Analysis and Design of Metabolic Reaction Networks

Thesis by

Vassily Hatzimanikatis

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1997

(Submitted 24 April, 1997)

Acknowledgements

Trying to write down these acknowledgements, I realize that there is a big number of people that have contributed to the completion of this work. People that shaped my scientific profile and my personal *character*.

First of all, I wish to acknowledge Jay Bailey for being more than an advisor to me. He taught me how to identify and address the important problems and his continuous challenging and criticism of my work, along with the freedom he provided to me, although they have been dangerously frustrating, have been the source of inspiration and the ground for my scientific growth.

On the other hand, Chris Floudas taught me how to postulate and attack the problems. Working with Chris was not always easy. But it was educational, challenging, and, in retrospective, joyful. Chris, thanx for everything.

I was very lucky to have Mary Lidstrom teaching me microbial physiology and Steve Wiggins teaching me applied mathematics. They have been two of my excellent teachers at Caltech, along with John Brady, with his i's and j's, and George Gavalas, with his entropies, enthalpies, and fugacities.

Speaking about teachers, I cannot help remembering and acknowledging my undergratuate advisors and teachers: Stavros Pavlou, George Papatheodorou, Alkis Payatakes, Costas Vayenas, and Gerry Lyberatos. I am most thankful to them for my solid educational background and their support and confidence in me.

Alex Seressiotis has been of precious help in the scientific and technical aspects of this work, and a great activator when frustration was saturating. His regulatory action deserves special acknowlegement.

Special, sincere thanks to Christos Frouzakis for his patience listening to my problems, for thinking about them, for making key suggestions, and for being a (Greek) friend of rare quality.

Speaking about friends. Looking back to these past five years two names flash:

Zhenya and Kelvin. Zhenya, aka Yevgeny Yurkovetsky, has been my first friend at Caltech, and one of the best, since then. If I survived the first year at Caltech, it is because Zhenya was around. Stoic, with his cynical humor (he prefers to call it "Odessan"), going down the list of the violins when depression was overtaking, taking care of me when homeworks were overloading, meditating with me over the uncountable infinity of foliations, sharing together the Fellinian emotions. Thanx a lot Zhenya.

Then it was Kelvin. A great friend in Caltech and in Zurich, and an excellent scientific companion. He was my coffee-breaks relief and the Boccalino-evenings relaxation. Sharing our frustrations and working together on the development of the Graphic Arts department. Being more than a friend when my father was dying and teaching me that a green-card-holder-to-be is neither a Democrat nor a Republican but an American. My days—and nights—in Zurich would have been unbearable without Kelvin. All the best Kelvin, because we will never be done.

Nikos Bekiaris and Chrisa Economou have been my first and best Greek friends at Caltech. Their help with starting my life in the States, the Greek fests we organized together, and the nice moments we shared will be gratefully remembered.

Marcel (*fratello mio*), Wil (*the sleepless*), Marjan, Feike, Maarten and the Dutch mafia at IBT helped me more than they can even imagine. Especially during the first two years, when this work was still an impossible possibility.

And when the results started coming, Uwe and Wolfgang R. were there to question them and drive me crazy. Because of them I have learned how I, a theoretical engineer, can fruitfully interact with applied scientists - as they would like to call themselves, and as I would like to call them. The real scientist though was sharing the same office with me: Wolfgang K. When my programs were not running and my e-mailbox was empty, discussing with him was pleasant, educational enrichment.

Pablo, Phil, Wilfred, and Dana have been invaluable friends and valuable groupmates—"brother" chemical engineers in the midst of hard-core biologists.

Petra, Sigi, Becky, Kyle, Julie, and Henriette were showing their lovely faces around thus making a significant difference.

Bernard Witholt, Erika and Helena were keeping the institute up and running and they have been always of great help and provided a congenial atmosphere that made life at IBT comfortable.

Adrianna and Conny provided me the support when I needed it and they will never be forgotten. Bigi, Ivo, Claire, Miquel, Lisa, Jan, Michele, Marc, Nicolas, Andrew, Roberto, Markus, Lucas, Christian, Peter: thanx a lot!

Nobody else from IBT will be forgotten, even if they have not been named here. Everybody, in their way, made it easier or harder for me to successfully work on this thesis.

And, of course, I have not forgotten Ellen but I want to thank her ending my acknowlegements to IBT people remembering the way we have been ending our days beer-washing our lab-problems and personal obsessions at the Bonnie Prince and at the Odeon, till four in the morning. Her early departure from the Institute (*leaving engineering for science*) meant loneliness to me.

In Zurich though I had my *familia*: Giovanna (*cara sorellina mia!*) and Annalaura, who took care of me, provided me with the mostly needed family warmth, believed in me (thanx a lot for the Verona nights, the Pisa days, and the Italian or...ms!), and made Zurich my hometown. Giuvi and Nuccia, *i miei piú sinceri ringraziamenti!* And of course it was Secondo, member of the family, with our long after hours wine sessions, the spring-fests at his *atelier*, and his explosive creations on canvas.

Sophia and Thomas, thank you very much for your friendship, for Angelika and Jannis, and for the *Gemütlichkeit.*

However, a significant part of this work was carried out at Princeton and a bunch of wonderful people there ought to be acknowleged.

I must thank Ioannis Androulakis, my office- and housemate at Princeton for his help, friendship, patience with me, and the great, unforgettable time we had together. I would also like to thank Costas Maranas for being of significant help with my work. Ioannis and Costas have taught me the basic principle in programming: "if it works ..."

Cordial thanks to Claire and Omar for their sincere friendship and the pleasant

time we have had together.

Ioannis, Claire, Omar, Carl, Costas, and Janna made me feel nice and comfortable in Princeton creating the environment I needed to be efficient and productive. That's how they have significantly contributed to this work.

Jette, Lydia, Marjan, and I were close while I was finishing writing this thesis. Their company made the hardest part of this work a lot easier. Thank you, girls!

I will close these long acknowledgements thanking Chianti Classico, Gordon's, Parisienne, Red Door Café, Boccalino, Menza, Gnädinger, Sprüngli, Odeon, Rosengarten, and Eulenspiegel for the inspiration and for being there when everything was falling apart.

Pasadena, April 1996

PS: Financial support was generously provived by the Swiss Priority Program in Biotechnology.

Abstract

Different mathematical methods can be used for the analysis of metabolic systems and the subsequent engineering of metabolism. The available experimental information dictates the most appropriate mathematical framework for such studies. Several approaches for metabolic system analysis and design are developed in this thesis. It is shown that for several model systems, a (log)linear model shows excellent agreement with the corresponding nonlinear model. The (log)linear model which is developed here describes the dynamical and steady-state responses of the logarithmic deviations of the metabolic variables and functions with respect to a change of the metabolic parameters around a corresponding reference state. The parameters of the (log)linear model are quantities easily estimated from experimental and theoretical tools developed within metabolic control analysis (MCA). A significant advantage of the newly developed (log)linear model is the linearity with respect to logarithms which makes computational analysis easier as compared to the correponding nonlinear model. A second approach introduces a novel, production-oriented optimization framework. Maximizing the performance of a metabolic reaction pathway is treated as a mixed-integer linear programming (MILP) formulation when a (log)linear model of the pathway is available and as a mixed-integer nonlinear programming (MINLP) formulation when a nonlinear model is available. The objective of the MILP and MINLP formulation is to identify changes in regulatory structure and strength, and in cellular content of pertinent enzymes, which should be implemented in order to optimize a particular metabolic process. A *regulatory superstructure* is proposed that contains all alternative regulatory structures that can be considered for a given pathway. The proposed approach is followed in order to find the optimal regulatory structure for maximization of phenylalanine selectivity in the microbial aromatic amino acid synthesis pathway. The solution suggests that, from the 8 feedback inhibitory loops in the original regulatory structure of this pathway, inactivation of at least three loops

and overexpression of three enzymes will increase phenylalanine selectivity by 42%. Moreover, novel regulatory structures with only two loops, none of which exists in the original pathway, could result in a selectivity of up to 95%.

Contents

Acknowledgements iii

Abstract vii

1 Prologue 1

1.1 Dealing with Metabolic Complexity 1

1.2 Mathematical and Computational Methods for Metabolic Analysis and Design 3

2 Analysis of Metabolic Reaction Rates 7

2.1 Introduction 8

2.2 From a Net Metabolic Reaction to a Network of Metabolic Reactions 9

2.2.1 Aerobic Growth Without Product Formation: A Net Reaction Rate 10

2.2.2 Aerobic Growth Without Product Formation: A Metabolic Reaction Network 13

2.2.3 Determining the Underdetermined: Assumptions and Solutions 18

2.3 Example 27

2.3.1 ATP Mass Balance 28

2.3.2 NADPH Mass Balance 29

2.3.3 Reducing Equivalents Mass Balance 31

2.4 Conclusions 33

2.5 References 36

2.6 Figures 40

3 Effects of Spatiotemporal Variations on Metabolic Control: Approximate Analysis Using (Log)Linear Kinetic Models 47

3.1 Introduction 48
3.2 Mathematical Description of Metabolic Reaction Networks 49
3.2.1 Development of a (Log)Linear Kinetic Model 49
3.2.2 Analytical Solution of the (Log)Linear Model 54
3.3 Accuracy of the (Log)Linear Metabolic Model 55
3.3.1 Linear Pathway 56
3.3.2 Branched Pathway 56
3.3.3 Glycolytic Pathway 57
3.4 Averaged Metabolic Functions and Control Coefficients for Periodic Parameter Variations 58
3.4.1 Mathematical Framework 60
3.4.2 Quasi-Steady-State Approximation: a Limiting Case 63
3.4.3 Example 65
3.5 Concluding Remarks 67
3.6 References 69
3.7 Appendices 72
3.8 Figures 78

4 Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization **88**
4.1 Introduction 89
4.2 Problem Statement 93
4.3 Mathematical Description of Metabolic Reaction Networks 94
4.4 Analysis and Synthesis Problems 96
4.5 Mathematical Formulation 98
4.5.1 *The Objective Function* 100
4.5.2 *Constraints* 101
4.6 Computational Studies 105
4.6.1 *Problem 1* 107
4.6.2 *Problem 2* 110

4.6.3 *Problem 3* 112
4.6.4 *Problem 4* 113
4.7 Discussion 115
4.8 Conclusions 117
4.9 References 118
4.10 Appendices 121
4.11 Tables 128
4.12 Figures 134

5 Optimization of Regulatory Architectures in Metabolic Reaction Networks 141

5.1 Introduction 142
5.2 Mathematical Framework 144
5.2.1 Mathematical Modeling of Biochemical Systems 144
5.2.2 Mixed-Integer Linear Optimization 147
5.2.3 Regulatory Superstructure 155
5.2.4 Consistency Checks 162
5.3 Example 167
5.4 Concluding Remarks 172
5.5 References 174
5.6 Appendix 178
5.7 Tables 182
5.8 Figures 188

6 Multiple Steady States in Metabolic Reaction Networks 195

6.1 Introduction 196
6.2 A Kinetic Model for a Glycolysis Prototype 197
6.3 The Mathematical Model 199
6.4 Steady-State Multiplicities 204
6.5 Discussion 210
6.6 References 212

6.7 Figures 214

7 Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Nonlinear Optimization 223

7.1 Introduction 224

7.2 Problem Statement and Formulation 224

7.2.1 Problem Statement 224

7.2.2 Mathematical Description of Metabolic Reaction Networks . . 225

7.2.3 Analysis Problem 226

7.2.4 Synthesis Problem 228

7.3 Mathematical Formulation 230

7.3.1 Reaction Rates and Metabolic Outputs 232

7.3.2 The Objective Function 235

7.3.3 Constraints 235

7.4 Algorithmic Procedure 241

7.4.1 Generalized Benders Decomposition, **GBD** 242

7.4.2 Initialization Schemes 246

7.5 Computational Study 248

7.6 Linear vs. Nonlinear Framework 250

7.7 References 251

7.8 Appendices 257

7.9 Figures 261

8 Epilogue 263

8.1 Future Prospects 264

List of Figures

2.1 The reaction network of the central carbon pathways for aerobic growth without product formation of *B. subtilis* considered in the analysis. The metabolites for which the mass balances were used are underlined. Shaded arrows indicate fluxes to biosynthesis. Double-underlined numbers indicate irreversible reaction steps. 41

2.2 Flux analysis results for various dilution rates without considering the CO_2 mass balance and under the assumption that $2v_1 + v_{13} = \phi\mu Y_{NADPH}$, where: dotted line ($\phi = 0.8$), dashed line ($\phi = 1$), and dashed-dotted line ($\phi = 1.2$). Solid line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production. 42

2.3 Flux analysis results for various dilution rates including the CO_2 mass balance and under the assumption that $2v_1 + v_{13} = \phi\mu Y_{NADPH}$, where: dotted line ($\phi = 0.8$), dashed line ($\phi = 1$), and dashed-dotted line ($\phi = 1.2$). Solid line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production. 43

2.4 Flux analysis results comparison for various dilution rates under the assumption that $2v_1 + v_{13} = \mu Y_{NADPH}$, including the CO_2 mass balance (solid line) and without considering the CO_2 mass balance (dashed line). Dotted line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production. 44

2.5 Flux analysis results comparison for various dilution rates without considering $FADH_2$ in the mass balance for the reducing equivalents and for "corrected" value for the oxygen uptake rate including the CO_2 mass balance (solid line) and without considering the CO_2 mass balance (dashed line). Dotted line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production. 45

2.6 The values of the ratio $(2v_1+v_{15})/(\mu Y_{NADPH})$ for various dilution rates, as it was estimated from the analysis without considering $FADH_2$ in the mass balance for the reducing equivalents and for "corrected" value for the oxygen uptake rate without (dashed line) and with (solid line) the constraint for the CO_2 mass balance. 46

3.1 Linear pathway. 79

3.2 Dynamic responses of the flux through the linear pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A. Responses to step changes of the input flux. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 1 - 0.2sin\left(\frac{t}{\pi}\right)$. 80

3.3 Branched pathay. Solid arrows indicate reaction steps. Dashed arrows indicate activation (plus sign) or inhibition (minus sign). 81

3.4 Dynamic responses of the ratio of fluxes through the branched pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A. Responses to step changes of the input flux. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 3(1 + 0.2sin\,(\pi t))$. . . . 82

3.5 Anaerobic fermentation pathway of the yeast *Saccharomyces cerevisiae* under nitrogen starvation, with glucose as the sole carbon source. Enzyme/pathway steps: **in**, glucose uptake; **HK**, hexokinase; $\mathbf{K_1}$, equilibrium step; **PFK**, phosphofructokinase; **GAPD**, glyceraldehyde 3–phopshate dehydrogenase; $\mathbf{K_2}$, equilibrium step; **PYK**, pyruvate kinase; **GRO**, glycerol production; *POL*, polysaccharide production; **ATPase**, net ATP consumption; **AK**, adenylate kinase. Solid arrows indicate reaction steps, dashed arrows indicate activation and dotted arrows indicate inhibition. 83

3.6 Dynamic responses of the ethanol specific production rate of the yeast glycolytic pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A & C. Responses to step changes of the glucose uptake. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 27.3(1 + 0.2sin\,(2\pi t))$. 84

3.7 Pulsed periodic variation of parameter p_1. 85

3.8 The percentage difference of the average ethanol specific production rate, i.e. $\frac{\overline{J}_{EtOH} - J_{EtOH,ref}}{J_{EtOH,ref}}$ as a function of the period, T, of the pulsed variation of the glucose uptake rate. I: $\epsilon = 0.5$, $\delta = 0.1$ (solid line) and $\delta = 0.2$ (dashed line). II: $\delta = 0.2$, $\epsilon = 0.5$ (dashed line) and $\epsilon = 0.8$ (dashed–dotted line). 86

3.9 The control coefficients of the average ethanol specific production rate, $\overline{J}_{EtOH}$, with respect to: A. phosphofructokinase, B. pyruvate kinase, and C. ATPase, as functions of the period, T, of the pulsed variation of the glucose uptake rate. Solid lines: $\epsilon = 0.5$ and $\delta = 0.1$. Dashed lines: $\epsilon = 0.5$ and $\delta = 0.2$. Dashed–dotted lines: $\epsilon = 0.8$ and $\delta = 0.2$. 87

4.1 The aromatic amino acid synthesis pathway. Solid arrows indicate reactions and dashed arrows indicate feedback inhibition loops. Chemical species: G6P = glucose-6-phosphate; PEP = phosphoenolpyruvate; ATP = adenosine triphosphate; ADP = adenosine diphosphate; DAHP = 3-deoxy-D-arabino-heptulosonate-7-phosphate; CHR = chorismate; PHP = prephenate; PHE = phenylalanine; TYR = tyrosine; TRP = tryptophan. 135

4.2 The eight best solutions from problem 1. Solid arrows indicate reactions, dashed arrows indicate inhibitory loops, and thick solid arrows indicate enzyme overexpression for the respective reaction. In solutions b-h the reaction numbering has been omitted for clarity. 136

4.3 The four best solution from problem 2. Solid, dashed, and thick arrows as in Figure 2. In solutions b-d the reaction numbering has been omitted for clarity. 137

4.4 Twelve of the 37 best solutions from problem 3. Solid and thick arrows as in Figure 2. Dashed arrows indicate activation. In solutions b-l the reaction numbering has been omitted.(Figure continuous on the next page). 138

4.5 The six best solutions from problem 4 that feature activation and inhibition. Solid and thick arrows as in Figure 2. Dashed arrows indicate regulation, (+) indicates activation and (−) inhibition. 140

5.1 Linear pathway wih feedback inhibition. 189

5.2 Linear pathway with regulatory superstructure around it. Dashed lines denote inhibition and dashed-dotted lines denote activation. Three different levels of strength are considered for each regulatory loop. . 190

5.3 The first two optimal regulatory structures for the linear pathway. A. Both regulatory loops correspond to inhibition with strength -0.5. B. Dashed line denotes inhibition with strength -0.5 and dashed-dotted line denotes activation with strength 0.5. 191

5.4 Branched biosynthetic pathway. 192

5.5 Xanthine monophosphate (XMP) and guanosine monophosphate (GMP) sythesis pathway. Solid lines denote reaction steps, dotted lines denote dependency on the corresponding parameters (P_l), dashed lines denote inhibition, and dashed-dotted lines denote activation. Numbers in circles identify the reaction steps as they are referenced in the text. 193

5.6 The changes in regulatory structure and in enzyme expression levels of the XMP and GMP synthesis pathway which maximize XMP and GMP (X_4) concentration. Signs in circles next to the dotted lines indicated increase (+) or decrease (−) in the expression level of the corresponding enzymes (P_l). 194

6.1 The glycolysis pathway in bacterial cells. Abbreviations: G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; FdP: fructose 1,6-diphospate; DAHP: dihydroxyacetone phosphate; GA3P: glyceraldehyde 3-phosphate; DPG: dipsphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; Pyr: pyruvate; ATP: adenosine triphosphate; ADP: adenosine diphosphate; AMP: adenosine monophosphate; **PTS**: phosphotransferase system; **PGI**: phosphoclucose isomerase; **PFK**: phosphofructokinase; **ALD**: aldolase; **TPI**: triosephosphate isomerase; **GAPD**: glyceraldehyde phosphate dehydrogenase; **DPGK**: diphosphoglycerate kinase; **PGM**: phosphoglycerate mutase; **ENO**: enolase; **PYK**: pyruvate kinase; **AK**: adenylate kinase; **ATPase**: ATPase; **ATS**: ATP synthesis; **AD**: adenylate degredation; **AS**: adenylate synthesis. Bold-face abbreviations indicate enzymes and reaction steps. 215

6.2 The glycolysis prototype pathway. 216

6.3 Qualitative graph of the dependency of v_1 (solid line) and v_4 (dashed line) on ϵ for different combiantions of the parameters α and θ. . . . 217

6.4 The multiplicity regions in the α–θ parameter space. The latin numbers correspond to the scenarios shown in Figure 3. In region I two steady states are possible, in region III three steady states are possible, and in region V only steady state is possible. 218

6.5 The dependency of ϕ_1 (convex lines) and ϕ_2 (concave lines) on ϕ_3 for different values of q_a : 0.01 (solid line), 0.1 (dotted line), 1 (dashed lines), 10 (long-dashed lines), and 100 (dashed-dotted line). 219

6.6 Qualitative diagram of the dependency of $f(\phi_3)$ on ϕ_3 for different values of the kinetic parameters ψ, ω, λ, and ξ. 220

6.7 The multiplicity regions in the $\delta - \lambda$ parameter space for $q_a = 1$ and $\xi = 1$. The latin numbers correspond to the scenarios shown in Figure 6. In the parameter-value regions I, III, V-VI, and VIII, three, seven, nine or ten, and four steady states possibly exist for the system. . . 221

6.8 The multiplicity regions in the δ–λ parameter space for different values of q_a and ξ. Numbers on the arrows are the values considered for q_a; A: $\xi = 0.01$, B: $\xi = 0.1$, C: $\xi = 1$. 222

7.1 The two optimal solutions of the MINLP problem. Solid arrows indicate enzyme overexpression for the respective reaction, dotted arrows indicate reactions with enzyme levels at the reference state, and dashed arrows indicate inhibitory loops. In solution b the reaction numbering has been omitted for clarity. 262

List of Tables

4.1 The values for the continuous variables **z** and **q** for the first four best solutions of Problem 1. 129
4.2 The values for the continuous variables **z** and **q** for the second four best solutions of Problem 1. 130
4.3 The values for the continuous variables **z** and **q** for the second four best solutions of Problem 2. 131
4.4 The number of alternative regulatory structures for each of the 12 combinations of enzyme overexpression from Problem 3. 132
4.5 The values of the continuous variables **z** and **q** for six solutions of Problem 3 (cases a-f from Figure 4). 133
4.6 The values of the continuous variables **z** and **q** for six solutions of Problem 4 (cases a-f from Figure 5). 134

5.1 The values for the objective function, enzyme expression levels and the concentrations of the metabolites when one, two, three, four, five, and six enzymes change simultaneously, within ±10% of their reference value and without any modifications in the regulatory strucure. Bold type, underlined numbers indicate the changes in the enzyme levels. 183
5.2 The concentration control coefficients for the four metabolites with respect to the six enzymes. 184
5.3 The values for the objective function, enzyme expression levels and the concentrations of the metabolites when one, two, three, four, five, and six enzymes change simultaneously, between ±20% and ±500% of their reference value and without any modifications in the regulatory strucure. Bold type, underlined numbers indicate the changes in the enzyme levels. 185

5.4 The optimal values for the objective function, enzyme expression levels and the concentrations of the metabolites when six enzymes change simultaneously, between ±20% and ±500% of their reference value and modifications in the regulatory structure are considered. The corresponding optimal regulatory structure is presented in Figure 6. 186

5.5 The optimal values for the objective function, enzyme expression levels and the concentrations of the metabolites when six enzymes change simultaneously, between ±20% and ±500% of their reference value and inactivation of only one regulatory loop is considered. 187

7.1 The metabolite concentration levels and the enzyme levels for the two optimal solutions of the MINLP problem. The subscript i denotes the corresponding reaction steps for the V_m variables, and for the metabolite concentrations: x_1 =[DAHP], x_2 =[CHR], x_3 =[PHP], x_4 =[PHE], x_5 =[TYR], x_6 =[TRP]. 260

Chapter 1 Prologue

1.1 Dealing with Metabolic Complexity

Current knowledge of biochemical systems is composed of a vast set of data that accumulate with an increasing rate. Advances in analytical methods and development of sophisticated techniques and instrumentation have provided the tools that allow us to know more than we can understand. However, it is well-understood that living organisms are characterized by high complexity. This complexity increases from unicellular organisms to isolated tissue cells and multicellular structures, such as tissues and organs.

Recombinant DNA and genetic technology have been major advances in molecular biology that allow the introduction of precise changes in many aspects of cell function at the molecular level and thus, the engineering of metabolic activities for novel and/or improved functions. The application of molecular biology and engineering tools for the useful manipulation of cellular processes is defined as metabolic engineering. This is a very neat definition since it identifies the engineering virtues of the manipulation of metabolism: putting together the available information, employing intuition, and tackling the problem. However, as in many engineering approaches, solution of the problem has not been always achieved. Moreover, the discovery of some of the most profitable biotechnological applications was serendipitous.

The current approaches and methodologies, while providing with the understanding of isolated cellular processes and subsystems, do not allow understanding of the simultaneous contributions of these subsystems to the overall cellular metabolism. It is the organization of these subsystems, which are themselves relatively small but elaborate networks, in large complex networks where intuition and piece-wise knowledge are failing as tools for rational metabolic engineering (and, similarly, for accomplishing molecular biology and molecular medicine). Moreover, the high dimensionality

of these networks is offering an almost infinite number of alternative approaches toward the achievement of a goal. Most of the metabolic networks considered for engineered manipulation are composed of enzyme-catalyzed reactions with complicated stoichiometry, nonlinear kinetics, and superimposed regulatory structures. These regulatory structures are interactions of enzymes with compounds in the reaction networks, which often are not reactants or products of the reaction catalyzed by the enzymes in question, but which modulate the catalytic activity of the enzymes.

Unsuccessful attempts to engineer cellular metabolism by simply manipulating the amount of various enzymes suggest that engineering of the regulatory characteristics of the enzymes in a metabolic network offers a great potential for the achievement of desired metabolic properties. In many examples of small reaction networks with simple regulatory structures, changes in the regulation of one or two enzymes improved product formation significantly. In large metabolic networks the regulatory structures tend to be more elaborate. Moreover, the experimental difficulties in modifying regulatory interactions add another degree of complexity, since the common trial-and-error experimental approach is infeasible.

The development of tools and frameworks that will organize the available biological knowledge and will help in the analysis and design of metabolic networks is of immediate importance. These tools should be able to screen efficiently through an almost unlimited set of cellular modifications (realizable by genetic engineering technology) and report a small set of most promising options that can be further tested for their potential to lead to the development of new products and improved bioprocesses.

In this thesis a set of mathematical and computational methods are proposed as tools for the accelaration of this iterative cycle of metabolic engineering. Mathematical models of the cell have been used successfully in the past for the analysis of cellular processes, and they have provided useful initial directions for genetic improvements of the process of interest. The realization that mathematical models are the only way that net consequences of simultaneous, coupled, and often counteracting processes can be evaluated consistently and quantitatively, has led to the growth of

mathematical modeling in many biological and biotechnological areas.

The analysis of these mathematical models has been mainly based on simulating consequences of genetic and environmental changes using the models as surrogate organisms for *in silico* studies, the same way the living cells would be studied *in vivo* in the laboratory. However, simulations of alternative approaches do not really address the problem in a systematic and consistent way. The mathematical frameworks presented in this thesis do not use mathematical models for simulating experiments. They use the models in order to suggest experiments for the optimization of a biotechnological objective without enumerating a very large set of alternative approaches.

1.2 Mathematical and Computational Methods for Metabolic Analysis and Design

Mathematical models of different classes, with different information bases and different inputs and outputs are necessary to organize and apply data on metabolic networks towards the ultimate goal of effective redirection of metabolism. Associated with each class of models are special methods of mathematical analysis and computational algorithms. In this thesis a set of model types and mathematical methods is presented, focusing in particular on the influence of metabolic regulation on the performance of metabolism.

Fluxes of nutrients and metabolites into and out of the cell can be estimated directly from measurements of the concentrations of these components in the medium. If we wish to know how these external fluxes change in response to a particular genetic modification or change in the cellular environment, this change can be directly determined. However, this tells us little about how the cell's metabolism functions, and provides no guidance about expected effects of other changes in fluxes. More insight into the workings of metabolism can be gained by analyzing external flux data using a stoichiometric model of the pertinent intracellular metabolic pathways. The metabolic stoichiometry is employed to formulate quasi-steady state mass balances on

metabolic intermediates. Typically this set of equations is underdetermined: there are more unknown fluxes than linear independent mass balances. The various approaches used for addressing this problem are discussed in Chapter 2, their limitations are illustrated using an example, and a novel approach is proposed.

Prediction of the changes in metabolic reaction rates which will occur after a change in any parameter affecting the metabolism requires a kinetic representation relating fluxes to metabolite concentrations and metabolic parameters. In general such models are not available, and the kinetic model must therefore be developed from measurements of reaction rates and metabolite concentrations. A variety of formulations are possible. A (log)linear kinetic model for metabolic reaction networks is introduced in Chapter 3. The model does not require detailed information about the kinetic mechanisms of the reactions. It simply employs experimental knowledge about the strength of interaction of the various metabolites, substrates and regulators, with the rates of enzyme-catalyzed reactions of the metabolic network. Comparative studies between (log)linear models and nonlinear models based on common reaction mechanisms are presented, demonstrating the satisfactory accuracy of (log)linear models in approximating the dynamic responses of metabolic networks to changes in metabolic parameters.

A production-oriented optimization framework is developed in Chapter 4. The framework considers the optimization of the performance of a metabolic pathway with respect to changes in the amounts of enzymes in the pathway and in modifications in the regulatory characteristics of those enzymes. The optimization study is undertaken using the kinetic description provided by the (log)linear kinetic model, used in this case to define constraints on the optimization in the form of steady-state mass balances of intracellular metabolites. Consideration of changes in enzyme regulation which abolish initially present inhibition or activation introduces binary decision variables into the optimization, resulting in a mixed-integer programming problem. Computational studies applying this optimization framework to a prototype mathematical model of bacterial aromatic amino acid production are presented. The results of such studies provide useful qualitative guidance for promising targets

for metabolic and protein engineering to achieve a preferred flux distribution.

The experimental information used in building the (log)linear model can be also used to build approximate dynamic nonlinear models of metabolic reaction networks. This class of models, called S-system models, at steady-state provide with a linear relation between the logarithms of metabolite concentrations and the logarithms of the metabolic parameters. In Chapter 5, the optimization framework introduced in Chapter 4 is developed for S-system models. The computational studies for two prototype pathways suggest that significant improvements in the performance of metabolism can be achieved if, additionally to changes in the amounts of the enzymes, modifications in the regulatory characteristics are considered, especially when metabolic constraints are taken into account.

The (log)linear kinetic model introduced in Chapter 3 and the S-system models of Chapter 4 provide representation of a limited class of nonlinear kinetic expressions, but these representations do not in general describe flux-substrate-effector-parameter relationships for large deviations from the reference state. Formulating more complete, more broadly valid nonlinear kinetic models for cellular processes will typically require more complete information which may come from a combination of more extensive experimental studies and literature information. One of the most important types of mathematical analyses which can be accomplished given a general nonlinear kinetic model of metabolism concerns steady-state multiplicities. While (log)linear and S-system models feature a unique steady state, nonlinear models can exhibit multiple steady states; i.e., for a given set of metabolic parameters there exist more than one time-invariant metabolic state. Analysis of a prototype model for the bacterial glycolytic pathway, considering the simplest nonlinear kinetics, is performed in Chapter 6. The analysis indicates that up to ten steady states can exist for certain parameter values, suggesting the complexity of metabolism and the difficulties that arise when nonlinear models are considered.

The optimization framework developed in Chapter 4 is further extended in Chapter 7 for studies of nonlinear kinetic models. As is expected, use of nonlinear models requires a more complicated formulation of the framework and advanced algorithmic

procedures for the solution of the optimization problem. The prototype mathematical model of bacterial aromatic amino acid production is studied again as an example, and the significant improvement in the objective of interest suggested by the solution illustrates the power of the optimization framework as a tool for rational metabolic engineering.

Chapter 2 Analysis of Metabolic Reaction Rates

2.1 Introduction

Mathematical descriptions of metabolic reaction networks have been widely used for better understanding of metabolism. The information obtained from such mathematical descriptions can be used to design genetically engineered organisms with desired properties. The basic information needed, prior to construction of any *mathematical* model of metabolism, is the *biochemical* model of the network of interest; i.e., the stoichiometry of the reactions that make up the network. Determination of the stoichiometry of biochemical reactions has been the subject of intensive studies in biochemistry, and as a result the stoichiometry governing the intermediary metabolism of many organisms is well known (Gottschalk, 1986; Neidhardt *et al*, 1987; Sonenshein *et al.*, 1993; Stryer, 1988; Wood, 1985).

Once the stoichiometry of a bioreaction network is known, the equations that describe the mass balances for the metabolites in the network can be formulated. This is the first step towards a mathematical description of the metabolism. Depending on the available information and on the purposes of the mathematical analysis, one can further develop the mathematical model by obtaining information ranging from reaction rates in the network to the dynamic and steady-state responses of metabolite concentrations to manipulations of process and/or genetic parameter.

In this chapter a study of metabolic reaction networks using metabolic flux balancing will be presented. This technique, although it is almost twenty years old, has been recently revived in response to the widening spectrum of metabolic engineering applications as well as to the need for optimized growth medium formulation and process design (Holms *et al.*, 1990; Varma and Palsson, 1994). However, most published studies of metabolic fluxes for various organisms are based upon assumptions that are not easily justified and the results obtained can be very sensitive to these assumptions. The sensitivity of the estimated metabolic fluxes to these assumptions will be discussed here using the metabolic reaction network for aerobically growing *Bacillus subtilis* as an example model.

A new framework for the calculation of the fluxes in a given metabolic network

is presented. The framework minimizes the number of assumptions with respect to cellular energetics. This framework can be used for the formulation of mathematical models for the design and control of bioprocesses. The formulation of the flux balancing problem as a constrained quadratic or linear programming problem allows analysis of the system of interest with respect to uncertainties in the measurements and to physiological limits on the bioreaction rates.

2.2 From a Net Metabolic Reaction to a Network of Metabolic Reactions

Every living organism is composed of chemical species that mainly consist of four chemical elements: carbon, C, hydrogen, H, oxygen, O, and nitrogen, N. If the elemental composition of a particular strain growing under particular conditions is known, the ratios of subscripts in the empirical cell formula $C_\theta H_\alpha O_\beta N_\delta$ are easily determined. In order to establish a unique cell formula and corresponding molecular weight, it is convenient to employ a formula which contains one gram-atom of carbon. That is, θ can be set equal to 1, and then α, β, and δ can be calculated from the known relative elemental weight content of the cells. One *C-mole of cells* is by definition the quantity of cells containing one gram-atom (12.011 grams) carbon, and corresponds to the cell formula weight with the carbon subscript θ set to unity.

Next, I will consider next the aerobic growth of bacterial cells without product formation, and how this simple bioprocess can be described as a *net metabolic reaction.* The elemental mass balances will be used to derive relations among the stoichiometric coefficients. These relations will be used to derive relations between intracellular metabolic fluxes and extracellular physiological characteristics of the cells, such as O_2 uptake rate, specific growth rate, and CO_2 production rate.

2.2.1 Aerobic Growth Without Product Formation: A Net Reaction Rate

During the aerobic growth of bacterial cells in the absence of product formation the only products of the growth reaction are cells, CO_2 and H_2O. Writing the carbon source and nitrogen source chemical formulas as CH_xO_y and $H_lO_mN_n$, respectively, the growth reaction equation is

$$a'CH_xO_y + b'O_2 + c'H_lO_mN_n \longrightarrow CH_\alpha O_\beta N_\delta + d'H_2O + e'CO_2 \tag{2.1}$$

Balances on the four elements in equation (2.1) provide four relationships among the five unknown stoichiometric coefficients a', b', c', d', and e':

$$C \;:\; a' = 1 + e' \tag{2.2}$$

$$H \;:\; a'x + c'l = \alpha + 2d' \tag{2.3}$$

$$O \;:\; a'y + 2b' + c'm = \beta + d' + 2e' \tag{2.4}$$

$$N \;:\; c'n = \delta \tag{2.5}$$

An additional relationship can be derived by using the experimentally determined *respiratory quotient*, or RQ, for the growth reaction. The respiratory quotient is defined as the molar ratio of CO_2 formed to O_2 consumed:

$$\text{Respiratory Quotient} = RQ = \frac{\text{moles } CO_2 \text{ formed}}{\text{moles } O_2 \text{ consumed}} \tag{2.6}$$

and for the growth reaction (2.1) it can be written:

$$RQ = \frac{e'}{b'} \tag{2.7}$$

If the RQ is known, equations (2.2) to (2.5) and (2.7) can be solved for the five

unknown stoichiometric coefficients as follows:

$$a' = \frac{(4n - \delta l + \alpha n + 2\delta m - 2\beta n)\,\mathrm{RQ} - 4n}{(nx - 2ny + 4n)\,\mathrm{RQ} - 4n} \tag{2.8}$$

$$b' = \frac{-xn - \delta l + \alpha n + 2yn + 2\delta m - 2\beta n}{(nx - 2ny + 4n)\,\mathrm{RQ} - 4n} \tag{2.9}$$

$$c' = \frac{\delta}{n} \tag{2.10}$$

$$d' = \frac{(\delta mx + 2\delta l - 2\alpha n + 2xn - y\delta l + y\alpha n - \beta nx)\,\mathrm{RQ} - 2xn - 2\delta l + 2\alpha n}{(nx - 2ny + 4n)\,\mathrm{RQ} - 4n} \tag{2.11}$$

$$e' = \frac{\mathrm{RQ}\,(-xn - \delta l + \alpha n + 2yn + 2\delta m - 2\beta n)}{(nx - 2ny + 4n)\,\mathrm{RQ} - 4n} \tag{2.12}$$

Using the above relations and the on-line determination of RQ, bioprocess control and monitoring is possible. In a series of papers, Stephanopoulos and coworkers studied the applicability of the respiratory quotient as a measurement for on-line bioreactor identification and control (Grosz *et al.*, 1984; Stephanopoulos and San, 1984). They found that singularities can exist in the relations used, and they derived general rules for identifying conditions that may cause singularities.

The correlation of RQ with the process yield will become clear if we notice that for the biomass yield on the substrate (grams biomass formed per moles substrate consumed) can be written as:

$$Y_{X/S} = \frac{MW_c n_s}{a'} \tag{2.13}$$

where MW_c is the molecular weight of the cells and n_s is the number of the carbon atoms per molecule substrate. Combined with equation (2.8) this becomes:

$$Y_{X/S} = MW_c n_s \frac{(nx - 2ny + 4n)\,\mathrm{RQ} - 4n}{(4n - \delta l + \alpha n + 2\delta m - 2\beta n)\,\mathrm{RQ} - 4n} \tag{2.14}$$

or in general form:

$$Y_{X/S} = \frac{A_x RQ - B_x}{\Gamma RQ - 4} \tag{2.15}$$

It be should mentioned here that for aerobic growth on glucose as carbon source A_x will be always equal to B_x independent of the cellular composition. Similarly for

the molar yield factor of the moles of CO_2 produced per mole substrate consumed, $Y_{CO2/S}$, we can derive:

$$Y_{CO2/S} = n_s \frac{\mathrm{RQ}\,(-xn - \delta l + \alpha n + 2yn + 2\delta m - 2\beta n)}{(4n - \delta l + \alpha n + 2\delta m - 2\beta n)\,\mathrm{RQ} - 4n} \tag{2.16}$$

or, in general

$$Y_{CO_2/S} = \frac{A_{co} RQ}{\Gamma RQ - 4} \tag{2.17}$$

The analysis will be continued for a generalized bacterium with experimentally determined chemical formula $CH_2N_{0.25}O_{0.5}$ ($MW_c = 25.5$) which grows on glucose ($C_6H_{12}O_6$; $n_s = 6$) as a carbon source and ammonia (NH_3) as a nitrogen source. The following values can be determined for such a system:

$$A_x = 612$$

$$B_x = 612$$

$$A_{co} = 1.5$$

$$\Gamma = 4.25$$

These values permit us to identify limits on the respiratory quotient to ensure that the yields will be positive, finite numbers:

$$RQ > 1 \tag{2.18}$$

The problem of the sensitivity of the yields to uncertainties in the RQ determination becomes apparent when the scaled sensitivity of the yields is considered:

$$\frac{dlnY_{X/S}}{dlnRQ} = \frac{RQ(\Gamma B_x - 4A_x)}{(\Gamma RQ - 4)(A_x RQ - B_x)} = \frac{RQ}{(RQ - 1)(17RQ - 16)} \tag{2.19}$$

and

$$\frac{dlnY_{CO_2/S}}{dlnRQ} = \frac{-4}{\Gamma RQ - 4} = \frac{-16}{17RQ - 16} \tag{2.20}$$

For the values used here, the sensitivity of the yield $Y_{X/S}$ to RQ becomes infinite as RQ approaches the lower limit. More detailed analysis of the sensitivity issue can be found elsewhere (Grosz *et al.*, 1984).

The expressions derived above for the rates of the reactions given by the general stoichiometric equation (2.1) have be used extensively for bioprocess identification and control (Roels, 1983), but they do not provide any information about reaction rates in the intracellular reaction networks.

2.2.2 Aerobic Growth Without Product Formation: A Metabolic Reaction Network

Metabolite Mass Balancing

The stoichiometric metabolic networks that can be formulated for various organisms are, in general, dependent on the conditions under which the organism is growing, since different conditions cause expression of different enzymes and use of different metabolic pathways for the catabolism of external nutrients and cell growth. If we consider the bacterium *B. subtilis*, the metabolic stoichiometry is well-known and stoichiometric models have been formulated (Sonenshein *et al.*, 1993). For aerobically growing *B. subtilis*, the reaction network of the central carbon pathways are presented in Figure 1. The experimentally determined fluxes are the specific glucose uptake rate, V_n and the fluxes

$$\{V_i, V_k, V_g, V_h, V_j, V_l, V_m, V_f, V_e, V_a, V_b\}$$

from the precursor metabolites

$$\{G6P, F6P, R5P, E4P, T3P, PGA, PEP, Pyr, ACoA, OGA, OAA\}$$

to biosynthesis. These fluxes can be calculated by the following formula:

$$V_{precursor} = \frac{1}{x}\frac{dx}{dt}Y_{precursor} = \mu Y_{precursor} \tag{2.21}$$

where x denotes the biomass concentration in grams per unit volume bioreactor, μ is the specific growth rate of the cells, and $Y_{precursor}$ is the biosynthetic requirement of the corresponding precursor (moles of precursor required per gram cells produced). These precursor yields have been reported for *E. coli* (Neidhardt, 1987), *B. subtilis* (Sauer *et al.*, 1996), and many other organisms of industrial importance, and their experimental estimation is a rather standard technique (Sauer *et al.*, 1996).

Mass balances around the various metabolites can easily be constructed. Consider, for example, the mass balance for $G6P$:

$$\frac{d[G6P]}{dt} = V_n - V_1 - V_2 - V_i - \mu[G6P] \tag{2.22}$$

where the subscripts correspond to the reactions as labeled in Figure 1, and the term $\mu[G6P]$ corresponds to the effects of the dilution due to cell growth. The quasi-steady state assumption is the basic assumption of the metabolic flux balancing. This assumption is based on the fact that metabolic transients are typically rapid compared to cellular growth rates and changes in the environmental conditions. Based on this assumption, equation (2.22) becomes:

$$0 = V_n - V_1 - V_2 - V_i - \mu[G6P] \tag{2.23}$$

The reaction rates V_n and V_i can be experimentally determined as follows:

$$V_n = \frac{1}{x}\frac{d[S]}{dt} \tag{2.24}$$

and

$$V_i = \mu Y_{G6P} \tag{2.25}$$

where $[S]$ is the concentration of the extracellular carbon source, which in this case is glucose.

If we normalize all of the reaction rates in the network with respect to the specific

uptake rate (V_n) we can write for the Equation (2.23):

$$0 = 1 - v_1 - v_2 - Y_{X/S}Y_{G6P} - Y_{X/S}[G6P] \tag{2.26}$$

where the lower-case letters denote reaction rates normalized, with respect to V_n, rates. In Equation (2.26) the use of the yield coefficient of the biomass on the substrate results from its definition:

$$Y_{X/S} = \frac{dx}{d[S]} = \frac{\mu}{V_n} = \frac{\frac{1}{x}\frac{dx}{dt}}{\frac{1}{x}\frac{d[S]}{dt}} \tag{2.27}$$

This yield can be estimated experimentally. The value of the term $Y_{X/S}[G6P]$ in the mass balance equation is the normalized, with respect to V_n, dilution term and since it has a mcu smaller value with respect to the main fluxes in the pathway considered here, it will be omitted from the formulation of the mass balances.

For the mass balances of the metabolites from $G6P$ to $T3P$ we can write:

$$G6P: \quad 0 = \quad 1 - v_1 - v_2 - Y_{X/S}Y_{G6P} \tag{2.28}$$

$$Ru5P: \quad 0 = \quad v_1 - v_4 - v_5 \tag{2.29}$$

$$R5P: \quad 0 = \quad v_5 - v_6 - Y_{X/S}Y_{R5P} \tag{2.30}$$

$$X5P: \quad 0 = \quad v_4 + v_3 - v_6 \tag{2.31}$$

$$F6P: \quad 0 = \quad v_2 - v_3 + v_6 - v_7 - Y_{X/S}Y_{F6P} \tag{2.32}$$

$$E4P: \quad 0 = \quad v_3 + v_6 - Y_{X/S}Y_{E4P} \tag{2.33}$$

$$T3P: \quad 0 = \quad -v_3 + 2v_7 - v_8 - Y_{X/S}Y_{T3P} \tag{2.34}$$

Here we can identify the first major problem in metabolic flux balancing. These are seven linear relations between the fluxes, but there are eight uknown fluxes. Therefore, the system is underdetermined.

For the rest of the metabolites considered in the network we can formulate the following mass balance equations:

$$PGA: \quad 0 = \quad v_8 - v_9 - Y_{X/S}Y_{PGA} \tag{2.35}$$

$$PEP: \quad 0 = \quad -1 + v_9 - v_10 - Y_{X/S}Y_{PEP} \tag{2.36}$$

$$Pyr: \quad 0 = \quad 1 + v_10 - v_{11} - v_{12} - Y_{X/S}Y_{Pyr} \tag{2.37}$$

$$ACoA: \quad 0 = \quad v_{12} - v_{13} - Y_{X/S}Y_{AcCoA} \tag{2.38}$$

$$OGA: \quad 0 = \quad v_{13} - v_{14} - Y_{X/S}Y_{OGA} \tag{2.39}$$

$$OAA: \quad 0 = \quad v_{11} - v_{13} + v_{16} - Y_{X/S}Y_{OAA} \tag{2.40}$$

$$Suc: \quad 0 = \quad v_{14} - v_{15} \tag{2.41}$$

$$Mal: \quad 0 = \quad v_{15} - v_{16} \tag{2.42}$$

The above equations for the metabolic fluxes in the central carbon pathways for aerobically growing *B. subtilis* introduces a problem common in the metabolic flux balances: the unknown reaction rates are more numerous than are the metabolites being balanced, i.e., the number of the equations that can be formulated is less than the number of the unknown fluxes. Most of the metabolic systems are likewise *underdetermined* (Bonarius *et al.*, 1996; Sauer *et al.*, 1996; Savinell and Palsson, 1992a-c; Varma and Palsson, 1995). In studies that appeared before now in the literature, the investigators introduce a series of assumptions in order to circumvent this problem. Some of these will be examined here. Many research efforts have been devoted to the experimental determination of one or more of the unknown fluxes, so that these assumptions can be avoided or validated. Experimental techniques such as NMR, tracing of radioactive labels, and mass spectroscopy have been successfully used in order to define exact values or strict bounds for certain fluxes such as the flux from $G6P$ to $Ru5P$ (reaction step 1 in Figure 1) and the fluxes in the tricarboxylic acid cycle (TCA) (reaction steps 13 to 16 in Figure 1) (Mancuso *et al.*, 1994; Reitzer *et al.*, 1980; Walsh and Koshland, 1984).

Algorithmic Procedure

The general procedure for estimation of the unknown fluxes will be summarized next. The problem will be formulated as a nonlinear programming problem which, when solved, will provide an estimate for the fluxes that will satisfy the mass balances with minimal error. This objective can be mathematically formulated by defining a new set of variables, $\mathbf{r}$. The number of these variables will be equal to the number of the mass balances. For a metabolic system with n metabolites and m unknown fluxes, the general matrix expression for the mass balances can be written as:

$$\mathbf{Nv} + \mathbf{b} + \mathbf{r} = \mathbf{0} \tag{2.43}$$

where $\mathbf{N}$ is the $n \times m$ stoichiometric matrix, $\mathbf{v}$ is the m-dimensional vector of the unknown fluxes, $\mathbf{b}$ is the n-dimensional vector of the total sum of the known fluxes for each metabolite mass balance, and $\mathbf{r}$ is the n-dimensional vector of the residuals from the mass balances. The variables of the problem are the fluxes $\mathbf{v}$ and the residuals $\mathbf{r}$. The objective of the problem can be mathematically formulated as follows:

$$minimize \sum_{j=1}^{n} r_j^2 \tag{2.44}$$

with respect to the fluxes $v_1, \ldots, v_m$, subject to the following constraints:

I. Mass Balances

The mass balance equations (2.43) will define a set of equality constraints for the reaction rates and the residuals.

