5	(Cp)	Essential for WT growth
b3892	<i>fdoI</i>	-0.04	4	2	(NO2d > 0)	No shift
b3893	<i>fdoH</i>	0.61	5	5	((O2d > 0) OR ((NOT (O2d > 0) AND (NO2d > 0))))	Correct
b3894	<i>fdoG</i>	0.90	5	5	((O2d > 0) OR ((NOT (O2d > 0) AND (NO2d > 0))))	Correct
b3916	<i>pflA</i>	-1.06	5	5	(NOT (O2d > 0))	No knockout exhibited abolished shift
b3919	<i>fplA</i>	-0.56	5	5	(ON)	Essential for WT growth
b3952	<i>pflC</i>	0.12	4	2	(ON)	Essential for WT growth
b3956	<i>ppc</i>	-0.45	5	5	(ON)	Essential for WT growth
b4151	<i>frdD</i>	-2.23	5	5	(NOT (O2d > 0) AND (DcaR OR NOT (Nat1)))	No knockout exhibited abolished shift
b4152	<i>frdC</i>	-0.88	5	5	(NOT (O2d > 0) AND (DcaR OR NOT (Nat1)))	No knockout exhibited abolished shift
b4153	<i>frdB</i>	-2.31	5	5	(NOT (O2d > 0) AND (DcaR OR NOT (Nat1)))	No knockout exhibited abolished shift
b4154	<i>frdA</i>	-0.80	5	5	(NOT (O2d > 0) AND (DcaR OR NOT (Nat1)))	No knockout exhibited abolished shift

Figure C.1, continued.

		L2R	Old	New	New Rule	Comments
	Gene					
bnum	<i>fadd</i>	-2.14	4	2	(NOT (F _{add2})) ((NOT (Stringen > 0)) OR (Stringen < 0)) OR CpxR OR RpeE OR FadP2))	No shift
b1805	<i>fabB</i>	-0.53	5	5	((NOT (Stringen > 0)) OR (Stringen < 0)) OR CpxR OR RpeE OR FadP2))	Essential, small shift.
b2323	<i>fabB</i>	0.46	5	5	(ON)	Essential for WT growth.
b4160	<i>psc</i>	0.59	5	5	(ON)	Essential for WT growth.
b2207	<i>yefB</i>	0.77	5	5	(NOT (F _{adP2}) OR NOT (A _{cA})) ((C _a) AND (C _{2a1} > 0))	Correct.
b2221	<i>fadF</i>	0.98	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b1702	<i>ppf</i>	0.16	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b2040	<i>rbuD</i>	0.46	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b2308	<i>hisQ</i>	0.32	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b2463	<i>maeB</i>	0.80	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b2530	<i>iscS</i>	0.80	5	5	(NOT (Lysat > 0) AND NOT (A _{cA} AND F _{r1})) (O _{2xi} > 0))	No knockout exhibited abolished shift
b2676	<i>nrdF</i>	0.77	5	5	(NOT (F _{rr} OR A _{cA}) AND ((NOT(G _{cR}) AND G _{cA}) OR L _p OR NOT PurR)))	No knockout exhibited abolished shift
b2904	<i>gcvH</i>	0.77	5	5	(NOT (G _{cR}) AND ((NOT(G _{cR}) AND G _{cA}) OR L _p OR NOT PurR)))	No knockout exhibited abolished shift
b2905	<i>gcvT</i>	1.38	5	5	(NOT (A _{cA}) AND (G _{cA})) (O _{2xi} > 0))	Correct.
b2976	<i>glcB</i>	0.81	5	5	(NOT (A _{cA}) OR NOT (C _{a1})) OR (NOT (A _{cA})) (NOT (I _{cr}) OR NOT (C _{a1}) OR NOT (A _{cA}))	AND, OR change
b4014	<i>aceB</i>	0.77	5	5	(NOT (I _{cr}) OR NOT (C _{a1}) OR NOT (A _{cA})) (O _{2xi} > 0))	AND, OR change
b4015	<i>aceA</i>	1.33	5	5	((C _a AND NOT (F _{r1})) OR F _{r1}) (O _{2xi} > 0))	Correct.
b4139	<i>aspA</i>	-1.05	5	5	(NOT (A _{cA}) AND O _{2xi}) (O _{2xi} > 0))	No knockout exhibited abolished shift
b4232	<i>fhp</i>	0.46	5	5	(NOT (A _{cA}) AND O _{2xi}) (O _{2xi} > 0))	No knockout exhibited abolished shift
b0033	<i>carB</i>	0.92	5	5	(ON)	Essential for WT growth.
b0888	<i>trxB</i>	0.52	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR))	Essential for WT growth.
b0945	<i>pyrD</i>	-0.89	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR))	Essential for WT growth.
b1062	<i>pyrC</i>	0.16	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR)) (NOT (A _{cA}))	Essential for WT growth.
b2234	<i>nrdA</i>	1.46	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR)) (NOT (A _{cA}))	Essential for WT growth.
b2235	<i>nrdB</i>	0.73	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR)) (NOT (A _{cA}))	Essential for WT growth.
b2476	<i>purC</i>	0.48	5	5	((NOT (C _T TSx > 0)) OR (G _{MH} > 0)) OR NOT PurR))	Essential for WT growth.
b2518	<i>ndk</i>	2.84	5	5	none	Complex rule, 10000 A _{cA} and not F _{r1} .
b3431	<i>udp</i>	-0.53	5	5	(NOT (G _{cR}) OR C _p) (NOT (G _{cR}) > 0))	Essential for WT growth.
b4238	<i>nrdD</i>	-0.77	5	5	(NOT (G _{cR}) > 0))	Correct.
b1656	<i>sodB</i>	-0.20	5	5	(NOT (A _{cA} OR F _{r1}) OR (M _{AA} OR Rob OR SodS)) (O _{2xi} > 0))	No knockout exhibited abolished shift
b3908	<i>soda</i>	4.36	5	5	((F _{r1} AND A _{cA}) OR C _p AND NOT Nat) (NOT (Surplus FDP))	Correct.
b0034	<i>caif</i>	-1.37	5	5		Essential, again activity and transcription seem opposite
b0080	<i>fruR</i>	-0.73	5	5		

Figure C.1, continued.