II. Bounds on the rates

Many of the reaction rates in any metabolic work are reversible, i.e. they can proceed in both directions. However, there are reactions that are irreversible. Therefore, if the mass balance network is constructed in such a way that the irreversible reactions will be positive when they proceed only in the allowable direction, then the following inequality constraint for the reaction rate is imposed

for every i-th irreversible reaction step.:

$$v_i \geq 0 \tag{2.45}$$

The above nonlinear optimization problem can have either a unique solution or multiple solutions depending on the constraints. In general, when the number of reactions is larger than the number of mass balance equations, then multiplicities can occur. This is true provided that none of the inequality constraints are active, meaning none of the irreversible reaction rates are zero. In the general case that the rank of the stoichiometric matrix $\mathbf{N}$, is smaller than its smaller dimension, or $Rank(\mathbf{N})<m$, the system will be called *underdetermined.* In the case that $m \leq Rank(\mathbf{N}) < n$ the system will be called *overdetermined.* Due to the existence of branching pathways, cycles, and various interdependencies within the metabolic network, the number of reactions is greater than the number of metabolite mass balances in most metabolic systems. Therefore, in most cases, the metabolic network will be underdetermined.

2.2.3 Determining the Underdetermined: Assumptions and Solutions

In order to overcome the problem of the underdetermined system many applications of flux balancing use assumptions to make the system determined or overdetermined. The most common assumptions used concern:

1. flux distribution at network nodes, such as the $G6P$ node, and
2. cellular energetics (i.e., mass balances for ATP and/or $NAD(P)H$).

Least square methods in connection with the pseudoinverse algorithm have also been used. These methods, although they solve the underdetermined system by minimizing the Euclidean norm of the residual of the errors in the mass balance equations. However, they find from the infinite possible solutions for the fluxes the unique set of fluxes with the minimum Euclidian norm for the values of the fluxes (Bonarius *et al.*, 1996; Savinell and Palsson, 1992a-c; Stewart, 1973).

As it will be shown next, different assumptions result in different solutions for the fluxes. It is not clear which is the "correct" approach. A primary criterion will be agreement with experimental results. However, agreement with experimental results is not proof of the validity of the metabolic description. Further experimental data are required, such as enzyme assays that would validate the presence or absence of a reaction step, and NMR measurements that would validate the calculated values for the intracellular fluxes (Marx *et al.*, 1996). However, an approach that uses the minimum number of assumptions and parameters is preferable, since fewer assumptions lead to a less biased solution which is easier to validate experimentally.

The G6P node

Based on experimental data of tracing radioactive labels it had been suggested that, in *E. coli* growing aerobically on glucose, the flux from $G6P$ to $Ru5P$ (reaction step 1 in Figure 1) is equal to 30% of the glucose uptake rate, V_n (Gottschalk, 1986). Since then, many metabolic flux analyses have used the assumption that this flux is proportional to, usually 30% of, the glucose uptake rate. However, this assumption strongly biases the solution since flux distribution may be dependent on the growing conditions, it may vary among different organisms, and this assumption does not take into account the contribution of this reaction to $NADPH$ synthesis. Therefore, this assumption will not be considered here.

ATP Mass Balance

One of the most common approaches to overcoming the problem of the underdetermined systems is to introduce an additional mass balance equation for ATP:

$$\begin{aligned} ATP: \quad 0 \;=\; & -v_7 + v_8 + v_{10} + v_{14} - Y_{X/S}Y_{ATP,\mu} \\ & -Y_{ATP,m} + \frac{(P/O)}{2} Y_{O/S}Y_{X/S} \end{aligned} \tag{2.46}$$

where, $Y_{ATP,m}$ is the so-called *ATP maintenance yield*, , $Y_{O/S}$ is the yield of oxygen on the substrate, and (P/O) is the stoichiometric coefficient for the ATP production

via the respiratory chain (oxidative level phosphorylation). The maintenance yield accounts for all energy-requiring processes that are not linked to biomass synthesis, such as maintenance of gradients and electrical potential, futile cycles, and turnover of macromolecules. $Y_{O/S}$ is primarily dependent on the substrate, and indicates how many oxygen molecules are required to metabolize one substrate molecule. The (P/O) ratio is necessary because different amounts of ATP can be produced for each oxygen molecule consumed, depending on the conditions.

However this approach introduces three very significant assumptions:

i. A value for the (P/O) ratio is assumed;

ii. A value for the maintenance requirement is assumed; and

iii. Both of these values are assumed to be independent of the growth rate.

Assigning values for two stoichiometric parameters will, as expected, bias the solution. Experimental methods have been proposed for the determination of these paramaters (van Gulik and Heijen, 1995). However, the experimental methods to determine these coefficients rely on an additional seriously questionable assumption that the values of these parameters are invariable under varying operating conditions such as changes in the specific growth rate (Nielsen and Villadsen, 1994). These assumptions fail when genetically engineered organisms are considered (Tsai *et al.*, 1996). In a following section the sensitivity of calculated fluxes to these assumptions that dominate the literature will be illustrated.

It has been shown, for example, that the solution of a flux balancing problem for the bacterium *E. coli* is very sensitive to the assumed value for (P/O) and less sensitive to that of the maintenance yield (Varma and Palsson, 1995). The uncertainty of these parameters and their dependency on growth conditions suggest that fluxes calculated using such assumptions should be carefully considered, and their sensitivity with respect to the assumed values should be examined.

NADPH Mass Balance

Another approach formulates the mass balance for NADPH and assumes that the biosynthetic requirement for NADPH is entirely fulfilled by the NADPH production through reaction steps 1 (pentose phosphate pathway) and 13 (TCA cycle):

$$2v_1 + v_{13} = Y_{X/S}Y_{NADPH} \tag{2.47}$$

where Y_{NADPH} is the biosynthetic requirement of NADPH (moles of NADPH required per gram cells produced). This approach is preferable to the previous one as it introduces only one new parameter: the biosynthetic requirement for NADPH, which can be readily estimated from the biomass composition. However, it is possible to over- or underestimate the value of this parameter, and this approach does not consider the reactions that could potentially exchange reducing equivalents between NADH and NADPH. One such reaction is the transhydrogenase-catalyzed reaction (Gottschalk, 1986; reaction $\underline{\underline{T}}$ in Figure 1) :

$$NADH + NADP^+ \rightleftharpoons NAD^+ + NADPH$$

This possibility can be taken into account by examining the level of the objective function, i.e. the level of the residuals for the mass balances, both for lower and higher bounds on Y_{NADPH}. The former corresponds to production of NADPH from NADH, whereas the latter corresponds to production of NADH from NADPH. The "proper" Y_{NADPH} should be considered the one that results to a lower level of the residuals. This is essentially a parameter-fitting process, but it is preferable to the balance around ATP, since it only requires the fitting of a single parameter.

Another problem with this approach is that for low growth, $Y_{X/S}$ approaches zero and the equality constraint (2.47) forces the sum $2v_1 + v_{13}$ to be equal to zero. As neither reaction is reversible, both v_1 and v_{13} then must take zero value. This does not reflect the true physiological situation since, for a nongrowing cell (i.e., $Y_{X/S} = 0$), these fluxes can have nonzero values.

Reducing Equivalents Mass Balance

Here a different approach will be introduced that considers a mass balance on reducing equivalents ($2H^+$). This approach considers the transfer of protons through the cellular membrane from intracellular NADH, NADPH, and $FADH_2$ to the extracellular medium and the oxygen consumption associated with that process. The extruded protons will enter back into the cell through translocation mechanisms. The mass balance for the reducing equivalents can be written as:

$$\begin{aligned} 0 &= 2v_1 + v_8 + v_{12} + v_{13} + v_{14} + \rho v_{15} \\ &\quad + v_{16} - 2Y_{O/S}Y_{X/S} - Y_{X/S}(Y_{NADH} + Y_{NADPH}) \end{aligned} \tag{2.48}$$

where ρ is the stoichiometric coefficient for the $FADH_2$ production. If we consider that $FADH_2$ is exactly equivalent to NAD(P)H with respect to the numbers of protons extruded for each of their molecules that oxidized, then, $\rho = 1$. However, this makes the linear constraint (2.48) linearly dependent on the rest of the mass balance equations and the problem remains underdetermined. In formulating the mass balance equation (2.48) we have considered every reaction that regenerates NADPH from NADH and NADH from NADPH. However, since the net contribution of these reactions in the mass balance on reducing equivalents is zero, they do not appear in the final equation. Thus, no assumption for these fluxes is involved, except to consider that they do exist. Two approaches to overcome this problem can be considered based on an iterative scheme.

In a first approach we will ignore the production of $FADH_2$ via reaction step 15 by setting $\rho = 0$. This means that the amount of oxygen considered in mass balance equation (2.48) for each molecule of substrate consumed, $Y_{O/S}$, is overestimated since a portion of this is actually used for the oxidation of the "ignored" $FADH_2$. Since the production of $FADH_2$ is coupled to NADH production, its maximum value is equal to 10% of oxygen consumption. In the first step, we can solve the problem ignoring $FADH_2$ production and using the experimentally determined amount of oxygen consumed. Then the calculated value for reaction step 15 will be subtracted from the

amount of oxygen considered, since the value for v_{15} is a first approximation of the excess oxygen that is used for oxidation of $FADH_2$. Then we solve the problem again, using the "corrected" oxygen yield, and calculate a new set of fluxes. The calculated value for reaction step 15 will be compared with the observed oxygen yield, and, if it is within the experimental error the iterative procedure can stop. The last solution is then accepted as giving the estimated flux values. For the systems that have been examined until now, two iterations were sufficient for the estimated values to converge to a final value for v_{15} within the error limits for the oxygen uptake measurements.

In a second approach we can assume $\rho = 0.5$. This value comes from an analysis of the relative reducing power of NADH and $FADH_2$. For every molecule of NADH that is oxidized through the respiratory chain, is assumed there are four protons extruded, whereas, for every molecule of $FADH_2$ that is oxidized through the respiratory chain, only two protons are extruded (Gottschalk, 1986). The number of the protons extruded depends on the terminal oxidase of the respiratory chain and it is different between various organisms and growth conditions. However, we assume this assumption here as a first approximation. Therefore, from the proton mass balance we see that two $FADH_2$ molecules are equivalent to one NADH molecule. However, this approach essentially assumes that for every two molecules of $FADH_2$ that are oxidized, one molecule of O_2 is consumed. The validity of this assumption can be examined by comparing the estimated value of v_{15} with the value of $Y_{O/S}$. If half of the value of v_{15} is greater than the experimental error of $Y_{O/S}$ then it should be "corrected" as described in the previous paragraph.

Least Squares Method and the Minimum Norm Condition

In many cases in which the system is underdetermined, the pseudoinverse of the stoichiometric matrix has been used to estimate the fluxes (Savinel and Pallson, 1992a-c). The pseudoinverse algorithm solves the system by minimizing the sum of the squares of the residuals of the mass balances. In other words, it minimizes the same objective function as the one suggested for the general algorithmic procedure (Equation (2.44)). However, among the multiple possible solutions due to the underdetermined nature

of the system, the pseudoinverse chooses the one with the minimum Euclidean norm of the reaction rates.

This last condition is a purely mathematical criterion, and it is reasonable in the absence of additional consideration. However, it has recently been proposed that Darwin's principle of evolution by natural selection suggests that there are optimal regimes of operation along metabolic pathways in a biological system (Torres, 1991). Torres (1991) suggested that one should characterize "fitness" in thermodynamic terms, since a fundamental tendency of nonequilibrium systems towards stationary states of maximal organization and minimal dissipation constitutes a potentially solid bridge between thermodynamics and Darwin's principle. Therefore, he proposed some thermodynamic criteria for optimality: maximal efficiency, maximal power, minimal rate of entropy production, and minimal loss of available energy.

Here I will consider the last two criteria. According to the theory of nonequilibrium thermodynamics (Prigogine, 1961) for the rate of entropy production we can write:

$$\frac{dS}{dt} = \frac{1}{T}\sum_{j=1}^{m} A_j v_j \tag{2.49}$$

and for the energy dissipation function, Ψ:

$$\Psi = \sum_{j=1}^{m} A_j v_j \tag{2.50}$$

where A_j is the *affinity* of reaction step j defined as:

$$A_j = -\sum_{k} \nu_k \mu_k \tag{2.51}$$

where k covers the range of species that participate in the reaction step j, ν_k is the stoichiometric coefficient of the k-th species and is positive for the products and negative for the reactants, and μ_k is the chemical potential of species k.

In order to find the functional dependency of v_j on the corresponding affinity we will use the same procedure followed by Torres (1991). He chose the rate expressions introduced and used within Biochemical Systems Theory (Savageau, 1976) (see also

further discussion in Chapter 5 for details on this theory) which employ a power-law representation:

$$v_j = k_r \prod_{i_r} x_{i_r}^{g_{i_r}} - k_p \prod_{i_p} x_{i_p}^{h_{i_p}} \tag{2.52}$$

where x_{i_r} and x_{i_p} are the concentrations of the reactants and the products, respectively, and k_r, k_p, g_{i_r} and h_{i_p} are phenomenological parameters that generalize the kinetic constants and the stoichiometric coefficients of the reactants (subscript r) and the products (subscript p).

Torres went on to conjecture that the kinetic orders g_{i_r} and h_{i_p} reduce to the absolute values of the corresponding stoichiometric coefficients $\mid \nu_n \mid$ when enzyme concentrations go to zero, and he proposed that they can be approximated by the following equations:

$$g_{i_r} \approx b \mid \nu_{i_r} \mid \tag{2.53}$$

and

$$h_{i_p} \approx b \mid \nu_{i_p} \mid \tag{2.54}$$

where b depends on the enzyme concentrations and goes to unity as they approach zero. After these assumptions, v_j can be written in the form:

$$v_j = v_{mj} \left(1 - exp\left(-\frac{bA_j}{RT}\right)\right) \tag{2.55}$$

with $v_{mj} = k_{j_r} \prod_{i_r} x_{i_r}^{g_{i_r}}$, which is the maximum forward rate, and A_j:

$$A_j = RT \left[ln\left(\prod_{i_r}\left(\frac{x_{i_r}}{x_{i_r,e}}\right)^{\nu_{i_r}}\right) - ln\left(\prod_{i_p}\left(\frac{x_{i_p}}{x_{i_p,e}}\right)^{\nu_{i_p}}\right)\right]$$

The subscript e denotes the equilibrium value of the corresponding concentration. Torres applied this theoretical development to study the efficiency of glycolysis as described by a single reaction for anaerobic glycolysis:

$$glucose + 2ADP + 2P^{+} \rightleftharpoons 2lactate + 2ATP \tag{2.56}$$

and for the aerobic case:

$$6O_2 + glucose + 36ADP + 36P^+ \rightleftharpoons 6CO_2 + 6H_2 + 36ATP \qquad (2.57)$$

In both cases his results were very close to the experimental observations.

In a subsequent paper Angulo-Brown *et al.* (1995) studied Torres' suggestion for thermodynamic optimization criteria in biochemical reactions. They assumed, following Prigogine's (1961) suggestion, that for most of the biological reactions

$$\mid A_j \mid /RT << 1$$

This allowed them to use the approximation

$$exp\left(-\frac{bA_j}{RT}\right) \approx 1 - \frac{bA_j}{RT}$$

to simplify the rate expression (2.55) to:

$$v_j = v_{mj}\left(1 - exp\left(-\frac{bA_j}{RT}\right)\right) \approx v_{mj}\frac{bA_j}{RT} \qquad (2.58)$$

Note that these approximations hold only if b is not a large parameter.

Angulo-Brown *et al.* (1995) used this approximation to study the same systems that Torres studied. They arrived at similar results, suggesting that approximation (2.58) is valid.

Using approximation (2.58), for the entropy production (2.49) we can write :

$$\frac{dS}{dt} = \frac{1}{T}\sum_{j=1}^{m} A_j v_{mj}\frac{bA_j}{RT} \qquad (2.59)$$

This implies that minimization of entropy production is equivalent to minimization of the sum of the square of the affinities over all of the m reaction steps:

$$\sum_{j=1}^{m} A_j^2$$

The least squares approach that solves the flux balance problem by choosing the minimum Euclidean norm of the fluxes seems to be consistent with the above thermodynamic considerations. If for every flux we can use the approximate rate expression (2.58), then for the Euclidean norm of the fluxes we can write:

$$\| \mathbf{v} \|_2 = \sum_{j=1}^{m} v_j^2 = \sum_{j=1}^{m} v_{mj}^2 \frac{b^2 A_j^2}{R^2 T^2} \tag{2.60}$$

Therefore, selecting the solution for the fluxes with the minimum Euclidean norm is equivalent to choosing the solution with the minimum sum of the squares of the affinities, and thus the one that corresponds to the minimum entropy production rate.

The only problem with this approach appears to be the correct choice of the fluxes that should be taken into account in the calculation of the entropy production rate. This approach should also be examined with respect to each of the thermodynamic criteria that have been suggested. Moreover, the application of this approach should take into account the bounds imposed by the irreversibility of certain reactions. This cannot be implemented using the pseudoinverse algorithms. The possibility of connecting the minimum Euclidean norm criterion with thermodynamic criteria and evolutionary objectives suggests that this approach is worth further investigation.

2.3 Example

We will study here the metabolic fluxes in glucose-limited cultures of aerobically growing *B. subtilis*. The metabolic network that will be used (Figure 1), based on the available biochemical and biological knowledge, has been presented in Section 2.2.2. The assumptions concerning ATP, NADPH, and reducing equivalents will be examined. I will evaluate the three approaches discussed above by comparing the estimated from the flux analysis value of the CO_2 production rate with the experimentally determined value. In general, a separate mass balance for CO_2 can be included in the mass balances. This mass balance equation does not affect the rank of the stoichiometric matrix since it is linearly dependent on mass balances already

present in the network. We will consider as experimental data the value of RQ as it is determined from the net reaction (2.1) and equation (2.15) for the "generalized" bacterium composition. The equation that will be considered is:

$$RQ_{exp} = 16\frac{Y_{X/S} - 153}{17Y_{X/S} - 2448} \tag{2.61}$$

where $Y_{X/S}$ is expressed in grams of cell dry weight per moles glucose consumed. The RQ calculated from equation (2.61) will be compared with the RQ estimated from the flux analysis:

$$RQ_{calc} = \frac{v_1 - v_{11} + v_{12} + v_{13} + v_{14} - Y_{X/S}Y_{CO_2}}{Y_{O/S}} \tag{2.62}$$

where Y_{CO_2} is the yield of CO_2 from biosynthesis.

The specific precursor, cofactor, and CO_2 requirements for *B. subtilis* have recently been determined experimentally (Sauer *et al.*, 1996). The experimental data that will be used in this example are from Sauer *et al.* (1996) for aerobically growing *B. subtilis* in a chemostat and they include the dilution rate (which is equal to the specific growth rate), specific uptake rate (specific glucose consumption) and specific oxygen consumption rate. From parameter fitting of the experimental data the following relations have been determined for the two physiological parameters, $Y_{X/S}$ and $Y_{O/S}$, as functions of the dilution rate (D):

$$Y_{X/S} = 1000\frac{D}{0.45 + 12.3D} \tag{2.63}$$

and

$$Y_{O/S} = \frac{3 + 26.15D}{0.45 + 12.3D} \tag{2.64}$$

2.3.1 ATP Mass Balance

I first examined the system by utilizing the ATP mass balance. However, it was impossible to find any set of values for the parameters $Y_{ATP,m}$ and the (P/O) ratio that

could be kept constant and still result in low values for the residuals over the whole range of dilution rates considered. This clearly suggests that one or both of these parameters are changing as the dilution rate (specific growth rate) changes. Previous experimental studies support the concept that $Y_{ATP,m}$ is an increasing function of the specific growth rate (Nielsen *at al.*, 1991; Nielsen and Villadsen, 1994). Moreover, (P/O) ration is also function of the specific growth and the operating conditions (Senior, 1988; Nielsen and Villadsen, 1994). Therefore, without considering functional dependence of $Y_{ATP,m}$ and the (P/O) ratio on the specific growth, we cannot estimate metabolic fluxes using the ATP mass balance as a constraint.

2.3.2 NADPH Mass Balance

I next considered various assumptions on the NADPH mass balance; i.e., the estimation of the fluxes using the NADPH mass balance equation (2.47). I also considered two cases: in the first the mass balance for the CO_2 was taken into account and the calculated RQ (2.62) was compared with the experimental value from equations (2.61) and (2.63). In the second case the mass balance for the CO_2 was also considered and the experimental specific production rate of CO_2 was estimated from the RQ_{exp} (2.61). The mass balance equation for the CO_2 is:

$$0 = v_1 - v_{11} + v_{12} + v_{13} + v_{14} - Y_{X/S}Y_{CO_2} - RQ_{calc}Y_{O/S} - r_{CO_2} \tag{2.65}$$

where r_{CO_2} is the residual variable.

The analysis considered three different physiological scenarios. It was considered that the sum:

$$2v_1 + v_{13} = \phi\mu Y_{NADPH} \tag{2.66}$$

fulfils the biosynthetic requirements for NADPH exactly ($\phi = 1$), or that it fulfils 80% ($\phi = 0.8$) or 120% ($\phi = 1.2$)of the same requirement. In the latter two cases it was assumed that there are reactions to exchange the required reducing equivalents between NADPH and NADH.

The results of the analysis without considering the CO_2 mass balance are presented in Figure 2. The calculated value for the RQ is very different from the experimental one suggesting that calculations based on these assumptions are not suitable. Moreover, most of the fluxes vary in a non-monotonic way as the dilution rate incereases (Figures 2.C, 2.D, and 2.E). The trend of the fluxes changes when the flux from $G6P$ to $Ru5P$ (v_1) changes from zero to a positive value. This physiologically unexplainable behavior of the system further suggests that this approach has some inherent deficiencies. Flux v_1 is also very sensitive to the assumed value of ϕ in Equation (2.66). Depending on the value assumed, the onset of flux from $G6P$ to $Ru5P$ is estimated at different dilution rates. The most dramatic effects are observed for the assumption that $\phi = 0.8$ of the biosythetic requirements for NADPH.

In a second study the mass balance for CO_2 (2.65) was also included in solving for the fluxes. In this case, the solution was better with respect to RQ values. This was expected since the mass balance for CO_2 constrained the CO_2 production and consumption fluxes. The changes in the fluxes were monotonic over the whole range of the dilution rates. The sensitivity of v_1 with respect to the assumed value of ϕ was significantly reduced. Moreover, the calculated net ATP production from substrate level phosphorylation (Figure 3.F) was clearly dependent on this assumption, as expected, since the assumption concerns cell energetics. Notice that in the first solution (Figure 2.F) it is not clear how the assumption for the value of ϕ influences the calculated net ATP production rate.

The two solutions obtained here, one without and one with the additional constraint of the CO_2 mass balance, are compared in Figure 4 under the assumption that $\phi = 1$. Although for dilution rates higher than $0.2h^{-1}$ both solutions follow the same trend, the qualitative difference between the two solutions is significant. For a 20% difference in the RQ values, a difference of up to 50% is observed for some of the fluxes.

The above analysis leads to two important conclusions. First, the constraint of the CO_2 mass balance should be always considered. Second, the approach that considers the $NADPH$ mass balance (2.47) should be used only for obtaining qualitative

conclusions (i.e., conclusions regarding relative values and trends of the fluxes) about the metabolic system. The sensitivity of the solution to the assumption for the value of ϕ should be examined before any final conclusion can be made.

2.3.3 Reducing Equivalents Mass Balance

The approach studied next was the one that considers the mass balance of the reducing equivalents from $NADH$ and $NADPH$ (equation (2.48)). Two different studies were again performed: one without and one with the constraint of the CO_2 mass balance. The results are presented in Figure 5. The oxygen uptake rate used for the mass balance equation (2.48) as a first estimate was the same as the experimental one. This first solution suggested that the value of the flux v_{15} was approximately 15% of twice the value of the oxygen uptake rate. The problem was solved again using a value for the oxygen uptake rate of 85% of the experimentally determined one. This second solution gave again a value for the v_{15} that was approximately 15% of twice the value of the experimentally determined oxygen uptake rate over the whole range of the dilution rates. Therefore, the iteration stopped and the "corrected" value for the oxygen uptake was used in both cases (without and with the constraint for the CO_2 mass balance).

Even in the absence of the constraint of the CO_2 mass balance, the estimated RQ is very close to the experimental value (Figure 5.A). This is in contrast to the previous approaches with which, even when the constraint for the CO_2 mass balance included, the estimated RQ was significantly different from the experimental value (Figure 3.A). In general, both cases were found to be in excellent agreement with each other (Figures 5.C-5.F). However, the flux v_1 was estimated to be zero over the whole range of the dilution rates when the constraint for the CO_2 mass balance was not included, but it had a positive value when the CO_2 mass balance was considered. This difference strongly suggests that CO_2 mass balance should be included in any flux analysis. The value for v_1 in this second case was found to be approximately equal to 0.1 over the whole range of dilution rates (Figure 5.B), a value that is very close

to the biosynthetic requirements for $R5P$ and $E4P$. Therefore, it appears that the the $R5P$ and $E4P$ required for biosynthesis are coming from $G6P$ via reaction step 1. However, this solution is qualitatively very different from the one found with the approach using the NADPH mass balance. There, for low dilution rates v_1 was zero below a certain value for the dilution rate, and above that it increased monotonically up to 0.25. The fact that no assumption concerning the value of ϕ in the sum (2.66) was made in this last approach suggests that this last result is more reliable.

In order to evaluate this approach with respect to the previous one, the ratio:

$$\frac{2v_1 + v_{13}}{\mu Y_{NADPH}} = \phi$$

was calculated based on the estimated fluxes. A value for this ratio had been assumed and used as the sum (2.66) constraint in the previous approach. The calculated ratio is presented in Figure 6. For low dilution rates the ratio is higher than one suggesting that an excess of NADPH is produced. This is probably recycled back to NADH. However, for dilution rates higher than $0.15h^{-1}$, this ratio is lower than unity suggesting that the NADPH produced does not fulfill the biosynthetic requirements and that the additional amount needed is provided from the excess NADH produced. Interestingly enough, when this ratio is equal to or higher than 0.8, the value for the flux v_1 is positive, while the previous approach suggested that, as this ratio decreases, flux v_1 becomes zero.

The Euclidean norm of the estimated fluxes was also calculated for every case in the two approaches (Figures 2.D, 3.D, 4.D, and 5.D). It appears that the last approach has the lowest norm, especially when the constraint of the CO_2 mass balance is included, forcing the fluxes considered to produce the minimum amount of entropy. If our thermodynamic suggestion is right, then the last approach also satisfies also the thermodynamic optimality criteria. However, the main uncertainty of this thermodynamic consideration still holds; i.e., the question of which fluxes should be considered in calculating the Euclidean norm. Another interesting difference between the Euclidean norms of the fluxes estimated from the two approaches is their trend

with respect to dilution rates. In the first approach the norm decreases with increasing dilution rate (Figures 2.D and 3.D), whereas, in the last approach, the norm increases with increasing dilution rates. In general, the latter is more reasonable, since for increasing dilution rates the specific growth rate, the specific glucose and oxygen uptake rates all increase, indicating that the biocatalytic machinery of the cell should operate at higher rates.

Similar results were also obtained when the stoichiometric coefficient ρ for reaction 15 was considered (see equation (2.48)). The oxygen uptake rate used in the mass balance was "corrected" following the same procedure as in the last approach. For a value of ρ equal to 0.5 the oxygen uptake rate was reduced by 7%. That corresponds closely to the 16% correction for the last approach when ρ was equal to zero.

In conclusion, this last approach is more attractive since it requires the fewest assumptions. No assumption has been made with respect to energetics of the cell. The correction of the oxygen uptake rate used in the mass balance equation (2.48) was not based on a parameter-fitting approach but simply on the consistency of the value of one flux, v_{15}, with respect to the corrected value of the oxygen uptake rate.

2.4 Conclusions

Metabolic flux analysis has enjoyed a lot of attention over the last five years. Various metabolic systems have been analyzed, and useful insights resulted from those analyses. However, because most metabolic systems are underdetermined (more metabolic reactions than metabolic species), various assumptions have been used to make the systems determined, and little attention has been paid on the effects of these assumptions to the final conclusions.

In this chapter flux analyses for the bacterium *B. subtilis* were performed. A systematic algorithmic procedure was proposed that can take metabolic constraints into account. Some of the commonly used assumptions were considered and the effects of these assumptions on the resulting estimated fluxes have been studied. Two main conclusions were drawn for aerobically grown bacterial systems:

1. Any flux analysis that employs assumptions about cell energetics should be considered as a qualitative description of the *trends* of the fluxes under different growing conditions. A sensitivity analysis of the results with respect to the assumptions should always be performed.

2. The mass balance of CO_2 should be included in the analysis even though it is not linearly independent from the rest of the mass balances. It appears to improve the estimation of the fluxes by making the system less sensitive to assumptions and by integrating additional experimental information with the analysis.

A novel procedure for flux analysis has been suggested and applied to the example metabolic system. This procedure does not employ any assumptions regarding the energetics of the cell and therefore does not bias the results. Application of the procedure to the example system and comparison of the results with those of approaches that use assumptions regarding cell energetics have shown that this approach is more advantageous.

It has been suggested that the common approach of minimizing the Euclidean norm of the fluxes as an additional criterion for choosing an estimate for the metabolic fluxes from an infinite number of possible solutions when the metabolic system is underdetermined is related to thermodynamic optimality criteria based on the evolution theory. However, further investigation is required before a definite connection between these criteria and flux analysis is asserted.

Finally, it should be stressed that flux analysis is a mathematical modeling method that integrates the available biochemical knowledge in order to provide further insight on the behavior of biological systems. It provides an estimate for the values of the metabolic fluxes and can be used to compare relative changes in the fluxes under different conditions that do not significantly influence the stoichiometry of the metabolic networks, such as different growth rates or changes in enzyme amounts arising from mutation or genetic engineering. As with every mathematical method used in biotechnology, it should be used in an iterative way: the initial information it provides based on preliminary experimental data will suggest the next experimental approach. The

results of this new experiment will be used for a second mathematical analysis or even for possible reformulations of the stoichiometric model, that will again suggest the next experimental approach.

2.5 References

Aiba, S., Matsuoka, M. 1979. Identification of a Metabolic Model: Citrate Production from Glucose by *Candida lipolytica*. Biotechnol. Bioeng. **21**: 1373-1386.

Angulo-Brown, F., Santillán, M., Calleja-Quevedo, E. 1995. Thermodynamic Optimality in Some Biological Reactions. Nuovo Cimento D **17**: 87-90.

Bailey, J. E., Ollis, D. F. 1986. Biochemical Engineering Fundamentals. 2nd ed., McGraw-Hill, New York.

Bonarius, H. P. J., Hatzimanikatis, V., Meesters, K. P. H., de Gooijer, C. D., Schmid, G., Tramper, J. 1996. Metabolic Flux Analysis of Hybridoma Cells in Different Culture Media Using Mass Balances. Biotechnol. Bioeng. In press.

Gottschalk, G. 1986. Bacterial Metabolism. 2nd edition, Springer-Verlag, New York.

Grosz, R., Stephanopoulos, G., San, K.-Y. 1984. Studies on On-Line Bioreactor Identification. III. Sensitivity Problems with Respiratory and Heat Evolution Measurements. Biotechnol. Bioeng. **26**: 1198-1208.

Holms, W. H. 1986. The central metabolic pathways of *Escherichia coli*: Relationship Between Flux and Control at a Branch Point, Effeciency of Conversion to Biomass, and Excretion of Acetate. Curr. Topics Cell. Reg. **28**: 69-105.

Holms, W. H., Hamilton, I. D., Mousdale, D. 1990. Application of Flux Analysis to Increase Productivity of Fermentation Processes, pp. 1057-1062. In: C. Christiansen, L. Munck, and J. Villadsen (eds.), 5th European Congress on Biotechnology, Copenhagen, Proceedings Vol. 2. Munksgaard, Copenhagen.

Madron, F. 1979. Material Balance Calculations of Fermentation Processes. Biotechnol. Bioeng. **21**, 1487-1490.

Mancuso, A., Sharfstein, S. T., Tucker, S. N., Clark, D. S., Blanch, H. W. 1994. Examination of Primary Metabolic Pathways in Murine Hybridoma with Carbon-13 Nuclear Magnetic Resonance Spectroscopy. Biotechnol. Bioeng. **44**: 563-585.

Marx, A., de Graff, A. A., Wiechert, W., Eggeling, L., Sahm, H. 1996. Determination of the Fluxes in the Central Metabolism of *Corynebacterium glutamicum* by

Nuclear Resonance Spectroscopy Combined with Metabolic Balancing. Biotechnol. Bioeng. **49**: 111-129.

Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., Umbarger, H. E. 1987. *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, vols 1 and 2. American Society for Microbiology, Washington, D.C.

Nielsen, J., Villadsen, J. 1994. Bioreaction Engineering Principles. Plenum Press, New York.

Nielsen, J., Nikolajsen, K., and Villadsen, J. 1991. Structured Modeling of a Microbial System II. Experimental Verification of a Structured Lactic Acid Fermentation Model. Biotechnol. Bioeng. **38**: 11-23.

Noorman, H. J., Heijnen, J. J., Luyben, K. Ch. A. M. 1991. Linear Relations in Microbial Reaction Systems: A General Overview of Their Origin, Form, and Use. Biotechnol. Bioeng. **38**:603-618.

Papoutsakis, E. T. 1984. Equations and Calculations for Fermentations of Butyric Acid Bacteria. Biotechnol. Bioeng. **26**: 174-187.

Prigogine, I. 1961. Thermodynamics of Irreversible Processes. John Wiley & Sons, New York.

Reitzer, L. J., Wice, B. M., Kennel, D. 1980. The Pentose Cycle. J. Biol. Chem. **255**: 5616-5262.

Roels, J. A. 1983. Energetics and Kinetics in Biotechnology. Elsevier Biomedical Press, Amsterdam.

Sauer, U., Hatzimanikatis, V., Hohmann, P., Manneberg, M., van Loon, A. P. G. M., Bailey, J. E. 1996. Determination of Physiological Parameters and Analysis of Metabolic Fluxes in Wild-Type and Riboflavin-Producing *Bacillus subtilis*. Submitted for publication.

Savageau, M. A. 1976. Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Addison–Wesley, Reading, MA.

Savinell, J. M., Palsson, B. O. 1992a. Network Analysis of Intermediary Metabolism Using Linear Optimization. II. Interpretation of Hybridoma Cell Metabolism. J.

Theor. Biol. **154**: 455-473.

Savinell, J. M., Palsson, B. O. 1992b. Optimal Selection of Metabolic Fluxes for *In vivo* Measurement. I. Development of Mathematical Methods. J. Theor. Biol. **155**: 201-214.

Savinell, J. M., Palsson, B. O. 1992c. Optimal Selection of Metabolic Fluxes for *In vivo* Measurement. II. Application to *Escherichia coli* and Hybridoma Cell Metabolism. J. Theor. Biol. **155**: 215-242.

Senior, A. E. 1988. ATP Synthesis by Oxidative Phosphorylation. Physiol. Rev. **68**: 177-231.

Sonenshein, A. L., Hoch, J. A., Losick, R. 1993. *Bacillus subtilis* and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology, Washington, D.C.

Stephanopoulos, G., San, K.-Y. 1984. Studies on On-Line Bioreactor Identification. I. Theory. Biotechnol. Bioeng. **26**: 1176-1188.

Stephanopoulos, G., Vallino, J. J. 1991. Network Rigidity and Metabolic Engineering in Metabolite Overproduction. *Science*: 1675-1681.

Stewart, G. W. 1973. Introduction to Matrix Computations. Academic Press, San Diego.

Stryer, L. 1988. Biochemistry. 3rd edition. W. H. Freeman, New York.

Torres, J.-L. 1991. Natural Selection and Thermodynamic Optimality. Nuovo Cimento D **13**: 177-185.

Tsai, P. S., Hatzimanikatis, V., Bailey, J. E. 1996. Effect of *Vitroescilla* Hemoglobin Dosage on Microaerobic *Escherichia coli* Carbon and Energy Metabolism. Biotechnol. Bioeng. **49**: 139-150.

van Gulik, W. M., Heijnen, J. J. 1995. A Metabolic Network Stoichiometry Analysis of Microbial Growth and Product Formation. Biotechnol. Bioeng. **48**: 681-698.

Varma, A., Palsson, B. O. 1994. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Bio/Technology **12**: 994-998.

Varma, A., Palsson, B. O. 1995. Parametric Sensitivity of Stoichiometric Flux Balance Models Applied to Wild-Type *Escherichia coli* Metabolism. Biotechnol. Bioeng.

45: 69-79.

Walsh, K., Koshland, Jr., D. E. 1984. Determination of Flux Through the Branch Point of Two Metabolic Cycles. J. Biol. Chem. **257**: 1189-1195.

Wood, T. 1985. The Pentose Phosphate Pathway. Academic, Orlando, FL.

2.6 Figures

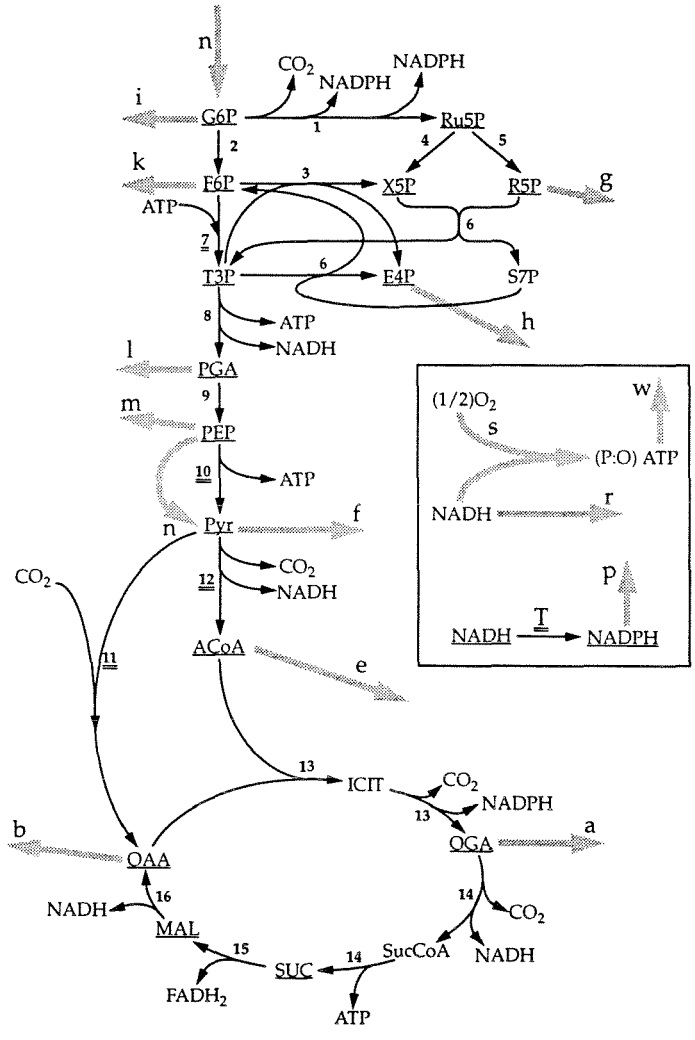

Figure 2.1: The reaction network of the central carbon pathways for aerobic growth without product formation of *B. subtilis* considered in the analysis. The metabolites for which the mass balances were used are underlined. Shaded arrows indicate fluxes to biosynthesis. Double-underlined numbers indicate irreversible reaction steps.

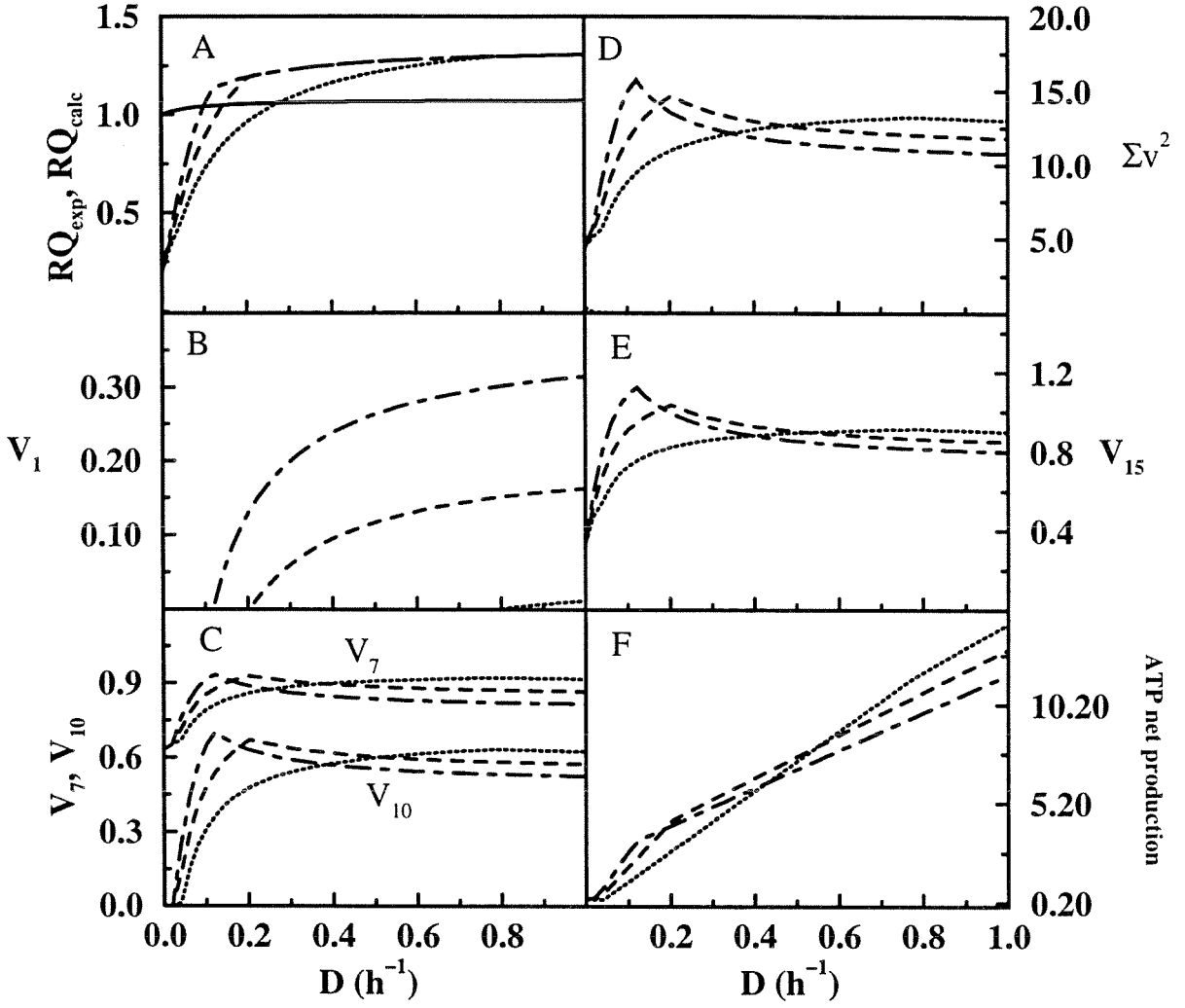

Figure 2.2: Flux analysis results for various dilution rates without considering the CO_2 mass balance and under the assumption that $2v_1 + v_{13} = \phi\mu Y_{NADPH}$, where: dotted line ($\phi = 0.8$), dashed line ($\phi = 1$), and dashed-dotted line ($\phi = 1.2$). Solid line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production.

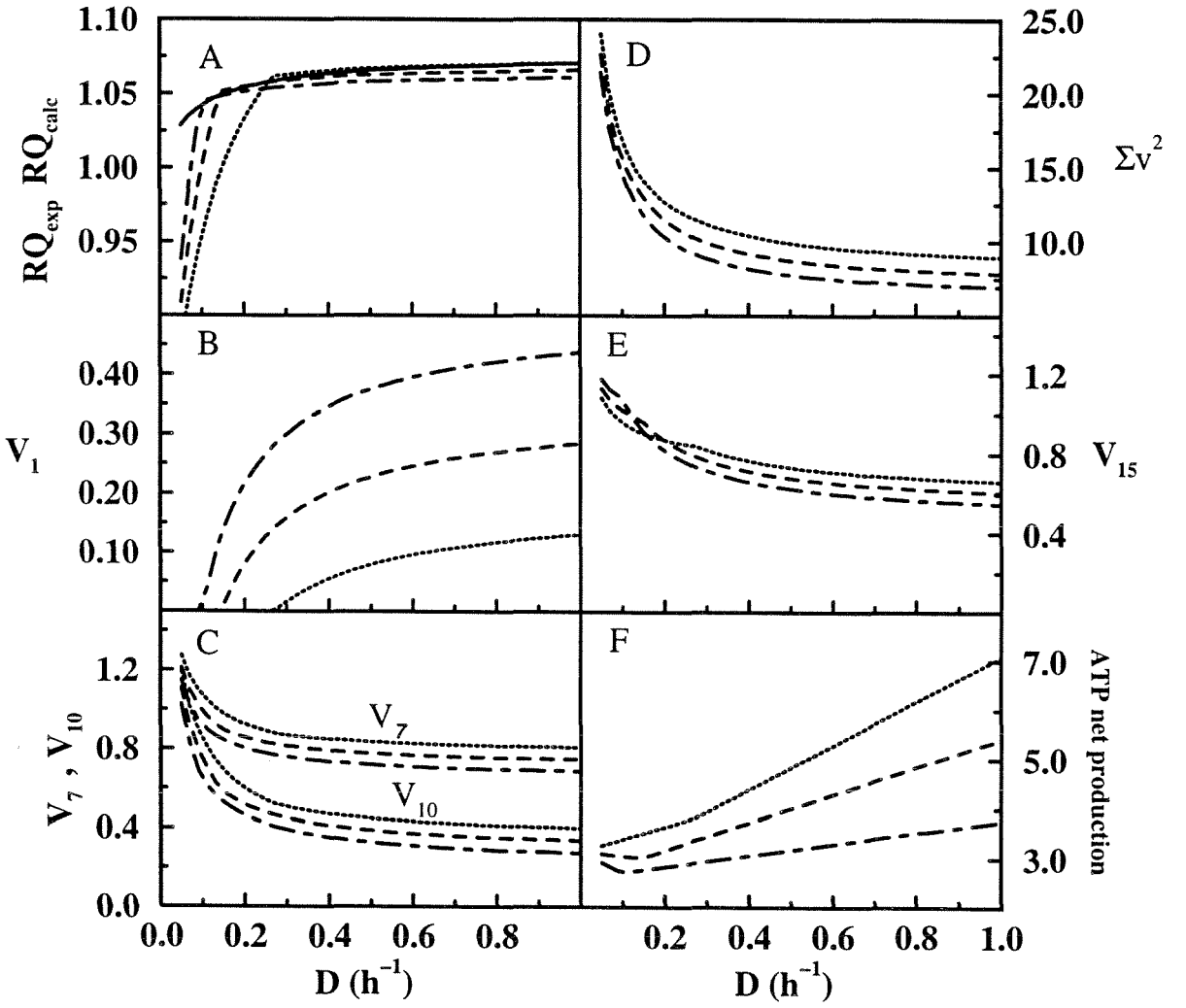

Figure 2.3: Flux analysis results for various dilution rates including the CO_2 mass balance and under the assumption that $2v_1 + v_{13} = \phi\mu Y_{NADPH}$, where: dotted line ($\phi = 0.8$), dashed line ($\phi = 1$), and dashed-dotted line ($\phi = 1.2$). Solid line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production.

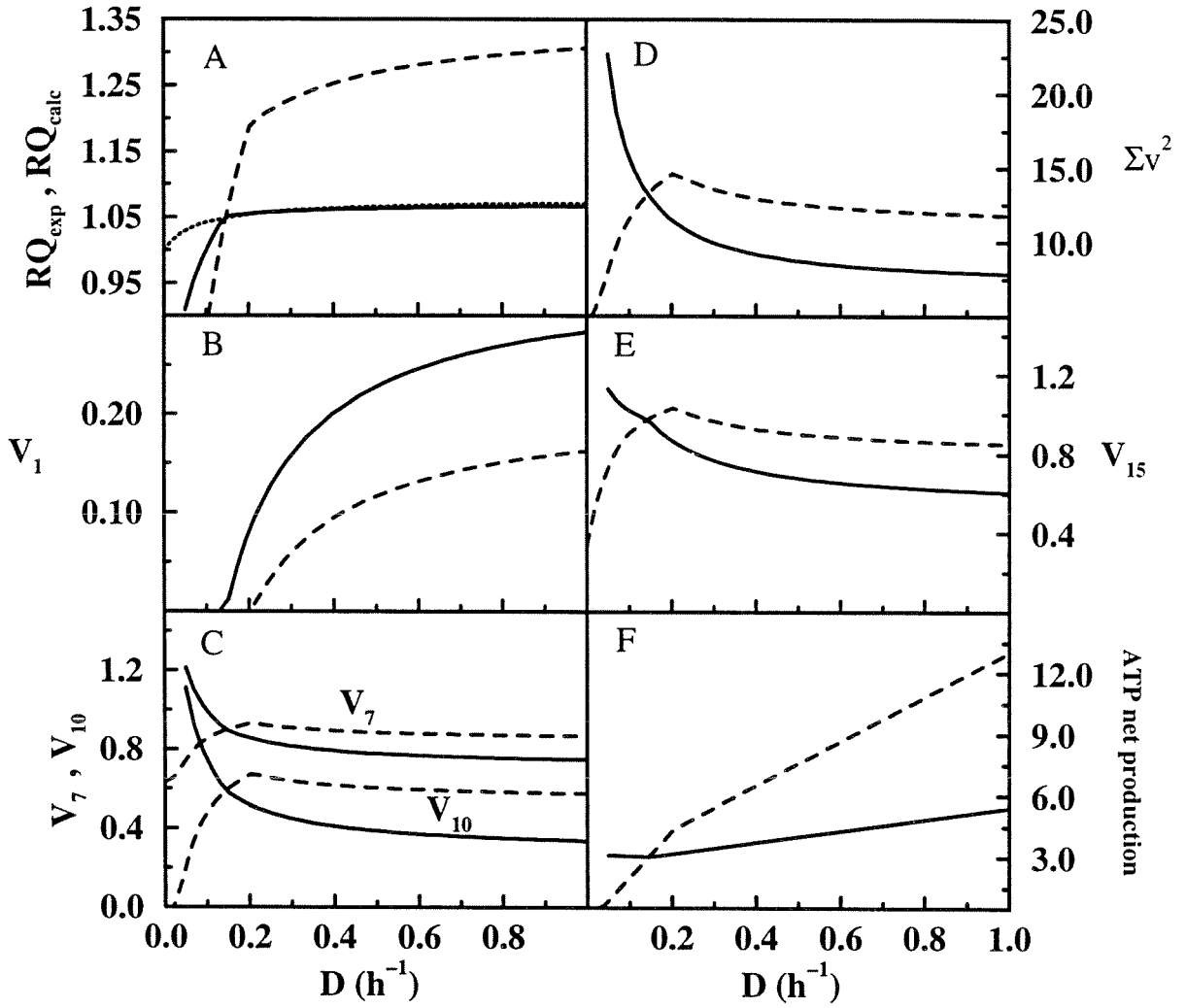

Figure 2.4: Flux analysis results comparison for various dilution rates under the assumption that $2v_1 + v_{13} = \mu Y_{NADPH}$, including the CO_2 mass balance (solid line) and without considering the CO_2 mass balance (dashed line). Dotted line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production.

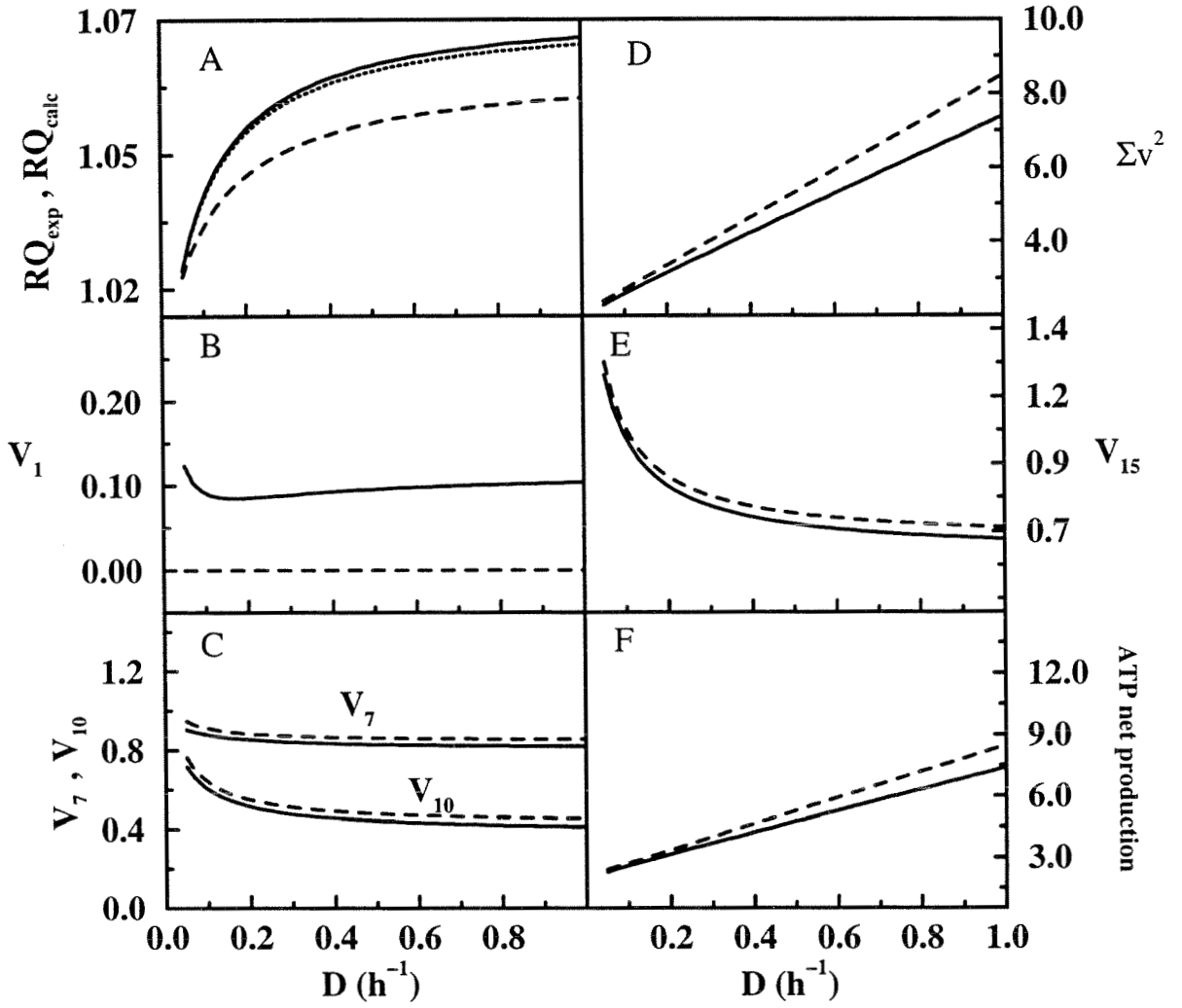

Figure 2.5: Flux analysis results comparison for various dilution rates without considering $FADH_2$ in the mass balance for the reducing equivalents and for "corrected" value for the oxygen uptake rate including the CO_2 mass balance (solid line) and without considering the CO_2 mass balance (dashed line). Dotted line in A for the experimental value of RQ. A: The experimental value for RQ and the value based on the estimated fluxes, B: Estimated values for reaction step 1, C: Estimated values for reaction steps 7 and 10, D: The Euclidean norm of the estimated fluxes, E: Estimated values for reaction step 15, and F: Estimated values for net ATP production.

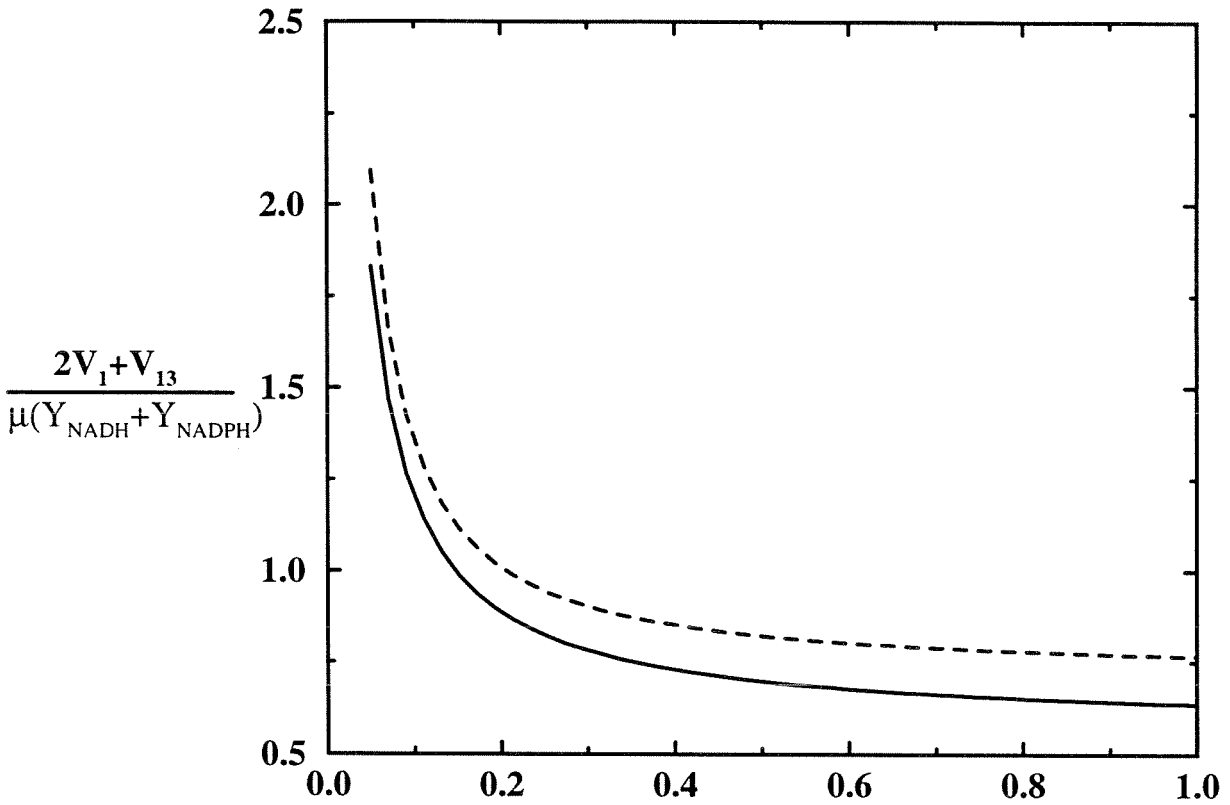

Figure 2.6: The values of the ratio $(2v_1+v_{15})/(\mu Y_{NADPH})$ for various dilution rates, as it was estimated from the analysis without considering $FADH_2$ in the mass balance for the reducing equivalents and for "corrected" value for the oxygen uptake rate without (dashed line) and with (solid line) the constraint for the CO_2 mass balance.

Chapter 3 Effects of Spatiotemporal Variations on Metabolic Control: Approximate Analysis Using (Log)Linear Kinetic Models

3.1 Introduction

The analysis and study of the responses of metabolic systems to process and to genetic manipulations have been the primary focus of numerous experimental and theoretical studies (Cornish-Bowden and Cárdenas, 1990; Bailey, 1991; Stephanopoulos and Vallino, 1991; Fell, 1992). The theoretical studies have mostly focused on the development of quantitative descriptions of metabolism and to associated theory and analytical frameworks. Metabolic control analysis (MCA), a sensitivity analysis framework, is perhaps the most developed of these methods for the quantitative description of metabolism and microbial physiology (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Kell and Westerhoff, 1986; Schlosser and Bailey, 1990; Cornish-Bowden and Cárdenas, 1990; Rutgers *et al.*, 1991; Fell, 1992; Brown, 1992; Schlosser *et al.*, 1993). Accordingly, several experimental methodologies have been developed to allow determination of quantitative indices which are defined by MCA, such as control coefficients and elasticities. The description of metabolic systems by MCA data is commonly used because rarely is sufficient information available to formulate a nonlinear mathematical description based on detailed enzyme kinetics. Furthermore, when such a nonlinear model is available, it can be linearized and well–studied within the same MCA framework.

One of the basic assumptions embedded in MCA is that the metabolic system under study is at steady-state. Every parameter of the system (for example, enzyme expression levels, external substrate concentrations and independent effector concentrations or activities) is assumed to remain time-invariant. However, this assumption is not valid when spatiotemporal variations in the parameters occur. Such spatiotemporal variations can arise, for example, in bioreactors when mixing is nonideal, a situation common in large-scale systems. Circulation of cells through spatially inhomogeneous fields of dissolved oxygen and substrate concentration, pH, temperature, and hydrodynamic conditions drives unsteady-state responses in metabolism which have profound effects on scale-up nad large-scale bioreactor prerformance. These phenomena have been modelled for bioreactors and the effects of nonideal mixing on

the metabolism and growth of microorganisms have also been studied. Spatiotemporal variations in the parameters or inputs of a system are known to result in a performance quantitatively and qualitatively different from the performance of the system with space- and time-invariant parameters (Bailey, 1974, 1977). For example, it has been recently reported that control coefficients for cells growing in a nonideally mixed bioreactor are different from the control coefficients calculated for the ideal well-mixed case.

In this chapter we develop a *(log)linear kinetic model* of metabolic systems based on MCA data. This model can be used to simulate *dynamic* responses of the system with spatiotemporal variations in its parameters. Interconnection of MCA with alternative modeling frameworks allows dynamic response analysis. An important prior formulation of metabolic kinetics is the S-system representation developed within the biochemical systems theory (BST). However the S-system representation is a nonlinear system and BST explicitly employ MCA parameters. The major advantages of the (log)linear model presented in this chapter are the linearity (in terms of logarithms) of the model and the specification of the system dynamics using the same parameters as employed in MCA. Exploiting the linear nature of the model, we present a simple procedure for the identification of the effects of the period and the waveform of a periodic spatiotemporal variation of the parameters on the average metabolic functions and their control coefficients.

3.2 Mathematical Description of Metabolic Reaction Networks

3.2.1 Development of a (Log)Linear Kinetic Model

Consider a metabolic system consisting of n metabolites and m enzymatically-catalyzed reactions. We are interested in studying how modifications of the expression levels and of the properties of the enzymes that catalyze these reactions affect the time response characteristics of metabolic functions of the system, such as metabolite concentra-

tions, fluxes, and the specific growth rate. The mass balances on the metabolites of the system may be written as:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{v}(\mathbf{x};\mathbf{p}),\mathbf{x};\mathbf{p}) \tag{3.1}$$

where $\mathbf{x}$ is the n-dimensional metabolite concentration vector, $\mathbf{f}$ is a function determined by the mass balances, $\mathbf{v}$ is the m-dimensional reaction rate vector, and $\mathbf{p}$ is the s-dimensional manipulated parameter vector (e.g., extracellular substrate concentration). In addition to metabolite reaction rates, the mass balance equations also include terms that account for other processes by which concentrations of metabolites change (such as the dilution brought about by increases in the biomass volume (Fredrickson, 1976) and transport through the cell wall envelope).

If we consider temporal variations in the parameters then $\mathbf{p}$ will be a function of space coordinates and time. In many systems with spatial gradients (for example, a bioreactor with internal fluid circulation, or a plug-flow bioreactor), description of system changes in space can be transformed into a description in terms of a circulation or residence time, again resulting in ordinary differential equations in the form of Equation (3.1). Here the temporal dependence of the parameters is indicated explicitly:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{v}(\mathbf{x};\mathbf{p}(t)),\mathbf{x};\mathbf{p}(t),t) \tag{3.2}$$

In addition, consider the r-dimensional vector of metabolic outputs, $\mathbf{h}$, which we can be written in general:

$$\mathbf{h} = \mathbf{h}(\mathbf{v}(\mathbf{x};\mathbf{p}(t)),\mathbf{x};\mathbf{p}(t),t) \tag{3.3}$$

In equation (4.2) $\mathbf{h}$ is a function of the rates of interest, of the metabolite concentration, and of the parameters. Let $\mathbf{x}_o$ be a steady state – or one of the steady states in the case of steady state multiplicity – that corresponds to the given parameter vector $\mathbf{p}_o$, and let $\mathbf{x}_o$ be nonzero with positive elements. Linearization around this steady

state results in the linear system:

$$\frac{d(\mathbf{x}-\mathbf{x}_o)}{dt} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}(\mathbf{x}-\mathbf{x}_o) + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o}(\mathbf{p}-\mathbf{p}_o) \tag{3.4}$$

for the mass balances, and

$$\mathbf{h} = \mathbf{h}(\mathbf{v}_o(\mathbf{x}_o;\mathbf{p}_o),\mathbf{x}_o;\mathbf{p}_o,0) + \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o,0}(\mathbf{x}-\mathbf{x}_o) + \left.\frac{\partial \mathbf{h}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o,0}(\mathbf{p}(t)-\mathbf{p}_o) \tag{3.5}$$

for the metabolic outputs.

If we define the matrices $\mathbf{X}_o$ and $\mathbf{P}_o$ to be diagonal matrices with diagonal elements $X_{o,ii} = x_{o,i}$ and $P_{o,ii} = p_{o,i}$, respectively, then, we can write:

$$\mathbf{X}_o^{-1}\frac{d(\mathbf{x}-\mathbf{x}_o)}{dt} = \mathbf{X}_o^{-1}\left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o\mathbf{X}_o^{-1}(\mathbf{x}-\mathbf{x}_o) + \mathbf{X}_o^{-1}\left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{P}_o\mathbf{P}_o^{-1}(\mathbf{p}-\mathbf{p}_o) \tag{3.6}$$

for the linear system.

Now it is useful to define new variables which are the logarithmic deviations of the systems state variables and parameters:

$$z_i = ln\left(\frac{x_i}{x_{o,i}}\right) \tag{3.7}$$

and

$$q_i = ln\left(\frac{p_i}{p_{o,i}}\right) \tag{3.8}$$

Given the above definitions we further observe that:

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o = \left.\frac{\partial \mathbf{f}}{\partial(\mathbf{x}-\mathbf{x}_o)}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o \Rightarrow \left.\frac{\partial f_i}{\partial(x_j-x_{o,j})}x_{o,j}\right|_{\mathbf{x}_o,\mathbf{p}_o} = \left.\frac{\partial f_i}{\partial z_j}\right|_{\mathbf{x}_o,\mathbf{p}_o} \tag{3.9}$$

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{P}_o = \left.\frac{\partial \mathbf{f}}{\partial(\mathbf{p}-\mathbf{p}_o)}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{P}_o \Rightarrow \left.\frac{\partial f_i}{\partial(p_j-p_{o,j})}p_{o,j}\right|_{\mathbf{x}_o,\mathbf{p}_o} = \left.\frac{\partial f_i}{\partial q_j}\right|_{\mathbf{x}_o,\mathbf{p}_o} \tag{3.10}$$

A key approximation used in the development of (log)linear models is the following Taylor expansion of the logarithm about a reference value at y_o, which, to first-order

terms, is::

$$ln(y) \cong ln(y_o) + \frac{y - y_o}{y_o} \Rightarrow ln\left(\frac{y}{y_o}\right) \cong \frac{y - y_o}{y_o} \tag{3.11}$$

Finally, by defining

$$g_i = \frac{f_i}{x_{o,i}} \tag{3.12}$$

we can write the linearized system for the logarithmic deviations, noting that $[\mathbf{z}_o, \mathbf{q}_o] = [\mathbf{0}, \mathbf{0}]$:

$$\frac{d\mathbf{z}}{dt} = \left.\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right|_{0,0} \mathbf{z} + \left.\frac{\partial \mathbf{g}}{\partial \mathbf{q}}\right|_{0,0} \mathbf{q} \tag{3.13}$$

Similarly, by linearizing equation (4.2) around the same steady state and by defining

$$w_l = ln(h_l/h_{l,o}) \tag{3.14}$$

we can write the following linear equation for the logarithmic deviations of the metabolic outputs:

$$\mathbf{w} = \left.\frac{\partial \mathbf{w}}{\partial \mathbf{z}}\right|_{0,0} \mathbf{z} + \left.\frac{\partial \mathbf{w}}{\partial \mathbf{q}}\right|_{0,0} \mathbf{q} \tag{3.15}$$

In order to describe the dependence of the metabolic system explicitly on the reaction rates and on the rest of the metabolic processes the following final representation is adopted:

$$\frac{d\mathbf{z}}{dt} = \mathbf{N}\mathcal{E}\mathbf{z} + \mathcal{K}\mathbf{z} + \mathbf{N}\Pi\mathbf{q} + \mathbf{\Lambda}\mathbf{q} \tag{3.16}$$

$$\mathbf{w} = \Xi\mathcal{E}\mathbf{z} + \mathcal{H}\mathbf{z} + \Xi\Pi\mathbf{q} + \Theta\mathbf{q} \tag{3.17}$$

where, $\mathbf{z}$, $\mathbf{q}$, and $\mathbf{w}$, are the logarithmic deviations of the metabolite concentrations, the enzyme levels, and the metabolic outputs, respectively:

$$z_i = ln(x_i/x_{i,o})$$

$$q_k = ln(p_k/p_{k,o})$$

$$w_l = ln(h_l/h_{l,o})$$

and $\mathbf{N}$, $\mathbf{\Xi}$, $\mathcal{K}$, $\mathbf{\Lambda}$, $\mathbf{H}$, $\mathbf{\Theta}$, $\mathcal{E}$, and $\mathbf{\Pi}$, are matrices, defined as:

$$\mathbf{N} = \{n_{i,j} \mid n_{i,j} = \frac{v_{j,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial v_j}\right)_{\mathbf{x}_o,\mathbf{p}_o} \} \quad , \quad \mathbf{\Xi} = \{\xi_{l,j} \mid \xi_{l,j} = \frac{v_{j,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial v_j}\right)_{\mathbf{x}_o,\mathbf{p}_o} \},$$

$$\mathcal{K} = \{\kappa_{i,k} \mid \kappa_{i,k} = \frac{x_{k,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial x_k}\right)_{\mathbf{x}_o,\mathbf{p}_o} \} \quad , \quad \mathbf{\Lambda} = \{\lambda_{i,k} \mid \lambda_{i,k} = \frac{p_{k,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o} \},$$

$$\mathbf{H} = \{\eta_{l,i} \mid \eta_{l,i} = \frac{x_{i,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial x_i}\right)_{\mathbf{x}_o,\mathbf{p}_o} \} \quad , \quad \mathbf{\Theta} = \{\theta_{l,k} \mid \theta_{l,k} = \frac{p_{k,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o} \},$$

$$\mathcal{E} = \{\epsilon_{j,i} \mid \epsilon_{j,i} = \frac{x_{i,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_i}\right)_{\mathbf{x}_o,\mathbf{p}_o} \} \quad , \quad \mathbf{\Pi} = \{\pi_{j,k} \mid \pi_{j,k} = \frac{p_{k,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o} \} \tag{3.18}$$

At steady-state, the solution of (3.16) and (3.17) yields:

$$\mathbf{w} = \mathbf{C}\mathbf{q} \tag{3.19}$$

where

$$\mathbf{C} = -\left(\mathbf{\Xi}\mathcal{E} + \mathcal{H}\right)\left(\mathbf{N}\mathcal{E} + \mathcal{K}\right)^{-1}\left(\mathbf{N}\mathbf{\Pi} + \mathbf{\Lambda}\right) + \mathbf{\Xi}\mathbf{\Pi} + \mathbf{\Theta} \tag{3.20}$$

with

$$\mathbf{C} = \{c_{l,k} \mid c_{l,k} = \frac{p_{k,o}}{h_{l,o}} \left(\frac{dh_l}{dp_k}\right)_{\mathbf{p}_o} \}$$

The mathematical description presented above depends explicitly on the same information as that employed within the framework of metabolic control analysis (MCA) (Reder, 1988; Schlosser and Bailey, 1990). Matrices $\mathcal{E}$ and $\mathbf{\Pi}$ are the elasticity matrices with respect to metabolites and to parameters, respectively. The matrix $\mathbf{C}$ is the control coefficient matrix of the metabolic functions $\mathbf{h}$ with respect to parameters $\mathbf{p}$.

We should notice here that the final representation (equations (3.16) and (3.17)) allows the explicit description of the system with quantities that are characteristic of each enzyme. Therefore, we could study the effects of modifications of the catalytic properties of enzyme i with respect to its substrate (or regulatory effector) j, by changing the value of the corresponding elasticity ϵ_{ij}. Henceforth the above

description of metabolic systems will be called a *(log)linear metabolic model.*

The (log)linear metabolic model is an approximate *linear* description of nonlinear models for metabolic systems that shares common properties with the approximate *nonlinear* S-system representation. While they both give the same steady-state solution and they have the same local stability characteristics at the reference steady states, their dynamic responses are in general different, and the S-system dynamic representation is nonlinear, and therefore they can be solved only numerically, whereas, the (log)linear models are linear and thus, as it will be shown in the following section, they can be analytically integrated. With respect to their parameters, the (log)linear model uses explicitly, the same parameters as used in MCA, while S-system models use different parameters, which can be derived from MCA parameters.

3.2.2 Analytical Solution of the (Log)Linear Model

A crucial advantage of the (log)linear model is the linearity of its unsteady-state equations which allows relatively straightforward analysis. In particular, the solution of equation (3.16) may be written explicitly (Seinfeld and Lapidus, 1974):

$$\mathbf{z}(t) = e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-t_o)}\mathbf{z}_o + \int_{t_o}^{t} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-\tau)}(\mathbf{N}\Pi + \Lambda)\mathbf{q}(\tau)d\tau \tag{3.21}$$

where

$$\mathbf{z}_o = \mathbf{z}(t_o)$$

The time dependence of the logarithmic deviations of the metabolic outputs can then be calculated directly from equations (3.17) and (3.21)

$$\mathbf{w}(t) = (\Xi\mathcal{E}+\mathcal{H})(e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-t_o)}\mathbf{z}_o + \int_{t_o}^{t} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-\tau)}(\mathbf{N}\Pi + \Lambda)\mathbf{q}(\tau)d\tau) + (\Xi\Pi + \Theta)\mathbf{q}(t) \tag{3.22}$$

Numerical calculation of the integral in equations (3.21) and (3.22) is simple and rapid.

For the time-dependent metabolite concentrations and metabolic outputs, we sim-

ply calculate the inverse transformations of the equations (3.7) and (3.14):

$$x_i(t) = x_{o,i}e^{z_i(t)} \tag{3.23}$$

and

$$h_l(t) = h_{o,l}e^{w_l(t)} \tag{3.24}$$

which can also be written as

$$h_l = h_{o,l}\prod_{k=1}^{n}\left(\frac{x_k(t)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=1}^{s}\left(\frac{p_j(t)}{p_{o,j}}\right)^{\beta_{lj}} \tag{3.25}$$

where

$$\alpha_{lk} = (\sum_{i=1}^{n}\xi_{li}\epsilon_{ik} + \eta_{lk}) \tag{3.26}$$

and

$$\beta_{lj} = (\sum_{i=1}^{n}\xi_{li}\pi_{il} + \theta_{lj}) \tag{3.27}$$

Equations (3.21)-(3.27) account for the time dependence of metabolites and metabolic outputs based on MCA data and thus substantially extend the scope of applications of MCA data.

3.3 Accuracy of the (Log)Linear Metabolic Model

In this section three examples will be presented in which the dynamic responses of small metabolic pathways are described using alternative modeling representations. The following examples should be considered as illustrations of how accurately the (log)linear model can describe metabolic systems, but they are not presented for a strict quantitative comparison. The good agreement between the (log)linear model and the full nonlinear model available for these examples will suggest that, in the absence of a nonlinear model and when MCA data can be experimentaly determined or estimated, the (log)linear model can be used as a good first approximation for analysis of the dynamic response characteristics of metabolic systems.

3.3.1 Linear Pathway

The pathway considered in this example is a simple linear pathway (Figure 1). Michaelis-Menten kinetics were considered for every reaction rate and the parameters used are presented in Appendix A. The (log)linear model parameters were derived as described above. We studied the dynamic responses of the flux through step 4, to the final product P, for step changes (Figure 2.A) and for sinusoidal variation (Figures 2.B, 2.C, and 2.D) of the input flux, v_o.

As shown in Figures 2.A and B, the time response characteristics of the flux through the pathway are in very good agreement between the original nonlinear model and the approximate (log)linear model. The final steady-state differences for step responses (Figure 2.A) are the differences between the MCA calculations of new steady states and steady states of the original system. The question of the sensitivity of a system either to changes in its parameters of to fluctuations of the concentrations of cellular metabolites has been addressed and studied using the simple linearized model of the system; i.e., the system described by equations (3.4) and (3.5) (Kahn and Westerhoff, 1993). However, the (log)linear model is in much better agreement with the original system model than is a simple linear molel, especially with respect to response to periodic inputs (Figure 2.B, 2.C and 2.D). The time response characteristics of the concentrations of the metabolites were also found to be in excellent agreement (results not shown). Extensive analysis for different parameters showed that the (log)linear model can perform very satisfactory in approximating the linear pathway with reactions which follow nonlinear kinetics (results not shown).

3.3.2 Branched Pathway

The branched pathway depicted in Figure 3 was also considered. A regulatory structure, typical for branched pathways, was included (see the dashed arrows in Figure 3). The kinetics assumed for the pathway are presented in Appendix A.II. It is interesting to notice that the regulatory structure introduces nonlinearities which make the model more complicated than the one for the linear pathway. The metabolic output

considered in the simulation studies is the ratio of the flux through step 2 divided by the sum of the fluxes through steps 2 and 3:

$$R = \frac{v_2}{v_2 + v_3} \tag{3.28}$$

The dynamic responses of the ratio for step changes and for sinusoidal variation of the input flux, v_o, are presented in Figures 4.A and 4.B, respectively.

The agreement between the (log)linear model and the nonlinear model is very good, especially if we consider the complexity arising from the branching of the input flux and from the regulatory coupling. The differences between steady states, especially for large input changes, are indicative of the limits of MCA for calculating new steady states after large parameter changes.

3.3.3 Glycolytic Pathway

As a last example, the glycolytic pathway from yeast (Figure 6) was studied. The kinetics of the enzymes of the pathway are known to exhibit strong nonlinearities (see Appendix A.III for the kinetics considered). The flux to ethanol was the metabolic output considered, and responses of this flux to changes in glucose uptake rate were studied as in the previous examples. The simulation results, presented in Figures 7.A and 7.B, illustrate the excellent agreement between the nonlinear and the (log)linear models. As we can see from the kinetics in Appendix A.III, the nonlinearities are among the most complicated in enzyme kinetics, and the agreement between the two models is still excellent. Moreover, the (log)linear model is able to describe the overshooting (undershooting) initial response of the flux, as presented in Figure 7.C.

In conclusion, the (log)linear model is an attractive representation of metabolic pathways since it can be easily constructed directly from MCA data, it has an analytical solution which makes computation easier, and it can describe quite accurately the dynamic response characteristics of the pathways. Moreover, the analytical solution of the (log)linear model allows the derivation of indices such as the regulatory strength and the homeostatic strength (Kahn and Westerhoff, 1993) by simple manipulations

of equations (3.21)-(3.25).

3.4 Averaged Metabolic Functions and Control Coefficients for Periodic Parameter Variations

Suppose that $\mathbf{h}_o$ is the metabolic output corresponding to the steady state with parameters $\mathbf{p}_o$. If the parameters change with time in a periodic fashion, with period T, i.e.:

$$\mathbf{p}(t) = \mathbf{p}(t+T) \tag{3.29}$$

the metabolite concentrations and the metabolic outputs will usually, after some time, closely approximate periodic functions with the same period:

$$\mathbf{x}(t) = \mathbf{x}(t+T) \tag{3.30}$$

and

$$\mathbf{h}(t) = \mathbf{h}(t+T) \tag{3.31}$$

Suppose that $\mathbf{p}$ also varies with time so that its average value, $\overline{\mathbf{p}}$, is the same as $\mathbf{p}_o$:

$$\overline{\mathbf{p}} = \mathbf{p}_o \tag{3.32}$$

Then, if the metabolic system is nonlinear, the time-average value of the output

$$\overline{h} \equiv \frac{1}{T}\int_{nT}^{(n+1)T} \mathbf{h}(\tau)d\tau \tag{3.33}$$

will, in general, be *different from* $\mathbf{h}_o$:

$$\overline{\mathbf{h}}_o \neq \mathbf{h}_o \tag{3.34}$$

Extensive theoretical, computational, and experimental studies showing this have appeared in the chemical reaction engineering literature (Douglas, 1967; Bailey and

Horn, 1971; Bailey, 1974, 1977; Guardabassi, 1974; Lyberatos and Svoronos, 1987).

This implies that, if some parameter, say p_1, of the metabolic system varies periodically with time with mean equal to its reference steady-state value:

$$p_{o,1} = \overline{p}_1 = \frac{1}{T}\int_{nT}^{(n+1)T} p_1(\tau)d\tau \tag{3.35}$$

the effect of another parameter, say p_2, on the time-average metabolic outputs will, in general, be *different from* the effect of that parameter on the steady-state metabolic outputs. Stated in terms of control coefficients, this means that the average output control coefficient (AOCC), defined as:

$$C_{p_2}^{\overline{h}_l} \equiv \frac{p_{o,2}}{\overline{h}_l}\frac{d\overline{h}_l}{dp_2} \tag{3.36}$$

are, in a system subjected to time-periodic inputs, different from the corresponding steady state control coefficients $C_{p_r}^{h_l}$.

When we study metabolic systems subject to periodic variation of their parameters, many questions arise: How does the average metabolic function of interest depend on the period and the waveform of the variation? What is the optimal variation pattern? How are the responses to genetic manipulations of time-invariant parameters affected by the time variation of another parameter? The (log)linear model developed above can be used to find an approximate answer to these questions.

Equation (3.36) can be derived by differentiation of equation (3.33) with respect to p_2:

$$C_{p_2}^{\overline{h}_l} = \frac{p_{o,2}}{\overline{h}_l T}\int_0^T \left[\sum_i \frac{\partial h_l(\tau)}{\partial x_i}\frac{\partial x_i(\tau)}{\partial p_2} + \frac{\partial h_l(\tau)}{\partial p_2}\right] d\tau \tag{3.37}$$

which after rearrangement becomes:

$$C_{p_2}^{\overline{h}_l} = \frac{1}{T}\int_0^T \left[\sum_i \underline{\frac{x_{o,i}}{\overline{h}_l}\frac{\partial h_l(\tau)}{\partial x_i}}\left(\underline{\frac{p_{o,2}}{x_{o,i}}\frac{\partial x_i(\tau)}{\partial p_2}}\right) + \underline{\frac{p_{o,2}}{\overline{h}_l}\frac{\partial h_l(\tau)}{\partial p_2}}\right] d\tau \tag{3.38}$$

The first and third underlined terms in Equation (3.38) can be approximated using equations (3.21)-(3.27) of the (log)linear model. The second underlined term can be

approximated from the concentration control coefficient with respect to p_2 by differentiation of $\mathbf{z}(t)$ (Equation (3.21)) and, as shown in Appendix B, it is time-invariant and equal to the concentration control coefficient of the time-invariant steady-state, regardless of the periodic input.

Lyberatos and Svoronos (1987) developed a method for the study and optimization of pulsed periodically forced (Figure 7) linear systems. Their method has been shown to be very useful for the study of forced chemical reaction systems and the estimation of the optimum period and waveform of forcing (Hatzimanikatis *et al.*, 1993). Their method seems to be appropriate for the study of metabolic systems since the (log)linear model developed in the previous section has been shown to capture many of the most important dynamic and steady-state response characteristics of the nonlinear metabolic model.

3.4.1 Mathematical Framework

We will consider again a metabolic system consisting of n metabolites, $\mathbf{x}$, m enzymatically catalyzed reactions, $\mathbf{v}$ and s parameters, $\mathbf{p}$. One of the parameters, p_1, is varying in a pulsed periodic mode (Figure 7):

$$p_1(t) = \overline{p}_1 + \begin{cases} \delta & t \in [jT, (j+\epsilon)T] \\ \rho = \frac{\epsilon\delta}{\epsilon - 1} & t \in [(j+\epsilon)T, (j+1)T] \end{cases} \quad j = 0, 1, 2, \ldots \tag{3.39}$$

The magnitude of ρ is set at $\epsilon\delta/(\epsilon - 1)$ so that the p_1 average is $\overline{p}_1$, i.e.

$$\frac{1}{T}\int_0^T p_1(\tau)d\tau = \overline{p}_1 + \frac{\delta\epsilon T}{T} + \frac{\rho(1-\epsilon)T}{T} = \overline{p}_1 \tag{3.40}$$

Treating $p_1(t)$ as piecewise constant, system (3.16) can be written as

$$\frac{d\mathbf{z}}{dt} = \begin{cases} (\mathbf{N}\mathcal{E} + \mathcal{K})\mathbf{z} + (\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{q}(\delta) & t \in [jT, (j+\epsilon)T] \\ (\mathbf{N}\mathcal{E} + \mathcal{K})\mathbf{z} + (\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{q}(\rho) & t \in [(j+\epsilon)T, (j+1)T] \end{cases} \quad j = 0, 1, 2, \ldots \tag{3.41}$$

where $\mathbf{q}(\delta)$ and $\mathbf{q}(\rho)$ are the same as in system (3.16) with their first elements $ln(\delta/p_{o,1})$ and $ln(\rho/p_{o,1})$, respectively.

Integrating equation (3.41) we can find that the periodic solution is

$$\mathbf{z}(t) = \begin{cases} exp((\mathbf{N}\mathcal{E}+\mathcal{K})(t-jT))\mathbf{z}^o \\ -[\mathbf{I}-exp((\mathbf{N}\mathcal{E}+\mathcal{K})(t-jT))](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}(\mathbf{N}\mathbf{\Pi}+\mathbf{\Lambda})\mathbf{q}(\delta) \\ \qquad\qquad t \in [jT,(j+\epsilon)T] \\ exp((\mathbf{N}\mathcal{E}+\mathcal{K})(t-(j+\epsilon)T))\mathbf{z}^\epsilon \\ -[\mathbf{I}-exp((\mathbf{N}\mathcal{E}+\mathcal{K})(t-(j+\epsilon)T))](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}(\mathbf{N}\mathbf{\Pi}+\mathbf{\Lambda})\mathbf{q}(\rho) \\ \qquad\qquad t \in [(j+\epsilon)T,(j+1)T] \end{cases} \tag{3.42}$$

where

$$\mathbf{R} = exp((\mathbf{N}\mathcal{E}+\mathcal{K})(1-\epsilon)T)$$

$$\mathbf{D} = exp((\mathbf{N}\mathcal{E}+\mathcal{K})\epsilon T)$$

$$\mathbf{z}^o = -[\mathbf{I}-\mathbf{RD}]^{-1}$$
$$\{[\mathbf{R}-\mathbf{RD}](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}\mathbf{q}(\delta)+[\mathbf{I}-\mathbf{R}](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}\mathbf{q}(\rho)\}$$

$$\mathbf{z}^\epsilon = -[\mathbf{I}-\mathbf{DR}]^{-1}$$
$$\{[\mathbf{D}-\mathbf{DR}](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}\mathbf{q}(\rho)+[\mathbf{I}-\mathbf{D}](\mathbf{N}\mathcal{E}+\mathcal{K})^{-1}\mathbf{q}(\delta)\}$$

and $\mathbf{I}$ is the unitary $n \times n$ matrix.

Equation (3.42) enables the simple computation of the periodic solution using simple matrix calculations. The computed periodic solution can be used in order to integrate the nonlinear expression for the metabolic output (3.25), and calculate its average value:

$$\begin{aligned} \overline{h}_l = \; & \frac{h_{o,l}}{T}\int_0^{\epsilon T}\left(\frac{\delta}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lj}}d\tau \\ & +\frac{h_{o,l}}{T}\int_{\epsilon T}^{T}\left(\frac{\rho}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lj}}d\tau \end{aligned} \tag{3.43}$$

As expected, the average value of every metabolic function depends on ϵ and δ,

i.e. on the profile of the variation, on the period, T, of the variation, and on the value of the parameters p_j, if a time-invariant change in any of them is considered also. Equations (3.42) and (3.43) provide a fast and simple method to analyze and, when possible, design the spatiotemporal variation profiles of the operating parameters.

The analytical expression for $\overline{h}_l$ (equation (3.43)) can be used in order to calculate the control coefficient of $\overline{h}_l$ with respect to any of the parameters p_j according to Equation (3.38):

$$\begin{aligned}\frac{p_{o,r}}{\overline{h}_l}\frac{d\overline{h}_l}{dp_r} = & \frac{h_{o,l}}{\overline{h}_l T}\int_0^{(1-\epsilon)T}\left(\frac{\delta}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\phi_{lr}(\tau)d\tau \\ & + \frac{h_{o,l}}{\overline{h}_l T}\int_{(1-\epsilon)T}^{T}\left(\frac{\rho}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\phi_{lr}(\tau)d\tau\end{aligned} \tag{3.44}$$

where

$$\phi_{lr}(\tau) = \sum_{k=2}^{n}\alpha_{lk}\left(\frac{x_{o,k}}{x_k(\tau)}\right)C_{p_r}^{x_k} + \beta_{lr} \tag{3.45}$$

where $C_{p_r}^{x_k}$ is the concentration control coefficient of the metabolite x_k with respect to parameter p_r, as it has been determined at steady-state and for $p_1 = p_{o,1}$, and it appears from the approximation

$$\frac{d(x_k(t)/x_{o,k})}{d(p_r/p_{o,r})} \approx \frac{dz_k(t)}{dq_r}$$

in Equation (3.38) discussed above. Moreover, as shown in the Appendix B:

$$\frac{dz_k(t)}{dq_r} = C_{p_r}^{x_k} \quad \forall t$$

for any periodic input. Therefore, equation (3.44) can be written as:

$$\frac{p_{o,r}}{\bar{h}_l}\frac{d\bar{h}_l}{dp_r} =$$

$$\frac{h_{o,l}}{\bar{h}_l T}\int_0^{(1-\epsilon)T}\left(\frac{\delta}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\left(\sum_{k=2}^{n}\alpha_{lk}\left(\frac{x_{o,k}}{x_k(\tau)}\right)C_{p_r}^{x_k}+\beta_{lr}\right)d\tau$$

$$+\frac{h_{o,l}}{\bar{h}_l T}\int_{(1-\epsilon)T}^{T}\left(\frac{\rho}{p_{o,1}}\right)^{\beta_{l1}}\prod_{k=1}^{n}\left(\frac{x_k(\tau)}{x_{o,k}}\right)^{\alpha_{lk}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\left(\sum_{k=2}^{n}\alpha_{lk}\left(\frac{x_{o,k}}{x_k(\tau)}\right)C_{p_r}^{x_k}+\beta_{lr}\right)d\tau \tag{3.46}$$

3.4.2 Quasi-Steady-State Approximation: a Limiting Case

The Quasi-Steady-State Approximation (QSSA) considers the average performance of the metabolic system when the characteristic time scale for changes of the varying parameter (typically the period T, or some quantity scaled by T) is much greater than τ_c^{max} the *maximum characteristic response time* of the system (Bailey, 1974). For the (log)linear model (3.16), τ_c^{max} is the inverse of the minimum eigenvalue of the matrix $\mathbf{N}\mathcal{E}+\mathcal{K}$. Any further increase in the period in this range will not change the average performance of the metabolic system significantly. For pulsed periodic parameter variations the QSSA condition may be stated in the form

$$min\{(1-\epsilon)T, \epsilon T\} \gg \tau_c^{max} \tag{3.47}$$

In quasi-steady state operation, for the time period equal to $(1-\epsilon)T$ the system will be during most of this time interval very near the steady-state for $p_1 = \delta$ at which

$$\left(\frac{x_k}{x_{o,k}}\right) = \left(\frac{\delta}{p_{o,1}}\right)^{C_{p_1}^{x_k}}$$

and for most of a time period equal to ϵT the system will be near the steady-state

for $p_1 = \rho$ at which

$$\left(\frac{x_k}{x_{o,k}}\right) = \left(\frac{\rho}{p_{o,1}}\right)^{C_{p_1}^{x_k}}$$

Therefore, time-trajectories of the metabolite concentrations may be well approximated by:

$$\left(\frac{x_k}{x_{o,k}}\right) = \begin{cases} \left(\frac{\delta}{p_{o,1}}\right)^{C_{p_1}^{x_k}} & jT \leq t < (j-\epsilon)T \\ \\ \left(\frac{\rho}{p_{o,1}}\right)^{C_{p_1}^{x_k}} & (j-\epsilon)T \leq t < (j+1)T \end{cases} \tag{3.48}$$

and the corresponding time-average value of the metabolic function is

$$\overline{h}_l = h_{o,l}\left[\epsilon\left(\frac{\delta}{p_{o,1}}\right)^{\beta_{l1}+\sum_{k=1}^{s}\alpha_{lk}C_{p_1}^{x_k}} + (1-\epsilon)\left(\frac{\rho}{p_{o,1}}\right)^{\beta_{l1}+\sum_{k=1}^{s}\alpha_{lk}C_{p_1}^{x_k}}\right]\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}} \tag{3.49}$$

For $p_j = p_{o,j}$ and from the definition of the control coefficient

$$C_{p_1}^{h_l} = \beta_{l1} + \sum_{k=1}^{s}\alpha_{lk}C_{p_1}^{x_k}$$

we have

$$\overline{h}_l = h_{o,l}\epsilon\left(\frac{\delta}{p_{o,1}}\right)^{C_{p_1}^{h_l}} + h_{o,l}(1-\epsilon)\left(\frac{\rho}{p_{o,1}}\right)^{C_{p_1}^{h_l}} = h_l^{\delta} + h_l^{\rho} \tag{3.50}$$

This last equation shows that knowledge of the control coefficient of the metabolic function with respect to the varying parameter, p_1, determined at $p_1 = p_{o,1}$, can provide us with a first prediction about the improvement or the deterioration of the metabolic function for different waveforms. We should also notice that the experimental determination of the control coeffiecient, $C_{p_1}^{h_l}$, is relatively simple by measuring the metabolic function at different values of p_1, which is usually a manipulated environmental condition, such as nutrient concentration or dissolved oxygen.

Similarly, we can find an analytical expression for the control coefficient, $C_{p_r}^{\overline{h}_l}$, for the limiting case of the QSSA:

$$\frac{p_{o,r}}{\overline{h}_l}\frac{d\overline{h}_l}{dp_r} =$$

$$\frac{h_{o,l}}{\overline{h}_l}\epsilon T\left(\frac{\delta}{p_{o,1}}\right)^{C_{p_1}^{h_l}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\left(\sum_{k=2}^{n}\alpha_{lk}\left(\frac{p_{o,1}}{\delta}\right)^{C_{p_1}^{h_l}}C_{p_r}^{x_k}+\beta_{lr}\right)d\tau \tag{3.51}$$

$$+\frac{h_{o,l}}{\overline{h}_l}(1-\epsilon)\left(\frac{\rho}{p_{o,1}}\right)^{C_{p_1}^{h_l}}\prod_{j=2}^{s}\left(\frac{p_j}{p_{o,j}}\right)^{\beta_{lm}}\left(\sum_{k=2}^{n}\alpha_{lk}\left(\frac{p_{o,1}}{\rho}\right)^{C_{p_1}^{h_l}}C_{p_r}^{x_k}+\beta_{lr}\right)d\tau$$

For $p_j = p_{o,j}$ and after the proper rearrangements we obtain

$$C_{p_r}^{\overline{h}_l} = C_{p_r}^{h_l} + \frac{h_l^{\delta}\Sigma^{\delta} + h_l^{\rho}\Sigma^{\rho}}{h_l^{\delta} + h_l^{\rho}} \tag{3.52}$$

where

$$\Sigma^{\delta} = \sum_{k=2}^{n}\alpha_{lk}\left[\left(\frac{p_{o,1}}{\delta}\right)^{C_{p_1}^{h_l}} - 1\right]C_{p_r}^{x_k} + \beta_{lr}$$

and

$$\Sigma^{\rho} = \sum_{k=2}^{n}\alpha_{lk}\left[\left(\frac{p_{o,1}}{\rho}\right)^{C_{p_1}^{h_l}} - 1\right]C_{p_r}^{x_k} + \beta_{lr}$$

As equation (3.52) suggests, the control coefficient, $C_{p_r}^{\overline{h}_l}$, for the limiting case of the QSSA, can be experimentally determined for various waveforms by performing a steady-state MCA at different values of p_1. Given this information, we can estimate the effects of periodic variation of p_1 on the resulting time-average control coefficients.

3.4.3 Example

The glycolytic system studied in Section 2 will be analyzed here with respect to the effects of time-variation of glucose uptake on the ethanol production rate and its control coefficients. Such a variation can be the result of variation in the external

glucose concentration, or temperature (Ribeiro *et al.*, 1994). In what follows we will use the (log)linear representation of the glycolytic pathway, derived around the steady state given at Appendix A.III. The specific production rate of ethanol at the reference steady-state is

$$J_{EtOH} = 37.61 mM/L_{cell} min$$

and its (steady-state) control coefficients with respect to glucose uptake, phosphofructokinase, pyruvate kinase, and ATPase are

$$[C^J_{Vin}, C^J_{pfk}, C^J_{pyk}, C^J_{ATPase}] = [0.6762, 0.3184, 0, 0.0455]$$

The specific glucose uptake rate was assumed to vary in a pulsed periodic fashion. The effects of the period and the waveform of the variation on the average value of the specific rate of ethanol production are presented in Figure 9.A and 9.B. The average specific rate of ethanol production decreases monotonically as the period of the variation decreases, independently of the waveform as described by different values in δ and ϵ. Moreover, as δ and/or ϵ increase, the average specific rate decreases. In general, any periodic variation on the glucose uptake will probably result in lower ethanol production rates.

The variation of the glucose uptake also has a dramatic effect on the average-flux control coefficients. Figures 10.A.I to 10.C.II illustrate the significance of this effect. Several conclusions can be drawn from these figures. First, the differences between invariant control coefficients and average-flux control coefficients can be orders of magnitude. Second, they are not necessarily monotonic functions of the period. On the contrary, as we can see from the control coefficient with respect to ATPase (Figure 10.C.I and 10.C.II), for high periods of variation the average-flux control coefficient is higher than for low periods. However, there exists a range of period for which the average-flux control coefficient is much lower than both the extreme values. As ϵ and δ increase, the AFCC's for high and low period also increase, and the difference between them is also increasing. The most interesting effect is observed in Figure 10.C.II, where the AFCC can be either positive or negative, depending on the period, and on

the waveform parameters ϵ and δ. This observation is particularly significant because it implies that decisions taken at a certain stage of a scale-up process, with respect to which enzyme in a pathway should be genetically manipulated, could have no effect or even be counterproductive. This example analysis illustrates the importance of the consideration of spatiotemporal variations and the usefulness of methods that can estimate the effects of such variations.