Bnum	Gene	L2R	Old	New	Comments
b0113	<i>pdfR</i>	0.95	5	(NOT (Surplus_PiR) OR (NOT (AcA) OR NOT (Fm)))	-
b0313	<i>betI</i>	1.98	5	(NOT (AcA) OR (CHOx1 > 0))	No ApPi-dependent genes were detected (all NA)
b0564	<i>appY</i>	-1.87	5	(NOT CIB)	OvR and SocS did not exhibit abolished shift
b0683	<i>fur</i>	0.26	5	((IF2x1 > 0) OR (NOT (Fm OR AcG)))	No knockout exhibited abolished shift
b0993	<i>torS</i>	-0.97	5	((TMAO>0) OR (NOT (Ox1 > 0)))	Essential for WT growth
b1187	<i>fadR</i>	0.54	5	((GLCx1 > 0 OR NOT (Ac1 > 0))	Separated activity and transcription
b1221	<i>narl</i>	0.56	5	((NOT Fm) AND (NOT AcA))	No knockout exhibited abolished shift
b1323	<i>tyrR</i>	-0.62	5	((TRPA1 > 0) OR ((TRX1 > 0) OR (PHE1 > 0)) OR NOT (O241 > 0))	Transcription of frr is opposite to activity
b1334	<i>frr</i>	0.63	5	(ON)	Separated activity and transcription
b1531	<i>marA</i>	0.90	5	((NOT AcA OR NOT Fm) OR OxyR OR (SafC1 > 0))	-
b1827	<i>kdgR</i>	0.47	5	((AcA) AND (Frr) AND (NOT (KDGx2>0)) AND NOT(UVAAx0)))	Essential for WT growth
b2087	<i>galR</i>	1.29	5	((NOT (GalT1r1 > 0))	Separated activity and transcription
b2573	<i>rpoE</i>	-0.62	5	(NOT (OxyR))	-
b2707	<i>srfR</i>	-0.36	5	((Frr AND NOT (GLTx1 > 0))	-
b2731	<i>flhA</i>	0.97	4	((NOT (NO3rx1 > 1)) AND (NOT (NO2rx1 > 1)) AND (NOT (TMAOr1 > 1))	Threshold required: trace amounts
b3357	<i>cpr</i>	-0.16	5	AND (NOT (DMSOr1 > 1)) AND (FOROr1>1))	Essential for WT growth
b3423	<i>glpR</i>	0.13	5	((CRT negAC))	No knockout exhibited abolished shift
b3806	<i>cyaA</i>	-0.54	5	((NOT (Cpx)>0) AND (O2x1 > 0))	-
b4124	<i>dcuR</i>	0.69	4	((DCuS))	Threshold required: trace amounts
b4125	<i>dcuS</i>	0.17	4	((SUCCx1 > 1) OR (ASPx1 > 1) OR (FUMx1 > 1) OR (MALx1 > 1))	Correct - activity matches expression shift
b4401	<i>arcA</i>	-0.69	5	((NOT (O2ar1 > 0))	Bei transcription is opposite to activity
b0314	<i>betT</i>	0.51	5	(NOT (BetI))	No knockout exhibited abolished shift
b0336	<i>codE</i>	0.43	5	((NOT (PurR) OR (NRI<-hi)) AND OxyR)	-
b0401	<i>bmQ</i>	-0.65	5	((NOT (O2ar1 > 0))	No knockout exhibited abolished shift
b0653	<i>gfkK</i>	0.73	5	((NOT (GLCx1 > 0)) OR NOT (AcA AND Fm))	-
b0854	<i>pofF</i>	0.83	5	((NOT (AcA AND Fm))	No knockout exhibited abolished shift
b0864	<i>arfP</i>	-0.57	5	((NOT (O2ar1 > 0))	-
b2129	<i>yehX</i>	0.35	5	((NOT (AcA) AND Frr))	-
b2309	<i>hisJ</i>	0.44	5	((NOT AcA OR NOT Fm) OR OxyR AND NOT (LysA1 > 0))	-
b2344	<i>fadI</i>	0.98	5	((NOT (Ctp OR FadR OR OmpR) OR NOT (AcA)))	Essential for WT growth
b2423	<i>cysW</i>	-0.34	5	(CysB)	-
b2425	<i>cysP</i>	-0.35	5	(CysB)	Essential for WT growth

Figure C.1, continued.

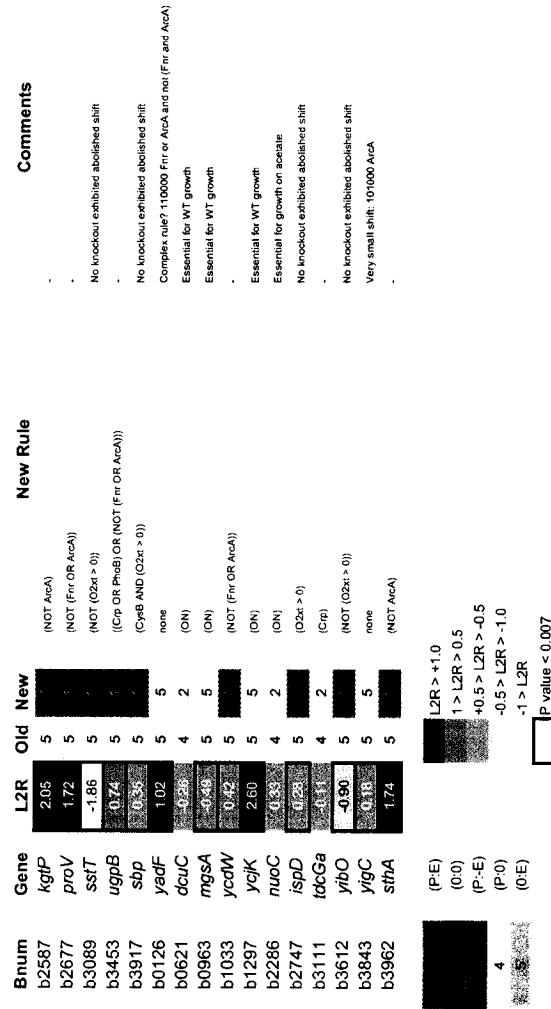


Figure C.1, continued.