As it has been shown in the branched pathway example, changes in the regulation of the pathway can alter its dynamic response characteristics. Therefore, it is expected that, by altering the regulation in the glycolytic pathway, the dependence of the AFCC's and the average ethanol production, will also be altered. Such analysis suggests genetic engineering approaches for the solution to the problem of performance deterioration under parameter variations. Moreover, useful conclusion with respect to the evolutionary design of pathways can be drawn if we consider that most of the microorganisms have been evolved to survive in a varying environment.

3.5 Concluding Remarks

Rigorous, general relationships between properties of nonlinear systems and properties of simple models used to approximate them are very rarely available. Nevertheless, the introduction of approximations to facilitate analysis and design of nonlinear systems is a widely used engineering method. Clearly, to preserve the greatest possible correspondence between results obtained by analysis of the approximate model and result from the original nonlinear system, one seeks approximate models which mimic closely the original nonlinear system (recognizing that again, general characterization of the extent, or quality, of such mimicry is typically not possible).

We have undertaken the formulation of (log)linear models in this spirit. The comparisons between responses of models of this class, and the exact responses of the nonlinear systems they are intended to approximate presented here for several different examples of significant complexity suggest engineering utility of our approach.

As an illustration of applications which can be effectively developed with such

an approximate linear representation in hand, the effects of periodic variations in a system parameter on metabolic outputs and their sensitivities have been examined. Here we have shown how indications of spatiotemporal parametric fluctuation effects on fluxes and sensitivities can be obtained based on steady-state experimental data. Our analytical results substantially extend and generalize the important observation from numerical simulations that average control coefficients can shift significantlly in response to spatiotemporal fluctuations in a large-scale bioreactors. Further applications of (log)linear models for other purposes in metabolic analysis and design will be presented in future publications.

Acknowledgement: This research was supported by the Swiss Priority Program in Biotechnology (SPP Biotech).

3.6 References

Bailey, J. E. (1974). Periodic Operation of Chemical Reactors: A Review. *Chem. Eng. Comm.* **1**, 111-124.

Bailey, J. E. (1977). Periodic Phenomena. In: *Chemical Reactor Theory. A Review.* (Lapidus, L. & Amundson, N. R., eds) pp. 758-813. Prentice-Hall, Englewood Cliffs, N. J.

Bailey, J. E., "Toward a Science of Metabolic Engineering", *Science*, **252**, 1668-1675 (1991).

Brown, G. C., "Control of Respiration and ATP Synthesis in Mammalian Mitochondria and Cells", *Biochem. J.*, **284**, 1-13 (1992).

Cornish-Bowden, A., and M.L. Cárdenas (Eds.), "Control of Metabolic Processes", NATO ASI Series A: Lifesciences Vol. 190, Plenum Press, New York (1990).

Douglas, J. M., "Periodic Reactor Operation", *Ind. Engng Chem. Proc. Des. Dev* **6**, 43-48 (1967).

Fell, D.A., "Metabolic Control Analysis - A Survey of its Theoretical and Experimental Development", *Biochem. J.*, **152**, 313-330 (1992).

Fredrickson, A. G. (1976). Formulation of Structured Growth Models. *Biotechnol. Bioeng.* **28**, 1481-1486.

Gallazo, J. L., and J. E. Bailey, "Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces cerevisiae*", *Enzyme Microb. Technol.*, **12**,162-172 (1990).

Guardabassi, G., A. Locatelli, and S. Rinaldi, "Status of Periodic Optimization of Dynamical Systems", *J. Optimization Theory Applic.*, **14**, 1-20 (1974).

Hatzimanikatis, V., G., Lyberatos, S. Pavlou, and S. Svoronos, "A Method for Pulsed Periodic Optimization of Chemical Reaction Systems", *Chem. Engng Sci.*, **48**(4), 789-797 (1993).

Heinrich, R., and T. A. Rapoport, "A Linear Steady-State Treatment of Enzymatic Chains", *Eur. J. Biochem.*, **42**, 89-95 (1974).

Kacser, H., and J. A. Burns, "The Control of Flux", *Symp. Soc. Exp. Biol.*, **27**,

65-104 (1973).

Kahn, D. & Westerhoff, H. V. (1993). Regulation and Homeostasis in Metabolic Control Theory: Interplay between fluctuations of variables and parameter changes. In: *Modern Trends in Biothermokinetics* (Schuster, S. *et al.*, eds) pp. 199-204. New York: Plenum Press.

Kell, D.B. & Westerhoff, H. V. (1986). Metabolic Control Theory: Its Role in Microbiology and Biotechnology. *FEMS Microbiol. Rev.* **39**, 305-320.

Lyberatos, G., and S. Svoronos, "Optimal Periodic Square-wave Forcing: A New Method", in *Proceedings of the American Control Conference*, Minneapolis, MN, pp. 257-262 (1987).

Reder, C., "Metabolic Control Theory: A Structural Approach", *J. theor. Biol.*, **135**, 175-201 (1988).

Ribeiro, M. J. S., Silva, J. T., and Panek, A. D., "Trehalose Metabolism in *Saccharomyces cerevisiae* During Heat-shock", *Biochim. Biophys. Acta*, **1200**(2), 139-147 (1994).

Rutgers, M., Van Dam, K. & Westerhoff, H. V. (1991). Control and Thermodynamics of Microbial Growth: Rational Tools for Bioengineering. *Crit. Rev. Biotech.* **11**, 367-395.

Sanwal, B. D., M. Kapoor, H. W. Duckworth, "The Regulation of Branched and Converging Pathways", *Curr. Top. Cell. Regul.*, **3**, 1-115 (1971).

Savageau, M. A., "Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology", Reading, MA, Addison–Wesley (1990).

Schlosser, P. M., and J. E. Bailey, "An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways", *Math. Biosci.*, **100**, 87-114 (1990).

Schlosser, P. M., T. Holcomb, and J. E. Bailey, "Determining Metabolic Sensitivity Coefficients Directly from Experimental Data", *Biotechnol. Bioeng.*, **41**, 1027-1038 (1993).

Seinfeld, J. H. & Lapidus, L. (1974). Process Modeling, Estimation and Identification. Prentice-Hall, Englewood Cliffs, N. J.

Stephanopoulos, G., and J. J. Vallino, "Network Rigidity and Metabolic Engi-

neering in Metabolite Overproduction", *Science*, **252**, 1675-1681 (1991). **252**, 1675 (1991).

3.7 Appendices

APPENDIX A

In this appendix the kinetics used for the nonlinear models in the examples of Section 3 will be presented. The parameters of the corresponding (log)linear models, i.e. the elasticities, can be derived from this information following standard definitions (**REFS**).

APPENDIX A.I

Linear Pathway

The linear pathway considered is presented in Figure 1. The input flux, v_o, was the manipulated variable. The kinetics used are the following (numbered according to the numbering of the steps in Figure 1):

$$v_1 = 2\frac{s_1}{0.3333 + s_1}$$

$$v_2 = 1.7\frac{s_2}{0.6667 + s_2}$$

$$v_3 = 1.5\frac{s_3}{0.6429 + s_3}$$

$$v_4 = 1.3\frac{s_4}{0.1875 + s_4}$$

For the reference steady state we set

$$v_o = 1$$

and for the steady state metabolite concentrations we had

$$[s_1, s_2, s_3, s_4] = [0.3333, 0.9524, 1.2858, 0.6250]$$

APPENDIX A.II

Branched Pathway

The branched pathway considered and the regulatory structure around it is presented in Figure 3. The input flux was again the manipulated variable. The kinetic

expressions used for the reaction rates numbered according to Figure 3.

$$v_1 = 10\frac{s_1}{0.3333 + s_1}$$

$$v_2 = 4.1667\frac{s_2}{0.6667\left(1 + \frac{s_3}{1 + \frac{s_1}{2}}\right) + s_2}$$

$$v_3 = 7.5\frac{s_2}{0.6429\left(1 + \frac{s_4}{7\left(1 + \frac{s_3}{1.5}\right)}\right) + s_2\left(1 + \frac{s_4}{3.5\left(1 + \frac{s_3}{1.5}\right)}\right)}$$

For the reference steady state we set

$$v_o = 3$$

and for the steady state metabolite concentrations we had

$$[s_1, s_2, s_3, s_4] = [0.1429, 0.2425, 0.0393, 0.4001]$$

APPENDIX A.III

Yeast Glycolytic Pathway

The kinetics for the yeast glycolytic pathway have been taken from Schlosser et al. (**REFS**).

$$v_{in} = v_{in}^o - 0.5([G6P] - 2.7)$$

$$v_{HK} = 68.5\left(\frac{0.00062}{[G^{in}][ATP]} + \frac{0.11}{[G^{in}]} + \frac{0.1}{[ATP]} + 1\right)^{-1}$$

and

$$v_{POL} = 15\left(\left(1 + \left(\frac{2}{G6P}\right)^{8.25}\right)\left(\frac{1.1}{[G6P][UDPG]} + \frac{1}{[UDPG]} + 1\right)\right)^{-1}$$

where the concentration of UDPG was fixed at [UDPG]=0.7 mM.

$$v_{PFK} = 5283\frac{[F6P][ATP]R}{R^2 + L_o L^2 T^2}$$

where

$$R = 1 + [F6P] + 16.67[ATP] + 166.7[F6P][ATP]$$

$$T = 1 + 0.0005[F6P] + 16.67[ATP] + 0.0083[F6P][ATP]$$

$$L = \frac{1 + 0.76[AMP]}{1 + 40[AMP]}$$

and

$$L_o = e^{(4.17pH - 20.42)} - 1658$$

$$v_{GAPD} = 49.9\left(1 + \frac{0.0025}{[G3P]} + \frac{0.18S}{[NAD^+]}\left(1 + \frac{0.0025}{[G3P]}\left(1 + 3333[NADH]\right)\right)\right)^{-1}$$

where

$$S = 1 + 0.9091[AMP] + 0.6667[ADP] + 0.4[ATP]$$

$$v_{PK} = 68.8\frac{[PEP][ADP]\left(R_{PK}K_I + 0.2L_{o,PK}L_{PK}^2T_{PK}\right)}{\left(R_{PK}^2 + L_{o,PK}L_{PK}^2T_{PK}^2\right)\left(1 + 10^{(pH - 8.02)}\right)}$$

where

$$R_{PK} = 1 + K_I[PEP] + 0.2[ADP] + 0.02K_I[PEP][ADP]$$

$$T_{PK} = 1 + 0.02[PEP] + 0.2[ADP] + 0.004[PEP][ADP]$$

$$K_I = 17.92 + 10^{(pH - 4.907)}$$

and

$$L_{o,PK} = ab - a^2$$

with

$$a = 1.713 + 10^{(pH - 6.306)}$$

and

$$b = \frac{0.1192 + 10^{(pH - 7.084)}}{11.83 + 1.61pH - 8.722\sqrt{pH}}$$

$$v_{GOL} = 15\frac{[FdP]}{24 + [FdP]}$$

The equilibrium step, $G6P \rightleftharpoons F6P$, was described by $[F6P] = 0.3[G6P]$, and the equilibrium step $G3P \rightleftharpoons FdP$, was described by $[G3P] = 0.01[FdP]$. For the adenylate kinase (AK) an equilibrium constant $q_{AK} = 1$ was considered and the total adenylate pool, $[AN] = [ATP]+[ADP]+[AMP]$, was treated as fixed parameters at a value $[AN] = 2.8$ mM. The ratio $[NADH]/[NAD^+]$ and the sum $[NADH]+[NAD^+]$ were also treated as fixed at values 0.03 and 2.5 mM, respectively.

The intracellular pH was assumed to be a function of [ATP], and the following relationship was used:

$$pH = 7.11 + 0.113([ATP] + [ATP]_o)$$

where $[ATP]_o = 0.967$ mM.

APPENDIX B

In this appendix we will show that for the special case of pulsed periodic variation of a parameter, the derivative

$$\frac{d(x_k(t)/x_{o,k})}{d(p_r/p_{o,r})}$$

which can be approximated by the X_k concentration control coefficient and is independent of time

$$\frac{d(x_k(t)/x_{o,k})}{d(p_r/p_{o,r})} \approx \frac{dz_k(t)}{dq_r} = C_{p_r}^{x_k} \quad \forall t \tag{3.53}$$

If we differentiate $\mathbf{z}(t)$, as it is given by equation (3.21) we obtain:

$$\frac{d\mathbf{z}(t)}{dq_r} = e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-t_o)}\frac{d\mathbf{z}_o}{dq_r} + \int_{t_o}^{t} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-\tau)}(\mathbf{N}\mathbf{\Pi} + \mathbf{\Lambda})\mathbf{e}_r d\tau \tag{3.54}$$

where

$$\frac{d\mathbf{q}(\delta)}{dq_r} = \frac{d\mathbf{q}(\rho)}{dq_r} = \mathbf{e}_r \tag{3.55}$$

is a vector of size equal to the size of $\mathbf{q}$ and with zero elements except the r-th element which is equal to one.

For the periodic solution of $\mathbf{z}(t)$ we will have

$$\mathbf{z}(t_o) = \mathbf{z}(t_o + T)$$

where T is the period of variation of the varying parameter q_j $(j \neq r)$. From Equation (3.21) we have for the periodic solution

$$\begin{aligned} \mathbf{z}(t_o + T) &= e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T}\mathbf{z}(t_o) + \int_{t_o}^{t_o+T} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t_o+T-\tau)}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{q}(\tau)d\tau \Rightarrow \\ \mathbf{z}(t_o) &= (\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T})^{-1} \int_{t_o}^{t_o+T} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t_o+T-\tau)}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{q}(\tau)d\tau \end{aligned} \quad (3.56)$$

from which, by differentiation with respect to q_r, we can calculate the term $d\mathbf{z}_o/dq_r$ in the Equation (3.54):

$$\frac{d\mathbf{z}(t_o)}{dq_r} = \left[\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T}\right]^{-1} \int_{t_o}^{t_o+T} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t_o+T-\tau)}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r d\tau \quad (3.57)$$

For the two integrals in Equations (3.54) and (3.57) we can solve analytically:

$$\int_{t_o}^{t} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-\tau)}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r d\tau = -\left[\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t-t_o)}\right](\mathbf{N}\mathcal{E} + \mathcal{K})^{-1}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r \quad (3.58)$$

and

$$\int_{t_o}^{t_o+T} e^{(\mathbf{N}\mathcal{E}+\mathcal{K})(t_o+T-\tau)}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r d\tau = -\left[\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T}\right](\mathbf{N}\mathcal{E} + \mathcal{K})^{-1}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r \quad (3.59)$$

Substituting the integral from Equation (3.59) into Equation (3.57) we obtain

$$\begin{aligned} \frac{d\mathbf{z}(t_o)}{dq_r} &= -\left[\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T}\right]^{-1}\left[\mathbf{I} - e^{(\mathbf{N}\mathcal{E}+\mathcal{K})T}\right](\mathbf{N}\mathcal{E} + \mathcal{K})^{-1}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r \\ &= -(\mathbf{N}\mathcal{E} + \mathcal{K})^{-1}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r \end{aligned} \quad (3.60)$$

Finally, from equations (3.60), (3.58), (3.54) the expression for $d\mathbf{z}(t)/dq_r$ can be

written as:

$$\frac{d\mathbf{z}(t)}{dq_r} = -(\mathbf{N}\mathcal{E} + \mathcal{K})^{-1}(\mathbf{N}\Pi + \mathbf{\Lambda})\mathbf{e}_r \tag{3.61}$$

which, by definition, is the vector of the metabolites concentration control coefficients with respect to parameter q_r, and is independent of time and of the varied parameter.

3.8 Figures

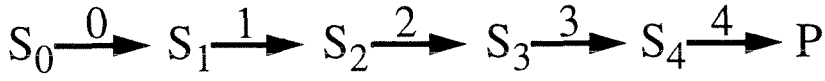

Figure 3.1: Linear pathway.

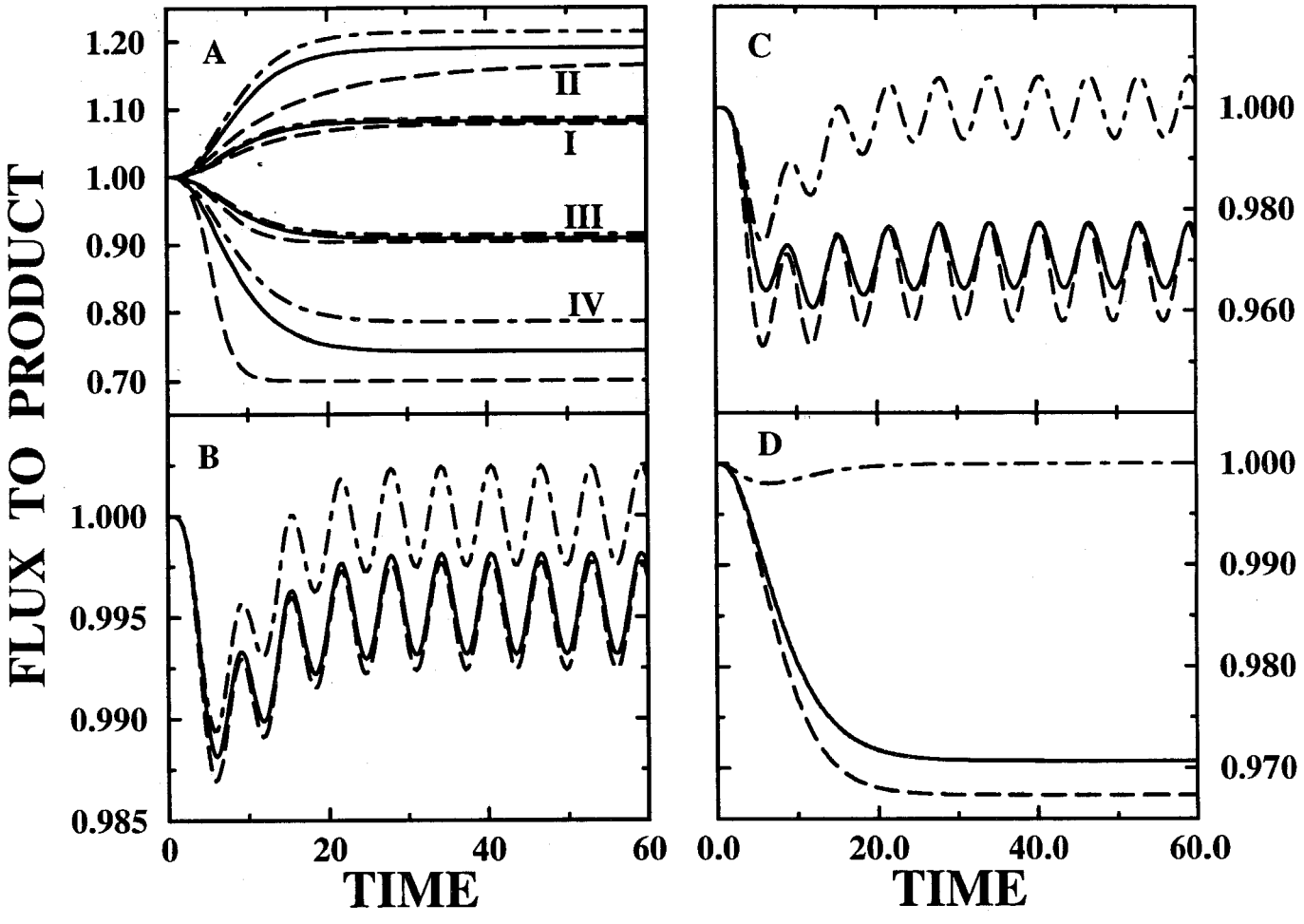

Figure 3.2: Dynamic responses of the flux through the linear pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A. Responses to step changes of the input flux. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 1 - 0.2sin\left(\frac{t}{\pi}\right)$.

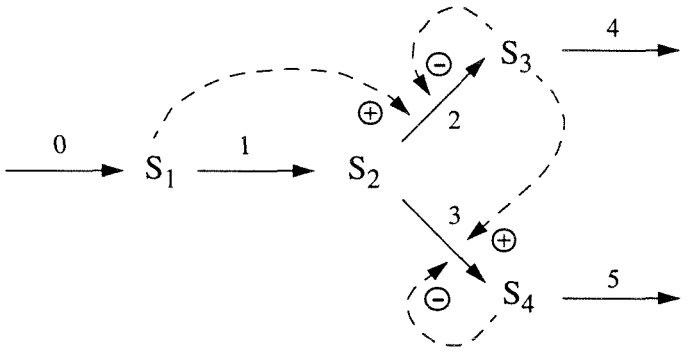

Figure 3.3: Branched pathay. Solid arrows indicate reaction steps. Dashed arrows indicate activation (plus sign) or inhibition (minus sign).

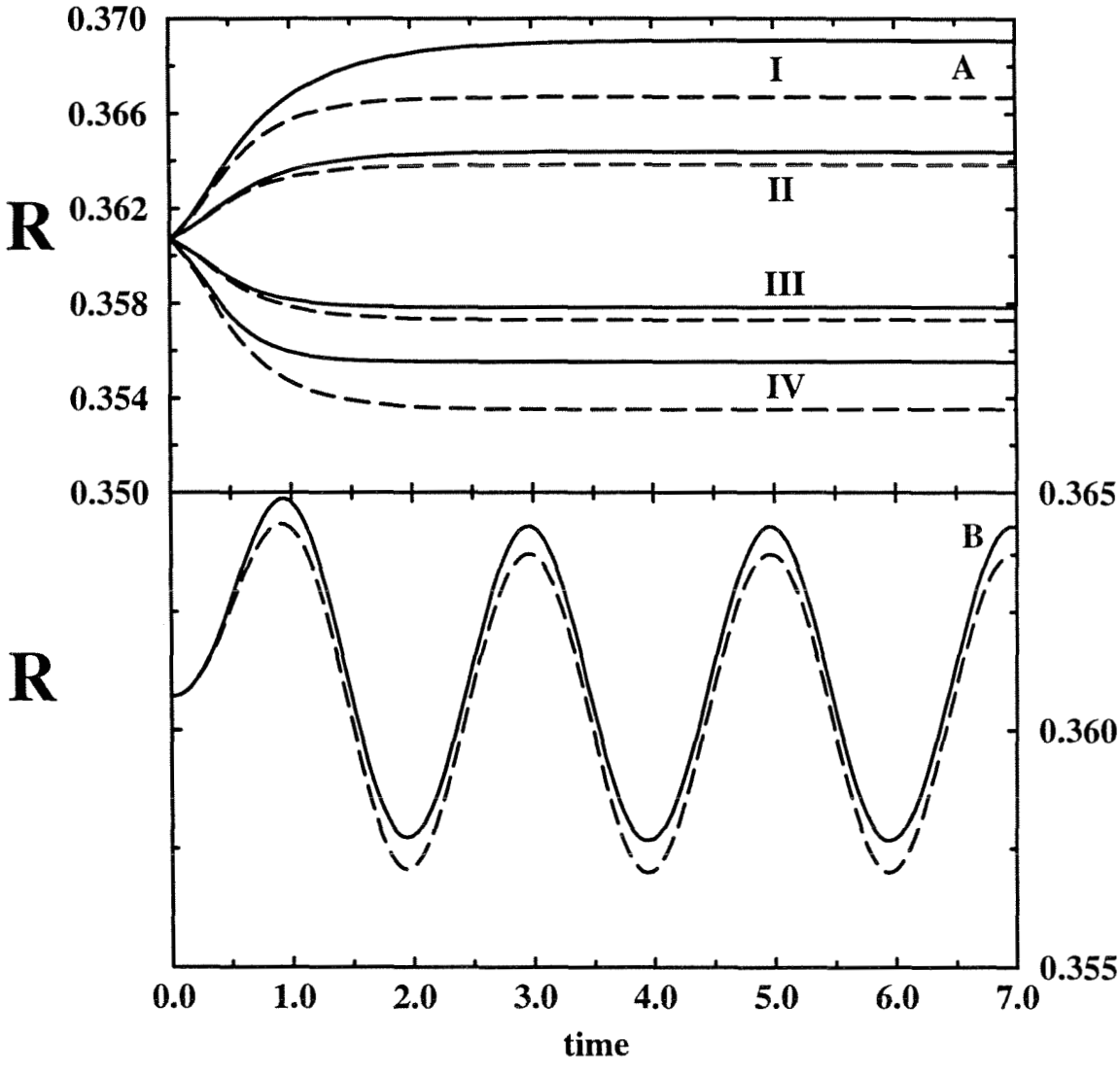

Figure 3.4: Dynamic responses of the ratio of fluxes through the branched pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A. Responses to step changes of the input flux. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 3(1 + 0.2sin\,(\pi t))$.

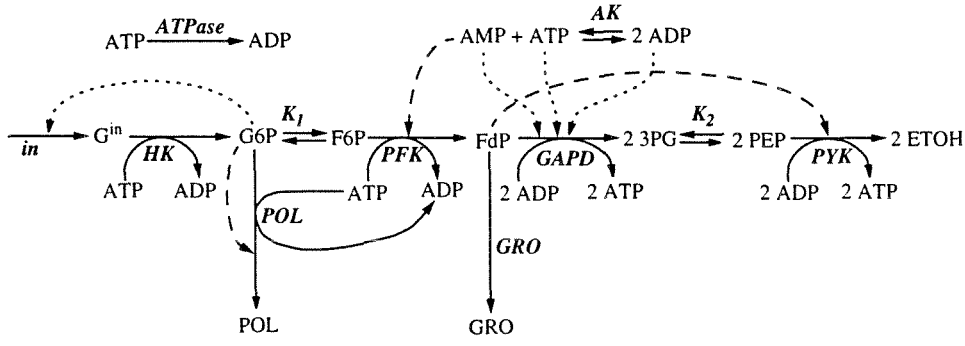

Figure 3.5: Anaerobic fermentation pathway of the yeast *Saccharomyces cerevisiae* under nitrogen starvation, with glucose as the sole carbon source. Enzyme/pathway steps: **in**, glucose uptake; **HK**, hexokinase; $\mathbf{K_1}$, equilibrium step; **PFK**, phosphofructokinase; **GAPD**, glyceraldehyde 3–phopshate dehydrogenase; $\mathbf{K_2}$, equilibrium step; **PYK**, pyruvate kinase; **GRO**, glycerol production; *POL*, polysaccharide production; **ATPase**, net ATP consumption; **AK**, adenylate kinase. Solid arrows indicate reaction steps, dashed arrows indicate activation and dotted arrows indicate inhibition.

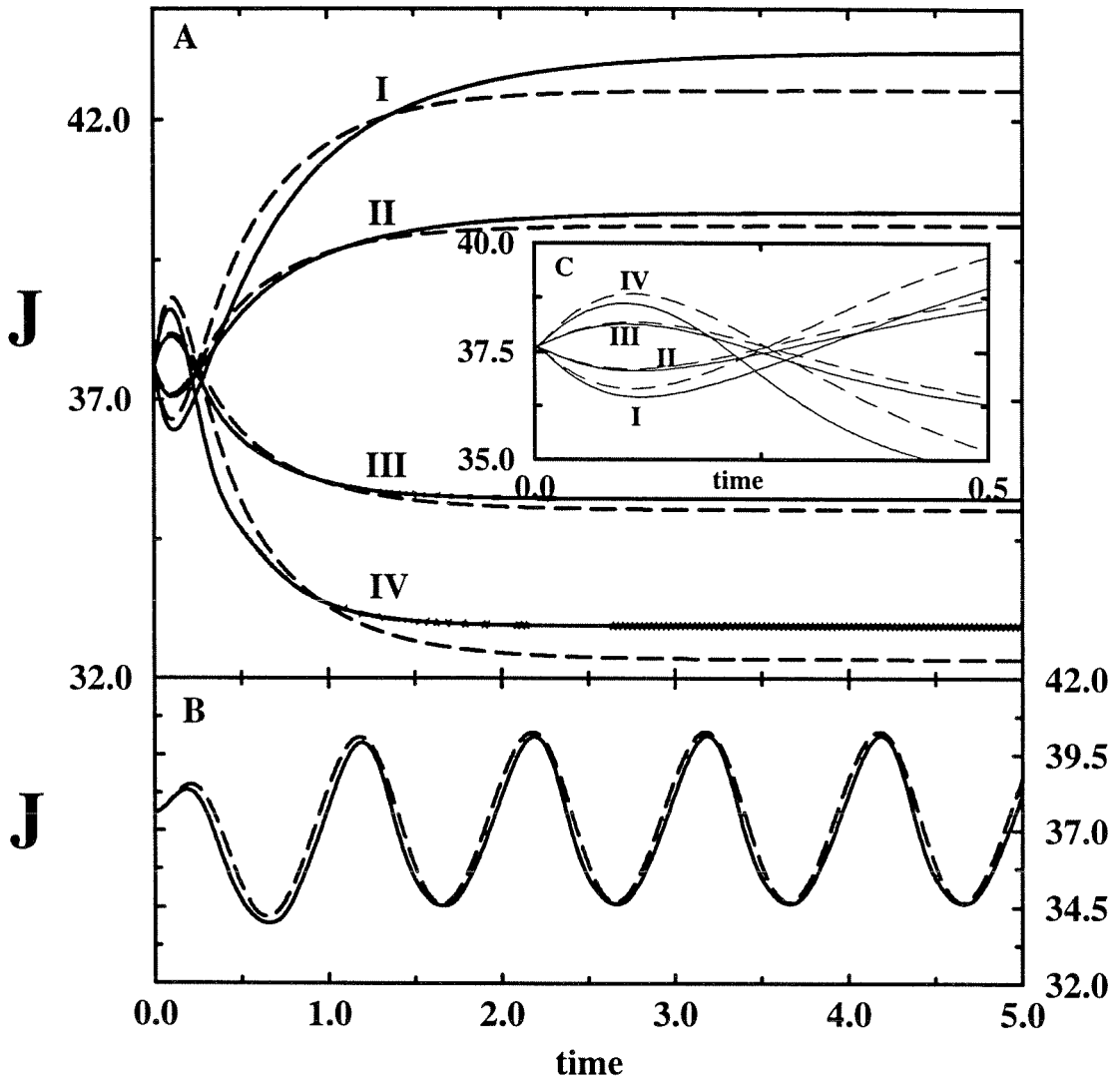

Figure 3.6: Dynamic responses of the ethanol specific production rate of the yeast glycolytic pathway. Solid lines correspond to the nonlinear model and dashed lines correspond to the (log)linear model. A & C. Responses to step changes of the glucose uptake. I: +20%; II: +10%; III: -10%; IV: -20%. B. Responses to sinusoidal variation of the input flux: $v_o = 27.3(1 + 0.2sin\,(2\pi t))$.

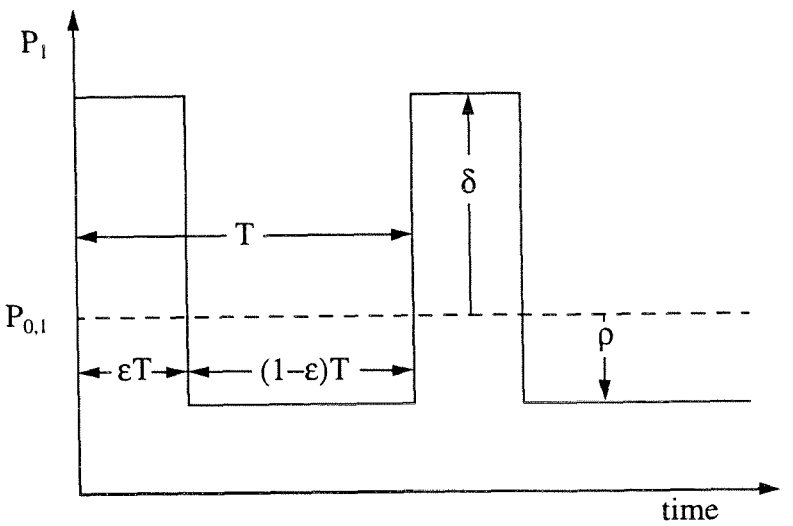

Figure 3.7: Pulsed periodic variation of parameter p_1.

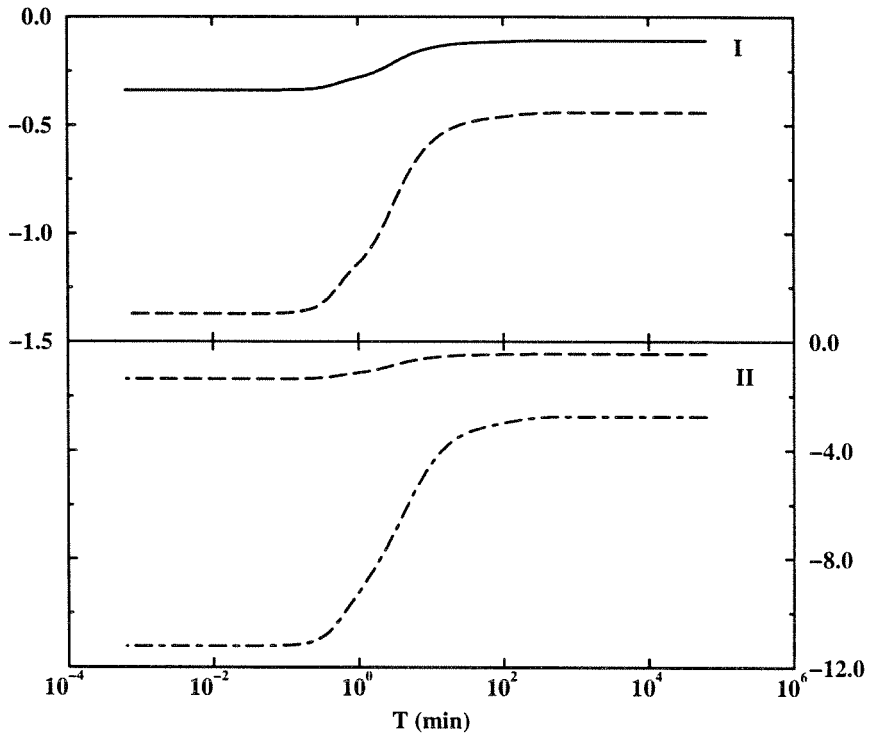

Figure 3.8: The percentage difference of the average ethanol specific production rate, i.e. $\frac{\overline{J}_{EtOH} - J_{EtOH,ref}}{J_{EtOH,ref}}$ as a function of the period, T, of the pulsed variation of the glucose uptake rate. I: $\epsilon = 0.5$, $\delta = 0.1$ (solid line) and $\delta = 0.2$ (dashed line). II: $\delta = 0.2$, $\epsilon = 0.5$ (dashed line) and $\epsilon = 0.8$ (dashed–dotted line).

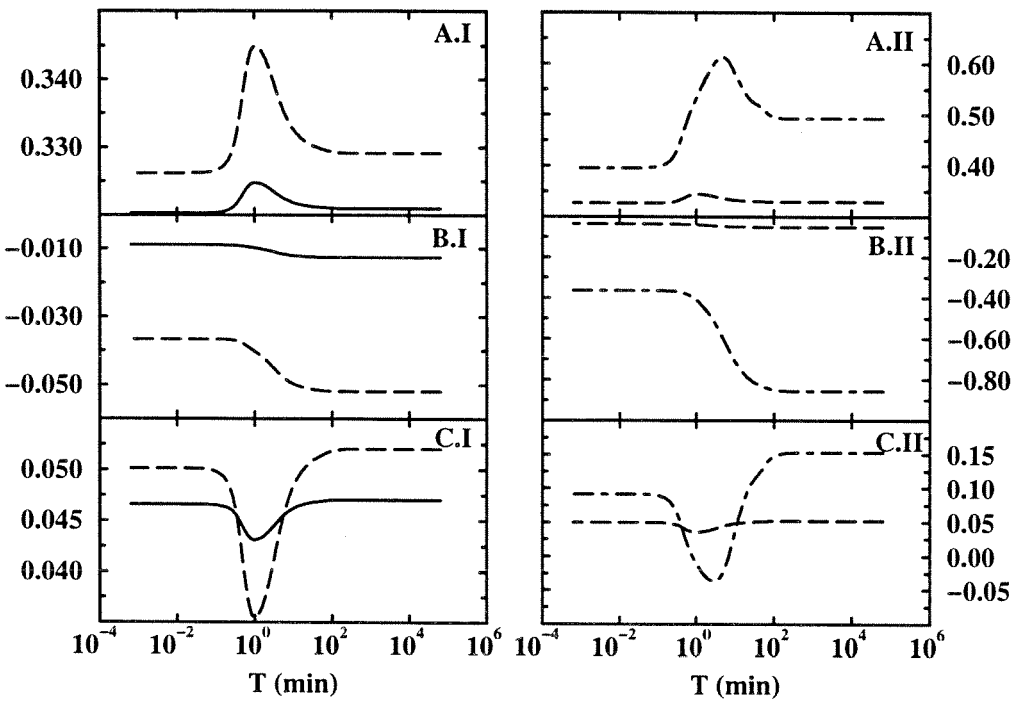

Figure 3.9: The control coefficients of the average ethanol specific production rate, $\overline{J}_{EtOH}$, with respect to: A. phosphofructokinase, B. pyruvate kinase, and C. ATPase, as functions of the period, T, of the pulsed variation of the glucose uptake rate. Solid lines: $\epsilon = 0.5$ and $\delta = 0.1$. Dashed lines: $\epsilon = 0.5$ and $\delta = 0.2$. Dashed–dotted lines: $\epsilon = 0.8$ and $\delta = 0.2$.

Chapter 4 Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization

4.1 Introduction

Improvements in the product yield, rate of production, and final product concentration are common goals in achieving more efficient and cost-effective bioprocesses. These improvements can be achieved by two main approaches: genetics and process. Process improvements involve the adjustment of the environment of the organisms and the optimization of parallel and downstream processes in order to achieve the best possible performance. Genetic improvements are based on the use of organisms with altered DNA such that their functional characteristics are enhanced. Traditionally, the latter approach has been based on the introduction of random changes in the DNA of a population of organisms and the subsequent selection of an improved organism from the resulting heterogeneous population. However, recent advances in recombinant DNA technology make targeted modifications in the DNA of an industrial microorganism possible. Moreover, recombinant DNA methods enable the introduction into an organism of DNA fragments from other organisms with the possibility of creating hybrid metabolic networks combining features from metabolic networks in different organisms.

All the major cellular pathways are subject to a collection of natural independent control loops with different signals and different loci of action (Sanwal et al. 1971; Savageau, 1976; Stephanopoulos and Vallino, 1991). These mechanisms of metabolic regulation operate at essentially two different levels. Genetic-level controls regulate the expression of genes, thereby determining which enzymes are present and in what quantity. Protein-level controls regulate the activity of particular enzymes, and other proteins, in the cell. With respect to protein-level controls, each enzyme can be classified as having no such control, or as having activity modulation by one or more particular metabolites in the cell.

The metabolic control structures in a native, wild-type organism have evolved through natural selection and are therefore configured to maximize the probability of survival of the species, at least insofar as evolution has progressed. While explicit formulation of the objective function for natural metabolism is nontrivial, prior in-

vestigators have proposed maximization of growth rates or most efficient utilization of cellular energetic and chemical resources as the objective function for evolution of natural metabolism (Savageau, 1976; Ramkrishna, 1983; Heinrich et al., 1987; Marr, 1991; Schuster and Heinrich, 1991). However, in chemical and pharmaceutical manufacturing that utilize cultivated microorganisms, it is desirable to identify a different configuration of fluxes which directs raw materials to products efficiently at high rates and in the presence of high concentrations of product. A production-oriented evolution is needed to achieve these goals.

Through currently available genetic engineering technology, it is possible to modify both genetic and protein-level regulation. Thus, the amount of a particular enzyme which is expressed under a particular process condition can be altered by changing the genetic information of the organism. Similarly, by changing the gene which codes for a particular enzyme, the response of that enzyme to metabolites which influence its activity can be altered.

Prior research and industrial practice have clearly shown that very large increases in process performance can be realized by genetic modifications of metabolic control systems (Bailey, 1991; Katsumata and Ikeda, 1993). Past improvements in the performance of a process by modification of the control structures were mainly based on trial and error methods and on well-understood, relatively simple pathways. As the complexity of a set of pathways of interest increases, intuitive and trial and error methods are increasingly ineffective. Modifying the regulatory characteristics of an enzyme is presently a much more difficult experimental challenge than changing the ammount of enzyme present in the cell. Therefore, guidance as to what changes in regulation might be of greatest benefit to improve the network is important. To this end, a systematic, multilevel, multiparametric methodology for evolving effective control structures is needed.

In an attempt to achieve a quantitative understanding and rational metabolic engineering of biochemical reaction pathways, mathematical descriptions of metabolic systems have been developed in many cases and the expected responses of pathways to changes at individual reactions or within certain pathway segments have been cal-

culated. A population of living cells is an extremely complex system, so complicated that some people doubt the possibility of a credible mathematical description, and therefore of quantitative engineering design, of any of its attributes. Two different bodies of experience contradict this view. First, several complex biological phenomena have been well described by mathematical models which are based on the essential molecular mechanism. Examples include regulated gene expression (Lee and Bailey, 1984b,c), replication of DNA (Lee and Bailey, 1984a), growth of bacterial cells (Shuler and Domach, 1983), animal cell cycle regulation (Hatzimanikatis et al., 1995), and receptor trafficking (Starbuck and Lauffenberger, 1992). Second, design and control of most industrial chemical processes, ranging from catalytic cracking to olefin polymerization are based upon mathematical models which are known to be only crude approximations of physical reality. Most chemical engineering applications involve partially understood, approximately described complex physical systems. Useful engineering has been achieved in many facets of chemical engineering endeavor in spite of this. There is no reason to expect a different outcome in the engineering of complicated, imperfectly known cellular processes.

The modeling approaches previously used can be classified into two kinds: linear and nonlinear. Linear models can be accessed through analysis of input-output relations and certain stimulus-response experiments by applying advanced regression analysis (Schlosser et al., 1993) or other experimental methods developed within and around the metabolic control analysis (MCA) framework (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Cornish-Bowden and Cardenas, 1990; Fell, 1992). Nonlinear models, on the other hand, can be constructed when detailed kinetic expressions for each step in the reaction pathway are known or can be estimated (Joshi and Palsson, 1989; Gallazo and Bailey, 1990). Because of the greater data requirements for nonlinear model formulation and validation linear (or log–linear, see below) models will often be the only practically accessible description.

Using the available mathematical description of a biochemical system, various analytical and computational techniques can be used for analysis and optimization of the system. Optimization techniques have been used in the analysis of biochemi-

cal systems using mainly stoichiometry information and metabolic requirements for growth (Majewski and Domach, 1990; Stephanopoulos and Vallino, 1991; Varma and Palsson, 1994). These approaches do not require kinetic data; therefore, they cannot be used in order to quantify the effects of genetic modifications of enzyme levels and of regulatory structures. Optimization of metabolic pathways on the basis of a kinetic model developed using experimental data represented mathematically using the S-system formalism of biochemical systems theory (BST) has been presented using linear programming (Voit, 1992; Regan et al., 1993). These studies provide information only about optimum manipulation of the external inputs to the system (such as independent effectors and external substrates) and do not address the problem of optimizing the regulatory structure of the metabolic network.

The aim of this work is to provide a mathematical framework for determining changes in regulatory structure and strength which should be considered to optimize a particular metabolic process. A mixed-integer linear programming (MILP) formulation is proposed for the general case of linear model optimization. The solution of the MILP formulation provides information on which enzymes should be present at different levels, the extent of such changes needed, and the accompanying modifications in the regulatory structure that will optimize the process.

Any mathematical description of cellular processes is an approximation, and, genetic manipulation of the cell may cause secondary responses which were not considered in the mathematical (or the conceptual) model. Therefore, the output which the metabolic engineer seeks, and all that can be expected, from such optimization calculations are reasonable suggestions for changes in the metabolic network which might give useful improvements in cellular performance. Strategic quidance, not fine quantitative rules, is the intended outcome of these type of calculations. A large body of prior experience with engineering of other complex chemical systems clearly indicates the value of such an approach, even when models are crude approximations, relative to completely ad hoc approaches which are, of course, necessarily based on more crude mental optimization of much more crude mental models.

4.2 Problem Statement

In this chapter we address the following general problem:

A mathematical description of a metabolic pathway with a postulated number of regulatory loops is given. These regulatory loops are classified as either activation (increase the activity of regulatory enzyme) or inhibition (decrease the activity of the regulatory enzyme) loops. The objective is to determine (i) which of the regulatory loops should be retained, and (ii) the number, type, and level of manipulation of amounts of enzymes, in order to optimize a certain function of the outputs of the metabolic pathway (e.g., production of primary or secondary metabolites, growth, selectivity, etc.).

Many metabolic pathways are common to many organisms. However, enzymes that catalyze the same reaction in different organisms are not necessarily the same in their catalytic and regulatory properties. As discussed in the introduction, recombinant DNA methods enable the introduction into an organism of the DNA from other organisms with the possibility then of combining regulatory features from the metabolic pathways present in these different organisms. Moreover, protein engineering methods allow modifications of the properties of natural enzymes and design of enzymes with novel regulatory characteristics. Therefore, the possible number of regulatory loops for a certain pathway in an organism ranges from the number of the existing loops to this number plus the number of additional, different regulatory loops which can be introduced into the same pathway by genetic engineering.

Most generally, we can consider a *regulatory superstructure* in which every metabolite in the system can potentially regulate any enzyme in that system. The mathematical formulation of such a general regulatory superstructure leads to a large combinatorial problem. Its solution will provide the maximum or minimum performance achievable for a given system, and thus provide valuable insight for protein and metabolic engineering. This problem addresses the question of how regulation and catalyst levels should be chosen *de vove* in order to maximize the performance of the metabolic network.

4.3 Mathematical Description of Metabolic Reaction Networks

We consider here a linear model description for metabolic systems. This is the most common situation because of limitations in the available information for most systems. Furthermore, when a nonlinear model is available, it can be linearized and studied within the same framework. In what follows we describe the linearization procedure in a way similar to that presented by Reder (1988).

Consider a metabolic system consisting of n metabolites and m enzymatically-catalyzed reactions. We are interested in studying how modifications of the expression levels and of the properties of the enzymes that catalyze these reactions affect metabolic functions of the system, such as metabolite concentrations, fluxes, and specific growth rate. The mass balances on the metabolites of the system may be written:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{v}(\mathbf{x};\mathbf{p}),\mathbf{x};\mathbf{p}) \tag{4.1}$$

where $\mathbf{x}$ is the n-dimensional metabolite concentration vector, $\mathbf{f}$ is a function determined by the mass balances, $\mathbf{v}$ is the m-dimensional reaction rate vector, and $\mathbf{p}$ is the s-dimensional manipulated parameter vector (e.g., enzyme concentrations). In addition to metabolite reaction rates, the mass balance equations also include terms that account for other processes by which concentrations of metabolites change (such as the dilution brought about by increases in the biomass volume (Fredrickson, 1976) and transport through the cell wall envelope).

In addition, consider the r-dimensional vector of metabolic outputs, $\mathbf{h}$, for which we have:

$$\mathbf{h} = \mathbf{h}(\mathbf{v}(\mathbf{x};\mathbf{p}),\mathbf{x};\mathbf{p}) \tag{4.2}$$

In equation (4.2) $\mathbf{h}$ is a function of the rates of interest, of the metabolite concentration, and of the parameters. Linearization of equations (3.2) and (3.3) around a steady-state, $(\mathbf{x}_o, \mathbf{p}_o)$ results in the following linear system for the logarithmic devia-

tions (see Appendix A for explanation of the transformation):

$$\frac{d\mathbf{z}}{dt} = \mathbf{N}\mathcal{E}\mathbf{z} + \mathcal{K}\mathbf{z} + \mathbf{N}\Pi\mathbf{q} + \Lambda\mathbf{q} \tag{4.3}$$

$$\mathbf{w} = \Xi\mathcal{E}\mathbf{z} + \mathcal{H}\mathbf{z} + \Xi\Pi\mathbf{q} + \Theta\mathbf{q} \tag{4.4}$$

where, $\mathbf{z}$, $\mathbf{q}$, and $\mathbf{w}$, are the logarithmic deviations of the metabolite concentrations, the enzyme levels, and the metabolic outputs, respectively:

$$z_i = ln(x_i/x_{i,o})$$
$$q_k = ln(p_k/p_{k,o})$$
$$w_l = ln(h_l/h_{l,o})$$

and $\mathbf{N}$, Ξ, $\mathcal{K}$, Λ, $\mathbf{H}$, Θ, $\mathcal{E}$, and Π, are matrices, defined as:

$$\mathbf{N} = \{n_{i,j} \mid n_{i,j} = \frac{v_{j,o}}{x_{i,o}}\left(\frac{\partial f_i}{\partial v_j}\right)_{\mathbf{x}_o,\mathbf{p}_o}\} \quad , \quad \Xi = \{\xi_{l,j} \mid \xi_{l,j} = \frac{v_{j,o}}{h_{l,o}}\left(\frac{\partial h_l}{\partial v_j}\right)_{\mathbf{x}_o,\mathbf{p}_o}\},$$
$$\mathcal{K} = \{\kappa_{i,k} \mid \kappa_{i,k} = \frac{x_{k,o}}{x_{i,o}}\left(\frac{\partial f_i}{\partial x_k}\right)_{\mathbf{x}_o,\mathbf{p}_o}\} \quad , \quad \Lambda = \{\lambda_{i,k} \mid \lambda_{i,k} = \frac{p_{k,o}}{x_{i,o}}\left(\frac{\partial f_i}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o}\},$$
$$\mathbf{H} = \{\eta_{l,i} \mid \eta_{l,i} = \frac{x_{i,o}}{h_{l,o}}\left(\frac{\partial h_l}{\partial x_i}\right)_{\mathbf{x}_o,\mathbf{p}_o}\} \quad , \quad \Theta = \{\theta_{l,k} \mid \theta_{l,k} = \frac{p_{k,o}}{h_{l,o}}\left(\frac{\partial h_l}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o}\},$$
$$\mathcal{E} = \{\epsilon_{j,i} \mid \epsilon_{j,i} = \frac{x_{i,o}}{v_{j,o}}\left(\frac{\partial v_j}{\partial x_i}\right)_{\mathbf{x}_o,\mathbf{p}_o}\} \quad , \quad \Pi = \{\pi_{j,k} \mid \pi_{j,k} = \frac{p_{k,o}}{v_{j,o}}\left(\frac{\partial v_j}{\partial p_k}\right)_{\mathbf{x}_o,\mathbf{p}_o}\}$$

Here the subscript o indicates the reference steady-state about which the approximate model is developed. Thus, the right-hand side of equation (4.1) is zero when evaluated at $(\mathbf{x}_o, \mathbf{p}_o)$, and $\mathbf{h}_o$ denotes the value of $\mathbf{h}$ at $(\mathbf{x}_o, \mathbf{p}_o)$.