The text of this chapter, in part or in full, is a reprint of the material as it appears in "Integrating high-throughput data and computational models leads to *E. coli* network elucidation" which has been submitted for publication. I was the primary author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Bibliography

- [1] The *lac* and *gal* operons today. *Regulation of gene expression in Escherichia coli*. E. C. C. Lin and A. S. Lynch. Austin, Texas, R.G. Landes Company: 181-200, 1996.
- [2] Altschul, S. F., T. L. Madden, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-402, 1997.
- [3] Anderson, A. and R. A. Cooper. Gluconeogenesis in *Escherichia coli*: The role of triose phosphate isomerase. *FEBS Letters* 4: 19-20, 1969.
- [4] Aristidou, A. and M. Penttilä. Metabolic engineering applications to renewable resource utilization. *Current Opinion in Biotechnology* 11: 187-98, 2000.
- [5] Arkin, A., J. Ross, et al. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149: 1633-48, 1998.
- [6] Bailey, J. E. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology Progress* 14: 8-20, 1998.
- [7] Bailey, J. E. Complex biology with no parameters. *Nature Biotechnology* 19: 503-4, 2001.
- [8] Bar-Joseph, Z., G. K. Gerber, et al. Computational discovery of gene modules and regulatory networks. *Nature Biotechnology* 21: 1337-42, 2003.
- [9] Benjamini, Y. and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 289-300, 1995.
- [10] Blaine-Metting, F. and M. F. Romine. Microbial genomics: the floodgates open. *Trends in Microbiology* 5: 91-92, 1997.
- [11] Blakeslee, S. Searching for simple rules of complexity. *The New York Times*, December 26, 1995: C1.

- [12] Blattner, F. R., G. Plunkett, 3rd, et al. The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453-74,1997.
- [13] Bledig, S. A., T. M. Ramseier, et al. Frur mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*. *Journal of Bacteriology* 178: 280-3,1996.
- [14] Bolouri, H. and E. H. Davidson. Modeling transcriptional regulatory networks. *Bioessays* 24: 1118-29.,2002.
- [15] Bonarius, H. P. J., G. Schmid, et al. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology* 15: 308-314,1997.
- [16] Bono, H., H. Ogata, et al. Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Research* 8: 203-10,1998.
- [17] Boston, T. and T. Atlung. FNR-mediated oxygen-responsive regulation of the *nrdDG* operon of *Escherichia coli*. *Journal of Bacteriology* 185: 5310-3,2003.
- [18] Calhoun, M. W., K. L. Oden, et al. Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. *Journal of Bacteriology* 175: 3020-5,1993.
- [19] Campbell, J. W. and J. E. Cronan, Jr. *Escherichia coli* FadR positively regulates transcription of the *fabB* fatty acid biosynthetic gene. *Journal of Bacteriology* 183: 5982-90,2001.
- [20] Carrier, T. A. and J. D. Keasling. Investigating Autocatalytic Gene Expression Systems through Mechanistic Modeling. *Journal of Theoretical Biology* 201: 25-36,1999.
- [21] Chao, G., J. Shen, et al. Aerobic regulation of isocitrate dehydrogenase gene (*icd*) expression in *Escherichia coli* by the *arcA* and *fnr* gene products. *Journal of Bacteriology* 179: 4299-304,1997.
- [22] Charpentier, B., et al. The EII Glc protein is involved in glucose-mediated activation of *Escherichia coli* *gapA* and *gapB-pgk* transcription. *Journal of Bacteriology* 180(24):6476-6483, 1998.
- [23] Chartrain, M., P. M. Salmon, et al. Metabolic engineering and directed evolution for the production of pharmaceuticals. *Current Opinion in Biotechnology* 11: 209-14,2000.
- [24] Christensen, B. and J. Nielsen. Metabolic network analysis. A powerful tool in metabolic engineering. *Advances in Biochemical Engineering/Biotechnology* 66: 209-31,2000.

- [25] Compan, I. and D. Touati. Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Molecular Microbiology* 11: 955-64,1994.
- [26] Cortay, J.-C., et al. Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor. *EMBO Journal* 10(3):675-9, 1991.
- [27] Cotter, P. A., S. B. Melville, et al. Aerobic regulation of cytochrome d oxidase (*cydAB*) operon expression in *Escherichia coli*: roles of Fnr and ArcA in repression and activation. *Molecular Microbiology* 25: 605-15,1997.
- [28] Courtright, J. B. and U. Henning. Malate dehydrogenase mutants in *Escherichia coli* K-12. *Journal of Bacteriology* 102,1970.
- [29] Covert, M. W., C. H. Schilling, et al. Metabolic Modeling of Microbial Strains *in silico*. *Trends in Biochemical Sciences* 26: 179-186,2001.
- [30] Covert, M. W., C. H. Schilling, et al. Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology* 213: 73-88,2001.
- [31] Covert, M. W. and B. O. Palsson. Transcriptional Regulation in Constraints-based Metabolic Models of *Escherichia coli*. *Journal of Biological Chemistry* 277: 28058-64,2002.
- [32] Covert, M. and B. O. Palsson. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *Journal of Theoretical Biology* 221: 309-25,2003.
- [33] Creaghan, I. T. and J. R. Guest. Succinate dehydrogenase-dependent nutritional requirement for succinate in mutants of *Escherichia coli* K-12. *Journal of General Microbiology* 107: 1-13,1978.
- [34] Cronan Jr., J., E and D. Laporte. Tricarboxylate Acid cycle and glyoxylate bypass.*Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, DC., ASM Press. 1: 206-216, 1996.
- [35] Cunningham, P. R. and D. P. Clark. The use of suicide substrates to select mutants of *Escherichia coli* lacking enzymes of alcohol fermentation. *Molecular General Genetics* 205: 487-93,1986.
- [36] Cunningham, L., M. J. Gruer, et al. Transcriptional regulation of the aconitase genes (*acnA* and *acnB*) of *Escherichia coli*. *Microbiology* 143 (Pt 12): 3795-805,1997.
- [37] Cunningham, L. and J. R. Guest. Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology* 144 (Pt 8): 2113-23,1998.

- [38] Darwin, A.J., Li, J. and Stewart, V. Analysis of nitrate regulatory protein NarL-binding sites in the *fdnG* and *narG* operon control regions of *Escherichia coli* K-12. *Molecular Microbiology* 20(3): 621-632, 1996.
- [39] Datsenko, K. A. and B. L. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6640-5,2000.
- [40] Davies, S. J., P. Golby, et al. Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. *Journal of Bacteriology* 181: 5624-35, 1999.
- [41] Davidson, E. H., J. P. Rast, et al. A genomic regulatory network for development. *Science* 295: 1669-78.,2002.
- [42] Davidson, E. H., J. P. Rast, et al. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology* 246: 162-90.,2002.
- [43] De Reuse, H., Kolb, A. and Danchin, A. Positive regulation of the expression of the *Escherichia coli pts* operon: identification of the regulatory regions. *Journal of Molecular Biology* 226:623-635, 1992.
- [44] DeRisi, J., L. Penland, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics* 14: 457-460,1996.
- [45] DeRisi, J. L., V. R. Iyer, et al. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686,1997.
- [46] Edwards, J. and B. Palsson. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry* 274: 17410-17416,1999.
- [47] Edwards, J. S., R. Ramakrishna, et al. Metabolic Flux Balance Analysis. *Metabolic Engineering*. S. Y. Lee and E. T. Papoutsakis, Marcel Dekker: 13-57, 1999.
- [48] Edwards, J. S. and B. O. Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences* 97: 5528-5533,2000.
- [49] Edwards, J. S., R. U. Ibarra, et al. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology* 19: 125-130,2001.
- [50] Edwards, J. S., R. Ramakrishna, et al. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnology and Bioengineering* 77: 27-36.,2002.

- [51] Eisenberg, D., E. M. Marcotte, et al. Protein function in the post-genomic era. *Nature* 405: 823-6,2000.
- [52] Fell, D. *Understanding the Control of Metabolism*. London, Portland Press, 1996.
- [53] Fell, D. A. Increasing the flux in metabolic pathways: A metabolic control analysis perspective. *Biotechnology and Bioengineering* 58: 121-124,1998.
- [54] Fell, D. A. and A. Wagner. Structural properties of metabolic networks: implications for evolution and modelling of metabolism. *Animating the cellular map*. J.-H. S. Hofmeyr, J. M. Rohwer and J. L. Snoep. Stellenbosch, Stellenbosch University Press: 79-85, 2000.
- [55] Ferenci, T. Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *FEMS Microbiology Reviews* 18:301-317, 1996.
- [56] Farnyough, P. Quantification of mRNA levels using northern blotting. *Methods in Molecular Biology* 169: 53-63,2001.
- [57] Fotheringham, I. Engineering biosynthetic pathways: new routes to chiral amino acids. *Current Opinion in Chemical Biology* 4: 120-4,2000.
- [58] Fraenkel, D. G. and B. L. Horecker. Fructose-1, 6-diphosphatase and acid hexose phosphatase of *Escherichia coli*. *Journal of Bacteriology* 90: 837-42,1965.
- [59] Fraenkel, D. G. Glycolysis. *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, DC., ASM Press. 1: 189-198, 1996.
- [60] Gaasterland, T. and E. Selkov. Reconstruction of metabolic networks using incomplete information. *ISMB* 3: 127-35,1995.
- [61] Geanacopoulos, M., and Adhya, S. Functional Characterization of roles of GalR and GalS as regulators of the gal Regulon. *Journal of Bacteriology* 179(1):228-234, 1997.
- [62] Glasner, J. D., P. Liss, et al. ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Research* 31: 147-51,2003.
- [63] Golby, P., D. J. Kelly, et al. Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*. *Journal of Bacteriology* 180: 6586-96,1998.
- [64] Golby, P., S. Davies, et al. Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*. *Journal of Bacteriology* 181: 1238-48,1999.

- [65] Gombert, A. K. and J. Nielsen. Mathematical modelling of metabolism. *Current Opinion in Biotechnology* 11: 180-6,2000.
- [66] Gruer, M. J., A. J. Bradbury, et al. Construction and properties of aconitase mutants of *Escherichia coli*. *Microbiology* 143 (Pt 6): 1837-46,1997.
- [67] Guardia, M. J., A. Gambhir, et al. Cybernetic modeling and regulation of metabolic pathways in multiple steady states of hybridoma cells. *Biotechnology Progress* 16: 847-853,2000.
- [68] Gui, L., A. Sunnarborg, et al. Regulated expression of a repressor protein: FadR activates *iclR*. *Journal of Bacteriology* 178: 4704-9, 1996.
- [69] Gutierrez-Rios, R. M., D. A. Rosenblueth, et al. Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. *Genome Research* 13: 2435-43,2003.
- [70] Hatzimanikatis, V., C. Floudas, et al. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AICHE Journal* 1996: 1277-1292,1996.
- [71] Heinrich, R. and S. Schuster. *The regulation of cellular systems*. New York, Chapman & Hall, 1996.
- [72] Henderson, P. J. and M. C. Maiden. Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 326: 391-410, 1990.
- [73] Henriksen, C. M., L. H. Christensen, et al. Growth energetics and metabolism fluxes in continuous cultures of *Penicillium chrysogenum*. *Journal of Biotechnology* 45: 149-164,1996.
- [74] Herrgard, M. J., M. W. Covert, et al. Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Research* 13: 2423-34,2003.
- [75] Herrgard, M. J., M. W. Covert, et al. Reconstruction of microbial transcriptional regulatory networks. *Current Opinion in Biotechnology*, in press.
- [76] Hoffmann, A., A. Levchenko, et al. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* 298: 1241-5,2002.
- [77] Hutchison, C. A., S. N. Peterson, et al. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286: 2165-9,1999.
- [78] Ibarra, R. U., J. S. Edwards, et al. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 420: 186-9,2002.