It should be noted that a mathematical description linear in logarithms of the system variables is in fact a nonlinear (power law) representation, a functionality well suited to approximating closely the nonsingular rational polynomial kinetic expressions typical of metabolic processes. The quality of this form of approximate representation of metabolic kinetics were considered in detail in Chapter 3. Here the nature of this approximation will be tested for one of the examples presented by cal-

culating the consequences for the original nonlinear model (equations (4.1) and (4.2)) of the optimization strategy determined using the log–linear approximate model of equations (4.3) and (4.4).

At steady-state, solution of (4.3) and (4.4) yields:

$$\mathbf{w} = \mathbf{C}\mathbf{q} \tag{4.5}$$

where

$$\mathbf{C} = -\left(\Xi\mathcal{E} + \mathcal{H}\right)\left(\mathbf{N}\mathcal{E} + \mathcal{K}\right)^{-1}\left(\mathbf{N}\Pi + \Lambda\right) + \Xi\Pi + \Theta \tag{4.6}$$

with

$$\mathbf{C} = \{c_{l,k} \mid c_{l,k} = \frac{p_{k,o}}{w_{l,o}}\left(\frac{dw_l}{dp_k}\right)_{\mathbf{x}_o,\mathbf{p}_o}\}$$

The mathematical description presented above depends on the same information as that employed within the framework of metabolic control analysis (MCA) (Reder, 1988; Schlosser and Bailey, 1990). Matrices $\mathcal{E}$ and Π are the elasticity matrices with respect to metabolites and to parameters, respectively. The matrix $\mathbf{C}$ is the control coefficient matrix of the metabolic functions $\mathbf{h}$ with respect to parameters $\mathbf{p}$. Experimental determination of the parameters for this linear system has been the subject of several studies (Fell, 1992; Cornish-Bowden and Cardenas, 1990; Schlosser and Bailey, 1990) and in many cases it is the only available description of a metabolic system.

4.4 Analysis and Synthesis Problems

The regulatory structure of a metabolic network is typically deduced from experimental analysis of the integrated system or from the reported kinetic properties of the enzymes involved in the pathway. In this case the matrix $\mathcal{E}$ can be written as a sum of two matrices:

$$\mathcal{E} = \mathcal{E}^s + \mathcal{E}^r \tag{4.7}$$

where the elements in matrix $\mathcal{E}^s$ correspond to the substrate elasticities of the enzymes, that is, the sensitivities of enzyme activities with respect to their substrates, and the elements of matrix $\mathcal{E}^r$ correspond to the regulatory elasticities of the enzymes, that is, the sensitivities of enzyme activities with respect to regulatory metabolites. In this representation, the substrates themselves can also be considered as regulatory metabolites (in cases such as substrate inhibition):

$$\mathcal{E}^s = \{\epsilon_{j,i_s} = \frac{x_{i_s,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_{i_s}}\right)_{\mathbf{x}_o,\mathbf{p}_o} \mid x_{i_s} \text{ is a substrate for reaction } j\}$$

and

$$\mathcal{E}^r = \{\epsilon_{j,i_r} = \frac{x_{i_r,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_{i_r}}\right)_{\mathbf{x}_o,\mathbf{p}_o} \mid x_{i_r} \text{ is a regulator for reaction } j\}$$

Changes of the elements in matrix $\mathcal{E}^r$ from non-zero values to zero or vice-versa correspond to modifications in the regulatory structure of the system. In this study, which emphasizes the role of control structure, we define the *analysis problem* in the context of a given control structure, which we reasonably assume can be modified only by deleting certain control interactions:

Which of the existent regulatory loops should be inactivated, and what associated changes should be made in the manipulated variables (e.g., enzyme expression levels, environmental conditions, effectors external to the system), in order to optimize the performance of the metabolic network?

The *synthesis problem* considers the possibility of postulating a *regulatory superstructure* and addressing the problem of selecting among alternative regulatory structures for each enzyme. The regulatory superstructure embeds a set of alternative regulatory elasticities for each enzyme that correspond to different kinds of regulation by each metabolite. In particular, the *synthesis problem* addresses the following question:

What kind of regulation (i.e., activation or inhibition, by which metabolite and of what strength) should be assigned to each enzyme in the network, and what associated changes should be made in the manipulated parameters (e.g., enzyme expression levels, environmental conditions, effectors external to the system), in order to optimize the performance of the metabolic network?

The analysis and synthesis problems include discrete decisions concerning the regulatory elasticities of the system. Inactivation or activation of a regulatory loop is equivalent to elimination or introduction of non-zero terms in the $\mathcal{E}^r$ matrix. Moreover, the synthesis problem will typically be subject to some constraints such as the possible number of regulatory actions on each enzyme and the requirement that an enzyme cannot be activated and inhibited by the same metabolite.

On the other hand the continuously adjustable manipulated parameters can potentially be subject to discrete constraints, such as the maximum number of these parameters that we can manipulate simultaneously.

The mixed discrete and continuous nature of the problem and the linear description of the system lead to the formulation of the analysis and synthesis problems as MILP problems, solutions of which provide the optimal regulatory structure and the optimal parameter configuration of a metabolic reaction network.

4.5 Mathematical Formulation

The mathematical formulation for the synthesis problem as a MILP problem is presented in this section. To derive the mathematical formulation the following index sets and variables are introduced to characterize the postulated regulatory superstructure. The metabolites will be denoted by the index set $I = \{i\}$, the reaction rates by the index set $J = \{j\}$, the manipulated parameters by the index set $K = \{k\}$, and the metabolic outputs by the index set $L = \{l\}$. The following sets will be defined to establish the connections of the sets of metabolites with the reaction rates in the network:

$I_s^j = \{i_s \mid i_s \in I \text{ is a substrate for reaction } j, \ \ j \in J\}$

$I_r^j = \{i_r \mid i_r \in I \text{ is a regulator for reaction } j, \ \ j \in J\}$

The sets $M^+ = \{m^+\}$ and $M^- = \{m^-\}$ denote the indices for the activation and the inhibition elasticities, respectively, that can be applied to each enzyme by each metabolite. The regulatory elasticities of the postulated regulatory superstructure will be denoted as ϵ_{mji_r}, where m belongs to the index set $M = M^+ \cup M^-$. Therefore,

ϵ_{mji_r} will denote the regulatory elasticity of reaction j with respect to metabolite i and can be positive (for activation) or negative (for inhibition) with a fixed magnitude for each m.

The continuous variables of the model are the logarithmic deviations of the metabolite concentrations, z_i, the logarithmic deviations of the manipulated variables, q_k, the logarithmic deviations of the metabolic outputs, w_l, and the reaction rates, v_j.

A binary variable, y_{mji_r}, is associated with each regulatory elasticity, ϵ_{mji_r}. If a regulatory loop with an elasticity, ϵ_{mji_r}, is active in the network, y_{mji_r} is set to 1, otherwise it is zero. A binary variable, d_k, is associated with each manipulated variable, q_k. The introduction of these variables serves as a control on the number of the manipulated parameters that are allowed to vary. In many cases, practical experimental limitations allow only a limited number of simultaneous manipulations of enzyme activities.

The linearization procedure described earlier transforms the rate expressions, v_j, and metabolic output functions, h_l, to the following form:

$$v_j = v_{j,o}(1 + \sum_{i_s \in I_s^j} \epsilon_{ji}^s z_{i_s} + \sum_{m \in M} \sum_{i_r \in I_r^J} \epsilon_{mji_r}^r y_{mji_r} z_{i_r} + \sum_{k \in K} \pi_{jk} d_k q_k) \tag{4.8}$$

and

$$\begin{aligned} w_l = ln(h_l/h_{l,o}) \;=\; & \sum_{j \in J} \sum_{i_s \in I_s^j} \xi_{lj} \epsilon_{ji_s}^s z_{i_s} + \sum_{m \in M} \sum_{j \in J} \sum_{i_r \in I_r^j} \xi_{lj} \epsilon_{mji_r}^r y_{mji_r} z_{i_r} \\ & + \sum_{i \in I} \eta_{li} z_i + \sum_{k \in K} \sum_{j \in J} \xi_{lj} \pi_{jk} d_k q_k + \sum_{k \in K} \theta_{lk} d_k q_k \end{aligned} \tag{4.9}$$

The second and third terms on the right-hand side of equation (4.8) correspond to the dependence of the rates on the metabolites, and the fourth term corresponds to the dependence on the manipulated parameters. The first and the second terms in equation (4.9) correspond to the first term in equation (4.4) and the third, fourth, and fifth terms correspond to the second, third, and fourth terms in equation (4.4), respectively.

The presence of binary variables in the formulation of the problem introduces

bilinear products of continuous and binary variables in equations (4.8) and (4.9) which make the problem nonlinear. In order to circumvent these nonlinearities we follow the modeling approach applied in Psarris and Floudas (1990) using the idea proposed by Petersen (1971) and extended by Glover (1975). This is also discussed in chapter 7 of the book by Floudas (1995).

The basic idea is to introduce new continuous variables for each bilinear product:

$$u_{mji_r} = y_{mji_r} z_{i_r} \quad \forall (m, j, i_r) \tag{4.10}$$

and

$$g_k = d_k q_k \quad \forall k \tag{4.11}$$

and to introduce additional constraints, for each (m, j, i_r) and each k, which are described in Section 5.2.

4.5.1 *The Objective Function*

The process of interest which we optimize (maximize or minimize) can be any of the metabolic outputs or combination of them. Note that using equation (4.9) we can treat metabolite concentrations and rates as metabolic outputs. Therefore, in the case of having a single metabolic function w_l as objective we can express the objective function as:

$$\begin{aligned} f_{obj} = & \sum_{j \in J} \sum_{i_s \in I_s^j} \xi_{lj} \epsilon^s_{ji_s} z_{i_s} + \sum_{m \in M} \sum_{j \in J} \sum_{i_r \in I_r^j} \xi_{lj} \epsilon^r_{mji_r} u_{mji_r} \\ & + \sum_{i \in I} \eta_{li} z_i + \sum_{k \in K} \sum_{j \in J} \xi_{lj} \pi_{jk} g_k + \sum_{k \in K} \theta_{lk} d_k q_k \end{aligned} \tag{4.12}$$

Coupling between the various processes in the cell is sometimes a limitation in the performance of a pathway, even when classical metabolic engineering techniques are applied. However, we can study within the present framework the following questions: *If and how can we decouple the pathway of interest from other cellular*

processes by manipulating the regulatory structure of the pathway? If possible, how can this decoupling be achieved? The objective function that corresponds to this question will be formulated and a corresponding solution approach will be discussed in Section 6.

4.5.2 *Constraints*

The metabolic optimization problem is typically subject to one or more of the following types of constraints:

(i) Mass balance for each metabolite i

Equation (4.3) is the set of mass balance equations for the metabolites. Therefore, at steady-state, the left-hand side of equation (4.3) is set to zero, and, for each metabolite x_i, we have a constraint of the form:

$$\sum_{j\in J}\sum_{i_s\in I_s^j} n_{ij}\epsilon_{ji_s}z_{i_s}+\sum_{m\in M}\sum_{j\in J}\sum_{i_r\in I_r^j} n_{ij}\epsilon_{ji_r}u_{mji_r}+\sum_{i'\in I}\kappa_{ii'}z_{i'}+\sum_{j\in J}\sum_{k\in K} n_{ij}\pi_{jk}g_k+\sum_{k\in K}\lambda_{ik}g_k=0 \tag{4.13}$$

where the first two terms correspond to the first term in the right-hand side of equation (4.3), and the third, fourth, and fifth terms correspond to the second, third, and fourth terms in the right-hand side of equation (4.3), respectively.

(ii) Bounds on metabolites, manipulated variables, rates, and metabolic outputs

In modeling metabolic pathways it is extremely difficult to describe all relevant processes completely. Metabolic engineering of the pathway of interest will result in changes in metabolite concentrations, metabolic outputs of the pathway, and reaction rates. These changes will propagate into the rest of the cellular processes with unpredictable and, in many cases, undesirable effects.

The concentrations of metabolites should neither exceed toxicity levels nor be very low because it is possible to induce responses, such as stringent responses, that will alter qualitatively cellular activities that are not included in the model. Changes in the manipulated variables can have similar effects. Overexpression of enzymes can influence growth, and an excess of external substrate can result in toxic by-product

synthesis. Therefore, variables should be constrained within the bounds determined by their physiological ranges for the pathway of interest and by the available biological knowledge.

$$z_i^L \leq z_i \leq z_i^U \quad i \in I \tag{4.14}$$

$$q_k^L \leq q_k \leq q_k^U \quad k \in K \tag{4.15}$$

Note that variables z_i and q_k are the logarithmic deviations of the concentrations, x_i, and the logarithmic deviations of the parameters, p_k, respectively, from the steady-state value around which the linear model has been constructed. As a result, their lower bounds can take negative values.

The reaction rates of the pathway cannot be increased infinitely, and zero values for the fluxes are not physiologically acceptable in general. The minimum lower bound for a rate will be zero only if the metabolites produced by the corresponding reaction are provided externally. Reaction rate expressions, v_j, and metabolic outputs of the system, w_l, will generally be constrained within physiological bounds depending on the system under study. Therefore, the bounds of these variables will be of the form:

$$v_j^L \leq v_j \leq v_j^U \quad j \in J \tag{4.16}$$

$$w_l^L \leq w_l \leq w_l^U \quad l \in L \tag{4.17}$$

The variables w_l represent the logarithmic deviations of the metabolic outputs and, therefore, their lower bounds can be negative.

(iii) Constraints for the u_{mji_r} variables

The variables represented by u_{mji_r} are connected with the continuous variables, z_{i_r}, and the binary variables, y_{mji_r}, via the following conditions:

$$z_{i_r} - z_{i_r}^U(1 - y_{mji_r}) \leq u_{mji_r} \leq z_{i_r} - z_{i_r}^L(1 - y_{mji_r}) \tag{4.18}$$

$$i_r \in I_r^j,\ j \in J,\ m \in M$$

$$z_{i_r}^L y_{mji_r} \leq u_{mji_r} \leq z_{i_r}^U y_{mji_r} \tag{4.19}$$

$$i_r \in I_r^j,\ j \in J,\ m \in M$$

Note that the above constraints are linear in z_{i_r} and y_{mji_r}. It is interesting to examine the effect of these constraints.

If $y_{mji_r} = 1$, then they become:

$$z_{i_r} \leq u_{mji_r} \leq z_{i_r}$$
$$z_{i_r}^L \leq u_{mji_r} \leq z_{i_r}^U$$

and the first two constraints imply that $u_{mji_r} = z_{i_r}$ while the second two constraints simply provide bounds.

If $y_{mji_r} = 0$, then we have

$$z_{i_r} - z_{i_r}^U \leq u_{mji_r} \leq z_{i_r} - z_{i_r}^L$$
$$0 \leq u_{mji_r} \leq 0$$

and the second two constraints imply that $u_{mji_r} = 0$, while the first two constraints are relaxed since $z_{i_r} - z_{i_r}^U \leq 0$ and $z_{i_r} - z_{i_r}^L \geq 0$.

Similarly, for each of the g_k variables we have the following four inequality linear constraints:

$$q_k - q_k^U(1 - d_k) \leq g_k \leq q_k - q_k^L(1 - d_k) \quad k \in K \tag{4.20}$$

$$q_k^L d_k \leq g_k \leq q_k^U d_k \quad k \in K \tag{4.21}$$

for which a similar analysis holds.

(iv) Logical constraints

There are constraints based on the binary variables which are associated with the existence or nonexistence of various regulatory loops and the activation or deactivation of different continuously adjustable manipulated variables. An important logical constraint is one that forbids activation and inhibition of an enzyme by the same metabolite. Moreover, when we consider alternative loops with the same type of action (inhibition or activation) at different levels (e.g. low inhibition, high inhibition), only one of the values should be considered. These constraints appear in the

formulation for each (j, i_r):

$$\sum_{m^+ \in M^+} y_{m^+ j i_r} + \sum_{m^- \in M^-} y_{m^- j i_r} \leq 1 \quad \forall (j, i_r) \tag{4.22}$$

$$m^+ \in M^+,\ m^- \in M^-,\ j \in J,\ i_r \in I_r^j$$

The maximum number of the regulatory actions for each enzyme, $\mid I^j_{r,max} \mid$, and the maximum number of enzymes that each metabolite regulates, $\mid J^{i_r}_{max} \mid$, will, in general, impose one additional constraint for each j and one for each i:

$$\sum_{m \in M} \sum_{i_r \in I^j_{i_r}} y_{m j i_r} \leq \mid I^j_{r,max} \mid \quad j \in J \tag{4.23}$$

and

$$\sum_{m \in M} \sum_{j \in J} y_{m j i_r} \leq \mid J^{i_r}_{max} \mid \quad i_r \in I^j_{i_r} \tag{4.24}$$

The simultaneous manipulation of the variables q_k will be subject to the following constraint:

$$\sum_{k \in K} d_k \leq \mid K_{max} \mid \tag{4.25}$$

where $\mid K_{max} \mid$ is the maximum number of the manipulated variables that can be modified simultaneously. This constraint arises from practical and physiological limitations; and $\mid K_{max} \mid$ varies from system to system.

Integer cuts are also introduced such that by solving the proposed mathematical model in an iterative way we can exclude all the previous solutions so that we can calculate the next best solution. In this case, when we solve for the n-th best solution we have to include $n - 1$ constraints of the form:

$$\sum_{i \in B_i} y_i - \sum_{i \in NB_i} y_i \leq \mid B_i \mid -1 \tag{4.26}$$

$$B_i = \{i \mid y_i = 1\}$$

$$NB_i = \{i \mid y_i = 0\}$$

where $\mid B_i \mid$ is the cardinality of the set B_i (i.e. the number of the elements in the set).

The proposed mathematical formulation involves continuous and binary variables. The problem, in its initial formulation, features bilinearities as products of continuous and binary variables. By introducing a new continuous variable and four linear inequality constraints for each bilinear term, the final formulation involves only linear terms and, therefore, becomes a MILP formulation. The solution of the MILP will provide the desired optimal regulatory structure. Because of its linear nature, the problem is convex and a global solution is guaranteed. The model allows constraint flexibility in relation with the physical system it describes and is designed to encompass feasible manipulations within the set of current methods for metabolic and protein engineering.

4.6 Computational Studies

The proposed approach will be illustrated using the aromatic amino acid biosynthetic network in bacteria as an example system. Four specific problems will be postulated and solved with the proposed mixed integer linear optimization framework. The pathway and the original regulatory structure are presented in Fig. 1. The pathway has 8 regulatory loops all of which are feedback inhibitory loops. In order to derive a linear model for the pathway we started from the nonlinear model for this presented by Schlosser and Bailey (1990). The nonlinear model and the parameters for the linear model are presented in Appendix B.

In the following examples, the following bounds on the logarithmic deviations of the metabolite concentrations were imposed:

$$-2 \leq z_i \leq 2 \tag{4.27}$$

which implies that we allow the concentrations of the metabolites to vary between 13.5% and 639% of the reference steady state. These wide bounds might allow con-

centrations that can be toxic to the organism or that induce stress responses that will affect other cellular activities. We use such wide bounds in the interest of exploring the structure of this example and the characteristics of its optimal solution. If needed to accomodate physiological limitations, tighter constraints on metabolite concentrations can certainly be considered within this formalism.

The only physiological constraint which will be introduced constrains the specific growth rate, μ, to its reference steady value:

$$w_\mu = 0 \tag{4.28}$$

The only consideration in the model used in this example for the effects of manipulation of the aromatic amino acid biosynthesis network on the rest of the organism's metabolism is the dependence of the specific growth rate on the aromatic amino acid concentration levels (see Appendix B). By imposing the above constraint we essentially constrain the solutions to the ones that are consistent with other requisite coupled metabolic activities of the cell.

The first example deals with the modification of the existing regulatory and activity structure. The second, the third, and the fourth examples assume that the metabolic pathway has no regulation at the outset, and consider what regulatory connections, with what strength, should be introduced so as to optimize the objective.

The procedure for solving the examples, equally applicable to any other metabolic network optimization problem of the class formulated earlier, was implemented using the high-level modeling language GAMS (general algebraic modeling system), into an algorithmic procedure named METAOPT (METAbolic network OPTimization). The procedure accepts the mathematical model and the postulated regulatory superstructure as a set of matrices and is interfaced with CPLEX, a mixed-integer linear programming solver. At each solution, the optimal regulatory structure is used to form an integer cut constraint, and the problem is solved again for the next best structure. This way a sequence of several solutions is generated. The following com-

putational studies were run on a HP/730 workstation with a Unix-based operating system.

4.6.1 *Problem 1*

The question addressed in the first problem can be stated as follows:

Which of the existing regulatory loops should be inactivated and what should be the associated changes in the enzyme expression levels to maximize the phenylalanine selectivity?

The phenylalanine selectivity is defined here as the rate of phenylalanine production divided by the overall rate of all aromatic amino acids (phenylalanine, tyrosine, and tryptophan) production:

$$S_{phe} = \frac{v_4}{v_4 + v_5 + v_6} \tag{4.29}$$

The initial number of the regulatory loops is equal to eight, which can be either active or inactive and in any possible combination. Therefore, there are $2^8 = 256$ alternative regulatory structures. The six enzymes of the pathway are the continuously adjustable variables for which we set the bounds:

$$0 \leq q_k \leq ln(2) \quad k = 1, \ldots, 6 \tag{4.30}$$

The zero value for the lower bound means that enzyme downregulation is not feasible. On the other hand, the value $ln(2)$ for the upper bound allows enzyme overexpression up to twice the level of the reference state. In practice, overexpression of an enzyme can result in higher levels, but, since we study a linear model for which the enzyme levels are the inputs, we do not permit large deviations in the input values so as better to stay within range in which the linear model is a good approximation to the original nonlinear model. Only the six enzymes are considered as the continuously adjustable inputs, and the concentrations of the precursors (metabolites feeding into the reaction network considered here) remain constant at their reference values. Therefore, the

following constraints are needed to inactivate changes in the precursors:

$$d_k = 0 \quad k = 7, 8, 9 \tag{4.31}$$

The objective function to be maximized is the ratio of the phenylalanine production rate divided by the overall rate of the aromatic amino acids production. Equation (4.29) in linearized form is written as in equation (4.4):

$$w_{S_{phe}} = ln\left(\frac{S_{phe,opt}}{S_{phe,o}}\right) = \Xi_{S_{phe}}\mathcal{E}\mathbf{z} + \mathcal{H}_{S_{phe}}\mathbf{z} + \Xi_{S_{phe}}\mathbf{\Pi q} + \Theta_{S_{phe}}\mathbf{q} \tag{4.32}$$

where

$$\Xi_{S_{phe}} = [0,\ 0,\ 0,\ 0.572665,\ -0.439811,\ -0.132855]$$

and $\mathcal{H}_{S_{phe}}$ and $\Theta_{S_{phe}}$ are zero vectors. The selectivity for the reference state is:

$$S_{phe,o} = 0.427335.$$

No improvement in this value could be achieved only by enzyme overexperession, without having an effect on the growth rate.

The MILP optimization model discussed in section 5 was solved and four alternative regulatory structures were identified corresponding to the optimal value for the phenylalanine selectivity. The problem consisted of 65 variables (48 continuous, 17 binary) and 94 equations and the first optimal solution was found within 0.26 CPU s. The structures are presented in Figure 2 (cases a-d) and the value of the objective function for all four solutions is:

$$S_{phe,opt} = 0.605883$$

The overexpressed enzyme levels and metabolite concentration levels associated with each regulatory structure are presented in Table 1. The concentration of the metabolites are the same for all four solutions. The results suggest that the optimal selectivity can be achieved by inactivation of at least four regulatory connections (v_1 by

DAHP, v_3 by TYR, v_4 by PHE, and v_6 by TRP) and that three enzymes should be overexpressed, specifically the enzymes that catalyze reactions 1, 3 and 4.

Four more alternative regulatory structures correspond to a value of 0.601315 for the selectivity. These structures are also presented in Figure 2 (cases e-h) and the associated enzyme and metabolite levels are presented in Table 2. Again, the three enzymes that should be overexpressed correspond to reactions 1, 3, and 4. However, the minimum number of the regulatory loops that should be inactivated is three (v_1 by DAHP, v_3 by TYR, and v_4 by PHE). Case (f) among these results is attractive since implementing it requires less effort than other solutions presented from the genetic and protein engineering point of view.

Examination of the eight regulatory structures indicates that the phenylalanine selectivity can be significantly improved, while maintaining constant specific growth rate, by inactivating at least three regulatory structures and overexpressing three enzymes. While inactivation of the inhibition of v_4 by PHE is quite obvious, the rest of the manipulations, all considered subject to the constraint on the growth rate, cannot be easily anticipated *a priori*.

In this problem which begins with a specified nonlinear model including enzyme regulation, the metabolic design strategy determined by the MILP optimization of an approximate log–linear model can be tested using the original nonlinear metabolic model. Such calculations for the strategies designated a-d give $S_{phe} = 0.66730$, and $S_{phe} = 0.86544$ for the strategies designated e-h. These selectivities substantially exceed that for the reference state, indicating that, as desired, the approach described here provides useful guidance towards effective metabolic design. It should be noted that the full nonlinear model, when subjected to the modifications computed using the log–linear model, no longer displays all the properties of the log–linear model. In particular, all of the constraints imposed are no longer exactly satisfied. For example, specific growth rate is decreased somewhat when these strategies are applied to the original nonlinear model. The changes observed are well within an acceptable range from a biological and process point of view. The purpose of this work is to provide reasonable guidance for starting optimization of a metabolic system. Efforts to seek

an optimum for a real system will always require further experimental refinement.

4.6.2 *Problem 2*

In this problem the aromatic amino acid pathway is considered without any regulatory connections, and we postulate a regulatory superstructure such that any of the first six reactions can be inhibited by any of the six metabolites with an inhibition strength of -0.75 or -0.075. We allow only two regulatory connections for each enzyme and we allow three enzyme level manipulations. The questions addressed in this problem are the following:

In order to maximize the phenylalanine selectivity:

(i) which pair of metabolites should inhibit each reaction?

(ii) what are their inhibition strengths?

(iii) which three enzymes should be overexpressed?

and

(iv) what should be their expression levels?

The use of discrete values for the level of the inhibition allows preservation of system linearity and provides qualitative information about the order of magnitude of the regulatory loop strength. This problem can provide us with information about the maximum selectivity which can be achieved for this model.

In the formulation of the problem two regulatory elasticity matrices are introduced:

$$\mathcal{E}_1^r = \{\epsilon_{1ji_r} = -0.75 \mid 1 \leq j \leq 6 \text{ and } 1 \leq i_r \leq 6\}$$

$$\mathcal{E}_2^r = \{\epsilon_{2ji_r} = -0.075 \mid 1 \leq j \leq 6 \text{ and } 1 \leq i_r \leq 6\}$$

The problem was solved with the same bounds on the continuous variables. Moreover, we introduced three additional logical constraints. The first one takes the form:

$$\sum_{m^+ \in M^+} y_{m^+ji_r} + \sum_{m^- \in M^-} y_{m^-ji_r} \leq 1 \tag{4.33}$$

$$m^+ \in M^+,\ m^- \in M^-,\ j \in J,\ i_r \in I_r^j$$

and allows for only one level of inhibition chosen from the two different orders of magnitude allowed. The second one is

$$\sum_{m\in M}\sum_{i_r\in I_{i_r}^j} y_{mji_r} \leq 2 \tag{4.34}$$

and allows for only two regulatory connections for each enzyme. The third one is

$$\sum_{m\in M}\sum_{j\in J} y_{mji_r} \leq 6 \quad i_r \in I_{i_r}^j \tag{4.35}$$

and allows any of the six enzymes to be regulated by any of the six metabolites. Under these constraints the number of the alternative regulatory structures we can build around the pathway for this problem are 2^{47}.

The resulting mathematical formulation accounts for 193 variables (112 continuous, 81 binary) and 393 constraints. The problem was solved and the optimal solution, $S_{phe,opt}$, found in 19.56 CPU s, was 1.05938. This value is greater than the actual upper bound for the phenylalanine selectivity (4.29), because we used as objective function the logarithmic selectivity (4.32) which is a linearized approximation of equation (4.29), and it is not subject to any upper bound. On the other hand, the value for the optimal selectivity is 0.818672, if the calculated from the solution linearized rate expressions used in equation (4.29). This solution suggests that, for the parameters chosen for the system and with an optimized regulatory structure, we can increase the selectivity up to 95% by simply manipulating only three enzymes without affecting the specific growth rate.

The problem has multiple regulatory structures that result in the same optimal objective value. However, we are interested in the structures that achieve the optimal performance but have the minimum number of regulatory loops since the creation of these loops is very difficult. Therefore we formulate a new objective function:

$$min \sum_{m\in M}\sum_{j\in J}\sum_{i_r\in I_{i_r}^j} y_{mji_r} \tag{4.36}$$

with the additional equality constraint:

$$w_{phe} = ln\left(\frac{S_{phe,opt}}{S_{phe,o}}\right) \tag{4.37}$$

where $w_{phe} = 0.908115$

The solution to this problem resulted in only four alternative regulatory structures presented in Figure 3; the corresponding values for the continuous variables are presented in Table 3. The minimum number of loops is equal to two, the elasticity for each loop is -0.75, and the three enzymes that must be manipulated are the enzymes that catalyze reactions 1, 4, and 6. Moreover, we observe that the regulatory metabolites are only two, CHR and PHE, and the enzymes that should be regulated are the ones that catalyze reactions 5 and 6. The number of different structures is equal to the number of all possible combinations of two reactions regulated by two metabolites, with only one regulatory connection allowed for each reaction. Therefore, the solution suggests that the enzymes that catalyze reactions 5 and 6 in the pathway should be engineered, if possible, so that both will be inhibited by either CHR and PHE. Once this regulatory structure has been succesfully realized, the enzymes that catalyze reactions 1, 4, and 6 should be cloned and overexpressed simultaneously.

4.6.3 *Problem 3*

This problem is the same as the previous one except that we would like to design a regulatory structure for which only *enzyme activation* is allowed. We consider again that any of the first six reactions can be activated by any of the six metabolites with strength 0.75 or 0.075, and only two regulatory connections for every enzyme and three enzyme manipulations are allowed. The questions addressed are:

In order to maximize the phenylalanine selectivity:

(i) which pair of metabolites should activate each reaction?

(ii) what are their activation strengths?

(iii) which three enzymes should be overexpressed?

(iv) what should their expression levels be?

The mathematical formulation is the same as for problem 2 except that the elasticity matrices are:

$$\mathcal{E}_1^r = \{\epsilon_{1ji_r} = 0.75 \mid 1 \leq j \leq 6 \ \ and \ \ 1 \leq i_r \leq 6\}$$

$$\mathcal{E}_2^r = \{\epsilon_{2ji_r} = 0.075 \mid 1 \leq j \leq 6 \ \ and \ \ 1 \leq i_r \leq 6\}$$

The size of this problem is the same as the size of the previous problem, and it consists of 393 constraints and 193 variables (112 continuous, 81 binary). The optimal solution to this problem was found, within 249 CPU s, to be 1.05938 which is the same as before. The value for the optimal selectivity, calculated from the ratio of the linearized rate expressions was again 0.818672, and multiple regulatory structures were found to correspond to the optimal value. Therefore, we solved the problem with the objective function (4.36) and the additional equality constraint (4.37).

The minimum number of regulatory activation loops that correspond to the optimal selectivity is equal to 3, and 37 alternative regulatory structures were identified. The regulatory elasticities for every structure were equal to 0.75. Analysis of the alternative structures indicates that the enzymes that catalyze reactions 4, 5, and 6, should be activated (this is true in all 37 cases) and that these steps should be the target of any attempt to engineer the regulatory features of the pathway. Moreover, out of the 60 possible combinations only 12 combinations of enzyme levels should be manipulated. These combinations and the number of the number of the alternative regulatory structures for each combination are presented in Table 4. In Figure 4 the structures that correspond to the manipulation of enzyme levels for reactions 1, 2, and 4, are presented, and the associated changes in the expression levels and the metabolite concentrations are given in Table 5.

4.6.4 *Problem 4*

This last problem is a combination of Problems 2 and 3 and is formulated as follows:

To maximize the phenylalanine selectivity:

(i) which pair of metabolites should regulate each reaction?

(ii) what should be the type of regulation (i.e. activation or inhibition)?

(iii) what is the strength of the regulation?

(iv) which three enzymes should be overexpressed?

(v) what should their expression levels be?

We have four elasticity matrices:

$$\mathcal{E}_1^r = \{\epsilon_{1ji_r} = -0.75 \mid 1 \leq j \leq 6 \ \textit{and} \ 1 \leq i_r \leq 6\}$$

$$\mathcal{E}_2^r = \{\epsilon_{2ji_r} = -0.075 \mid 1 \leq j \leq 6 \ \textit{and} \ 1 \leq i_r \leq 6\}$$

$$\mathcal{E}_1^r = \{\epsilon_{3ji_r} = 0.75 \mid 1 \leq j \leq 6 \ \textit{and} \ 1 \leq i_r \leq 6\}$$

$$\mathcal{E}_2^r = \{\epsilon_{4ji_r} = 0.075 \mid 1 \leq j \leq 6 \ \textit{and} \ 1 \leq i_r \leq 6\}$$

and the rest of the constraints and the bounds are the same as in the two previous problems. The number of alternative regulatory structures are 2^{59}. The problem featured 681 constraints and 337 variables (184 continuous, 153 binary), and the first optimal solution was found in 156 CPU s.

As before, the value for the optimal selectivity was equal to 1.05938, and the ratio of the linearized rates was 0.818672 again. This optimal value corresponds to multiple regulatory structures and combinations for enzyme manipulations. Therefore, we solved the problem again in order to find the minimum number of the regulatory loops that correspond to this optimal value.

The minimum number of loops was 2, the regulatory elasticities were 0.75 and -0.75 for the activation and the inhibition loops respectively, and the number of structures with only two regulatory loops was 10. As expected, the four structures found in Problem 2 are also included in this set of structures. In Figure 5 the six new regulatory structures are depicted, and in Table 6 the corresponding values of the expression level of the enzymes are presented.

In all of the optimal regulatory structures involving activation and inhibition we

observe three main characteristics:

(i) Only the enzymes that catalyze reactions 5 and 6 are regulated;

(ii) The enzyme that catalyze reaction 5 is always inhibited and the enzyme that catalyzes reaction 6 is always activated;

(iii) DAHP, TRP, and TYR act as activators, and CHR and PHE act as inhibitors.

These observations clearly suggest that, if possible, the enzyme that catalyzes reaction 5 should be designed so that it is inhibited by CHR or PHE, and the enzyme for reaction 6 should be designed so that it is activated by DAHP or TRP or TYR. Moreover, if modification in the regulatory structure is accompanied by overexpression of the enzymes for reactions 1, 4, and 6, any combination of the just listed regulatory patterns will be successful, except for DAHP-activated reaction 6. If, on the other hand, the enzymes for reactions 1, 2, and 4 are overexpressed, the regulatory structure should be designed so that reaction 6 will always be activated by DAHP.

4.7 Discussion

Linear models have been used within MCA in order to characterize and identify the enzymes that limit the performance of metabolic pathways. Such linear models can provide, for each metabolic function, its control coefficients, defined as the fractional changes of the metabolic function expected for a unit fractional change in the amount of each enzyme or external effector participating in a given pathway. Experimental, theoretical, and computational analyses have shown that the existence of a single enzyme which limits a metabolic process should not be presumed, and overexpression of a limiting enzyme results in a shift of the limitation to other steps in the pathway. Many of these studies have attributed these responses to the coupling between different pathways through regulatory connections and the fact that they share metabolites (Kacser and Burns, 1973; Savageau, 1976; Cornish-Bowden and Cardenas, 1990; Bailey, 1991).

One effective way to manipulate metabolic pathways is to implement pathways that are desensitized and decoupled from other cellular activities, and limited by

a minimum number of enzymes. These objectives can be formulated and studied following the approach described above.

If in equations (4.3) and (4.4) we set a manipulated variable, q_k, equal to one and the rest of the manipulated variables equal to zero, the metabolic functions, $\mathbf{w}$, will be equal to the control coefficients of these functions with respect to this manipulated variable. Then, the problem of adjusting the control coefficients close to a desired value, $\mathbf{w}_f$, can be written as:

$$min(\mathbf{w} - \mathbf{w}_f)^T(\mathbf{w} - \mathbf{w}_f) \tag{4.38}$$

subject to the same constraints introduced in the above Mathematical Formulation. In the special case that $\mathbf{w}_f = \mathbf{0}$, we study the problem of decoupling functions $\mathbf{w}$ from other cellular processes for which q_k is an output. This problem is nonlinear in the objective function and combines both discrete and continuous variables and as a result can be formulated as a Mixed-Integer Nonlinear Programming (MINLP) problem. An optimization framework that can address the analysis and synthesis problem of metabolic pathways for nonlinear models is presently being developed.

The results presented above for the analysis and synthesis of the regulatory structure of the aromatic amino acid pathway do not take into account the stability and the dynamic characteristics of the network with alternative regulatory structures. Even if a system is stable, obtaining desirable transient responses associated with changes in the manipulated variables and the dynamic responses to fluctuations in the parameters of the system, have been proposed as criteria for optimization of metabolic processes (Torres, 1994). The approach introduced in this work cannot explicitly formulate objectives associated with such dynamic characteristics, but can allow the generation of a sequence of optimal alternative regulatory structures which can be reordered based on their dynamic performance using simulation analysis and process control tools.

4.8 Conclusions

In this chapter we present a novel approach to the analysis and synthesis of metabolic pathways. The problem of designing the regulatory structures built around a given metabolic reaction network was formulated as a MILP optimization problem. A synthesis approach has been proposed which assumes that the metabolic pathway of interest has no regulation, and considers which regulatory structure optimizes the objective. Assuming that a linear model for the pathway of interest is given, integer variables were introduced to denote the existence or non-existence of the postulated regulatory loops.

The approach, implemented in METAOPT, was applied to the study of the aromatic amino acid pathway in bacteria. The solution allows the identification of the regulatory structures and the associated changes in the enzyme levels that result in an optimal phenylalanine selectivity. Multiple regulatory structures were found to correspond to optimal solutions. The consistent patterns identified within these solutions helped in the postulation of design principles that were effective when applied to the full nonlinear model on which the first example is based. For the other examples, in which new patterns of enzyme regulation were considered as options, there is not a unique transformation from the log–linear model used here to a corresponding nonlinear model. The implications of this will be examined in the future employing the MINLP framework now being developed.

4.9 References

Bailey, J. E., "Toward a Science of Metabolic Engineering," *Science*, **252**, 1668-1675 (1991).

Cornish-Bowden, A., and M.L. Cardenas (Eds.), "Control of Metabolic Processes," NATO ASI Series A: Lifesciences Vol. 190, Plenum Press, New York (1990).

Fell, D.A., "Metabolic Control Analysis - A Survey of its Theoretical and Experimental Development," *Biochem. J.*, **152**, 313-330 (1992).

Floudas, C.A., "Nonlinear and Mixed Integer Optimization: Fundamentals and Applications," Oxford University Press, New York, in press (1995).

Fredrickson, A. G. "Formulation of Structured Growth Models," *Biotechnol. Bioeng.*, **28**, 1481-1486 (1976).

Gallazo, J. L., and J. E. Bailey, "Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces cerevisiae*," *Enzyme Microb. Technol.*, **12**,162-172 (1990).

Glover, F., "Improved Linear Programming Formulations of Nonlinear Integer Problems," *Mgmt. Sci.*, **22**, 445 (1975).

Hatzimanikatis, V., K. H. Lee, W. A. Renner, and J. E. Bailey, "A Mathematical Model for the G1/S Transition of the Mammalian Cell Cycle", *Biotech. Lett.* (in press).

Heinrich, R., H.-G. Holzhütter, and S. Schuster, "A Theoretical Approach to the Evolution and Structural Design of Enzymatic Networks," *Bull. Math. Biol.*, **49**, 539-595 (1987).

Heinrich, R., and T. A. Rapoport, "A Linear Steady-State Treatment of Enzymatic Chains," *Eur. J. Biochem.*, **42**, 89-95 (1974).

Joshi, A., and B.O. Palsson, "Metabolic Dynamics in the Human Red Cell. Part I – A Comprehensive Kinetic Model," *J. theor. Biol.*, **141**, 515-528 (1989).

Kacser, H., and J. A. Burns, "The Control of Flux," *Symp. Soc. Exp. Biol.*, **27**, 65-104 (1973).

Katsumata, R., and M. Ikeda, "Hyperproduction of Tryptophan in *Corynebac-*

terium glutamicum by Pathway Engineering," *Bio/Technology*, **11**, 921-925 (1993).

Lee, S. B., and J. E. Bailey, "A Matheamtical Model for Lambda-dv Plasmid Replication - Analysis of Wild-Type Plasmid," *Plasmid*, **11**, 151-165 (1984a).

Lee, S. L., and J. E. Bailey, "Genetically Structured Models for lac Promoter-Operator Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids - lac Operator Function," *Biotechnol. Bioeng.*, **26**, 1372-1382 (1984b).

Lee, S. L., and J. E. Bailey, "Genetically Structured Models for lac Promoter-Operator Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids - lac Promoter Function," *Biotechnol. Bioeng.*, **26**, 1383-1389 (1984c).

Majewski, R.A., and M.M. Domach, "Simple Constrained-Optimization View of Acetate Overflow in *E. coli*," *Biotechnol. Bioeng.*, **35**, 732-738 (1990).

Marr, A. G., "Growth rate of *Escherichia coli*", *Microbiol. Rev.*, **55**, 316-333 (1991).

Petersen, C.C., "A Note on Transforming the Product of Variables to Linear Forms in Linear Programs," Purdue University, Working paper (1971).

Psarris, P., and C.A. Floudas, "Improving Dynamic Operability in MIMO Systems with Time Delays," *Chem. Eng. Sci.*, **45**, 3505-3524 (1990).

Ramkrishna, D., "A Cybernetic Respective of Microbial Growth," *ACS Symp. Ser.*, **207**, 161-178 (1983).

Reder, C., "Metabolic Control Theory: A Structural Approach," *J. theor. Biol.*, **135**, 175-201 (1988).

Regan, L., D.L. Bogle, and P. Dunnill, "Simulation and Optimization of Metabolic Pathways," *Comp. Chem. Engng*, **17**, 627-637 (1993).

Sanwal, B. D., M. Kapoor, H. W. Duckworth, "The Regulation of Branched and Converging Pathways," *Curr. Top. Cell. Regul.*, **3**, 1-115 (1971).

Savageau, M. A., "Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology," Reading, MA, Addison–Wesley (1976).

Schlosser, P. M., and J. E. Bailey, "An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways," *Math. Biosci.*, **100**, 87-114 (1990).

Schlosser, P. M., T. Holcomb, and J. E. Bailey, "Determining Metabolic Sensitivity

Coefficients Directly from Experimental Data," *Biotechnol. Bioeng.*, **41**, 1027-1038 (1993).

Schuster, S., and R. Heinrich, "Minimization of Intermediate Concentrations as a Suggested Optimality Principle for Biochemical Networks I. Theoretical Analysis," *J. Math. Biol.*, **29**, 425-442 (1991).

Shuler, M. L., and M. M. Domach, "Mathematical Models of the Growth of Individual Cells – Tools for Testing Biochemical Mechanisms," *ACS Symposium Series*, **207**, 93-133 (1983).

Starbuck, C., and Lauffenburger, D. A., "Mathematical Model for the Effects of Epidermal Growth-Factor Receptor Trafficking Dynamics on Fibroblast Proliferation Responses," *Biotech. Prog.*, *8*, 132-143 (1992).

Stephanopoulos, G., and J. J. Vallino, "Network Rigidity and Metabolic Engineering in Metabolite Overproduction," *Science*, **252**, 1675-1681 (1991).

Torres, N. V., "Application of Transition Time of Metabolic Systems as a Criterion for Optimization of Metabolic Processes," *Biotechnol. Bioeng.*, **44**, 291-296 (1994).

Varma, A., and B.O. Palsson, "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use," *Bio/Technology*, **12**, 994-998 (1994).

Voit, E. O., "Optimization in Integrated Biochemical Systems," *Biotechnol. Bioeng.*, **40**, 572-582 (1992).