- [79] Ideker, T., T. Galitski, et al. A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics* 2: 343-72,2001.
- [80] Ideker, T., V. Thorsson, et al. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* 292: 929-934,2001.
- [81] Irani, M. H. and P. K. Maitra. Properties of *Escherichia coli* mutants deficient in enzymes of glycolysis. *Journal of Bacteriology* 132: 398-410,1977.
- [82] Ishizuka, H., et al. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. *EMBO Journal* 13(13):3077-3082, 1994.
- [83] Iuchi, S. and Lin, E.C.C. *arcA(dye)*, a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proceedings of the National Academy of Sciences of the United States of America* 85:1888-1892, 1988.
- [84] Iuchi, S., Cole, S.T., and Lin, E.C.C. Multiple regulatory elements for the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-dehydrogenase in *Escherichia coli*: further characterization of respiratory control. *Journal of Bacteriology* 172(1):179-84, 1990.
- [85] Jeong, H., B. Tombor, et al. The large-scale organization of metabolic networks. *Nature* 407: 651-4,2000.
- [86] Jorgensen, H., J. Nielsen, et al. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnology and Bioengineering* 46: 117-131,1995.
- [87] Joyce, C. Quantitative RT-PCR. A review of current methodologies. *Methods in Molecular Biology* 193: 83-92,2002.
- [88] Kaiser, D. Building a multicellular organism. *Annual Reviews in Genetics* 35: 103-23,2001.
- [89] Kanehisa, M., S. Goto, et al. The KEGG databases at GenomeNet. *Nucleic Acids Research* 30: 42-6.,2002.
- [90] Kang, H. Y., S. Song, et al. Priority of pentose utilization at the level of transcription: arabinose, xylose, and ribose operons. *Mol Cells* 8: 318-23, 1998.
- [91] Karp, P. D., M. Riley, et al. The EcoCyc and MetaCyc databases. *Nucleic Acids Research* 28: 56-9,2000.
- [92] Karp, P. D., M. Riley, et al. The EcoCyc Database. *Nucleic Acids Research* 30: 56-8.,2002.

- [93] Kather, B., K. Stingl, et al. Another unusual type of citric acid cycle enzyme in *Helicobacter pylori*: the malate:quinone oxidoreductase. *Journal of Bacteriology* 182: 3204-9,2000.
- [94] Kauffman, S. A. *The Origins of Order*. New York, Oxford University Press, 1993.
- [95] Kaufman, M., J. Urbain, et al. Towards a logical analysis of the immune response. *Journal of Theoretical Biology* 114: 527-61,1985.
- [96] Kitano, H. Systems biology: a brief overview. *Science* 295: 1662-4,2002.
- [97] Kolisnychenko, V., G. Plunkett, 3rd, et al. Engineering a reduced *Escherichia coli* genome. *Genome Research* 12: 640-7,2002.
- [98] Kompala, D. S., D. Ramkrishna, et al. Investigation of Bacterial Growth on Mixed Substrates. Experimental Evaluation of Cybernetic Models. *Biotechnology and Bioengineering* 28: 1044-1056,1986.
- [99] Kremling, A., K. Bettenbrock, et al. The organization of metabolic reaction networks. III. Application for diauxic growth on glucose and lactose. *Metabolic Engineering* 3: 362-79,2001.
- [100] Kumari, S., R. Tishel, et al. Cloning, characterization, and functional expression of *acs*, the gene which encodes acetyl coenzyme A synthetase in *Escherichia coli*. *Journal of Bacteriology* 177: 2878-86,1995.
- [101] Kumari, S., C. M. Beatty, et al. Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. *Journal of Bacteriology* 182: 4173-9, 2000.
- [102] Kuo, M. H. and C. D. Allis. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19: 425-33,1999.
- [103] Langley, D. and J. R. Guest. Biochemical genetics of the alpha-keto acid dehydrogenase complexes of *Escherichia coli* K12: isolation and biochemical properties of deletion mutants. *Journal of General Microbiology* 99: 263-76,1977.
- [104] Lee, B., J. Yen, et al. Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnology and Bioengineering* 62: 722-729.,1999.
- [105] Lendenmann, U. and T. Egli. Is *Escherichia coli* growing in glucose-limited chemostat culture able to utilize other sugars without lag? *Microbiology* 141: 71-8,1995.

- [106] Li, J., and Stewart, V. Localization of upstream sequence elements required for nitrate and anaerobic induction of *fdn* (Formate Dehydrogenase-N) operon expression in *Escherichia coli* K-12. *Journal of Bacteriology* 174(15):4935-4942, 1992.
- [107] Li, C. and W. H. Wong. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* 98: 31-6,2001.
- [108] Liao, J. C., S. Y. Hou, et al. Pathway analysis, engineering and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering* 52: 129-140,1996.
- [109] Liao, J. C. and M. K. Oh. Toward predicting metabolic fluxes in metabolically engineered strains. *Metabolic Engineering* 1: 214-23,1999.
- [110] Lin, E. C. C. Dissimilatory Pathways for Sugars, Polyols, and Carboxylates. *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, DC., ASM Press. 1: 307-342.
- [111] Lopilato, J. E., J. L. Garwin, et al. D-ribose metabolism in *Escherichia coli* K-12: genetics, regulation, and transport. *Journal of Bacteriology* 158: 665-73, 1984.
- [112] Lynch, A. S. and E. C. C. Lin. Responses to molecular oxygen. *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, D.C., ASM Press. 1: 1526-1538, 1996.
- [113] Lynd, L. R., C. E. Wyman, et al. Biocommodity Engineering. *Biotechnology Progress* 15: 777-793,1999.
- [114] Madan Babu, M. and S. A. Teichmann. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Research* 31: 1234-44,2003.
- [115] Mahadevan, R., J. S. Edwards, et al. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal* 83: 1331-40,2002.
- [116] Maki, Y., D. Tominaga, et al. Development of a system for the inference of large scale genetic networks. *Pacific Symposium on Biocomputing 2001*: 446-458.
- [117] Martinez-Antonio, A. and J. Collado-Vides. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinions in Microbiology* 6: 482-489,2003.
- [118] Mat-Jan, F., K. Y. Alam, et al. Mutants of *Escherichia coli* deficient in the fermentative lactate dehydrogenase. *Journal of Bacteriology* 171: 342-8,1989.

- [119] McAdams, H. H. and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 94: 814-819,1997.
- [120] McAdams, H. H. and A. Arkin. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure* 27: 199-224,1998.
- [121] McAdams, H. H. and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics* 15: 65-9,1999.
- [122] McAdams, H. H. and L. Shapiro. Circuit simulation of genetic networks. *Science* 269: 651-656,1995.
- [123] Membrillo-Hernandez, J. and E. C. Lin. Regulation of expression of the *adhE* gene, encoding ethanol oxidoreductase in *Escherichia coli*: transcription from a downstream promoter and regulation by *fnr* and RpoS. *Journal of Bacteriology* 181: 7571-9, 1999.
- [124] Mikulskis, A., A. Aristarkhov, et al. Regulation of expression of the ethanol dehydrogenase gene (*adhE*) in *Escherichia coli* by catabolite repressor activator protein Cra. *Journal of Bacteriology* 179: 7129-34,1997.
- [125] Millenium evening at the White House: informatics meets genomics (1999). <http://www.rand.org/scitech/stpi/ourfuture/Rosetta/millennium.html>, 6 Nov. 2003.
- [126] Moreno-Hagelsieb, G. and J. Collado-Vides. A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* 18 Suppl 1: S329-36,2002.
- [127] Nadis, S. Spoof Nobels take researchers for a ride. *Nature* 425: 550,2003.
- [128] Negre, D., et al. FruR-mediated transcriptional activation at the *ppsA* promoter of *Escherichia coli*. *Journal of Molecular Biology* 276(2):355-65, 1998.
- [129] Neidhardt, F. C., J. L. Ingraham, et al. *Physiology of the bacterial cell*. Sunderland, MA, Sinauer Associates, Inc., 1990.
- [130] Neidhardt, F. C., Ed. *Escherichia coli and Salmonella: cellular and molecular biology*. Washington, D.C., ASM Press, 1996.
- [131] Ogata, H., S. Goto, et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 27: 29-34,1999.
- [132] Oh, M.-K. and J. C. Liao. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnology Progress* 16: 278-286,2000.

- [133] Ostergaard, S., L. Olsson, et al. Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nature Biotechnology* 18: 1283-1286,2000.
- [134] Overbeek, R., N. Larsen, et al. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Research* 28: 123-125,2000.
- [135] Paigen K., Williams B. Catabolite repression and other control mechanisms in carbohydrate utilization. *Adv Microbiol Physiol* 4:251-324, 1970.
- [136] Palsson, B. O. The challenges of *in silico* biology. *Nature Biotechnology* 18: 1147-1150,2000.
- [137] Papin, J. A., N. D. Price, et al. The Genome-Scale Metabolic Extreme Pathway Structure in *Haemophilus influenzae* Shows Significant Network Redundancy. *Journal of Theoretical Biology* 215: 67-82,2002.
- [138] Park, S.-J., et al. Regulation of the Citrate Synthase (*gltA*) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the *arcA* gene product. *Journal of Bacteriology* 176(16):5086-5092, 1994.
- [139] Park, S.-J., and Gunsalus, R.P. Oxygen, Iron, Carbon, and Superoxide control of the Fumarase *fumA* and *fumC* genes of *Escherichia coli*: Role of the *arcA*, *fnr*, and *soxR* gene products. *Journal of Bacteriology* 177(21):6255-6262, 1995.
- [140] Park, S. J., G. Chao, et al. Aerobic regulation of the *sucABCD* genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream *sdhCDAB* promoter. *Journal of Bacteriology* 179: 4138-42,1997.
- [141] Patte, J.-C. (1996). Biosynthesis of Threonine and Lysine. *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, D.C., ASM Press. 1: 528-541.
- [142] Pearson, W. R., T. Wood, et al. Comparison of DNA sequences with protein sequences. *Genomics* 46: 24-36,1997.
- [143] Peterson, J. D., L. A. Umayam, et al. The comprehensive microbial resource. *Nucleic Acids Research* 29: 123-125,2001.
- [144] Pieper, D. H. and W. Reineke. Engineering bacteria for bioremediation. *Current Opinion in Biotechnology* 11: 262-70,2000.
- [145] Pope, A. Columbia World of Quotations, 1996. <http://www.bartleby.com/66/28/44928.html>, 14 Nov. 2003.