4.10 Appendices

APPENDIX A

In this appendix the logaritmic transformation used for the linear description of metabolic systems will be presented in detail. Consider the nonlinear dynamical system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x};\mathbf{p})$$

where $\mathbf{x}$ is the n-dimensional independent variable vector and $\mathbf{p}$ is the s-dimensional parameter vector. Let $\mathbf{x}_o$ be a steady state, or one of the steady states in case of steady state multiplicity, that corresponds to the parameter vector values $\mathbf{p}_o$, and with nonzero, positive elements. Linearization around this steady state will result in the linear system:

$$\frac{d(\mathbf{x}-\mathbf{x}_o)}{dt} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}(\mathbf{x}-\mathbf{x}_o) + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o}(\mathbf{p}-\mathbf{p}_o)$$

If we define the matrices $\mathbf{X}_o$ and $\mathbf{P}_o$ to be the diagonal matrices with diagonal elements $X_{o,ii} = x_{o,i}$ and $P_{o,ii} = p_{o,i}$, respectively. Then, for the linear system we can write:

$$\mathbf{X}_o^{-1}\frac{d(\mathbf{x}-\mathbf{x}_o)}{dt} = \mathbf{X}_o^{-1}\left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o\mathbf{X}_o^{-1}(\mathbf{x}-\mathbf{x}_o) + \mathbf{X}_o^{-1}\left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{P}_o\mathbf{P}_o^{-1}(\mathbf{p}-\mathbf{p}_o)$$

In the above equation we can redefine the following vectors for the scaled variables and parameters:

$$\mathbf{z} = \mathbf{X}_o^{-1}(\mathbf{x}-\mathbf{x}_o) \Rightarrow z_i = \frac{x_i - x_{o,i}}{x_{o,i}}$$

$$\mathbf{q} = \mathbf{P}_o^{-1}(\mathbf{p}-\mathbf{p}_o) \Rightarrow q_i = \frac{p_i - p_{o,i}}{p_{o,i}}$$

Given the above definitions we can also observe that:

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o = \left.\frac{\partial \mathbf{f}}{\partial(\mathbf{x}-\mathbf{x}_o)}\right|_{\mathbf{x}_o,\mathbf{p}_o}\mathbf{X}_o \Rightarrow \left.\frac{\partial f_i}{\partial(x_j - x_{o,j})}x_{o,j}\right|_{\mathbf{x}_o,\mathbf{p}_o} = \left.\frac{\partial f_i}{\partial z_j}\right|_{\mathbf{x}_o,\mathbf{p}_o}$$

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{p}}\right|_{\mathbf{x}_o,\mathbf{p}_o} \mathbf{P}_o = \left.\frac{\partial \mathbf{f}}{\partial(\mathbf{p}-\mathbf{p}_o)}\right|_{\mathbf{x}_o,\mathbf{p}_o} \mathbf{P}_o \Rightarrow \left.\frac{\partial f_i}{\partial(p_j - p_{o,j})} p_{o,j}\right|_{\mathbf{x}_o,\mathbf{p}_o} = \left.\frac{\partial f_i}{\partial q_j}\right|_{\mathbf{x}_o,\mathbf{p}_o}$$

On the other hand, for any logarithmic function of variable y we can write for up to first-order approximation the Taylor series around a reference value, y_o:

$$ln(y) = ln(y_o) + \frac{y - y_o}{y_o} \Rightarrow ln\left(\frac{y}{y_o}\right) = \frac{y - y_o}{y_o}$$

Therefore, for the scaled variables, $\mathbf{z}$, and the scaled parameters, $\mathbf{q}$, we can write:

$$z_i = ln\left(\frac{x_i}{x_{o,i}}\right)$$

and

$$q_i = ln\left(\frac{p_i}{p_{o,i}}\right)$$

Finally, by defining

$$g_i = \frac{f_i}{x_{o,i}}$$

we can write the linearized system for the logarithmic deviations, noting that $[\mathbf{z}_o, \mathbf{q}_o] = [\mathbf{0}, \mathbf{0}]$:

$$\frac{d\mathbf{z}}{dt} = \left.\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right|_{\mathbf{0},\mathbf{0}} \mathbf{z} + \left.\frac{\partial \mathbf{g}}{\partial \mathbf{q}}\right|_{\mathbf{0},\mathbf{0}} \mathbf{q}$$

In the equation (4.3) for example

$$\left.\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right|_{\mathbf{0},\mathbf{0}} = \mathbf{N}\mathcal{E} + \mathcal{K}$$

and

$$\left.\frac{\partial \mathbf{g}}{\partial \mathbf{q}}\right|_{\mathbf{0},\mathbf{0}} = \mathbf{N}\mathbf{\Pi} + \mathbf{\Lambda}$$

APPENDIX B

The rate expressions for the aromatic amino acid pathway are taken from Schlosser and Bailey (1990). Here we consider only the aromatic amino acid biosynthesis reactions as an isolated, model subsystem of overall metabolism. In particular the glucose catabolic reactions considered by Schlosser and Bailey (1990) are not included here. The parameter values used for the dissociation constants are the same as in Schlosser and Bailey (1990) where the references for the estimation of these parameters can be found. The values for $v_{j,max}$ have been adjusted to give steady-state values similar to those found in bacterial cells for $[G6P] = 0.8\ mM$, $[PEP] = 0.1\ mM$, $[ATP] = 2.5\ mM$, $[ADP] = 0.4271\ mM$, and $[AMP] = 0.0729\ mM$. The rate expressions for the 6 enzymatically catalyzed reactions in the pathway are:

$$v_1 = v_{m,1} \frac{\dfrac{0.79}{1+[PHE]/53} + \dfrac{0.2}{1+[TYR]/40} + \dfrac{0.01}{1+[TRP]/16}}{\left(\dfrac{0.0002}{[PEP][G6P]} + \dfrac{0.006}{[PEP]}\right)(1+50[DAHP]) + \dfrac{0.1}{[G6P]} + 1}$$

$$v_2 = v_{m,2} \frac{[DAHP][PEP][ATP]}{(2+[DAHP])\,(0.00867+[PEP])\,(0.9281+[ATP])}$$

$$v_3 = v_{m,3} \frac{[CHR]}{(2+[CHR])\,(1+[PHE]/50)\,(1+[TYR]/40)}$$

$$v_4 = v_{m,4} \frac{[PHP]}{(1+[PHP])\,(1+[PHE]/50)}$$

$$v_5 = v_{m,5} \frac{[PHP]}{(1+[PHP])}$$

$$v_6 = v_{m,6} \frac{[G6P][CHR][ATP]}{(1.269+[G6P])(2+[CHP])(0.9821+[ATP])(1+[TRP]/16)}$$

where:

$$\mathbf{V}_m^T = [710,\ 22,\ 474,\ 64,\ 10.5,\ 28]$$

We also have 3 expressions that account for the incorporation of the amino acids into biomass:

$v_1' = 54\mu, \quad v_2' = 131\mu, \quad v_3' = 176\mu$

and 6 expressions that account for the dilution brought about by increases in the biomass:

$$v_4' = \mu[DAHP], \quad v_5' = \mu[CHR], \quad v_6' = \mu[PHP]$$

$$v_7' = \mu[PHE], \quad v_8' = \mu[TYR], \quad v_9' = \mu[TRP]$$

The growth function, μ, used is:

$$\mu = 0.014 \frac{Y[PHE][TYR][TRP][PEP]}{(0.25+Y)(18+[PHE])(13+[TRY])(5+[TRP])(0.005923+[PEP])}$$

where

$$Y = \frac{[ATP]+0.5[ADP]}{[ATP]+[ADP]+[AMP]}$$

The mass balance equations for each of the metabolites in the aromatic amino acid pathway is given by

$$\begin{aligned}
DAHP: \quad 0 &= v_1 - v_2 - v_4' \\
CHR: \quad 0 &= v_2 - v_3 - v_6 - v_5' \\
PHP: \quad 0 &= v_3 - v_4 - v_5 - v_6' \\
PHE: \quad 0 &= v_4 - v_3' - v_7' \\
TYR: \quad 0 &= v_5 - v_2' - v_8' \\
TRP: \quad 0 &= v_6 - v_1' - v_9'
\end{aligned}$$

The stable stady-state at which the linear model was constructed is:

$$\mathbf{x}_o^T = [3.41404,\ 30.3097,\ 0.609277,\ 263.660,\ 321.496,\ 82.6866]$$

and the values for the rates at these values are:

$$\mathbf{v}_o^T = [\ 9.340355 \quad 9.310362 \quad 7.843243 \quad 3.862561 \quad 3.975329]$$

$$\mathbf{v}_o'^T = [\ 1.200839\ \ 0.474409\ \ 1.150880\ \ 1.546221\ \ 0.029994$$
$$0.266281\ \ 0.005353\ \ 2.316340\ \ 2.824449\ \ 0.726430]$$

For the matrix $\mathbf{N}$ in equation (4.1) we have

$$\mathbf{N} = \begin{bmatrix} 2.735867 & -2.727082 & 0 & 0 & 0 & 0 \\ 0 & 0.307175 & -0.258770 & 0 & 0 & -0.039619 \\ 0 & 0 & 12.873042 & -6.339585 & -6.524671 & 0 \\ 0 & 0 & 0 & 0.014650 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.012365 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.014523 \end{bmatrix}$$

The elasticity matrices at the steady-state are:

$$\mathcal{E}^s = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.369410 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.061901 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.621397 & 0 & 0 & 0 \\ 0 & 0 & 0.621397 & 0 & 0 & 0 \\ 0 & 0.061901 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\mathcal{E}^r = \begin{bmatrix} -0.899843 & 0 & 0 & -0.705837 & -0.126183 & -0.008709 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.840591 & -0.889349 & 0 \\ 0 & 0 & 0 & -0.840591 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.837871 \end{bmatrix}$$

For the matrix $\mathcal{K}$ in equation (4.3) we have:

$$\mathcal{K} = \begin{bmatrix} -0.008785 & 0 & 0 & -0.000561 & -0.000341 & -0.000501 \\ 0 & -0.008785 & 0 & -0.000561 & -0.000341 & -0.000501 \\ 0 & 0 & -0.008785 & -0.000561 & -0.000341 & -0.000501 \\ 0 & 0 & 0 & -0.009721 & -0.000569 & -0.000835 \\ 0 & 0 & 0 & -0.000790 & -0.009266 & -0.000705 \\ 0 & 0 & 0 & -0.000928 & -0.000564 & -0.009613 \end{bmatrix}$$

We consider the enzymes of the first six reactions as the manipulated variables. Moreover, the three precursor metabolites are also treated as manipulated variables. Therefore, we can write for the manipulated variables the vector:

$$\mathbf{p}^T = [v_{m,1},\ v_{m,2},\ v_{m,3},\ v_{m,4},\ v_{m,5},\ v_{m,6},\ [G6P],\ [PEP],\ [ATP]\ ]$$

and for the matrices Π and $\mathbf{\Lambda}$ we have:

$$\mathbf{\Pi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0.046747 & 0.905114 & 0 \\ 0. & 1 & 0 & 0 & 0 & 0 & 0 & 0.079740 & 0.270733 \\ 0. & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 0 & 0 & 1 & 0.613340 & 0 & 0.270733 \end{bmatrix}$$

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000819 & -0.001865 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000691 & -0.001574 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000812 & -0.001849 \end{bmatrix}$$

The parameters in equation (4.9) that correspond to specific growth rate are:

$$\Xi_\mu = \mathbf{0}$$

$$\mathcal{H}_\mu = [0,\ 0,\ 0,\ 0.063907,\ 0.038864,\ 0.057021]$$

$$\Theta_\mu = [0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0.055918,\ 0.127322]$$

4.11 Tables

	z_i	q_i			
i	cases a-d	case a	case b	case c	case d
1	-0.28302	0.13926	0.21329	0.13837	0.21418
2	-1.0	0.0	0.0	0.0	0.0
3	-0.67889	0.36149	0.36149	0.36149	0.36149
4	0.45238	0.69315	0.69315	0.69315	0.69315
5	-0.59375	0.0	0.0	0.0	0.0
6	-0.10233	0.0	0.0	0.0	0.0

Table 4.1: The values for the continuous variables $\mathbf{z}$ and $\mathbf{q}$ for the first four best solutions of Problem 1.

	z_i	q_i			
i	cases a-d	case e	case f	case g	case h
1	-0.31840	0.17594	0.09728	0.09766	0.17631
2	-1.0	0.0	0.0	0.0	0.0
3	-0.71270	0.31102	0.31102	0.31102	0.31102
4	0.41735	0.69315	0.69315	0.69315	0.69315
5	-0.62333	0.0	0.0	0.0	0.0
6	-0.04290	0.0	0.0	0.0	0.0

Table 4.2: The values for the continuous variables $\mathbf{z}$ and $\mathbf{q}$ for the second four best solutions of Problem 1.

	case a		case b		case c		case d	
i	z_i	q_i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08347	0.03100	0.08302	0.03083	0.08346	0.03100	0.08302	0.03083
2	1.34812	0.0	1.34606	0.0	1.34811	0.0	1.34606	0.0
3	0.66751	0.0	0.48125	0.0	0.66751	0.0	0.48125	0.0
4	1.50038	0.48497	1.50038	0.60071	1.50038	0.48497	1.50038	0.60071
5	-1.0	0.0	-1.0	0.0	-1.0	0.0	-1.0	0.0
6	-1.0	0.43690	-1.0	0.32129	-1.0	0.32270	-1.0	0.43703

Table 4.3: The values for the continuous variables $\mathbf{z}$ and $\mathbf{q}$ for the second four best solutions of Problem 2.

Overexpressed enzymes	1,2,4	3,4,6	1,4,5	1,4,6	1,2,5	1,2,6
Number of solutions	6	6	5	4	3	3
Overexpressed enzymes	1,5,6	2,3,4	2,3,6	3,5,6	3,4,5	1,2,3
Number of solutions	3	3	1	1	1	1

Table 4.4: The number of alternative regulatory structures for each of the 12 combinations of enzyme overexpression from Problem 3.

	case a		case b		case c	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	-0.91840	0.02832	-0.91830	0.02823	-0.18806	0.02872
2	1.35478	0.37064	1.35368	0.37051	1.33065	0.09889
3	1.27053	0.0	1.17193	0.0	-0.91641	0.0
4	1.50038	0.11026	1.50038	0.17153	1.50038	0.34393
5	-1.0	0.0	-1.0	0.0	-1.0	0.0
6	-1.0	0.0	-1.0	0.0	-1.0	0.0

	case d		case e		case f	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	-0.18806	0.02873	-0.91677	0.02674	-0.91677	0.02674
2	1.33066	0.09889	1.33506	0.36844	1.33504	0.36844
3	-0.91641	0.0	-0.51808	0.0	-0.51808	0.0
4	1.50038	0.47122	1.50038	0.09641	1.50038	0.22042
5	-1.0	0.0	-1.0	0.0	-1.0	0.0
6	-1.0	0.0	-1.0	0.0	-1.0	0.0

Table 4.5: The values of the continuous variables $\mathbf{z}$ and $\mathbf{q}$ for six solutions of Problem 3 (cases a-f from Figure 4).

	case a		case b		case c	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08302	0.030832	0.08346	0.03100	0.08346	0.03100
2	1.34606	0.0	1.34811	0.0	1.34811	0.0
3	0.48125	0.0	0.66751	0.0	0.66751	0.0
4	1.50038	0.60071	1.50038	0.48497	1.50038	0.48497
5	-1.0	0.0	-1.0	0.0	-1.0	0.0
6	-1.0	0.06174	-1.0	0.06161	-1.0	0.06161

	case d		case e		case f	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08302	0.03083	-0.91785	0.02779	-0.91768	0.02672
2	1.34606	0.0	1.34811	0.36989	1.34607	0.36967
3	0.48125	0.0	0.66751	0.0	0.48127	0.0
4	1.50038	0.60071	1.50038	0.48497	1.50038	0.60070
5	-1.0	0.0	-1.0	0.0	-1.0	0.0
6	-1.0	0.06174	-1.0	0.0	-1.0	0.0

Table 4.6: The values of the continuous variables $\mathbf{z}$ and $\mathbf{q}$ for six solutions of Problem 4 (cases a-f from Figure 5).

4.12 Figures

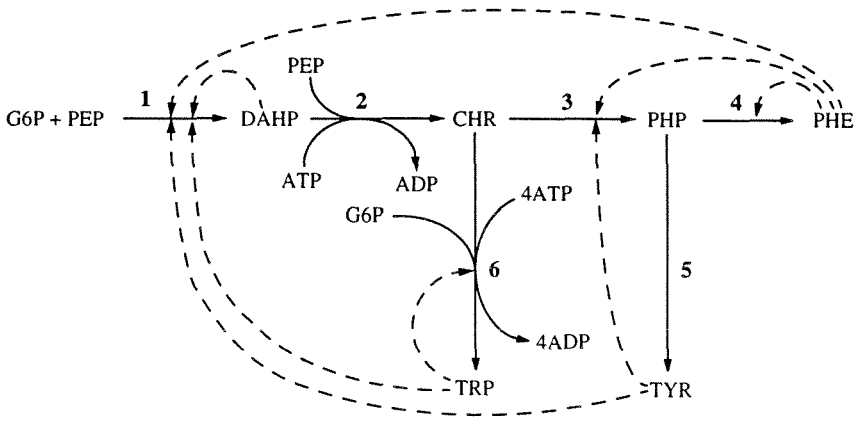

Figure 4.1: The aromatic amino acid synthesis pathway. Solid arrows indicate reactions and dashed arrows indicate feedback inhibition loops. Chemical species: G6P = glucose-6-phosphate; PEP = phosphoenolpyruvate; ATP = adenosine triphosphate; ADP = adenosine diphosphate; DAHP = 3-deoxy-D-arabino-heptulosonate-7-phosphate; CHR = chorismate; PHP = prephenate; PHE = phenylalanine; TYR = tyrosine; TRP = tryptophan.

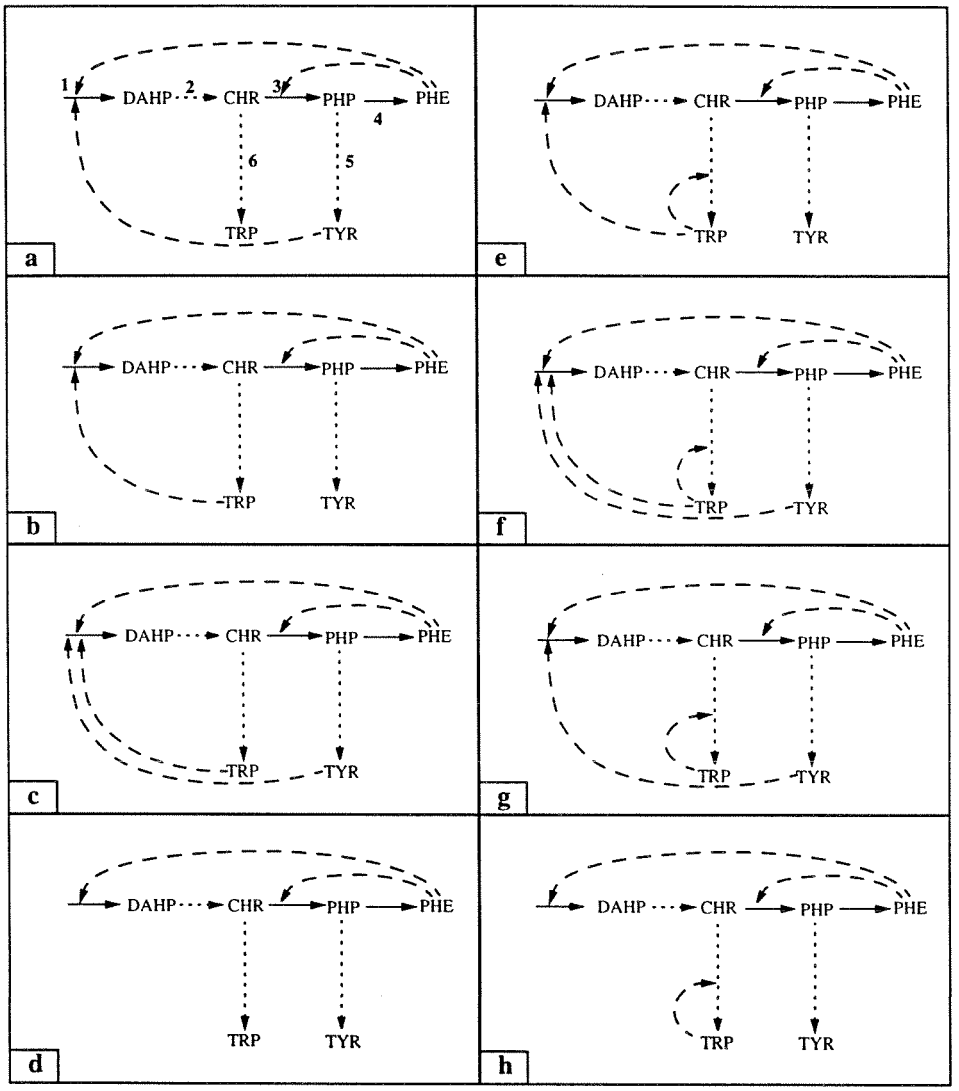

Figure 4.2: The eight best solutions from problem 1. Solid arrows indicate reactions, dashed arrows indicate inhibitory loops, and thick solid arrows indicate enzyme over-expression for the respective reaction. In solutions b-h the reaction numbering has been omitted for clarity.

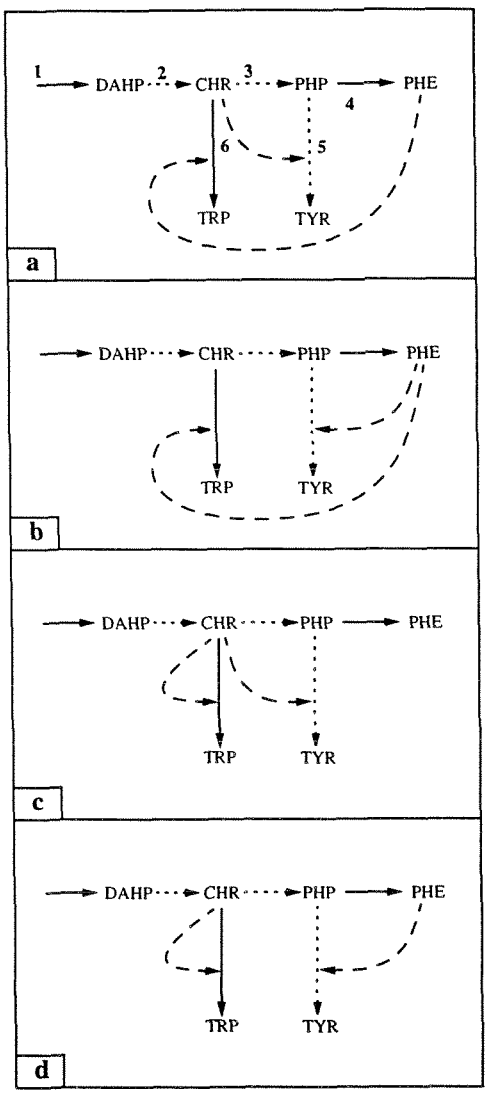

Figure 4.3: The four best solution from problem 2. Solid, dashed, and thick arrows as in Figure 2. In solutions b-d the reaction numbering has been omitted for clarity.

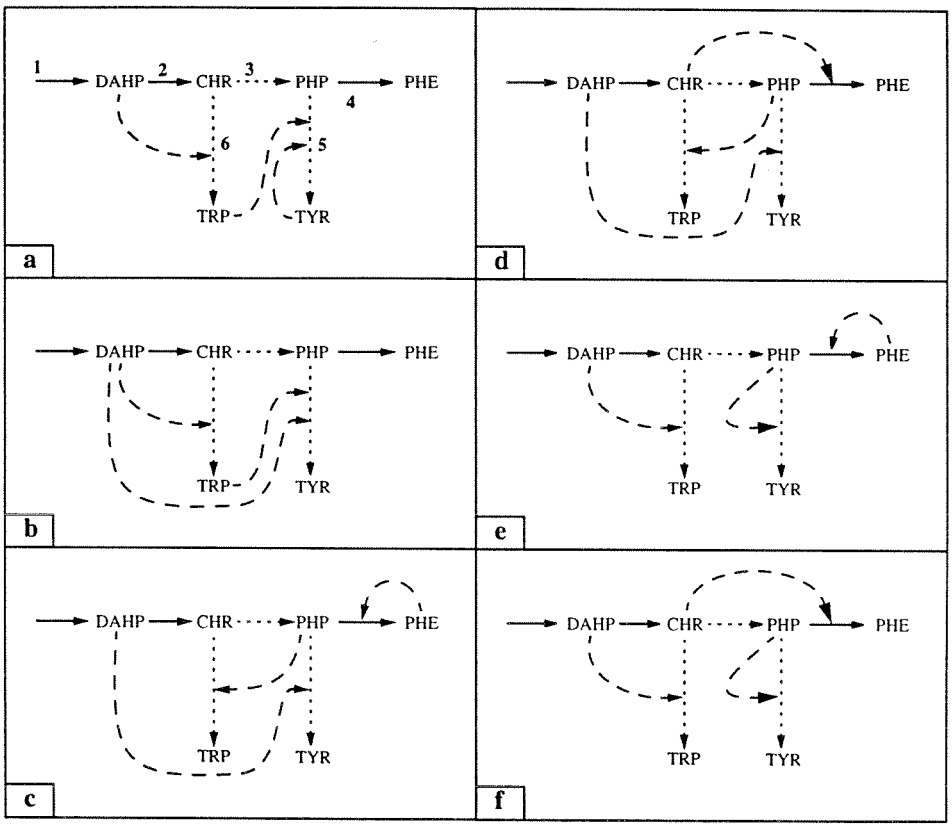

Figure 4.4: Twelve of the 37 best solutions from problem 3. Solid and thick arrows as in Figure 2. Dashed arrows indicate activation. In solutions b-l the reaction numbering has been omitted.(Figure continuous on the next page).

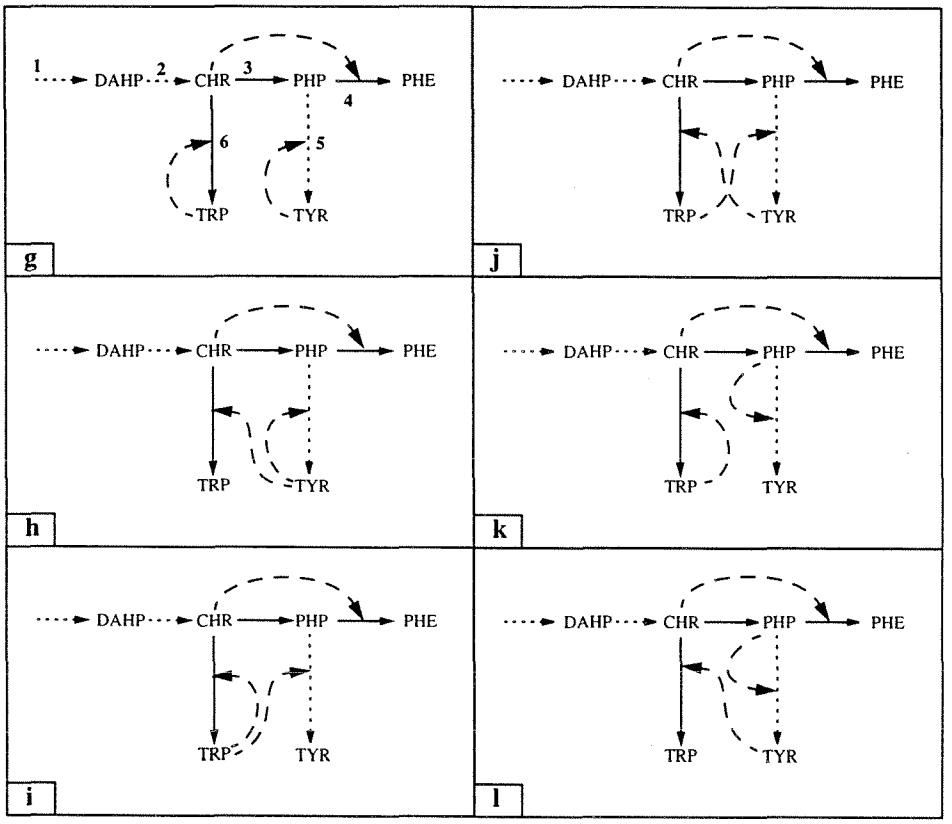

Figure 4.4 (continued)

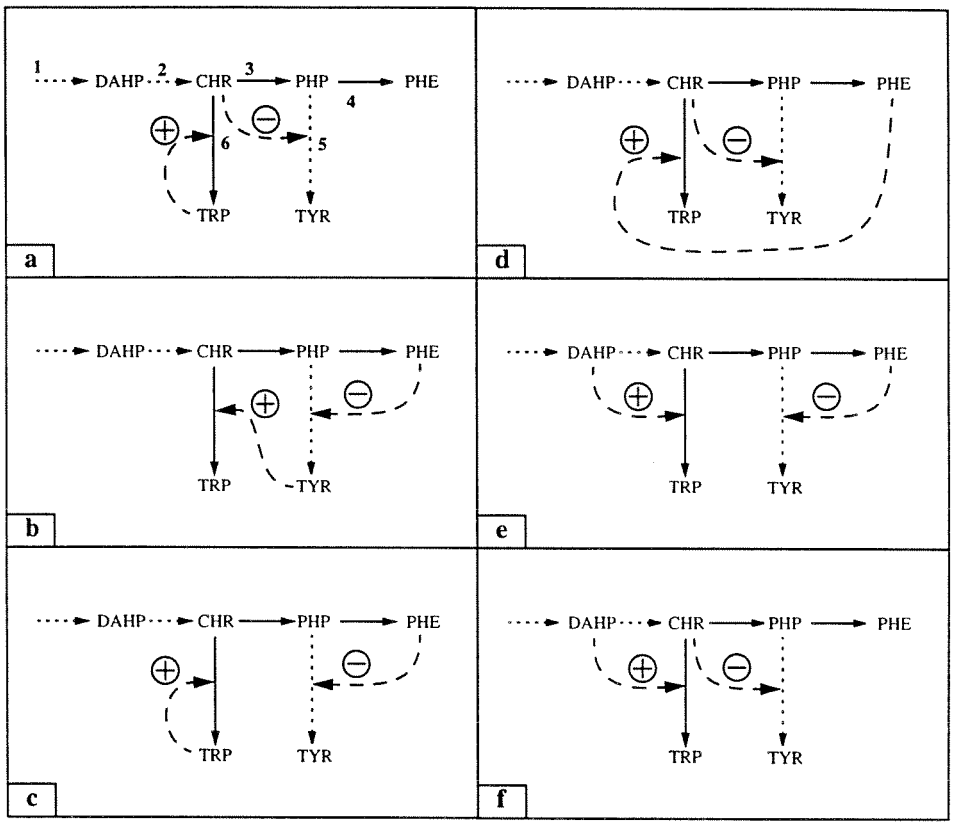

Figure 4.5: The six best solutions from problem 4 that feature activation and inhibition. Solid and thick arrows as in Figure 2. Dashed arrows indicate regulation, (+) indicates activation and (−) inhibition.

Chapter 5 Optimization of Regulatory Architectures in Metabolic Reaction Networks

5.1 Introduction

In chemical and pharmaceutical manufacturing that utilize cultivated organisms, improvements in the product yield, rate of production, and final product concentration can be achieved by two main approaches: genetics and process. The former approach is based on targeted modifications in the DNA of an industrial microorganism aiming at a configuration of metabolic fluxes that will direct raw materials to products efficiently at high rates and in the presence of high concentrations of product. These modifications in the DNA, enabled by recombinant DNA technology, range from overexpression of a single homologous enzyme of a pathway, to introduction into an organism of genes from other organisms with the possibility of creating hybrid metabolic pathways combining features from metabolic pathways in different organisms (Bailey, 1990; Zhang *et al.*, 1995).

Almost every metabolic reaction network is subject to a regulatory architecture built around it, which regulates the amount of the enzymes present in the network and/or the catalytic properties of the enzymes (Sanwal *et al.*, 1971; Stephanopoulos and Vallino, 1991). Significant improvements in the performance of bioprocesses have been realized by genetic modifications of regulatory structures (Bailey, 1990; Katsumata and Ikeda, 1993). These improvements were mainly based on trial and error methods applied to well-understood, relatively simple pathways. However, as the complexity of a set of pathways of interest increases, intuitive and trial and error methods are increasingly ineffective. As a result, a systematic methodology for effective targeted modifications of regulatory structures is needed.

The mathematical description of metabolic systems has been used successfully in the description, analysis and engineering of complex biochemical systems (Cornish-Bowden and Cárdenas, 1990; Gallazo and Bailey, 1990; Hatzimanikatis *et al.*, 1995; Joshi and Palsson, 1989; Lee and Bailey, 1984; Savageau, 1976; Shuler and Domach, 1983). The modeling approaches previously used can be classified into two kinds: linear and nonlinear. Linear models can be accessed through analysis of input-output relations and certain stimulus-response experiments by applying advanced regression

analysis (Schlosser *et al.*, 1993) or other experimental methods developed within and around the metabolic control analysis (MCA) framework (Cornish-Bowden and Cárdenas, 1990; Fell, 1992; Heinrich and Rapoport, 1974; Kacser and Burns, 1973; Reder, 1988; Schlosser and Bailey, 1990). Nonlinear models, on the other hand, can be constructed when detailed kinetic expressions for each step in the reaction pathway are known or can be estimated (Gallazo and Bailey, 1990; Hatzimanikatis *et al.*, 1995; Joshi and Palsson, 1989; Shuler and Domach, 1983; Starbuck and Lauffenburger, 1992).

The S-system representation developed within biochemical systems theory (BST) (Savageau, 1969a,b, 1970, 1972, 1976; Savageau *et al.*, 1987) allows the description of biochemical systems by nonlinear models of a power-law form. There are three main features that make this modeling approach attractive. First, the steady-state equations that describe the mass balances of the system become linear following a logarithmic transformation, and thus linear algebra methods can be applied for analysis of the system. Second, the parameters required to set up the nonlinear equations can be estimated from relatively simple experiments (Savageau, 1992), or they can be derived by appropriate manipulations of MCA data (Savageau *et al.*, 1987; see the Consistency checks below). Third, prior studies have shown S-system models to be very accurate over a wide range of variations in metabolite concentrations, enzyme levels and operating and physiological conditions (Shiraishi and Savageau, 1992; Voit and Savageau, 1987).

The linear nature of the S-system representation, after the logarithmic transformation, has been exploited for the steady-state optimization of biochemical systems using linear programming methods (Regan *et al.*, 1993; Voit, 1992). These studies provide information only about optimum manipulation of external inputs (such as independent effectors and external substrates) or enzyme levels and do not address the problem of optimizing the regulatory structure of the metabolic network. Recently, a mixed-integer linear programming (MILP) framework has been developed that addresses this problem when a (log)linear model, based on MCA data, is considered as the available description of the network (Hatzimanikatis *et al.*, 1995).

In this chapter, we present a MILP formulation for the general optimization of biochemical systems represented as S-systems. The approch provides information on which enzymes should be present at different levels, the extent of such changes needed, and the accompanying modifications in the regulatory structure that will optimize the metabolic function of interest. Three examples illustrate the multitude of questions that can be addressed within this framework and how one can apply the method and postulate the appropriate constraints depending on the questions addressed.

5.2 Mathematical Framework

5.2.1 Mathematical Modeling of Biochemical Systems

Consider a metabolic system consisting of N metabolites and S manipulated (external) parameters. For every metabolite in the system we can write the general form in S-system representation:

$$\frac{dX_i}{dt} = V_i^+(X_1, X_2, \ldots, X_N; P_1, P_2, \ldots, P_S) - V_i^-(X_1, X_2, \ldots, X_N; P_1, P_2, \ldots, P_S) \tag{5.1}$$

$$i = 1, 2, \ldots, N$$

where V_i^+ and V_i^- are net or aggregate rate laws desribing the processes that increase and decrease, respectively, the concentration of the metabolite i, and they are described by the following power-law form:

$$V_i^+ = \alpha_i \prod_{j=1}^{N} X_j^{g_{ij}} \prod_{l=1}^{S} P_l^{a_{il}} \tag{5.2}$$

and

$$V_i^- = \beta_i \prod_{j=1}^{N} X_j^{h_{ij}} \prod_{l=1}^{S} P_l^{b_{il}} \tag{5.3}$$

where X_j $(j = 1, \ldots, N)$ is the concentration of the metabolite j, and P_l $(l = 1, \ldots, S)$ is the level of the manipulated parameter l. The *rate constants* α_i and β_i are nonneg-

ative, and the *kinetic orders* g_{ij}, a_{il}, h_{ij} and b_{il}, defined as:

$$\begin{aligned} g_{ij} &= \frac{X_j}{V_i^+}\frac{\partial V_i^+}{\partial X_j} \\ h_{ij} &= \frac{X_j}{V_i^-}\frac{\partial V_i^-}{\partial X_j} \\ a_{il} &= \frac{P_l}{V_i^+}\frac{\partial V_i^+}{\partial P_l} \\ b_{il} &= \frac{P_l}{V_i^-}\frac{\partial V_i^-}{\partial P_l} \end{aligned}$$

are real. Each of the metabolites and the parameters that has an effect on any of the aggregate processes is associated with a nonzero kinetic order. Moreover, if a metabolite acts on a process as a substrate or activator, the corresponding kinetic order is positive; if it acts as inhibitor, the corresponding kinetic order is negative. The same applies to the kinetic orders of the manipulated parameters: a parameter which acts as activator (inhibitor) is associated with a positive (negative) kinetic order. Prior investigators of the BST and the S-system representation have extensively discussed the physical meaning and methods for experimental estimation of the parameters used in S-systems (Savageau, 1972, 1976).

Having constructed the S-system representation for the system of interest, the *steady-state* can be found after the appropriate logarithmic transformation:

$$\mathbf{y} = -\mathbf{L}^{-1}\mathbf{M}\mathbf{q} + \mathbf{L}^{-1}\mathbf{d} \tag{5.4}$$

where $\mathbf{y}$ is a N-dimensional vector with elements:

$$y_j = ln(X_j) \tag{5.5}$$

$\mathbf{q}$ is a S-dimensional vector with elements:

$$q_l = ln(P_l) \tag{5.6}$$

$\mathbf{d}$ is a N-dimensional vector with elements:

$$d_i = ln\left(\frac{\beta_i}{\alpha_i}\right) \tag{5.7}$$

$\mathbf{L}$ is a $(N \times N)$ matrix with elements:

$$L_{ij} = g_{ij} - h_{ij} \tag{5.8}$$

and $\mathbf{M}$ is a $(N \times M)$ matrix with elements:

$$M_{il} = a_{il} - b_{il} \tag{5.9}$$

However, when we describe or observe a biological system, there might be some metabolic outputs (i.e., ϕ_i) that are functions of the concentration of the metabolites and the parameters of the system. The i-th element of a set of K metabolic outputs can be also described by power-law functions:

$$\phi_i = \gamma_i \prod_{j=1}^{n} X_j^{f_{ij}} \prod_{l=1}^{s} P_l^{c_{il}} \qquad i = 1, 2, \ldots, K \tag{5.10}$$

where γ_i is real and f_{ij} and c_{il}, defined as

$$\begin{aligned} f_{ij} &= \frac{X_j}{\phi_i}\frac{\partial \phi_i}{\partial X_j} \\ c_{il} &= \frac{P_l}{\phi_i}\frac{\partial \phi_i}{\partial P_l} \end{aligned}$$

are also real. If we want to study how changes in the catalytic and regulatory characteristics of certain enzymes will affect the performance of the pathway under consideration we should examine how the performance depends on the kinetic orders as they quantify the affinity of the various enzymes for their substrates and their modulators. When a process V_i^+ is inhibited by metabolite X_j, then the corresponding kinetic order, g_{ij}, will have a negative real finite value. By setting this kinetic order equal to zero we can simulate and observe how the system would respond if we had

inactivated the inhibition of the process V_i^+.

Changes in the kinetic orders of the metabolites alter the elements of matrix **L** in equations (5.4) and (5.8). These changes determine the output values in a nonlinear fashion due to the inversion of matrix **L** in equation (5.4). Consequently, examination of the local sensitivity of the outputs with respect to kinetic orders assumes small changes in these parameters. On the contrary, any attempt to modify the affinity of an enzyme for its substrates and modulators will result in finite changes, often orders of magnitude different from the wild-type affinities. As a result, there is a need for a systematic framework that can optimize the regulatory structures selecting between alternative structures with kinetic orders that may differ from each other in order of magnitude.

5.2.2 Mixed-Integer Linear Optimization

The optimization of metabolic pathways by manipulation of the independent system parameters can be performed using linear programming methods as has been presented in previous studies (Regan, 1993; Voit, 1992). The framework developed in those studies did not address the problem of optimizing the regulatory structure of the metabolic network. The optimization of the regulatory structure involves discrete decisions concerning the regulatory loops that should be inactivated or the regulatory loops that should be introduced in the pathway (Hatzimanikatis *et al.*, 1995).

A number of algorithms and methods, developed in applied mathematics, exist that allow the optimization of linear and nonlinear mathematical models that include continuous and integer variables (Floudas, 1995; Luenberger, 1984; Nemhauser and Wolsey, 1988; Winston, 1995). The application of these methods requires formulation of the problem in a way that can be efficiently solved by the existing methods. In this section we will present the formulation of the problem of optimizing regulatory structures as a mixed-integer linear programming problem. The formulation will exploit the linearity of the steady-state equations of the S-system representation, as well as some additional simple transformations that will allow the introduction of

integer variables in a linear form. The mathematical formulation will be presented via the first example used by Voit (Voit, 1992).

Consider the simple linear pathway with two dependent metabolites, X_1 and X_2, with the second acting as inhibitor on the first reaction of the pathway (Figure 1). Four manipulated (external) variables are considered: the amount of the enzymes, P_1, P_2 and P_3, that catalyze the three reactions, and the amount of the effector, P_4, that activates the first and the second reactions of the pathway.

The S-system representation of the pathway is

$$\frac{dX_1}{dt} = \alpha_1 X_2^{g_{12}} P_1^{a_{11}} P_4^{a_{14}} - \beta_1 X_1^{h_{11}} P_2^{h_{12}} P_4^{b_{14}} \tag{5.11}$$

$$\frac{dX_2}{dt} = \alpha_2 X_1^{g_{21}} P_2^{a_{22}} P_4^{a_{24}} - \beta_2 X_2^{h_{22}} P_3^{b_{23}} \tag{5.12}$$

With the parameter values taken by Voit (Voit, 1992), the system becomes:

$$\frac{dX_1}{dt} = X_2^{-1} P_1 P_4^2 - 0.02 X_1^{0.5} P_2 P_4^2 \tag{5.13}$$

$$\frac{dX_2}{dt} = 0.02 X_1^{0.5} P_2 P_4^2 - 2 X_2 P_3 \tag{5.14}$$

We address the following questions:

In order to maximize the concentration of metabolite 2, X_2:

(i) which one of the manipulated parameters should be changed and what should be the level of the manipulated parameter?

and

(ii) should the feedback inhibition loop be inactivated?

The above questions are subject to the following constraints:

1. The system is at steady-state;
2. $X_{1,ss} \leq 500$;
3. $V_{3,ss} \leq 10$; and
4. For the three enzymes only overexpression is considered, that is $P_l \geq 1 \quad (l = 1, 2, 3, 4)$.

In the above constraints the subscript "ss" denotes the steady-state value of the corresponding variable. In general, metabolites, manipulated variables, reaction rates, and metabolic functions should be constrainted within bounds determined by their physiological ranges for the pathway of interest and by the available biological knowledge (Hatzimanikatis *et al.*, 1995; Voit, 1992).

We can introduce a set of new variables:

$$q_l^r + q_l = ln(P_l) \quad l = 1, 2, 3, 4 \tag{5.15}$$

where q_l^r denotes the logarithm of the reference value of the parameter l, and q_l denotes the logarithm of the factor by which the reference value is multiplied to give the value P_l. In the example studied here

$$q_l^r = ln(1) = 0$$

and

$$q_l = ln(1) = 0.$$

The questions addressed above involve two discrete decisions. The first one concerns the manipulated parameters (i.e., enzyme expression levels and external effector concentration level) that should be changed. It is a discrete decision since the answer should provide the one parameter that should be changed while the rest will be fixed to their reference value. The second discrete decision concerns the regulatory loop. Inactivation (or activation) of the feedback inhibition loop is equivalent to zeroing (or not) the non-zero (negative) kinetic order of V_1^+ with respect to X_2, g_{12}.

Such discrete decisions can be modeled by binary variables, that is, variables that can be zero or one, only. For the first discrete decision, we will introduce a set of binary variables, w_l, for which we will have:

$$q_l^r + w_l q_l = ln(P_l) \quad (l = 1, 2, 3, 4) \tag{5.16}$$

If w_l equals zero, we do not allow changes in the corresponding manipulated parameter, P_l, since from equation (5.16) we have that $ln(P_l) = q_l^r$ always. If w_l equals one, then (5.16) reduces to (5.15).

Moreover, by introducing the constraint:

$$w_1 + w_2 + w_3 + w_4 = 1 \tag{5.17}$$

we allow only one manipulated parameter to change since (5.17) should always hold.

The second discrete decision concerns the rate law V_1^+, in which we introduce the binary variable z_{12} in the exponent of X_2:

$$V_1^+ = \left(X_2^{z_{12}} \cdot g_{12}\right) P_1 P_4^2 \tag{5.18}$$

If z_{12} equals zero, V_1^+ does not depend on X_2 any more and, thus, we have modeled the inactivation of the feedback inhibition loop. If z_{12} equals one, then (5.18) reduces to

$$V_1^+ = X_2^{g_{12}} P_1 P_4^2$$

which is same as in equations (5.11) and (5.13).

At steady-state and after applying the logarithmic transformations the optimization problem can be mathematically formulated as follows:

$$maximize(y_2) \tag{5.19}$$

subject to

Mass balances

$$0.5y_1 + z_{12} \cdot y_2 + q_2^r + w_2 \cdot q_2 - q_1^r - w_1 \cdot q_1 = ln(1/0.02) \tag{5.20}$$

$$0.5y_1 - y_2 + q_2^r + w_2 \cdot q_2 + 2q_4^r + 2w_4 \cdot q_4 - q_3^r - w_3 \cdot q_3 = ln(2/0.02) \tag{5.21}$$

Bound on X_1

$$y_1 \leq ln(500) \tag{5.22}$$

Bound on V_1

$$-z_{12} \cdot y_2 + q_1^r + w_1 \cdot q_1 + 2q_4^r + 2w_4 \cdot q_4 \leq ln(10) \qquad (5.23)$$

Bound on enzyme level

$$P_1 \geq 1 \qquad (5.24)$$

$$P_2 \geq 1 \qquad (5.25)$$

$$P_3 \geq 1 \qquad (5.26)$$

Maximum number of parameter manipulations

$$w_1 + w_2 + w_3 + w_4 \leq 1 \qquad (5.27)$$

y_2, q_4 *unrestricted*

where

y_i $(i = 1, 2)$ and q_l $(l = 1, \ldots, 4)$ are real continuous variables,

q_l^r $(l = 1, \ldots, 4)$ are real parameters, and

z_{12} and w_l $(l = 1, \ldots, 4)$ are binary variables, i.e. they can take the values 0 or 1.

If w_1, w_2 and w_3 are equal to zero, w_4 is equal to one, and z_{12} is equal to 1, then the problem is the same as the one considered by Voit (Voit, 1992). Therefore, by solving the problem twice, once for z_{12} equal to one and once for z_{12} equal to zero we consider the case of maximization of X_2 with and without feedback regulation, respectively. However, we also want to find which of the four parameters can be more efficient in the presence or absence of regulation. In order to find the answer to the question we should solve the problem $2^4 = 16$ times. In general, if we had a pathway with m parameters and n regulatory loops the answer to the same questions would require the enumeration of $2^{(m+n)}$ solutions and the search among them for the best one. Therefore, the problem requires a computational method that will solve the problem efficiently and will ensure the global optimality of the desired solution. As mentioned above such methods exist for certain classes of problems that have discrete and continuous variables that participate either linearly or nonlinearly. In

order to apply these methods the problem is formulated in a mixed-integer linear form for which there are efficient and robust algorithms and software. The linear form implies that every continuous and binary variable will appear linearly in the objective function and in the constraints. Note that in the model (5.19)-(5.27), there are some bilinear products of continuous and binary variables, such as $(w_l \cdot q_l,\ l = 1,2,3,4)$ and $(z_{12} \cdot y_2)$, that introduce nonlinearities. These nonlinearities can be removed by the appropriate transformations (Floudas, 1995; Hatzimanikatis *et al.*, 1995). We will introduce a set of continuous variables, t_l $(l = 1,2,3,4)$ and s_{12}, that will replace the bilinear products:

$$t_l = w_l \cdot q_l; \quad l = 1,2,3,4$$

and

$$s_{12} = z_{12} \cdot y_2$$

For each of the new continuous variables we will include four linear constraints that will guarantee consistency between the bilinear product and the corresponding continuous variables. After the introduction of these new continuous variables and the linear constraints the problem can be formullated in the following mixed-integer linear form:

$$maximize(y_2) \tag{5.28}$$

subject to

Mass balances

$$0.5y_1 + s_{12} + q_2^r + t_2 - q_1^r - t_1 = ln(1/0.02) \tag{5.29}$$

$$0.5y_1 - y_2 + q_2^r + t_2 + 2q_4^r + 2t_4 - q_3^r - t_3 = ln(2/0.02) \tag{5.30}$$

Bound on X_1

$$y_1 \leq ln(500) \tag{5.31}$$

Bound on V_1

$$-s_{12} + q_1^r + t_1 + 2q_4^r + 2t_4 \leq ln(10) \tag{5.32}$$

Bound on enzyme level

$$P_1 \geq 1 \tag{5.33}$$

$$P_2 \geq 1 \tag{5.34}$$

$$P_3 \geq 1 \tag{5.35}$$

Linear transformation for z_{12}

$$y_2 - s_{12} + y_2^L z_{12} \geq y_2^L \tag{5.36}$$

$$y_2 - s_{12} + y_2^U z_{12} \leq y_2^U \tag{5.37}$$

$$z_{12} y_2^L - s_{12} \leq 0 \tag{5.38}$$

$$s_{12} - z_{12} y_2^U \leq 0 \tag{5.39}$$

Linear transformation for w_i $\quad (i = 1, \ldots, 4)$

$$q_i - t_i + w_i q_i^L \geq q_i^L \tag{5.40}$$

$$q_i - t_i + w_i q_i^U \leq q_i^U \tag{5.41}$$

$$w_i q_i^L - t_i \leq 0 \tag{5.42}$$

$$t_i - w_i q_i^U \leq 0 \tag{5.43}$$

Maximum number of parameter manipulation

$$w_1 + w_2 + w_3 + w_4 \leq 1 \tag{5.44}$$

where

y_i $(i = 1, 2)$, s_{12}, q_l and t_l $(l = 1, \ldots, 4)$ are real continuous variables,

q_l^T $(l = 1, \ldots, 4)$ are real parameters,

z_{12} and w_l $(l = 1, \ldots, 4)$ are binary variables, i.e. they can take the values 0 or 1, and

the superscripts U and L, in equations (5.36)-(5.43), denote the upper and lower bound, respectively, of the corresponding variables.

Equations (5.36)-(5.43) transform the bilinear terms of the initial formulation into linear inequality constraints. It is interesting to examine the effect of these constraints.

Let's consider equations (5.36)-(5.39). If $z_{12} = 1$, then (5.36) and (5.37) become

$$y_2 \leq s_{12} \leq y_2$$

and (5.38) and (5.39) become

$$y_2^L \leq s_{12} \leq y_2^U$$

and the first two constraints imply that $s_{12} = y_2$ while the second two constraints simply provide bounds. If $z_{12} = 0$, then we have from (5.36) and (5.37)

$$y_2 - y_2^U \leq s_{12} \leq y_2 - y_2^L$$

and from (5.38) and (5.39) we have

$$0 \leq s_{12} \leq 0$$

and the second two constraints imply that $s_{12} = 0$, while the first two constraints are relaxed since $(y_2 - y_2^U) \leq 0$ and $(y_2 - y_2^L) \geq 0$. A similar analysis holds for every set of inequalities that transform bilinear products of continuous and binary variables.

The problem described by equations (5.19)-(5.44) was solved using the high-level modeling language GAMS (General Algebraic Modeling System), which is interfaced with CPLEX, a mixed-integer linear programming solver. The only feasible optimal solution found was the same as the one reported by Voit:

$$X_1 = 100$$

$$X_2 = 5$$

$$P_l = 1 \quad l = 1, 2, 3$$

$$P_4 = 5\sqrt{2}$$

The only way to increase the concentration of X_2 is by the increase of P_4 in the

presence of the feedback regulatory loop. Every other parameter manipulation in the presence or after inactivation of the feedback loop cannot increase X_2 concentration without violating the inequality constraints.

Although the pathway is a simple linear pathway with three reactions, the result, that within the bounds considered there is only one way to optimize the objective even though changes in the regulation are considered, is not obvious, as it strongly depends on the numerical values of the various parameters and on the bounds of the various variables. The solution could have been identified by exhaustive enumeration of the alternative regulatory structures and the allowable manipulated parameters. However, such an approach is limited by the size of the pathway (i.e., the number of reactions and metabolites) and the questions we address, when they consider a large number of alternative regulatory structures and enzyme manipulation policies. In the following section, the same linear pathway is considered, while a superstructure of alternative regulatory structures is introduced.

5.2.3 Regulatory Superstructure

For every metabolic pathway we can consider a *regulatory superstructure* in which every metabolite in the pathway can potentially regulate any enzyme. Any attempt to optimize a certain metabolic function by genetic or protein engineering modifications of enzyme regulation would address the issue concerning the regulatory loops in this superstructure that could enhance the effects of enzyme expression level manipulations. Its mathematical formulation leads to a large combinatorial problem. Its solution will provide the best performance achievable for a given system, and thus provide valuable constructive insight for protein and metabolic engineering.

A regulatory superstructure can be also postulated as a combination of every regulatory characteristic found in nature for every enzyme in the pathway. The choice of the optimum combination of regulatory characteristics will essentially suggest for the appropriate combination of enzymes from different organisms that catalyze the same reaction but which have different regulatory characteristics.

As an illustration for the postulation and the analysis of the regulatory superstructure we will consider the same example presented in Section 2.2. We will consider that every reaction can be modulated by any of the two metabolites, X_1 and X_2, which will either inhibit or activate a reaction. This consideration results in the postulation of 12 regulatory loops (Figure 2). Moreover, for each loop we will consider $N_{reg} = 6$ alternative levels of regulatory strength and type of regulation: $\{-0.5, 0.5, -0.1, 0.1, -0.01, 0.01\}$. We will allow only two regulatory loops active in the pathway. The questions we will address are the following:

In order to maximize the concentration of X_2:

(i) which two regulatory loops should be active?

(ii) what should be the type of regulation (i.e., activation or inhibition) ?

(iii) what should be the strength of the regulation ?

(iv) which one of the manipulated parameters should be changed ?

and

(v) what should the level of the manipulated parameter be ?

The above questions are subject to the same constraints as before:

1. The system is at steady-state;
2. $X_{1,ss} \leq 500$;
3. $V_{3,ss} \leq 10$; and
4. For the three enzymes only overexpression is considered, that is $P_l \geq 1$ ($l = 1, 2, 3, 4$), and up to 10 times their reference value.

We will introduce the binary variables z_{ijm} and the parameters ϵ_{ijm}, with

$$m = 1, \ldots, N_{reg}$$

$$j = 1, \ldots, N_{rxn}$$

$$i = 1, \ldots, N_{met}$$

and N_{reg} is the number of the alternative strength and types of regulation for each regulatory loop in the superstructure, and N_{rxn} and N_{met} are the numbers of the reactions and metabolites, respectively, in the metabolic network. In this example, we have $N_{reg} = 6$, $N_{rxn} = 3$, and $N_{met} = 2$, and for the binary variables z_{ijm} and the parameters ϵ_{ijm} we have:

- z_{ij1} and z_{ij2} equal to 1, if reaction i is inhibited with strength $\epsilon_{ij1} = -0.5$ or activated with strength $\epsilon_{ij2} = 0.5$, respectively, from metabolite j;
- z_{ij3} and z_{ij4} equal to 1, if reaction i is inhibited with strength $\epsilon_{ij3} = -0.1$ or activated with strength $\epsilon_{ij4} = 0.1$, respectively, from metabolite j;
- z_{ij5} and z_{ij6} equal to 1, if reaction i is inhibited with strength $\epsilon_{ij5} = -0.01$ or activated with strength $\epsilon_{ij6} = 0.01$, respectively, from metabolite j;

An important constraint, when formulating regulatory superstructures, is the one that forbids activation and inhibition of an enzyme by the same metabolite since in the S-system representation they are indistinguishable. The S-system representation of the pathway, including the regulatory superstructure, is:

$$\frac{dX_1}{dt} = \prod_{m=1}^{6}\prod_{j=1}^{2} X_j^{z_{1jm}\epsilon_{1jm}} P_1 P_4^2 - 0.02 X_1^{0.5} \prod_{m=1}^{6}\prod_{j=1}^{2} X_j^{z_{2jm}\epsilon_{2jm}} P_2 P_4^2 \quad (5.45)$$

$$\frac{dX_2}{dt} = 0.02 X_1^{0.5} \prod_{m=1}^{6}\prod_{j=1}^{2} X_j^{z_{2jm}\epsilon_{2jm}} P_2 P_4^2 - 2 X_2 \prod_{m=1}^{6}\prod_{j=1}^{2} X_j^{z_{3jm}\epsilon_{3jm}} P_3 \quad (5.46)$$

We can introduce four sets of variables: y_i, (q_j^r, q_j), and t_l, the same as before, and $s_{mji} = z_{mji}\epsilon_{mji} y_i$. These variables will be used in the description of the steady-state equations after the logarithmic transformation. Similarly, we can write for the optimization problem the following mathematical formulation:

$$maximize(y_2) \quad (5.47)$$

subject to

Mass balances

$$-\sum_{m=1}^{6}\sum_{j=1}^{2} s_{1jm} + 0.5y_1 + \sum_{m=1}^{6}\sum_{j=1}^{2} s_{2jm} - q_1^r - t_1 + q_2^r + t_2 = ln\,(1/0.02) \quad (5.48)$$

$$0.5y_1 + \sum_{m=1}^{6}\sum_{j=1}^{2} s_{2jm} - y_2 - \sum_{m=1}^{6}\sum_{j=1}^{2} s_{3jm} + q_2^r + t_2 + 2q_4^r + 2t_4 - q_3^r - t_3 = ln(2/0.02) \quad (5.49)$$

Bound on X_1

$$y_1 \leq ln(500) \quad (5.50)$$

Bound on V_1

$$-s_{12} + q_1^r + t_1 + 2q_4^r + 2t_4 \leq ln(10) \quad (5.51)$$

Bound on enzyme level

$$P_1 \geq 1 \quad (5.52)$$

$$P_2 \geq 1 \quad (5.53)$$

$$P_3 \geq 1 \quad (5.54)$$

Linear transformation for z_{ijm}

$(i = 1, 2, 3;\ j = 1, 2;\ m = 1, \ldots, 6)$

$$\epsilon_{ijm}y_j - s_{ijm} + min(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm})z_{ijm} \geq min(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm}) \quad (5.55)$$

$$\epsilon_{ijm}y_j - s_{ijm} + max(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm})z_{ijm} \leq max(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm}) \quad (5.56)$$

$$z_{ijm}min(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm}) - s_{ijm} \leq 0 \quad (5.57)$$

$$s_{ijm} - z_{ijm}max(y_j^L\epsilon_{ijm}, y_j^U\epsilon_{ijm}) \leq 0 \quad (5.58)$$

Linear transformation for w_l $\quad (l = 1, \ldots, 4)$

$$q_l - t_l + w_i q_l^L \geq q_l^L \quad (5.59)$$

$$q_l - t_l + w_i q_l^U \leq q_l^U \quad (5.60)$$

$$w_l q_l^L - t_l \leq 0 \quad (5.61)$$

$$t_l - w_l q_l^U \leq 0 \tag{5.62}$$

Maximum number of parameter manipulation

$$w_1 + w_2 + w_3 + w_4 \leq 1 \tag{5.63}$$

Two regulatory loops active

$$\sum_{m=1}^{6} \sum_{j=1}^{2} \sum_{i=1}^{2} z_{mji} = 2 \tag{5.64}$$

One value of strength for each loop

$$(\ i = 1,2,3;\ j = 1,2\ )$$

$$\sum_{m=1}^{6} z_{ijm} \leq 1 \tag{5.65}$$

y_j $(j = 1,2)$, s_{ij} $(i = 1,2,3;\ j = 1,2)$, q_l and t_l $(l = 1,\ldots,4)$ are real continuous variables,

q_l^T, q_l^U and q_l^L $(l = 1,\ldots,S)$, y_j^U and y_j^L $(j = 1,2)$ are real parameters,

z_{ij} $(i = 1,2,3;\ j = 1,2)$ and w_l $(l = 1,\ldots,S)$ are binary variables, i.e. they can take the values 0 or 1, and

the superscripts U and L, in equations (5.55)-(5.62), denote the upper and lower bound, respectively, of the corresponding variables.

Equations (5.47)-(5.65) consist the mathematical formulation of the regulatory superstructure problem as a MILP problem. Equations (5.63), (5.64), and (5.65), are the *logical constraints* which allow us to define our policies with respect to the maximum number of the metabolic parameters that can be manipulated and the number of the active regulatory loops. Consider, for example, a network of N_{met} metabolites, N_{rxn} reactions, and N_{reg} loops of alternative strength and type of regulation. One constraint that can arise during the study of the network is the maximum number of regulatory loops, N_{loop}, on each enzyme. This physiological constraint can be mathematically formulated as follows:

$$\sum_{m=1}^{N_{reg}} \sum_{i=1}^{N_{met}} z_{mji} \leq N_{loop} \quad \text{for each } j = 1,\ldots,N_{rxn} \tag{5.66}$$

Another type of question that could be addressed within the same framework concerns the choice between two or more enzymes from different organisms. Consider for example that the reaction number 4 in the pathway we study is regulated from metabolites 2 and 3 with regulatory strengths ϵ_{242} and ϵ_{143}, respectively. However, there are two other organisms with the same enzyme but with different regulatory properties: one without any regulation, and one that is regulated by metabolites 1 and 6 with regulatory strengths ϵ_{141} and ϵ_{446}, respectively. In this case we can formulate a constraint that will not allow more than two of the loops to be present simultaneously:

$$z_{242} + z_{143} + z_{141} + z_{446} \leq 2 \tag{5.67}$$

and two constraints that will guarantee the existence of both loops of the same regulated enzyme or the introduction of the nonregulated enzyme:

$$z_{242} - z_{143} = 0 \tag{5.68}$$

and

$$z_{141} - z_{446} = 0 \tag{5.69}$$

Therefore, the formulation of the appropriate constraints for the binary variables allows the known regulatory features of alternative enzymes to be incorporated explicitly and unambiguously within the MILP framework.

The optimization problem described by equations (5.47)-(5.65) was solved as before and the best solutions found for the regulatory structure presented in Figure 3.A with continuous variable values:

$$X_1 = 500$$

$$X_2 = 125$$

$$P_l = 1 \quad l = 1, 3, 4$$

$$P_2 = 2.236$$

$$V_1 = 1$$

and:

$$z_{311} = z_{321} = 1$$

The objective function is strikingly increased from 5 to 125 with respect to the optimal solution of the first study. The manipulated parameter that should change is different and the upper bound for X_1 is the limiting factor.

In order to find the second best solution we can solve the problem again including an additional constraint that will exclude the previous solution:

$$2z_{131} + 2z_{132} - \sum_{m=1}^{6}\sum_{j=1}^{3}\sum_{i=1}^{2} z_{mji} \leq \sum_{m=1}^{6}\sum_{j=1}^{3}\sum_{i=1}^{2} z^1_{mji} - 1 \tag{5.70}$$

where z^1_{mji} denotes the binary variable solution of the first problem. The ability to formulate such constraints within the the MILP formulation allows the generation of alternative solutions in a declining, with respect to objective, order, providing with insight and alternative suggestions.

Solving the problem again and including constraint (5.70), we found the second best solution with continuous variables:

$$X_1 = 500$$

$$X_2 = 111.8$$

$$P_l = 1 \quad i = 2, 3, 4$$

$$P_1 = 10$$

$$V_1 = 10$$

and

$$z_{311} = z_{212} = 1$$

and the optimal regulatory structure is presented in Figure 3.B.

The third best can be found by solving again the original problem including equation (5.70) and the following equation that will exclude the second best solution:

$$2z_{131} + 2z_{221} - \sum_{m=1}^{6}\sum_{j=1}^{3}\sum_{i=1}^{2} z_{mji} \leq \sum_{m=1}^{6}\sum_{j=1}^{3}\sum_{i=1}^{2} z_{mji}^2 - 1 \tag{5.71}$$

Following this iterative approach we can obtain a sequence of high performance solutions to the problem. These solutions can be later evaluated with respect to their stability and their dynamic characteristics which are of significant importance for the performance of a pathway (Savageau, 1976; Torres, 1994; Voit, 1992) as well as with respect to the effort required for experimental implementation.

The vast improvement of the objective function when alternative regulatory structures were considered, demonstrates the effect of the regulation on the performance of metabolic patwhays. Similar impressive results of modifications of metabolic regulation have been reported in the literature. However, in the examples presented above and in the most of the sucessful experimental approaches, the pathways considered were relatively small, in terms of metabolites and enzymes, and stoichiometrically simple, such as linear and branched pathays, with a small number of regulatory loops. In Section 3 we will study a pathway with 10 regulatory loops and 6 manipulated variables under constraints, illustrating the efficiency of the MILP framework for patwhay with increased complexity.

Before we continue with the last example we will consider the linear transformations that correlate changes in the catalytic and regulatory properties characteristics of individual enzymes with changes in the kinetic orders of the aggreate rate laws of the S-system representation. These transformations are necessary for the consistency of the representation of the system under the changes.

5.2.4 Consistency Checks

The aggregation employed within S-systems results in kinetic orders that, although appearing to be independent from each other, are essentially dependent. The following example is taken from Savageau et al. (1987). Figure 4 represents a simplified model

of a branched biosynthetic pathway. For the mass balances of the three independent metabolites of the pathway we can write:

$$\begin{aligned} \frac{dX_1}{dt} &= V_1 - V_2 - V_3 \\ \frac{dX_2}{dt} &= V_2 - V_4 \\ \frac{dX_3}{dt} &= V_3 - V_5 \end{aligned} \tag{5.72}$$

The equations that describe this system in BST are:

$$\begin{aligned} \frac{dX_1}{dt} &= \alpha_1 X_0^{g_{10}} X_2^{g_{12}} X_3^{g_{13}} - \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_{h_{13}} \\ \frac{dX_2}{dt} &= \alpha_2 X_1^{g_{21}} X_2^{g_{22}} - \beta_2 X_2^{h_{22}} \\ \frac{dX_3}{dt} &= \alpha_3 X_1^{g_{31}} X_3^{g_{33}} - \beta_3 X_3^{h_{33}} \end{aligned} \tag{5.73}$$

As was pointed out by Savageau et al. (1992), the stoichiometry of the system leads to a correlation between the kinetic orders:

$$h_{11} = \frac{V_{2,o}}{V_{1,o}} g_{21} + \frac{V_{3,o}}{V_{1,o}} g_{31} \tag{5.74}$$

$$h_{12} = \frac{V_{2,o}}{V_{1,o}} g_{22} \tag{5.75}$$

$$h_{13} = \frac{V_{3,o}}{V_{1,o}} g_{33} \tag{5.76}$$

In the above equations g_{ij} is essentially equal to the so-called elasticity of enzyme i with respect to metabolite j:

$$\epsilon_{ij} = \frac{X_i}{V_j} \frac{\partial V_j}{\partial X_i} \tag{5.77}$$

Elasticities have been used as a measure of the effects of metabolites on individual enzymes (Fell, 1992; Kacser and Burns, 1973). Equations (5.74)-(5.76) show that, when we consider changes in the catalytic or regulatory properties of an enzyme, the kinetic orders of the S-system representation should change in a consistent way. Therefore, we will present next a linear transformation from elasticities to kinetic orders that

can be used within the MILP framework presented to guarantee consistency during the search for the optimal regulatory structure.

Consider a pathway consisting of N_{met} metabolites, N_{rxn} individual reaction steps, and S manipulated parameters. Other processes by which concentrations of metabolites change (such as dilution brought about by increases in the biomass volume (Fredrickson, 1976) and transport through the cell envelope) will not be considered here. The mass balances of the metabolites of the system may be written:

$$\frac{d\mathbf{x}}{dt} = \mathbf{N}\mathbf{v}(\mathbf{x}, \mathbf{p}) \tag{5.78}$$

where $\mathbf{N}$ is the $N \times M$ stoichiometric matrix, $\mathbf{v}$ is the M-dimensional reaction rate vector, $\mathbf{x}$ is the N-dimensional metabolite concentration vector, and $\mathbf{p}$ is the S-dimensional manipulated parameter vector (e.g., enzyme concentrations). In addition, consider the r-dimensional vector of metabolic outputs, ϕ, for which we have:

$$\phi = \phi(\mathbf{v}(\mathbf{x}, \mathbf{p}), \mathbf{x}, \mathbf{p}) \tag{5.79}$$

For the mass balance of metabolite i we will have, from equation (5.78):

$$\frac{dX_i}{dt} = \sum_{j=1}^{m} n_{ij} v_j \tag{5.80}$$

which in S-system representation will be written as:

$$\frac{dX_i}{dt} = V_i^+ - V_i^- = \sum_{j=1}^{m} n_{ij}^+ v_j - \sum_{j=1}^{m} n_{ij}^- v_j \tag{5.81}$$

where the superscripts "+" and "-" indicate the positive and the negative, respectively, elements of the i-th row of the stoichiometric matrix. For the kinetic orders we have by definition:

$$g_{ij} = \frac{X_j}{V_i^+} \frac{\partial V_i^+}{\partial X_j}$$

$$
\begin{aligned}
&= \sum_{l=1}^{m} \left(\frac{v_{l,o}}{\sum_{k=1}^{m} n_{ik}^{+} v_{k,o}} \left(\frac{x_j}{v_{l,o}} \frac{\partial v_l}{\partial x_j} \right) \right) \\
&= \sum_{l=1}^{m} \left(\frac{v_{l,o}}{\sum_{k=1}^{m} n_{ik}^{+} v_{k,o}} \epsilon_{lj} \right)
\end{aligned} \tag{5.82}
$$

$$
\begin{aligned}
h_{ij} &= \frac{X_j}{V_i^-} \frac{\partial V_i^-}{\partial X_j} \\
&= \sum_{l=1}^{m} \left(\frac{v_{l,o}}{\sum_{k=1}^{m} n_{ik}^{-} v_{k,o}} \left(\frac{x_j}{v_{l,o}} \frac{\partial v_l}{\partial x_j} \right) \right) \\
&= \sum_{l=1}^{m} \left(\frac{v_{l,o}}{\sum_{k=1}^{m} n_{ik}^{-} v_{k,o}} \epsilon_{lj} \right)
\end{aligned} \tag{5.83}
$$

$$
\begin{aligned}
a_{il} &= \frac{P_l}{V_i^+} \frac{\partial V_i^+}{\partial P_l} \\
&= \sum_{j=1}^{m} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{+} v_{k,o}} \left(\frac{p_l}{v_{j,o}} \frac{\partial v_j}{\partial p_l} \right) \right) \\
&= \sum_{j=1}^{m} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{+} v_{k,o}} \pi_{jl} \right)
\end{aligned} \tag{5.84}
$$

and

$$
b_{il} = \frac{P_l}{V_i^-} \frac{\partial V_i^-}{\partial P_l}
$$

$$= \sum_{j=1}^{m} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{-} v_{k,o}} \left(\frac{p_l}{v_{j,o}} \frac{\partial v_j}{\partial p_l} \right) \right)$$

$$= \sum_{j=1}^{m} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{-} v_{k,o}} \pi_{jl} \right) \tag{5.85}$$

Similarly for the kinetic orders of the i-th output function we will have (from (5.10)):

$$\begin{aligned} f_{ij} &= \sum_{l=1}^{m} \left(\left(\frac{v_{l,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial v_l} \right) \left(\frac{x_{j,o}}{v_{l,o}} \frac{\partial v_l}{\partial x_j} \right) \right) + \left(\frac{x_{j,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial x_j} \right) \\ &= \sum_{l=1}^{m} \left(\left(\frac{v_{l,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial v_l} \right) \epsilon_{lj} \right) + \left(\frac{x_{j,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial x_j} \right) \end{aligned} \tag{5.86}$$

and

$$c_{il} = \sum_{j=1}^{m} \left(\left(\frac{v_{j,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial v_j} \right) \pi_{jl} \right) + \left(\frac{p_{l,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial p_l} \right) \tag{5.87}$$

We observe that any change in the properties of individual enzymes has an effect on more than one kinetic order. When we study the effects of regulatory structures on the performance of a system, which is mathematically represented as an S-system, we should employ the above relations in order to be consistent. Therefore for the mass balance of the i-th metabolite at steady-state and after the logarithmic transformation we will have:

$$\begin{aligned} \sum_{t=1}^{N_{met}} (g_{it} - h_{it}) y_t + \sum_{l=1}^{S} (a_{il} - b_{il}) q_l^r - \sum_{t=1}^{N_{met}} \sum_{j=1}^{N_{rxn}} \sum_{m=1}^{N_{reg}} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{+} v_{k,o}} z_{mjt} \cdot \epsilon_{mjt} \cdot y_t \right) \\ + \sum_{t=1}^{N_{met}} \sum_{j=1}^{N_{rxn}} \sum_{m=1}^{N_{reg}} \left(\frac{v_{j,o}}{\sum_{k=1}^{m} n_{ik}^{-} v_{k,o}} z_{mjt} \cdot \epsilon_{mjt} \cdot y_t \right) + \sum_{l=1}^{S} (a_{il} - b_{il}) w_l \cdot q_l = ln \left(\frac{\beta_i}{\alpha_i} \right) \end{aligned} \tag{5.88}$$

where N_{reg}, defined as before, is the number of alternative values for the regulatory loops; these values can be positive (for activation) or negative (for inhibition). The first two terms on the left hand side of equation (5.88) correspond to the kinetic orders of the original system. The binary variable z_{mjt} in the third and forth terms determines the modifications of the elasticities and, consequently, the modifications of the kinetic orders of the original system. For example, when z_{235} is equal to one, then the elasticity of enzyme 3 with respect to metabolite 5 is reduced by an amount equal to ϵ_{235}. The binary variable w_l in the fifth term of the left hand side controls which manipulated parameter will change, as it has been discussed above.

For the i-th output we will similarly have:

$$\begin{aligned} & ln(\gamma_i) + \sum_{t=1}^{N_{met}} f_{it} y_t + \sum_{l=1}^{S} c_{il} q_l^r \\ + \sum_{t=1}^{N_{met}} \sum_{j=1}^{N_{rxn}} \sum_{m=1}^{N_{reg}} & \left(\frac{v_{j,o}}{\phi_{i,o}} \frac{\partial \phi_i}{\partial v_j} \right) z_{mjt} \cdot \epsilon_{mjt} \cdot y_t + \sum_{l=1}^{S} c_{il} w_l q_l \;=\; ln(\phi_i) \end{aligned} \tag{5.89}$$

Finally, for every binary value, when formulating the problem as an MILP problem, we should apply the transformations introduced earlier, and include four inequality constraints, as has been illustrated with the previous examples.

5.3 Example

As an example we will study a similar problem to that examined by Voit as Example 2.

Yield Optimization in XMP and GMP Production

A simplified form of the pathway is shown in Figure 5. Discussion and references on the biochemical characteristics of the pathway can be found in Voit's paper. The S-system representation of the pathway is:

$$\frac{dX_1}{dt} \;=\; 900 X_3^{-0.5} X_4^{-0.5} P_1 - 10 X_1^{0.5} X_2^{-0.1} X_3^{-0.2} X_4^{-0.2} P_2^{0.6} P_3^{0.4} \tag{5.90}$$

$$\frac{dX_2}{dt} = 7.34X_1^{0.308}X_2^{-0.062}X_3^{-0.162}X_4^{-0.1}P_2^{0.37}P_3^{0.245}P_4^{0.385}$$
$$-43.8X_2^{0.42}X_3^{-0.339}X_4^{-0.5}P_5^{0.4}P_6^{0.6} \quad (5.91)$$

$$\frac{dX_3}{dt} = 2.71X_2^{0.409}X_3^{-0.387}P_5^{0.455} - 0.036X_1^{0.14}X_3^{0.43}X_4^{-0.014}P_3^{0.28} \quad (5.92)$$

$$\frac{dX_4}{dt} = 13.03X_2^{0.041}X_4^{-0.339}P_6^{0.405} - 0.143X_3^{-0.026}X_4^{0.46}P_4^{0.26} \quad (5.93)$$

$$P_i = 1 \quad i = 1, \ldots, 6 \quad (5.94)$$

The kinetic orders were calculated from the rates and their elasticities with respect to their substrates and modulators, as they were qualitatively set to values presented in the Appendix, and by using the linear transformations presented in the previous section. The parameters $P_1 - P_6$ represent the concentration of the enzymes that catalyze the corresponding reactions.

The objective of the optimization is to maximize the steady-state concentration of metabolite X_4. The problem is subject to the following constraints:

- the concentration of the metabolites X_1, X_2, and X_3, should not deviate more than $\pm 10\%$ from their reference steady-state values;
- the concentration of the enzymes cannot increase or decrease beyond certain limits.

The mathematical formulation of the optimization problem is presented in the Appendix.

Let's consider first the following question:

Which one of the enzymes should be manipulated, within $\pm 10\%$ *of its reference value, in order to maximize* X_4 *?*

The answer to this question can be found by solving six times the optimization problem using a linear programming algorithm. However, if we use the MILP formulation of the problem presented in the Appendix, by including two additional constraints, we can found the answer by solving the MILP problem once. The additional

two constraints are on the binary variables:

$$w_1 + w_2 + w_3 + w_4 + w_5 + w_6 = 1 \tag{5.95}$$

and

$$z_{13} + z_{14} + z_{21} + z_{22} + z_{23} + z_{24} + z_{34} + z_{43} + z_{53} + z_{64} = 10 \tag{5.96}$$

where constraint (5.95) allows changes on the levels of only one of the enzymes, and constraint (5.96) does not allow any modification of the regulatory structure.

Similarly, the question:

Which two of the enzymes should be manipulated, within $\pm 10\%$ *of their reference value, in order to maximize* X_4 *?*

can be answer by changing constraint (5.95) to:

$$w_1 + w_2 + w_3 + w_4 + w_5 + w_6 = 2 \tag{5.97}$$

Following this procedure we addressed this question considering three, four, five, and six enzymes to change simultaneously. The optimal values for the objective function, the enzymes that should be changed, and the level of their change, are presented in Table I. One of the important conclusions that can be drawn from this Table is the one concerning the "controling enzymes". Enzyme 6 appears to have the highest effect on the concentration of X_4. It is interesting to note, however, that when this is the only manipulated enzyme, the optimum solution does not involve the maximal allowed increase in P_6. Only when another activity is changed does P_6 increase to its upper limit at the optimal solution. Moreover, there are only three enzymes (i.e., enzymes 1, 6, and 4) that significantly influence X_4 concentratio (within the bounds of the expression levels and of the concentration of the other metabolites), since simultaneous manipulation of four or more enzymes does not improve the concentration of X_4 much.

One could argue that this information might be obtained by examining the concentration control coefficents. In Table II the concentration control coefficients for

the four metabolites with respect to the six enzymes are presented. Although the indication for which single enzyme should be manipulated from the maximum X_4-concentration control coefficient is consistent with the optimization calculation result, there is (of course) no indication from the concentration control coefficient that an optimal increase in P_6 exists. Furthermore, the two enzymes that should be manipulated, according to the control coefficient values in Table II, are enzymes 4 and 6, which is not the case. This is due to the existence of constraints on metabolites and other physiological parameters and indicates the necessity of the MILP framework when constraints on metabolic functions are considered.

If we consider the same optimization problem and allow the enzyme changes between 20% and 500% of their reference value, the results are quite different (Table III). The optimal value of the concentration of X_4 can be achieved in two different ways when one and two enzyme manipulations are considered. The value for the objective function when we change one enzyme is the same as before because the limiting factor is in both cases the concentration of metabolite X_2 which in both cases is equal to the higher value allowed.

When two enzymes are changed in this second case we have two possible solutions, and the corresponding objective value is almost equal to the objective value found in the previous case (Table I) for the manipulation of three enzymes. Moreover, the combination of the enzymes for two and three enzyme manipulations are different for the two cases, that is, they depend on the bounds for the enzyme expression levels. This difference clearly arises from the nonlinear nature of the original system, as well as on the constrained character of the problem.

The maximum concentration of X_4 that can be achieved by changes in the six of the enzymes is 240.5% of the reference value. If we consider changes in the regulatory structure, by inactivating any or some of the existent loops, we might achieve an even higher value for the objective. Solving the MILP problem by allowing also changes in the binary values z_{ij}, we can find the maximum that can be achieved by both enzyme manipulation and regulatory structure modifications, within the constraints and for the given kinetic properties of the system as they are quantified by the elasticities

of the substrates and the modulators. The solution can be found by changing the constraint (5.96) to:

$$z_{13} + z_{14} + z_{21} + z_{22} + z_{23} + z_{24} + z_{34} + z_{43} + z_{53} + z_{64} \leq 10 \qquad (5.98)$$

which allows changes in the regulatory structure. The maximum concentration of X_4 was found to be equal to 55015.6, or, 11414% of the reference value, and multiple alternative combinations of regulatory structures and enzyme manipulations were found to correspond to this optimal value.

In modifying regulatory structures it is reasonable to look for the minimum number of loops that should be inactivated since experimental modifications of the regulatory characteristics of an enzyme are much more difficult to achieve than changing the amount of the enzyme. The MILP framework allowed us to find among the alternative optimal regulatory structures the one with the maximum number of regulatory loops active by reformulating the optimization problem. The objective function was the number of the regulatory loops and the objective was the maximization of this number:

$$maximize(z_{13} + z_{14} + z_{21} + z_{22} + z_{23} + z_{24} + z_{34} + z_{43} + z_{53} + z_{64}) \qquad (5.99)$$

subject to the equality constraint:

$$y_4 = 55015.6 \qquad (5.100)$$

The maximum number of regulatory loops that can remain active in the reaction network is equal to 5. The optimal regulatory structure is presented in Figure 6 and the levels of metabolites and enzymes are presented in Table IV.

In any attempt to modify the regulatory structure of a reaction network we are interested in identifying the first regulatory loop that we should try to inactivate. This question can be formulated mathematically by the formulation presented in the

Appendix, including the constraint:

$$z_{13} + z_{14} + z_{21} + z_{22} + z_{23} + z_{24} + z_{34} + z_{43} + z_{53} + z_{64} = 1 \tag{5.101}$$

The solution of this problem suggests two optimal modifications that result in a concentration for X_4 equal to 386.6% the reference value. The regulatory loops that should be inactivated are the feedback loops on enzyme 2 by metabolite 3 and by metabolite 4, referred as solution A and B in Table V, where the corresponding changes in enzymes and concentrations are presented.

Many similar questions can be addressed and answered using the MILP framework presented above and applied here to this example. A study of the values of the metabolites and the enzyme levels, as well as of any metabolic function of the network, at the optimal steady-states can provide metabolic engineer with useful insight prior to any attempt to manipulate enzyme levels and regulatory architectures of metabolic reaction pathways.

5.4 Concluding Remarks

The framework developed in this chapter considers the optimization of bioprocesses that can be described by an S-system representation. The novelty of the approach lies in its ability to find the optimal regulatory structure with respect to a metabolic function of a biochemical reaction network. The transformation of the nonlinear S-system representation of biochemical systems into linear systems and the introduction of binary variables and a set of key transformations allowed this problem in optimal manipulation of a biochemical system to be formulated as a MILP problem.

The examples presented illustrate the ability of the formulation to address various problems concerning analysis and understanding of metabolic pathways and, more important, problems concerning optimal combinations of regulatory structures. The postulation of a regulatory superstructure around a metabolic pathway, and the formulation of appropriate constraints allow design of regulatory architectures that can

optimize an objective. Although not pursued in this work, this approach can also be applied to suggest possible evolutionary criteria that gave rise to natural regulatory structures (Heinrich *et al.*, 1987; Majewski and Domach, 1990; Marr, 1991; Schuster and Heinrich, 1991).

A similar formulation for linear models has been also developed (Hatzimanikatis *et al.*, 1995), as well as for nonlinear models that are described by enzyme kinetic models based on rate laws of a generalized Michaelis-Menten type (Hatzimanikatis et al., manuscript in preparation). The formulation of a system described by nonlinear models includes nonlinear constraints and objective functions that are more difficult to solve, and the global optimality of the solution is not guaranteed. However, every kinetic description of metabolic systems can be transformed into an approximate S-system representation. Then, this MILP formulation can be applied to the transformed systems to suggest promising strategies for achieving a metabolic engineering objective.

The transfer of enzymes between different organisms, as well as protein engineering of homologous enzymes, can result in pathways with altered regulatory and catalytic properties. The corresponding changes in the catalytic properties can be mathematically formulated and studied within the MILP framework presented here. The linear transformations presented above that correlate elasticities of individual enzymes with kinetic orders, should always be applied when changes in the catalytic and regulatory properties of the enzymes are considered in order to preserve consistency in the change of the parameters and to avoid erroneous results.

The formulation of constraints that exclude previous solutions enables the generation of a series of solutions in a hierarchical order, with respect to the value of the objective function. These solutions can be analyzed with respect to their dynamic or other metabolic characteristics, and the ones that satisfy the set of desired criteria can be chosen for experimental implementation.

5.5 References

Bailey, J. E. 1990. Toward a Science of Metabolic Engineering. Science. **252**: 1668-1675.

Cornish-Bowden, A., Cárdenas M.L. 1990. Control of Metabolic Processes. NATO ASI Series A: Lifesciences Vol. 190. Plenum Press, New York.

Fell, D.A. Metabolic Control Analysis - A Survey of its Theoretical and Experimental Development. Biochem. J. **152**: 313-330.

Floudas, C.A. 1995. Nonlinear and Mixed Integer Optimization: Fundamentals and Applications. Oxford University Press, New York.

Fredrickson, A. G. 1976. Formulation of Structured Growth Models. Biotechnol. Bioeng. **28**: 1481-1486.

Gallazo, J. L., Bailey, J. E. 1990. Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces cerevisiae*. Enzyme Microb. Technol. **12**: 162-172.

Hatzimanikatis, V., Floudas, C. A., Bailey, J. E. 1995. Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization. AIChE J. (in press).

Hatzimanikatis, V., Lee, K. H., Renner, W. A., Bailey, J. E. 1995. A Mathematical Model for the G1/S Transition of the Mammalian Cell Cycle. Biotech. Lett. **17**(7): 669-674.

Heinrich, R., Holzhütter, H.-G., Schuster, S. 1987. A Theoretical Approach to the Evolution and Structural Design of Enzymatic Networks. Bull. Math. Biol. **49**: 539-595.

Heinrich, R., Rapoport, T. A. 1974. A Linear Steady-State Treatment of Enzymatic Chains. Eur. J. Biochem. **42**: 89-95.

Joshi, A., Palsson, B.O. 1989. Metabolic Dynamics in the Human Red Cell. Part I – A Comprehensive Kinetic Model. J. Theor. Biol. **141**: 515-528.

Kacser, H., Burns, J. A. 1973. The Control of Flux. Symp. Soc. Exp. Biol. **27**: 65-104.

Katsumata, R., Ikeda, M. 1993. Hyperproduction of Tryptophan in *Corynebacterium glutamicum* by Pathway Engineering. Bio/Technology. **11**: 921-925.

Lee, S. L., Bailey, J. E. 1984. A Matheamtical Model for Lambda-dv Plasmid Replication - Analysis of Wild-Type Plasmid. Plasmid. **11**: 151-165.

Luenberger, D. 1984. Linear and Nonlinear Programming. Addison-Wesley, Reading, MA.

Majewski, R.A., Domach, M.M. 1990. Simple Constrained-Optimization View of Acetate Overflow in *E. coli.* Biotechnol. Bioeng. **35**: 732-738.

Marr, A. G. 1991. Growth rate of *Escherichia coli.* Microbiol. Rev. **55**: 316-333.

Nemhauser, R., Wolsey, L. 1988. Integer and Combinatorial Optimization. John Wiley & Sons, San Diego.

Reder, C. 1988. Metabolic Control Theory: A Structural Approach. J. Theor. Biol. **135**: 175-201.

Regan, L., Bogle, D.L., Dunnill, P. 1993. Simulation and Optimization of Metabolic Pathways. Comp. Chem. Engng. **17**: 627-637.

Sanwal, B. D., Kapoor, M., Duckworth, H. W. 1971. The Regulation of Branched and Converging Pathways. Curr. Top. Cell. Regul. **3**: 1-115.

Savageau, M. A. 1969. Biochemical Systems Analysis, I. Some Mathematical Properties of the Rate Law for the Component Enzymatic Reactions. J. Theor. Biol. **25**: 365-369.

Savageau, M. A. 1969. Biochemical Systems Analysis, II. The Steady-State Solutions for an n-pool System Using a Power-Law Approximation. J. Theor. Biol. **25**: 370-379.

Savageau, M. A. 1970. Biochemical Systems Analysis, III. Dynamic Solutions Using a Power-Law Approximation. J. Theor. Biol. **26**: 215-226.

Savageau, M. A. 1972. The Behavior of Intact Biochemical Control Systems. Curr. Top. Cell Reg. **6**: 63-130.

Savageau, M. A. 1976. Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Addison–Wesley, Reading, MA.

Savageau, M. A. 1992. Critique of the Enzymologist's Test Tube, pp. 45-108. In:

E. E. Bittar (ed.), Fundamentals of Medical Cell Biology, Vol. 3A, Chemistry of the Living Cell. JAI Press Inc.

Savageau, M. A., Voit, E. O., Irvine, D. H. 1987. Biochemical Systems Theory and Metabolic Control Theory: 2. The Role of Summation and Connectivity Relationships. Math. Biosci. **86**: 147-169.

Schlosser, P. M., and J. E. Bailey. 1990. An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways. Math. Biosci. **100**: 87-114.

Schlosser, P. M., Holcomb, T., Bailey, J. E. 1993. Determining Metabolic Sensitivity Coefficients Directly from Experimental Data. Biotechnol. Bioeng. **41**: 1027-1038.

Schuster, S., Heinrich, R. 1991. Minimization of Intermediate Concentrations as a Suggested Optimality Principle for Biochemical Networks I. Theoretical Analysis. J. Math. Biol. **29**: 425-442.

Shiraishi, F., Savageau, M. A. 1992. The Tricarboxylic-acid Cycle in *Dictyostelium-discoideum.* 1. Formulation of Alternative Kinetic Representations. J. Biol. Chem. **267**(32): 22912-22918.

Shuler, M. L., Domach, M. M. 1983. Mathematical Models of the Growth of Individual Cells – Tools for Testing Biochemical Mechanisms. ACS Symposium Series. **207**: 93-133.

Starbuck, C., Lauffenburger, D. A., 1992 Mathematical Model for the Effects of Epidermal Growth-Factor Receptor Trafficking Dynamics on Fibroblast Proliferation Responses. Biotech. Prog. **8**: 132-143.

Stephanopoulos, G., Vallino, J. J. 1991. Network Rigidity and Metabolic Engineering in Metabolite Overproduction. Science. **252**: 1675-1681.

Torres, N. V. 1994. Application of Transition Time of Metabolic Systems as a Criterion for Optimization of Metabolic Processes. Biotechnol. Bioeng. **44**: 291-296.

Voit, E. O. 1992. Optimization in Integrated Biochemical Systems. Biotechnol. Bioeng. **40**: 572-582.

Voit, E. O., Savageau, M. A. 1987. Accuracy of Alternative Representations for Integrated Biochemical Systems. Biochemistry **26**(21): 6869-6880.

Winston, W.L. 1995. Introduction to Mathematical Programming: Applications and Algorithms. Duxbury Press, Belmont, CA.

Zhang, M., Eddy, C., Deanda, K., Finkenstein, M., Picataggio, S. 1995. Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic *Zymomonas mobilis*. Science. **267**(5195): 240-243.

5.6 Appendix

In this appendix the elasticities and the steady-state values of the metabolite concentrations and the reaction rates of the XMP and GMP production pathway will be given and the mathematical formulation of the optimization problem will be analytically presented.

The steady-state value of the concentration of the four metabolites of the pathway is:

$$\{X_1, X_2, X_3, X_4\} = \{5.42, 213, 2417, 482\}$$

and the corresponding steady-state values of the reaction rates are:

$$\{V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}\} =$$
$$\{0.83, 0.5, 0.33, 0.52, 0.54, 0.81, 0.86, 1.48, 0.65, 1.19\}$$

The elasticities of the system can be partitioned into two groups: the elasticities of the enzymes with respect to their substrates and the elasticities of the enzymes with respect to their modulators.

For the elasticities with respect to substrates we have:

$$\{\epsilon_{21}, \epsilon_{31}, \epsilon_{33}, \epsilon_{44}, \epsilon_{52}, \epsilon_{62}, \epsilon_{73}, \epsilon_{84}\} =$$
$$\{0.3, 0.5, 0.25, 0.061, 0.9, 0.1, 0.5, 0.6\}$$

and for the elasticities with respect to modulators we have:

$$\{\epsilon_{13}, \epsilon_{14}, \epsilon_{21}, \epsilon_{22}, \epsilon_{23}, \epsilon_{24}, \epsilon_{34}, \epsilon_{43}, \epsilon_{53}, \epsilon_{64}\} =$$
$$\{-0.5, -0.5, 0.2, -0.167, -0.5, -0.3, -0.05, -0.1, -0.85, -0.833\}$$

For the elasticities with respect to paramaters we have:

$$\pi_{il} = 1 \quad (i = 1, \ldots, 10; l = 1, \ldots, 6)$$

The optimization problem is mathematically formulated as follows:

$$maximize\ (y_4) \tag{5.102}$$

Mass Balances

$$\begin{aligned} -0.5y_1 + 0.1y_2 - 0.3y_3 - 0.3y_4 - s_{13} - s_{14} + 0.6s_{21} + 0.6s_{22} \\ +0.6s_{23} + 0.6s_{34} + 0.4s_{34} + t_1 - 0.6t_2 - 0.4t_3 &= -4.4998 \end{aligned} \tag{5.103}$$

$$\begin{aligned} 0.308y_1 - 0.482y_2 + 0.177y_3 + 0.4y_4 - 0.37s_{21} - 0.37s_{22} \\ -0.37s_{23} - 0.37s_{24} - 0.245s_{34} - 0.385s_{43} + 0.4s_{53} \\ +0.6s_{64} + 0.37t_2 + 0.245t_3 + 0.385t_4 - 0.4t_5 - 0.6t_6 &= 1.7863 \end{aligned} \tag{5.104}$$

$$\begin{aligned} -0.14y_1 + 0.409y_2 - 0.817y_3 - 0.014y_4 \\ -0.455s_{53} + 0.287s_{34} - 0.28t_3 + 0.455t_5 &= -4.3212 \end{aligned} \tag{5.105}$$

$$\begin{aligned} 0.041y_2 + 0.026y_3 - 0.799y_4 - 0.405s_{64} \\ +0.26s_{43} - 0.26t_4 + 0.405t_6 &= -4.5122 \end{aligned} \tag{5.106}$$

Bounds on X_j $\qquad j = 1, 2, 3$

$$ln(4.9) \geq y_1 \geq ln(6.0) \tag{5.107}$$

$$ln(192) \geq y_2 \geq ln(234) \tag{5.108}$$

$$ln(2176) \geq y_3 \geq ln(2660) \tag{5.109}$$

Bounds on P_l $\qquad l = 1, \ldots, 6$

$$ln(P_l^L) \geq P_l \geq ln(P_l^U) \tag{5.110}$$

where P_l^L and P_l^U are the lower and upper bounds, respectively, of the l-th enzyme

expression level. For each of the variables s_{ij} and t_l we need four inequality constraints similar to equations (5.36)-(5.39) and (5.40)-(5.43):

$$\textit{Linear transformation for } s_{ij} \tag{5.111}$$

$$\begin{aligned}(i,j) = \{&(1,3),(1,4),(2,1),(2,2),(2,3),\\ &(2,4),(3,4),(4,3),(5,3),(6,4)\}\end{aligned}$$

$$\epsilon_{ij}y_j - s_{ij} + min(y_j^L\epsilon_{ij}, y_j^U\epsilon_{ij})z_{ji} \geq min(y_j^L\epsilon_{ij}, y_j^U\epsilon_{ij}) \tag{5.112}$$

$$\epsilon_{ij}y_j - s_{ij} + max(y_j^L\epsilon_{ij}, y_j^U\epsilon_{ij})z_{ji} \leq max(y_j^L\epsilon_{ij}, y_j^U\epsilon_{ij}) \tag{5.113}$$

$$z_{ij}min(y_j^L\epsilon_{ij}, y_j^U\epsilon_{ij}) - s_{ij} \leq 0 \tag{5.114}$$

$$s_{ij} - z_{ij}max(y_i^L\epsilon_{ij}, y_j^U\epsilon_{ij}) \leq 0 \tag{5.115}$$

$$\textit{Linear transformation for } w_l \qquad (l = 1,\ldots,6)$$

$$q_l - t_l + w_l ln(P_l^L) \geq ln(P_l^L) \tag{5.116}$$

$$q_l - t_l + w_l ln(P_l^U) \leq ln(P_l^U) \tag{5.117}$$

$$w_l ln(P_l^L) - t_l \leq 0 \tag{5.118}$$

$$t_l - w_l ln(P_l^U) \leq 0 \tag{5.119}$$

where

$$y_j^L = ln(X_j^L) \quad j = 1,\ldots,4$$

and

$$y_j^U = ln(X_j^U) \quad j = 1,\ldots,4$$

and the superscripts U and L denote the upper and lower bound, respectively, of the corresponding variables. The elasticities used in equations (5.112)-(5.115

In equations (5.102)-(5.119) above

y_j $(j = 1,\ldots,4)$,

q_l and t_l $(l = 1,\ldots,6)$ and

$\{s_{13}, s_{14}, s_{21}, s_{22}, s_{23}, s_{24}, s_{34}, s_{43}, s_{53}, s_{64}\}$

are real continuous variables,

q_l^r, P_l^U, and P_l^L $(l = 1, \ldots, 6)$,

y_j^U and y_j^L $(j = 1, \ldots, 4)$, and

$\{\epsilon_{13}, \epsilon_{14}, \epsilon_{21}, \epsilon_{22}, \epsilon_{23}, \epsilon_{24}, \epsilon_{34}, \epsilon_{43}, \epsilon_{53}, \epsilon_{64}\}$

are real parameters, and

$\{z_{13}, z_{14}, z_{21}, z_{22}, z_{23}, z_{24}, z_{34}, z_{43}, z_{53}, z_{64}\}$ and

w_l $(l = 1, \ldots, 6)$

are binary variables, i.e. they can take the values 0 or 1,

5.7 Tables

X_4	P_1	P_2	P_3	P_4	P_5	P_6	X_1	X_2	X_3
103.84	100	100	100	100	100	**109.17**	98.62	90	95.14
104.84	**105.68**	100	100	100	100	**110**	110	96.92	96.92
107.83	**106.24**	100	100	**90**	100	**110**	110	90.11	93.49
108.25	**110**	**110**	100	**90**	100	**110**	103.75	94.97	96.96
108.78	**110**	**110**	100	**90**	**90**	**110**	106.31	105.27	95.89
108.82	**110**	**110**	**94.94**	**90**	**90**	**110**	110	105.27	97.04

Table 5.1: The values for the objective function, enzyme expression levels and the concentrations of the metabolites when one, two, three, four, five, and six enzymes change simultaneously, within $\pm 10\%$ of their reference value and without any modifications in the regulatory strucure. Bold type, underlined numbers indicate the changes in the enzyme levels.

$C_{P_l}^{X_j}$ $j = 1, \ldots, 4;$ $l = 1, \ldots, 6$						
	P_1	P_2	P_3	P_4	P_5	P_6
X_1	1.9993	-1.3427	-0.6569	0.1005	-0.2313	-0.1564
X_2	1.5040	0.0005	-0.0030	0.7687	-0.9773	-1.1981
X_3	0.4119	0.2305	-0.2318	0.3629	0.1065	-0.5657
X_4	0.0906	0.0075	-0.0077	-0.2742	-0.0467	0.4270

Table 5.2: The concentration control coefficients for the four metabolites with respect to the six enzymes.