- [146] Pramanik, J., P. L. Trelstad, et al. Development and validation of a flux-based stoichiometric model for enhanced biological phosphorus removal metabolism. *Water Research* 33: 462-476.,1999.
- [147] Price, N. D., J. A. Papin, et al. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends in Biotechnology* 21: 162-9,2003.
- [148] Prost, J. F., D. Negre, et al. Cra-dependent transcriptional activation of the *icd* gene of *Escherichia coli*. *Journal of Bacteriology* 181: 893-8,1999.
- [149] Qu, Z., J. N. Weiss, et al. Regulation of the mammalian cell cycle: a model of the G1-to-S transition. *American Journal of Physiology and Cell Physiology* 284: C349-64,2003.
- [150] Quail, M.A., Haydon, D.J., and Guest, J.R. The *pdhR-aceEF-lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex. *Molecular Microbiology* 12(1):95-104, 1994.
- [151] Quail, M. A. and J. R. Guest. Purification, characterization and mode of action of PdhR, the transcriptional repressor of the *pdhR-aceEF-lpd* operon of *Escherichia coli*. *Molecular Microbiology* 15: 519-29,1995.
- [152] Rao, C. V. and A. P. Arkin. Control motifs for intracellular regulatory networks. *Annual Review of Biomedical Engineering* 3: 391-419,2001.
- [153] Reed, J. L. and B. O. Palsson. Thirteen Years of Building Constraint-Based *In Silico* Models of *Escherichia coli*. *Journal of Bacteriology* 185: 2692-9,2003.
- [154] Reed, J. L., T. D. Vo, et al. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology* 4: R54.1-R54.12,2003.
- [155] Reich, J. G. and E. E. Sel'kov. *Energy Metabolism of the Cell*. New York, Academic Press, 1981.
- [156] Ren, B., F. Robert, et al. Genome-wide location and function of DNA binding proteins. *Science* 290: 2306-9.,2000.
- [157] Rivett, A. J. Regulation of intracellular protein turnover: covalent modification as a mechanism of marking proteins for degradation. *Current Topics in Cellular Regulation* 28: 291-337,1986.
- [158] Ronen, M., R. Rosenberg, et al. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America* 99: 10555-60.,2002.
- [159] Ryu, S., and Garges, S. Promoter switch in the *Escherichia coli pts* operon. *Journal of Biological Chemistry* 269(7):4767-4772, 1994.

- [160] Saier, M. H. J., T. M. Ramseier, et al. (1996). Regulation of carbon utilization. *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt. Washington, D.C., ASM Press. 1: 1325-1343.
- [161] Saier, M. H., Jr. Genome sequencing and informatics: new tools for biochemical discoveries. *Plant Physiology* 117: 1129-33,1998.
- [162] Salgado, H., A. Santos-Zavaleta, et al. RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Research* 29: 72-4,2001.
- [163] Savageau, M. A. Design of molecular control mechanisms and the demand for gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5647-51,1977.
- [164] Savageau, M. A. Are there rules governing patterns of gene regulation? *Theoretical Biology – Epigenetic and Evolutionary Order*. B. C. Goodwin and P. T. Saunders. Edinburgh, Edinburgh University Press: 42-66, 1989.
- [165] Savageau, M. A. Rules for the evolution of gene circuitry. *Pacific Symposium on Biocomputing* 7: 54-65,1998.
- [166] Savageau, M. A. Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *Biosystems* 47: 9-36,1998.
- [167] Savageau, M. A. Rules for the evolution of gene circuitry. *Pacific Symposium on Biocomputing* 7: 54-65,1998.
- [168] Sawers, G. Specific transcriptional requirements for positive regulation of the anaerobically inducible *pfl* operon by ArcA and FNR. *Molecular Microbiology* 10(4):737-747, 1993.
- [169] Schilling, C. H. and B. O. Palsson. The underlying pathway structure of biochemical reaction networks. *Proceedings of the National Academy of Sciences of the United States of America* 95: 4193-8,1998.
- [170] Schilling, C. H., J. S. Edwards, et al. Towards metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnology Progress* 15: 288-95,1999.
- [171] Schilling, C. H., S. Schuster, et al. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress* 15: 296-303,1999.
- [172] Schilling, C. H., J. S. Edwards, et al. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering* 71: 286-306,2000.

- [173] Schilling, C. H. and B. O. Palsson. Assessment of the Metabolic Capabilities of *Haemophilus influenzae* Rd through a Genome-scale Pathway Analysis. *Journal of Theoretical Biology* 203: 249-283,2000.
- [174] Schilling, C. H., M. W. Covert, et al. Genome-scale metabolic model of *Helicobacter pylori* 26695. *Journal of Bacteriology* 184: 4582-4593,2002.
- [175] Shuler, M. L. and M. M. Domach (1983). Mathematical Models of the Growth of Individual Cells. Foundations of Biochemical Engineering. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos. Washington, DC, American Chemical Society: 101.
- [176] Schuster, S., T. Dandekar, et al. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* 17: 53-60,1999.
- [177] Selkov, E., S. Basmanova, et al. The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucleic Acids Research* 24: 26-8,1996.
- [178] Selkov, E., Jr., Y. Greshkin, et al. MPW: the Metabolic Pathways Database. *Nucleic Acids Research* 26: 43-5,1998.
- [179] Selkov, E., R. Overbeek, et al. Functional analysis of gapped microbial genomes: amino acid metabolism of *Thiobacillus ferrooxidans*. *Proceedings of the National Academy of Sciences of the United States of America* 97: 3509-14,2000.
- [180] Serres MH, Gopal S, et al. A functional update of the *Escherichia coli* K-12 genome. *Genome Biology* 2: research0035.1-research0035.7,2001.
- [181] Shen-Orr, S. S., R. Milo, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31: 64-8,2002.
- [182] Smolen, P., D. A. Baxter, et al. Modeling transcriptional control in gene networks – methods, recent results, and future directions. *Bulletin of Mathematical Biology* 62: 247-292,2000.
- [183] Somogyi, R. and C. A. Sniegowski. Modeling the complexity of genetic networks: understanding multigenic and pleitropic regulation. *Complexity* 1: 45-63,1996.
- [184] Song, S. and C. Park. Utilization of D-ribose through D-xylose transporter. *FEMS Microbiology Letters* 163: 255-61,1998.
- [185] Sorensen, K. I. and B. Hove-Jensen. Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression. *Journal of Bacteriology* 178: 1003-1011,1996.

- [186] Sprenger, G. A. Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. *Arch Microbiol* 164: 324-30,1995.
- [187] Stephanopoulos, G., A. Aristodou, et al. *Metabolic Engineering*, Academic Press, 1998.
- [188] Sullivan, R. Sir Edmund Hilary: a visit with the world's greatest living adventurer. *Time: The Greatest Adventures of All Time*. <http://www.time.com/time/2003/adventures/interview.html>, 7 Nov. 2003.
- [189] Tanaka, Y., et al. Negative regulation of the *pts* by Mlc: mechanism underlying glucose induction in *Escherichia coli*. *Genes to Cells* 4:391-399, 1999.
- [190] Thieffry, D. and R. Thomas. Dynamical behaviour of biological regulatory networks-II. Immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology* 57: 277-97,1995.
- [191] Thieffry, D., A. M. Huerta, et al. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* 20: 433-40,1998.
- [192] Thieffry, D., H. Salgado, et al. Prediction of transcriptional regulatory sites in the complete genome sequence of *Escherichia coli* K-12. *Bioinformatics* 14: 391-400,1998.
- [193] Thomas, R. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology* 42: 563-85,1973.
- [194] Thomas, R. Regulatory networks seen as asynchronous automata: a logical description. *Journal of Theoretical Biology* 153: 1-23,1991.
- [195] Tran, Q. H., J. Bongaerts, et al. Requirement for the proton-pumping NADH dehydrogenase I of *Escherichia coli* in respiration of NADH to fumarate and its bioenergetic implications. *European Journal of Biochemistry* 244: 155-60,1997.
- [196] Tseng, C.P., Albrecht, J., and Gunsalus, R.P. Effect of microaerophilic cell growth conditions on expression of the aerobic (*cyoABCDE* and *cydAB*) and anaerobic (*narGHIJ*, *frdABCD*, and *dmsABC*) respiratory pathway genes in *Escherichia coli*. *Journal of Bacteriology* 178(4):1094-1098, 1996.
- [197] Tseng, C. P. Regulation of fumarase (*fumB*) gene expression in *Escherichia coli* in response to oxygen, iron and heme availability: role of the *arcA*, *fnr*, and *hemA* gene products. *FEMS Microbiology Letters* 157: 67-72,1997.
- [198] Varma, A. and B. O. Palsson. Metabolic Flux Balancing: Basic concepts, Scientific and Practical Use. *Bio/Technology* 12: 994-998,1994.

- [199] Varma, A. and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology* 60: 3724-3731,1994.
- [200] Varner, J. and D. Ramkrishna. Mathematical models of metabolic pathways. *Current Opinion in Biotechnology* 10: 146-50,1999.
- [201] Varner, J. D. Large-scale prediction of phenotype: concept. *Biotechnology and Bioengineering* 69: 664-78,2000.
- [202] Vaseghi, S., A. Baumeister, et al. In vivo Dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metabolic Engineering* 1: 128-140,1999.
- [203] Wackwitz, B., et al. Growth phase-dependent regulation of *nuoA-N* expression in *Escherichia coli* K-12 by the Fis protein: upstream binding sites and bioenergetic significance. *Molecular General Genetics* 262:876-883, 1999.
- [204] Wagner, R. *Transcription Regulation in Prokaryotes*. Oxford, Oxford University Press, 2000.
- [205] Wagner, A. and D. A. Fell. The small world inside large metabolic networks. *Proceeding of the Royal Society of London B: Biological Sciences* 268: 1803-10,2001.
- [206] Weissenborn, D.L., Wittekindt, N. and Larson, T.J. Structure and regulation of the *glpFK* operon encoding Glycerol Diffusion Facilitator and Glycerol Kinase of *Escherichia coli* K-12. *Journal of Biological Chemistry* 267(9):6122-6131, 1992.
- [207] Wiechert, W. and A. A. de Graaf. In vivo stationary flux analysis by ¹³C labeling experiments. *Advances in Biochemical Engineering/Biotechnology* 54: 109-54,1996.
- [208] Wiechert, W. ¹³C metabolic flux analysis. *Metabolic Engineering* 3: 195-206,2001.
- [209] Wildermuth, M. C. Metabolic control analysis: biological applications and insights. *Genome Biology* 1: REVIEWS1031,2000.
- [210] Wiley, H. S., S. Y. Shvartsman, et al. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends in Cell Biology* 13: 43-50,2003.
- [211] Wong, P., S. Gladney, et al. Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnology Progress* 13: 132-43,1997.

- [212] Yang, B. and Larson, T.J. Action at a distance for negative control of transcription of the *glpD* gene encoding sn-glycerol-3-dehydrogenase in *Escherichia coli* K-12. *Journal of Bacteriology* 178(24):7090-7098, 1996.
- [213] Yang, B. and Larson, T.J. Multiple promoters are responsible for transcription of the *glpEGR* operon of *Escherichia coli* K-12. *Biochimica et Biophysica Acta* 1396:114-126, 1998.
- [214] Zientz, E., S. Six, et al. Identification of a third secondary carrier (DcuC) for anaerobic C4-dicarboxylate transport in *Escherichia coli*: roles of the three Dcu carriers in uptake and exchange. *Journal of Bacteriology* 178: 7241-7, 1996.
- [215] Zubay, G. In vitro synthesis of protein in microbial systems. *Annual Review of Genetics* 7: 267-87, 1973.