X_4	P_1	P_2	P_3	P_4	P_5	P_6	X_1	X_2	X_3
103.84	100	100	100	100	100	**109.17**	98.62	90	95.14
103.84	100	100	100	**87.18**	100	100	98.62	90	95.14
107.95	100	100	100	100	**90.68**	**118.21**	99.62	90	90
107.95	100	100	100	**76.97**	**90.68**	100	99.62	90	90
226.45	**351.22**	**500**	100	100	100	**500**	90	110	107.79
232.64	**352.93**	**500**	100	**90.80**	100	**500**	110	90	94.68
240.47	**372.65**	**500**	**116.08**	**81.61**	100	**500**	110	90	90
240.47	**500**	**500**	**242.17**	**81.61**	**157.38**	**500**	110	90	90

Table 5.3: The values for the objective function, enzyme expression levels and the concentrations of the metabolites when one, two, three, four, five, and six enzymes change simultaneously, between ±20% and ±500% of their reference value and without any modifications in the regulatory strucure. Bold type, underlined numbers indicate the changes in the enzyme levels.

i	1	2	3	4	5	6
P_i	500	500	69.53	500	64.70	349.48
X_i	110	90	90	**11414**		

Table 5.4: The optimal values for the objective function, enzyme expression levels and the concentrations of the metabolites when six enzymes change simultaneously, between $\pm 20\%$ and $\pm 500\%$ of their reference value and modifications in the regulatory structure are considered. The corresponding optimal regulatory structure is presented in Figure 6.

	Solution A		Solution B	
i	P_i	X_i	P_i	X_i
1	479.83	110	460.97	90
2	189.43	110	38.88	110
3	20	110	20	110
4	20	386.57	20	386.57
5	39.87		37.48	
6	500		500	90

Table 5.5: The optimal values for the objective function, enzyme expression levels and the concentrations of the metabolites when six enzymes change simultaneously, between $\pm 20\%$ and $\pm 500\%$ of their reference value and inactivation of only one regulatory loop is considered.

5.8 Figures

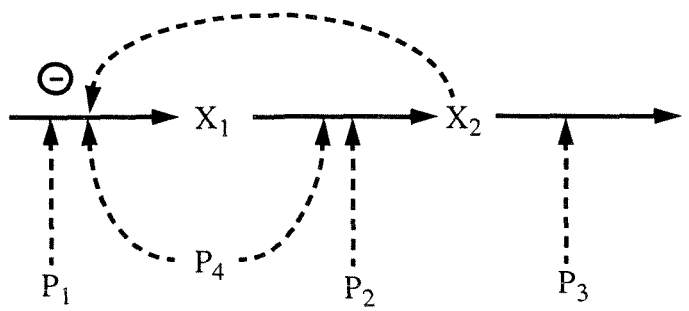

Figure 5.1: Linear pathway wih feedback inhibition.

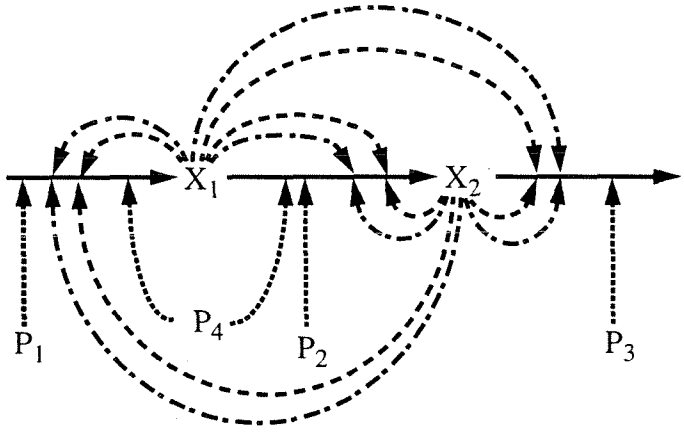

Figure 5.2: Linear pathway with regulatory superstructure around it. Dashed lines denote inhibition and dashed-dotted lines denote activation. Three different levels of strength are considered for each regulatory loop.

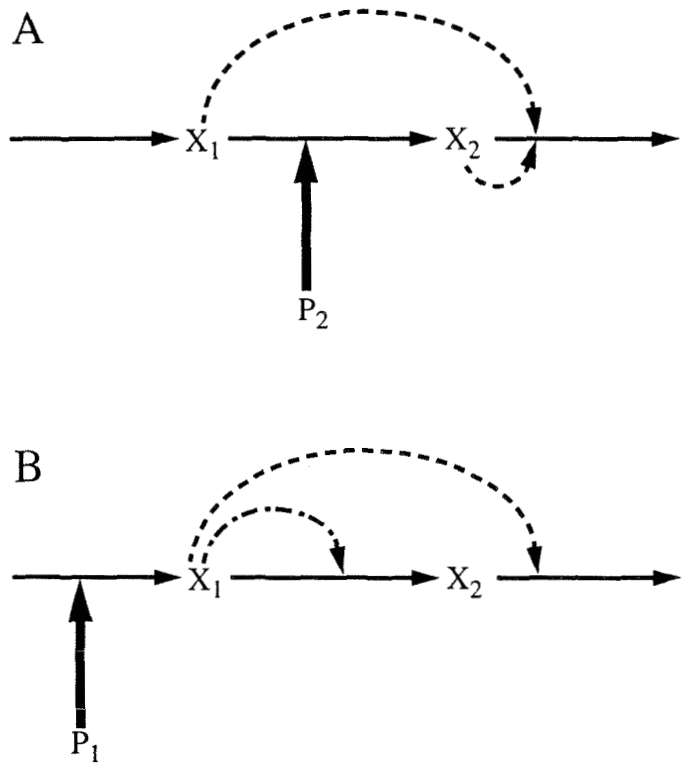

Figure 5.3: The first two optimal regulatory structures for the linear pathway. A. Both regulatory loops correspond to inhibition with strength -0.5. B. Dashed line denotes inhibition with strength -0.5 and dashed-dotted line denotes activation with strength 0.5.

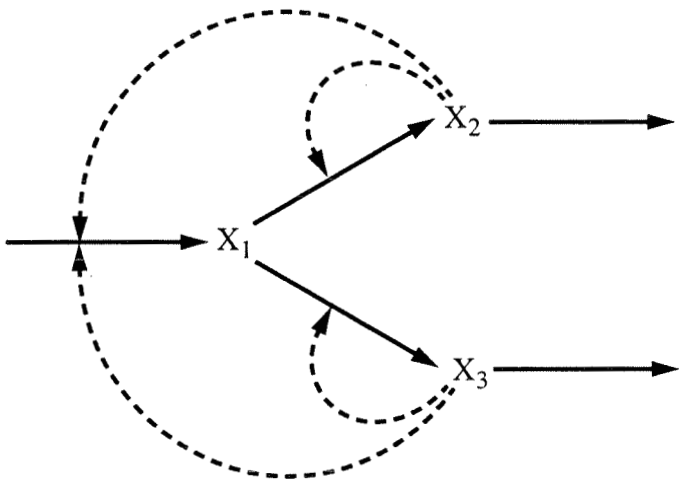

Figure 5.4: Branched biosynthetic pathway.

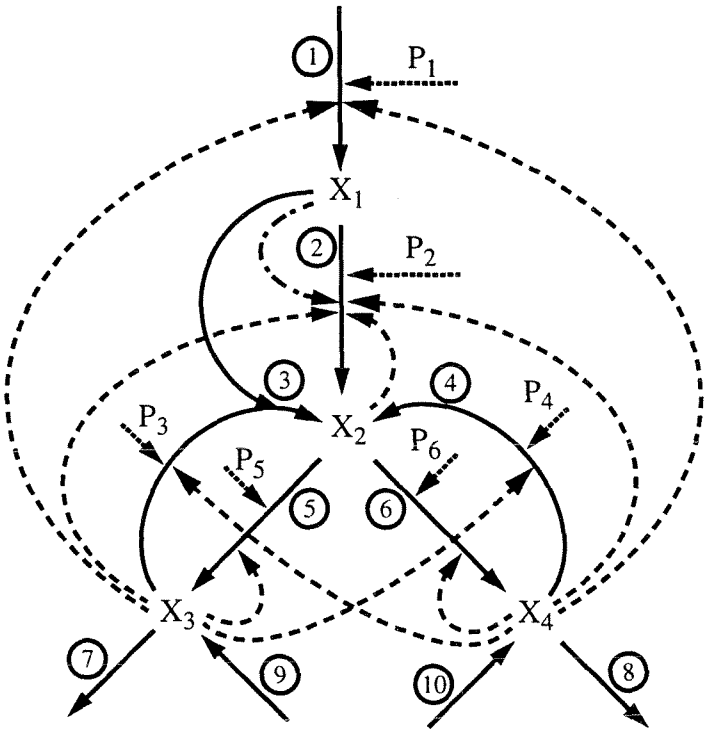

Figure 5.5: Xanthine monophosphate (XMP) and guanosine monophosphate (GMP) sythesis pathway. Solid lines denote reaction steps, dotted lines denote dependency on the corresponding parameters (P_l), dashed lines denote inhibition, and dashed-dotted lines denote activation. Numbers in circles identify the reaction steps as they are referenced in the text.

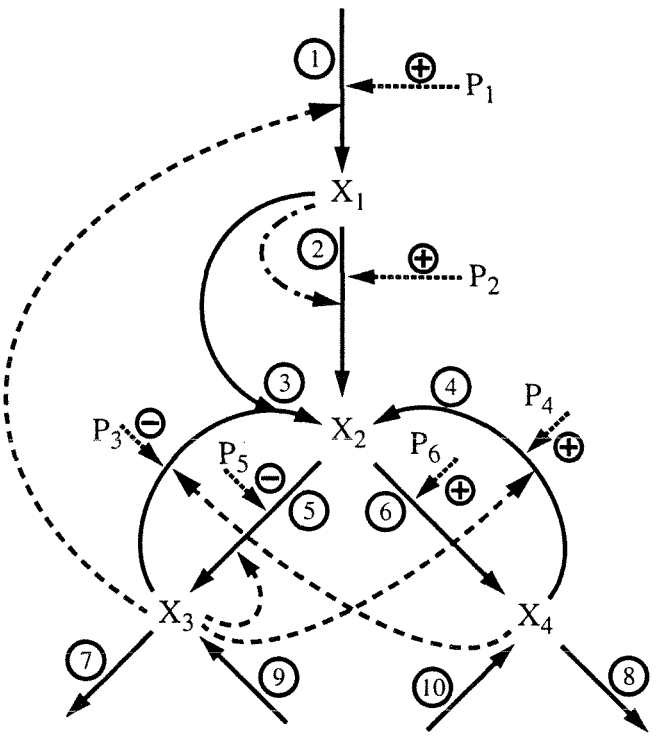

Figure 5.6: The changes in regulatory structure and in enzyme expression levels of the XMP and GMP synthesis pathway which maximize XMP and GMP (X_4) concentration. Signs in circles next to the dotted lines indicated increase (+) or decrease (−) in the expression level of the corresponding enzymes (P_l).

Chapter 6 Multiple Steady States in Metabolic Reaction Networks

6.1 Introduction

Discovery and technological development in several different fields of study have empowered the recent emergence of metabolic engineering, a powerful strategy for development of improved industrial organism. The interplay between increased definition of cellular phenomena at the molecular level and chemical reaction engineering provides the basic frameworks for analysis of cell function and underlies algorithms for qualitative guidance of metabolic design.

Because addressing the entire network of catalytic reactions and regulatory interactions at the level of the entire cell exceeds current knowledge, generally a subsystem of this network is selected for study. The mathematical descriptions of such sub-systems in steady state and for in transient conditions generally lead to coupled non-linear equations due to stoichiometric coupling among the reactions and kinetic nonlinearities for individual catalytic steps. Based upon studies of substantially simpler synthetic catalytic networks, generations of chemical reaction engineers have articulated the possibilities for steady-state multiplicity, limit-cycle oscillations, and more complex dynamic phenomena in non-linear chemical reaction system (Aris and Amundson, 1958; Burns *et al*, 1973; Uppal *et al*, 1974; Balakotaiah and Luss, 1983; Lyberatos *et al.*, 1985; Kevrekidis *et al.*, 1986; Farr and Aris, 1986; Aris and Cicarelli, 1995). This analysis has been built upon an expanding set of mathematical theory and efficient computational tools.

Prior investigations of steady-state multiplicity in man-made catalytic networks indicate several important motivations for characterizing steady-state multiplicity in metabolic systems. First, the possibility of steady-state multiplicity implies a complication in operating protocol since the steady state achieved can depend on earlier process history. Second, the existence of multiple steady states, when explored experimentally, can provide the engineer with important information about the system, including its governing equations and important parameters. Moreover, in a biological context, steady-state multiplicity can, in its most rudimentary form of two possible steady states, provide a biochemical mechanism for binary information storage, as

noticed in several previous investigations (Burns *et al.*, 1973; Kauffman, 1993).

If a biological system can operate in multiple steady states, it can, for a given configuration of protein activities, access different time-variant functional states, a capability which could be important from an evolutionary view point (Kauffman, 1993). In biotechnological applications, where cellular processes are manipulated by reconfiguration of protein activities, appearance of multiple steady states could lead to undesirable situations for bioprocess operation.

In spite of the technological and scientific significance of steady-state multiplicity in metabolic reaction systems, this phenomenon has been far less thoroughly investigated in biological context than in man-made catalytic process (Sel'kov and Betz, 1973; Heinrich *et al.*, 1977; Markus and Hess, 1984, 1990. Here a rudimentary description of glycolysis, a process by which a carbon source is digested by a cell, is formulated based upon pathways pertinent to bacteria. Considering rate expressions of minimal complexity, which do not include some important coupling and nonlinearity present in bacterial glycolysis, we demonstrate that, for certain values of the system parameters, as many as ten steady states can be achieved.

6.2 A Kinetic Model for a Glycolysis Prototype

The glycolysis pathway for bacterial cells such as *Escherichia coli* and *Bacillus subtilis* is presented in Figure 1 (Gottschalk, 1986). Although at first glance it appears to be a simple linear sequence of reactions, inspection of the stoichiometry of glycolysis reveals some interesting complications. These arise from participation of ATP and ADP in several reactions and from the involvement of PEP, an intermediate late in the pathway, in the first reaction. Moreover, the regulatory architecture enveloping this reaction network, which does not appear in the Figure 1, and which modulates the catalytic activities of certain reactions by compounds which are neither reactants (substrates) nor products of those reactions, introduces further complex coupling. Furthermore, if the organism in which this pathway operates is growing, then some of the metabolites are used for synthesis of new biomass and, additionally, the intracel-

lular concentration of every metabolite is affected due to the increasing cell volume (Fredrickson, 1976).

In this study we consider a glycolysis prototype; that is, a simpler reaction network that preserves the stoichiometric characteristics of the original pathway. The prototype used here is presented in Figure 2. It can be easily deduced, by comparing Figure 2 with Figure 1, that G corresponds to the pool of G6P and F6P, F corresponds to FdP, E corresponds to PEP, T corresponds to ATP, D to ADP, and M to AMP, and S and P correspond to extrecellular glucose and to pyruvate, respectively. The reactions from FdP to PEP, which are reversible, are lumped in reaction step 3 in Figure 2 for which, after the lumping, we have the following stoichiometry:

$$F + 2D \rightleftharpoons 2E + 2T$$

Reaction step 5 corresponds to the reversible reaction catalyzed by adenylate kinase (AK in Figure 1) and has the following stoichiometry:

$$T + M \rightleftharpoons 2D$$

The rest of the reaction steps are irreversible and have the following stoichiometries:

$$1 : E + S \rightarrow G + P$$

$$2 : G + T \rightarrow F + D$$

$$4 : E + D \rightarrow T + P$$

$$6 : D \rightarrow T$$

and

$$7 : T \rightarrow D$$

The last two reactions, 6 and 7, describe the production and consumption, respectively, of energy by cellular processes other than glycolysis. Reactions 8 and 9 corre-

spond to degradation (AS in Figure 1) and to synthesis (AD in Figure 1), respectively, of the adenylate nucleotides via M.

6.3 The Mathematical Model

Assuming negligible concentration gradients within the bacterial cell volume, unsteady-state mass balances on the components of the prototype pathway may be written as follows:

$$\frac{d[G]}{dt} = V_1 - V_2 - \mu Y_G - \mu[G] \tag{6.1}$$

$$\frac{d[F]}{dt} = V_2 - V_3^+ + V_3^- - \mu Y_F - \mu[F] \tag{6.2}$$

$$\frac{d[E]}{dt} = 2V_3^- - 2V_3^+ - V_1 - V_4 - \mu Y_E - \mu[E] \tag{6.3}$$

$$\frac{d[T]}{dt} = -V_2 + 2V_3^+ - 2V_3^- + V_4 - V_5^+ + V_5^- + V_6 - V_7 - \mu Y_T - \mu[T] \tag{6.4}$$

$$\frac{d[D]}{dt} = V_2 - 2V_3^+ + 2V_3^- - V_4 + 2V_5^+ - 2V_5^- - V_6 + V_7 - \mu Y_D - \mu[D] \tag{6.5}$$

$$\frac{d[M]}{dt} = -V_5^+ + V_5^- - V_8 + V_9 - \mu Y_M - \mu[M] \tag{6.6}$$

where the terms μY_j correspond to the usage of metabolite j for biosynthetic requirements and the last term in every mass balance, i.e. $\mu[j]$, represent dilution caused by increase in cell volume, V_c, at a specific rate:

$$\mu = \frac{dlnV_c}{dt} \tag{6.7}$$

The superscripts "+" and "–" denote the forward and backward direction, respectively, of the reversible reactions in Figure 2 as they have been conventionally assigned.

Two assumptions that are usually invoked about glycolysis by both modelers and biochemists are the following ones concerning the characteristic times of individual reaction steps (Reich, 1974; Heinrich, 1977; Gottschalk, 1985; Liao *et al.*, 1988; Fothergill-Gilmore and Michels, 1993):

A.I Reactions steps 3 and 5 in the prototype model (Figure 2), and the corre-

sponding reactions in Figure 1, are very "fast" with respect to the rest of the reactions; i.e., the metabolites that participate in these reactions as reactants and products are very close to thermodynamic equilibrium.

A.II The reactions that degrade and synthesize adenylates (V_8 and V_9, respectively), are very "slow" with respect to the rest of the reactions.

This last assumption can be mathematically described by the following equality:

$$V_8 - V_9 = \mu([T] + [D] + [M]) \tag{6.8}$$

which means that the net synthesis rate of the adenylate nucleotides via AMP (M) is equal to their dilution caused by cell volume increase.

Consideration of the above assumptions allows us to rewrite the mass balance equations as:

$$\frac{d[P_1]}{dt} = V_1 - V_2 - \mu Y_{P_1} - \mu[G] \tag{6.9}$$

$$\frac{d[P_2]}{dt} = V_2 + V_4 + V_6 - V_7 - \mu Y_{P_2} - \mu[P_2] \tag{6.10}$$

$$\frac{d[P_3]}{dt} = 2V_2 - V_1 - V_4 - \mu Y_{P_3} - \mu[P_3] \tag{6.11}$$

where

$$P_1 = G \tag{6.12}$$

$$P_2 = 2F + 2T + D \tag{6.13}$$

$$P_3 = 2F + E \tag{6.14}$$

and from assumptions A.I and A.II the following equlibrium and conservation relations hold:

$$q_t = \frac{[E]^2[T]^2}{[F][D]^2} \tag{6.15}$$

$$q_a = \frac{[D]^2}{[T][M]} \tag{6.16}$$

and

$$[T] + [D] + [M] = [A] \tag{6.17}$$

where $[A]$ is the concentration of the adenylate nucleotides which, after the assumption A.II, can be considered time-invariant. Equations (6.12) through (6.17) permit calculation of the concentrations of the prototype metabolites D, E, F, G, M, and T in terms of $[P_1]$, $[P_2]$, and $[P_3]$. In order to further simplify our analysis we will assume that

A.III There is no metabolic regulation of the activities of the various enzymes of the pathway.

Based on this assumption we will use the simplest possible enzymatic rate expressions to describe the kinetics of the enzyme-catalyzed reactions:

$$V_1 = V_{m,1}\frac{[E]}{K_{1,E} + [E]} \tag{6.18}$$

$$V_2 = V_{m,2}\left(\frac{[G]^2}{K_{G,2}^2 + [G]^2}\right)\left(\frac{[T]^2}{K_{T,2}^2 + [T]^2}\right) \tag{6.19}$$

$$V_4 = V_{m,4}\left(\frac{[E]^2}{K_{E,4}^2 + [E]^2}\right)\left(\frac{[D]^2}{K_{D,4}^2 + [D]^2}\right) \tag{6.20}$$

$$V_6 = V_{m,6}\frac{[D]}{K_{6,D} + [D]} \tag{6.21}$$

$$V_7 = V_{m,7}\frac{[T]}{K_{7,T} + [T]} \tag{6.22}$$

where $V_{m,j}$ is the maximum rate for reaction j, and $K_{i,j}$ is the dissociation constant of the enzyme that catalyzes reaction step j for the metabolite i. Parameter $V_{m,1}$ is also a function of external glucose concentration which in the following analysis will be fixed and considered as time-invariant. Notice also that V_2 and V_4 follow the Hill rate expression; i.e., sigmoidal kinetics, with a Hill coefficient equal to two, since it has been shown that the enzymes that catalyze these reactions (phosphofructokinase and pyruvate kinase, respectively) are composed of identical monomers, or identical pairs of monomers, and such enzymes typically follow Hill-type kinetics.

As was mentioned above the participation of T and D in several reactions, and the

effects of cell growth complicate the behavior of the system as described by equations (6.9)-(6.11). Therefore, we will further assume that:

A.IV The concentrations of T and D saturate the enzymes that catalyze reactions 2 and 4; i.e., $K_{T,2} \ll [T]$ and $K_{D,4} \ll [D]$; and

A.V The cells are not growing: $\mu = 0$.

Under these assumptions we can write the following dimensionless mass balance equations:

$$\frac{d\pi_1}{d\tau} = v_1 - v_2 \tag{6.23}$$

$$\frac{d\pi_2}{d\tau} = v_2 + v_4 + v_6 - v_7 \tag{6.24}$$

$$\frac{d\pi_3}{d\tau} = 2v_2 - v_1 - v_4 \tag{6.25}$$

where for the dimensionless concentrations we have:

$$\begin{array}{lll} \pi_1 = \gamma & \pi_2 = 2\sigma + 2\rho\phi_3 + \rho\phi_2 & \pi_3 = 2\sigma + \epsilon \\ \gamma = \dfrac{[G]}{K_{E,1}} & \sigma = \dfrac{[F]}{K_{E,1}} & \epsilon = \dfrac{[E]}{K_{E,1}} \\ \phi_3 = \dfrac{[T]}{[A]} & \phi_2 = \dfrac{[D]}{[A]} & \phi_1 = \dfrac{[M]}{[A]} \\ \tau = \dfrac{tV_{m,1}}{K_{E,1}} & & \rho = \dfrac{[A]}{K_{E,1}} \end{array} \tag{6.26}$$

For the equilibrium and conservation relations (6.15), (6.16), and (6.17) and the dimensionless rates expressions, we have:

$$\widetilde{q}_t = \frac{\epsilon^2\phi_3^2}{\sigma\phi_2^2} \tag{6.27}$$

$$q_a = \frac{\phi_2^2}{\phi_3\phi_1} \tag{6.28}$$

$$\phi_3 + \phi_2 + \phi_1 = 1 \tag{6.29}$$

$$v_1 = \frac{\epsilon}{1+\epsilon} \tag{6.30}$$

$$v_2 = \beta \frac{\gamma^2}{\eta^2 + \gamma^2} \tag{6.31}$$

$$v_4 = \alpha \frac{\epsilon^2}{\theta^2 + \epsilon^2} \tag{6.32}$$

$$v_6 = \psi \frac{\phi_2}{\xi + \phi_2} \tag{6.33}$$

$$v_7 = \omega \frac{\phi_3}{\lambda + \phi_3} \tag{6.34}$$

where the dimensionless parameters introduced above are defined as:

$$\begin{array}{lll} \tilde{q}_t = \dfrac{q_t}{K_{E,1}} & \beta = V_{m,2}/V_{m,1} & \eta = K_{G,2}/K_{E,1} \\ \alpha = V_{m,4}/V_{m,1} & \theta = K_{E,4}/K_{E,1} & \psi = V_{m,6}/V_{m,1} \\ \xi = K_{D,6}/[A] & \omega = V_{m,7}/V_{m,1} & \lambda = K_{T,7}/[A] \end{array} \tag{6.35}$$

Equations (6.23)-(6.34) describe the mathematical model of the simple glycolytic prototype presented in Figure 2, under the assumptions A.I-A.V. The assumptions used are not many and they do not interfere with the objectives of this analysis. Many of them are based on experimental studies (A.I and A.II), and they can be implemented experimentally easily (A.V). Moreover, metabolic engineering allows the consideration of a glycolytic pathway without metabolic regulation (A.III). Study of the prototype model in the absence of regulation will allow us to better understand the function of the existent regulatory structures by observing the changes in the system due to the introduction of the regulatory interactions. Previously studied prototypies, although they have been constructed based on more assumptions and simplifications were able to capture some of the qualitative features of glycolysis, such as oscillatory behavior and chaos (Higgins, 1967; Sel'kov and Betz, 1973; Heinrich *et al.*, 1977). However, the insight they provided was limited, because of the assumptions and their general character since they did not describe glycolysis from any particular organism.

Analysis of the glycolysis prototype will be performed next in order to identify if there can exist values for the kinetic parameters, realizable by natural mutations or genetic engineering, that could possibly result in multiple steady states and, thus, in

more than one phenotypes from a given genotype.

6.4 Steady-State Multiplicities

Previous models of glycolysis in various cells, such as yeast and erythrocytes and experimental data suggest that glycolysis can exhibit multiple steady-states (Heinrich *et al.*, 1977; Liao *et al.*, 1988; Joshi and Palsson, 1989). We will study the glycolytic prototype described by equations (6.23)-(6.29) in order to find necessary parameter value combinations for steady-state multiplicities to be observed, under the assumptions A.I-A.V, and the rate laws assumed (Equations (6.30)-(6.34)).

The steady-states are found by solving (6.23)-(6.25) for $d\pi_i/d\tau = 0$ $(i = 1, 2, 3)$. At any steady-state, after rearrangement of the equations, we have:

$$v_1(\epsilon) = v_4(\epsilon) \tag{6.36}$$

$$v_2(\sigma) = v_4(\epsilon) \tag{6.37}$$

$$2v_4(\epsilon) = v_7(\phi_3) - v_6(\phi_2(\phi_3)) \tag{6.38}$$

The function $\phi_2(\phi_3)$ appearing in Equation (6.38) is obtained by solution of the equilibrium equation (6.28) and the conservation equation (6.29), which yields:

$$\phi_2 = -\frac{1}{2}q_a\phi_3 + \frac{1}{2}\sqrt{(q_a - 4)q_a\phi_3^2 + 4q_a\phi_3} \tag{6.39}$$

Equation (6.36) can be solved explicitly for ϵ. The solutions will be the intersections of the curves for the rate laws for $v_1(\epsilon)$ and $v_4(\epsilon)$. The relationship between these two functions depends on the parameters α and θ in the formula for v_4 (Equation (6.32)) with possible cases schematically illustrated in Figure 3. The dashed lines correspond to different possible cases for v_4, and can intersect (or not) the solid line which corresponds to v_1. There are five qualitatively different ways that v_1 can be related to v_4. If v_4 follows kinetics that result in line I, then the system has two steady-state solutions, one zero and one nonzero. It is easy to see that this can occur

only when $V_{m,1} < V_{m,4}$, or when $\alpha < 1$. As α diminishes and attains $\alpha = 1$, then the v_1 and v_4 functions approach the same asymptote as $\epsilon \to \infty$ (line II in Figure 3) and we have just lost the case with two solutions. $\alpha = 1$ is a bifurcation point, since for $\alpha > 1$ there is only one solution with zero value (line V in Figure 3). However, depending on the value of θ, when α is less than one, we can move from a unique solution to three solutions (line III) through a bifurcation represented by line IV in Figure 3. The bifurcation diagram for the rate laws used is presented in Figure 4. The $\theta - \alpha$ paramater space is divided by lines II and IV in to three regions which correspond to a different number of steady states. For values of α and θ in region I v_4 follows kinetics that result in line I in Figure 3, and, therefore, two steady-state solutions exist. For values of α and θ in regions III and V v_4 follows kinetics that result in lines III and V, respectively, in Figure 3, and, thus, three and one steady-state solutions exist for parameter values in regions III and V, respectively.

The above conclusions are independent of the details of the kinetics. The only requirement for the possibility of three steady states under some conditions is that v_4 follows sigmoidal kinetics. The qualitative results are also independent of the value of the Hill coefficient. A higher value for the Hill coefficient will only increase area III in the bifurcation diagram (Figure 3) favoring multiplicity of steady states for $\alpha < 1$.

Solution of equation (6.37) provides a sufficient condition between α and β so that for every ϵ there is a solution for σ satisfying equation (6.37):

$$\beta \geq \alpha \tag{6.40}$$

In general, it is possible that solutions to equation (6.37) can exist even if inequality (6.40) does not hold, but the existence of a σ value satisfying equation (6.37) depends on the relative values of the other parameters involved; i.e., θ and η.

In the case of three ϵ values satisfying equation (6.36), one of them will always be zero, and the corresponding solution of equation (6.37) is $\sigma = 0$. This solution will be called *v-zero* solution since, at this solution, the reaction rates v_1, v_2 and v_4 will be equal to zero. The other two nonzero ϵ solutions of equation (6.36), and associated

solutions of equations (6.37) and (6.38), when they exist, will be positive and will be called *v-positive* solutions.

Equation (6.38) provides the steady-state relation of ϕ_3 with ϵ and σ satisfying equations (6.36) and (6.37). We consider next parameter values for which nonegative solutions of equation (6.38) exist.

For any allowed (i.e., nonengative) ϵ, v_4 is nonegative. Therefore, a sufficient condition for an *allowed* solution of equation (6.38) to exist is, for some ϕ_3 values between zero and one,

$$f(\phi_3) \equiv v_7(\phi_3) - v_6(\phi_2(\phi_3)) \geq 0 \tag{6.41}$$

This condition is sufficient since, if it is satisfied, then there exist combinations of ψ and ω that could satisfy equation (6.38). It should be noted here that $f(\phi_3)$ is a measure of the energetic state of the cells since it expresses the net ATP (ϕ_3) consumption rate.

From equation (6.39) we have $\phi_2 = 0$ at $\phi_3 = 0$ and at $\phi_3 = 1$. Therefore

$$f(0) = 0 \tag{6.42}$$

and

$$f(1) = \frac{\omega}{\lambda + 1} \tag{6.43}$$

and given the fact that f is a continuous function and that the maximum value of the left hand side of equation (6.38) is 2, then a necessary condition for the existence of at least one nonegative value of ϕ_3 satisfying (6.38) is:

$$2 \leq \frac{\omega}{\lambda + 1} \tag{6.44}$$

Numerical calculation of $f(\phi_3)$ for various combinations of values of the kinetic parameters ψ, ξ, ω, and λ revealed that $f(\phi_3)$ can have any of the qualitative forms indicated by the curves in Figure 6. For certain combinations of the kinetic parame-

ters, f possesses a local maximum with positive value for values of ϕ_3 between 0 and 1 (lines III, IV, V, and VI in Figure 6). Therefore, we could find a combination of values for the parameters ψ and ω such that $2v_4(\epsilon)$ would be lower than this local maximum. As a result a horizontal line, positioned at an ordinate value $2v_4(\epsilon)$, will intersect these f loci at three points resulting in three or one values of ϕ_3 which satisfy equation (6.38) (lines I and VIII).

If there are three solutions of equations (6.36) and (6.37) for ϵ and γ, then there could exist one solution ϕ_3 to equation (6.38) for each of them (line I), or one solution for the v-positive solutions and two solutions for the v-zero solution (line VIII), or three solutions for each of the v-positive solutions and one for the v-zero solution (line II), or three for the v-positive solutions and three, or four, for the v-zero solution (lines V and VI, respectively).

The lines II, IV and VIII correspond to limiting cases. Line II illustrates the transition from one solution of ϕ_3 to three possible solutions for each of the v-positive cases. This transition can be observed when there exists a concentration ϕ_3^* for which the following equalities hold:

$$\left.\frac{df}{d\phi_3}\right|_{\phi_3^*} = 0 \tag{6.45}$$

and

$$\left.\frac{d^2f}{d\phi_3^2}\right|_{\phi_3^*} = 0 \tag{6.46}$$

Line IV corresponds to the transition from one solution for the v-zero case to three (line V) or four possible solutions (line VI), while three solutions for the v-positive case are still possible. Such a transition can be observed when there exists a concentration ϕ_3^o at which:

$$\left.\frac{df}{d\phi_3}\right|_{\phi_3^o} = f(\phi_3^o) = 0 \tag{6.47}$$

These last equalities (6.47) hold also for line VII which illustrates the transition from three possible solutions for every v-positive case and three or four solutions for the v-zero case, to one solution for any of the two cases. The difference between these two transitions is that at the corresponding ϕ_3^o line II goes through a local minimum (i.e.

$(d^2f/d\phi_3)_{\phi_3^o} < 0$), and line VII goes through a local maximum (i.e. $(d^2f/d\phi_3)_{\phi_3^o} > 0$).

Considering the above conditions and following an approach similar to the one introduced by Regenass and Aris (1965) we can find in the parameter space regions for which multiplicities can arise. These regions will identify necessary sets of values for the kinetics parameters for the number of the solutions to possibly exceed the number of the solutions allowed by equation (6.36).

Although function f depends on five parameters, i.e. ψ, ξ, ω, λ, and q_a, the conditions postulated above depend on four essential parameters:

$$\xi,\ \lambda,\ q_a,\ \text{and}\ \delta = \frac{\omega}{\psi}$$

By fixing two of them, ξ and q_a, a two-parameter bifurcation diagram can be constructed (Figure 7). The numbers in the regions and on the bifurcation lines correspond to the multiplicity possibilities which are illustrated in Figure 6 and have been described above. Assuming that three solutions for σ and ϵ exist then for the overall system we will have:

SS.I Three steady-states for region I;

SS.II Seven steady-states for region III;

SS.III Nine or ten steady-states for region V-VI; and

SS.IV Four steady-states for region VIII.

The special nonlinear character of function f arises from the nonlinear dependece of ϕ_2 on ϕ_3, as described by equation (6.39). The equilibrium constant, q_a, is the only parameter appearing in equation (6.39) to have an effect on the value of ϕ_2 for a given value of ϕ_3. In order to understand better the dependency of ϕ_2 on ϕ_3 and q_a, ϕ_2 as a function of ϕ_3 is plotted in Figure 5 for different values of q_a. As we can see ϕ_2 is not a monotonic function of ϕ_3; it increases monotically as ϕ_3 increases from zero, it goes through a maximum and it decreases monotonically as ϕ_3 approaches one, which is the maximum value for ϕ_3. The maximum value of ϕ_2 and the corresponding value for ϕ_3 are functions of q_a. Solution of the equation $(d\phi_2/d\phi_3) = 0$ provides the

maximun value of ϕ_2, $\phi_{2,max}$, and the corresponding value of ϕ_3, $\phi_{3,max}$, as functions of q_a:

$$\phi_{2,max} = \frac{\sqrt{q_a}}{\sqrt{q_a}+2} \tag{6.48}$$

and

$$\phi_{3,max} = \frac{1}{2}\left(\frac{4-2\sqrt{q_a}}{4-q_a}\right) \tag{6.49}$$

which at this value is equal to ϕ_1.

Figure 5 further suggests that q_a has an important effect on the distribution of the various forms of the adenylate nucleotides (ϕ_i). For

$$\phi_3 > 1/(2q_a+1)$$

ϕ_2 will always be higher than ϕ_1, which implies that for low values of q_a, ATP (ϕ_3) and AMP (ϕ_1) will be the main components of the adenylate nucleotide pool, over a wide range of ATP concentration (ϕ_3), whereas, for high values of q_a, ATP (ϕ_3) and ADP (ϕ_2) will be the main components of the adenylate nucleotide pool.The dependency of $\phi_{2,max}$ and $\phi_{3,max}$ on q_a indicates that q_a could be an important bifurcation parameter.

The effects of the value of ξ and q_a have been further examined and the results are presented in Figure 8. For a given value of ξ, the area of the regions of multiplicity depend on the value of q_a. In general, as q_a increases, the area of the multiplicity regions in the $ln(\delta)-ln(\lambda)$ parameter space is decreasing in size. Whereas, for a given value of q_a, the area of these regions increase as ξ increases. The bifurcation lines II, IV, and VII, which define the multiplicity regions, have different sensitivities with respect to q_a. As q_a increases, line II moves upwards until a critical value of q_a beyond which it moves downwards as q_a increases. This critical value appears to be dependent on the value of ξ. On the other hand, lines IV and VII move monotonically upwards as q_a increases with a rate that depends on the value of ξ. All three, for fixed ξ, lines respond to changes in q_a without changing significantly their slopes in the $ln(\delta)-ln(\lambda)$ parameter space, which implies that q_a does not affect the (scaled) sensitivity of these bifurcation lines with respect to δ and λ.

This analysis has been based on some assumptions that concern the kinetics of the reactions (A.III and A.IV). The above results provide us with the necessary insight to the system regarding the important bifurcation parameters with respect to which further computational analysis can be conducted, since relaxation of the assumptions will result in a nonlinear system that can be analyzed only computationally.

6.5 Discussion

Construction of a prototype model for bacterial glycolysis presented in this chapter using certain simplifying assumptions and the simplest possible kinetics for the enzymes. Analysis of the number of the possible steady states with respect to the values of the kinetic parameters was performed suggesting that up to ten steady states are possible. Parent models, which features higher complexity due to elaborate kinetics of the enzymes and regulatory interactions will have different multiplicity characteristics which cannot be predicted based on the above analysis. Moreover, these models can be analyzed only computationally. However, the prototype model and the above results can serve as a starting point for further analysis of more complicated models.

From the above analysis, it can be concluded multiple steady states may arise in bacterial glycolysis because of three main stoichiometric and kinetic characteristics of the system: (i) the dependency of glucose uptake (v_1) on PEP concentration (ϵ), (ii) the energetic state of the cell, as it is expressed by the relative magnitude of the corresponding kinetic parameters (ψ, ω, λ, and ξ), and (iii) the reactions that result in the relationships between the three components of the adenylate nucleotides (ATP (ϕ_3), ADP (ϕ_2), and AMP (ϕ_1)).

The analysis of glycolysis in living organisms can be analyzed by constructing prototype mathematical models. However, these prototype models are complex and the postulation of certain assumptions is required so that the complexity of the models will be reduced without losing the essential characteristics of the system. The method presented above for bacterial glycolysis can be also applied for the glycolysis in other organisms, such as yeast and mammalian cells, that have different stoichiometry and

regulation. The choice of the assumptions have been shown to be an important step in the analysis of such prototypes.

The bifurcation diagram presented in Figure 4 suggests that overexpression of the enzyme that catalyzes the reaction 1 beyond a critical value for which $V_{m,1} > V_{m,4}$ will result in a unique steady state for the system with a value for all of the concentrations equal to zero. This suggestion implies that in an experiment that such an overexpression will be achieved it will not be observed since the cells that will express $V_{m,1} > V_{m,4}$ will not be able to survive and only the ones with $V_{m,1} < V_{m,4}$ will survive and be observed. This result though holds under the assumptions made. Qualitative similar results, i.e. maximum overexpression level that can be achieved, will be also obtained when the assumptions are removed. Similar information can obtained by studying the bifurcation diagrams in Figure 8.

The stability of the steady-states is another important problem that has not been addressed in the present study and is currently under investigation. The number of the multiple steady-states and the existence of the metabolite pools, π_j, do not allow the postulation of simple rules for stability analysis. Moreover, the existence of steady states with zero concentrations, which imply cell death, introduce questions like: What is the stability of the steady states at which one of the metabolite concentrations is zero? If they are stable, what is their basin of attraction? How do the regulatory structures affect their stability?

From a metabolic engineering point of view, the models and their analyses are more effective when one tries to identify wide parameter-value regions over which steady-state multiplicities are possible, since this could be an experimentally feasible scenario. On the other hand, multiplicity phenomena occuring over a narrow parameter-value range while mathematically interesting, have little chance to be implemented or observed. Given these complexities, mathematical models and analysis can provide imperative guidance for rational metabolic engineering.

6.6 References

Aris, R., and Amundson, N. R., 1958, An Analysis of Chemical Reactor Stability and Control, I, II, and III. *Chem. Engng. Sci.* **7**, 121-155.

Aris, R., 1976, How to Get the Most out of an Equation Without Really Trying. *Chem. Engng Educ.* **20**, 114-124.

Aris, R. and Cicarelli, P., 1995, Autocatalytic Continuous Reactions in a Stirred Tank: I. Multiplicity of Steady States. *Chem. Engng Sci.* **49**, 5307-5313.

Bailey, J. E., 1991, Toward a Science of Metabolic Engineering. *Science* **252**, 1668-1675.

Balakotaiah, V., and Luss, D., 1983, Multiplicity Criteria for Multi-Reaction Networks. *A.I.Ch.E. J.* **29**, 552-560.

Burns, D. D., Bailey, J. E. and Luss, D., 1973, Steady State Multiplicity and Stability of Enzymatic Reactions. *Biotech. Bioengng* **15**, 1131-1145.

Farr, W. W., and Aris, R., 1986, "Yet Who Would Have Thought the Old Man to Have so Much Blood in Him?"– Reflections on the Multiplicity of Steady States of the Stirred Sank Reactor. *Chem. Engng Sci.* **41**, 1385.

Fothergill-Gilmore, L. A. and Michels, P. A. M., 1993, Evolution of Glycolysis. *Prog. Biophys. Molec. Biol.* **59**, 105-235.

Fredrickson, A. G., 1976, Formulation of Structured Growth Models. *Biotechnol. Bioeng.* **28**, 1481-1486.

Gallazo, J. L. and Bailey, J. E., 1990, Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* **12**, 162-172.

Gottschalk, G., 1986, *Bacterial Metabolism*, Second Edition. Springer-Verlag, New York.

Heinrich, R., Rapoport, S. M. and Rapoport, T. A., 1977, Metabolic Regulation and Mathematical Models. *Prog. Biophys. Molec. Biol.* **32**, 1-82.

Higgins, J., 1967, The Theory of Oscillating Reactions. *Ind. Engng Chem.* **59**, 18-62.

Joshi, A. and Palsson, B. O., 1989, Metabolic Dynamics in the Human Red-Cell. I. A Comprehensive Kinetic-Model. *J. Theor. Biol.* **141**, 515-528.

Kauffman, S. A., 1993, *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford Univ. Press, New York.

Kevrekidis, I. G., Schmidt, L. D., and Aris, R., 1986, Some Common Features of Periodically Forced Reacting Systems. *Chem. Engng Sci.* **41**, 1263.

Liao, J. C., Lightfoot, E. N., Jolly, S. O. and Jacobson, G. K., 1988, Application of Characteristic Reaction Paths - Rate-Limiting Capability of Phosphofructokinase in Yeast Fermentation. *Biotech. Bioengng* **31**, 855-868.

Lyberatos, G., Kuszta, B. and Bailey, J. E., 1985, Bifurcation from the Potential Field Analog of Some Chemical Systems. *Chem. Engng Sci.* **40**, 1679-1687.

Markus, M. and Hess, B., 1984, Transitions Between Oscillatory Modes in a Glycolytic Model System. *Proc. Natl. Acad. Sci. USA* **81**, 4394-4398.

Markus, M. and Hess, B., 1990, Control of Metabolic Oscillations: Unpredictability, Critical Slowing down, Optimal Stability and Hysteresis. In *Control of Metabolic Processes* (edited by Cornish-Bowden, A. and Cárdenas, M. L.), pp. 303-313. NATO ASI Series,Plenum Press, New York.

Reich, J. G., 1974, Near-Equilibrium Reactions and the Regulation of Pathways. *Symp. Biol. Hung* **18**, 159-171.

Regenass, W. and Aris, R., 1965, Stability Estimates for the Stirred Tank Reactor. *Chem. Engng Sci.* **20**, 60-66.

Schaaff, I., Heinisch, J. and Zimmermann, F. K., 1989, Overproduction of Glycolytic-Enzymes in Yeast. *Yeast* **5**, 285-290.

Sel'kov, E. E. and Betz, A., 1973, On the Mechanism of Single-Frequency Glycolytic Oscillations. In *Biological and Biochemical Oscillators* (edited by B. Chance *et al.*), pp. 197-220. Academic Press, New York.

Uppal, A., Ray, W. H., and Poore, A. B., 1974, On the Dynamic Behavior of Continuous Stirred Tank Reactors. *Chem. Engng Sci.* **29**, 967-985.

Zimmermann, F. K., 1992, Glycolytic-Enzymes as Regulatory Factors. *J. Biotech* **27**, 17-16.

6.7 Figures

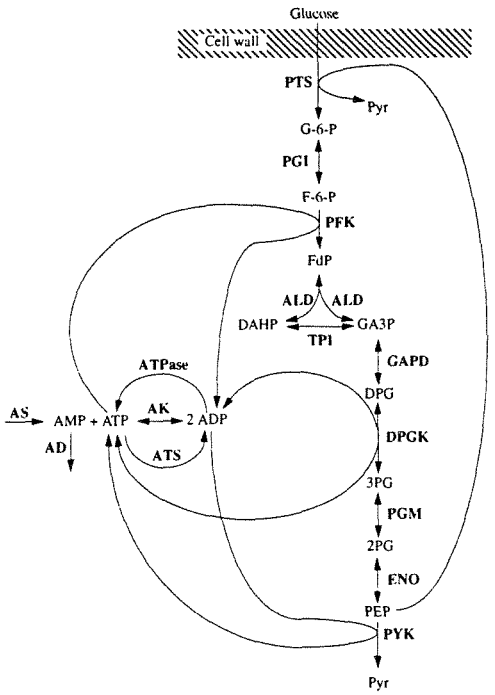

Figure 6.1: The glycolysis pathway in bacterial cells. Abbreviations: G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; FdP: fructose 1,6-diphospate; DAHP: dihydroxyacetone phosphate; GA3P: glyceraldehyde 3-phosphate; DPG: dipsphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; Pyr: pyruvate; ATP: adenosine triphosphate; ADP: adenosine diphosphate; AMP: adenosine monophosphate; **PTS**: phosphotransferase system; **PGI**: phosphoclucose isomerase; **PFK**: phosphofructokinase; **ALD**: aldolase; **TPI**: triosephosphate isomerase; **GAPD**: glyceraldehyde phosphate dehydrogenase; **DPGK**: diphosphoglycerate kinase; **PGM**: phosphoglycerate mutase; **ENO**: enolase; **PYK**: pyruvate kinase; **AK**: adenylate kinase; **ATPase**: ATPase; **ATS**: ATP synthesis; **AD**: adenylate degredation; **AS**: adenylate synthesis. Bold-face abbreviations indicate enzymes and reaction steps.

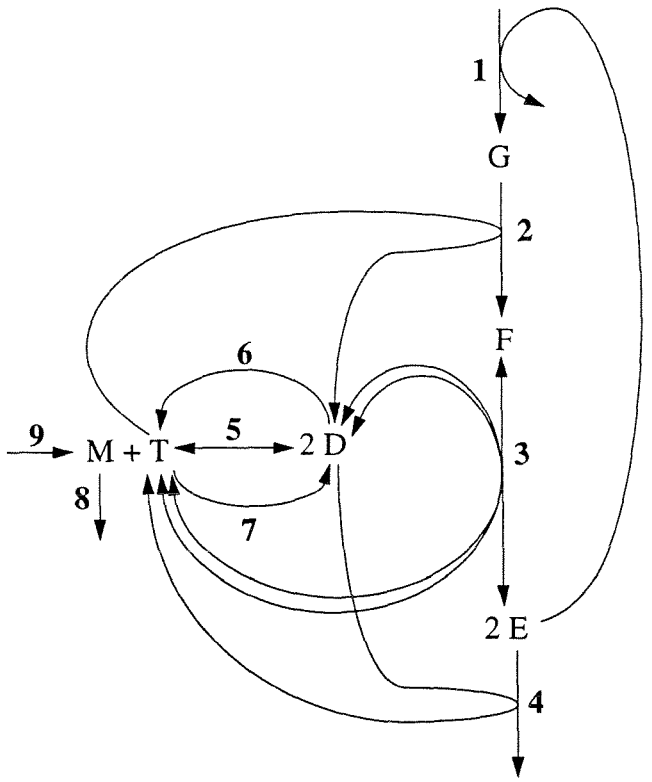

Figure 6.2: The glycolysis prototype pathway.

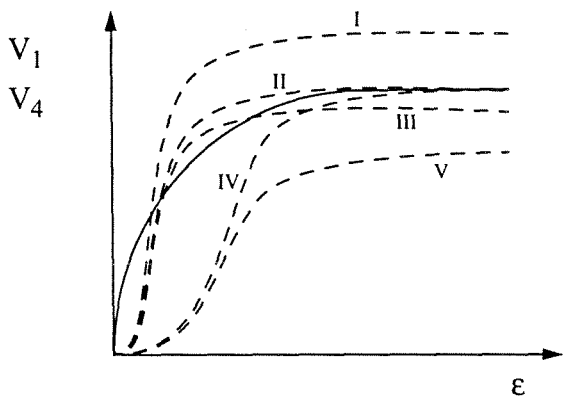

Figure 6.3: Qualitative graph of the dependency of v_1 (solid line) and v_4 (dashed line) on ϵ for different combiantions of the parameters α and θ.

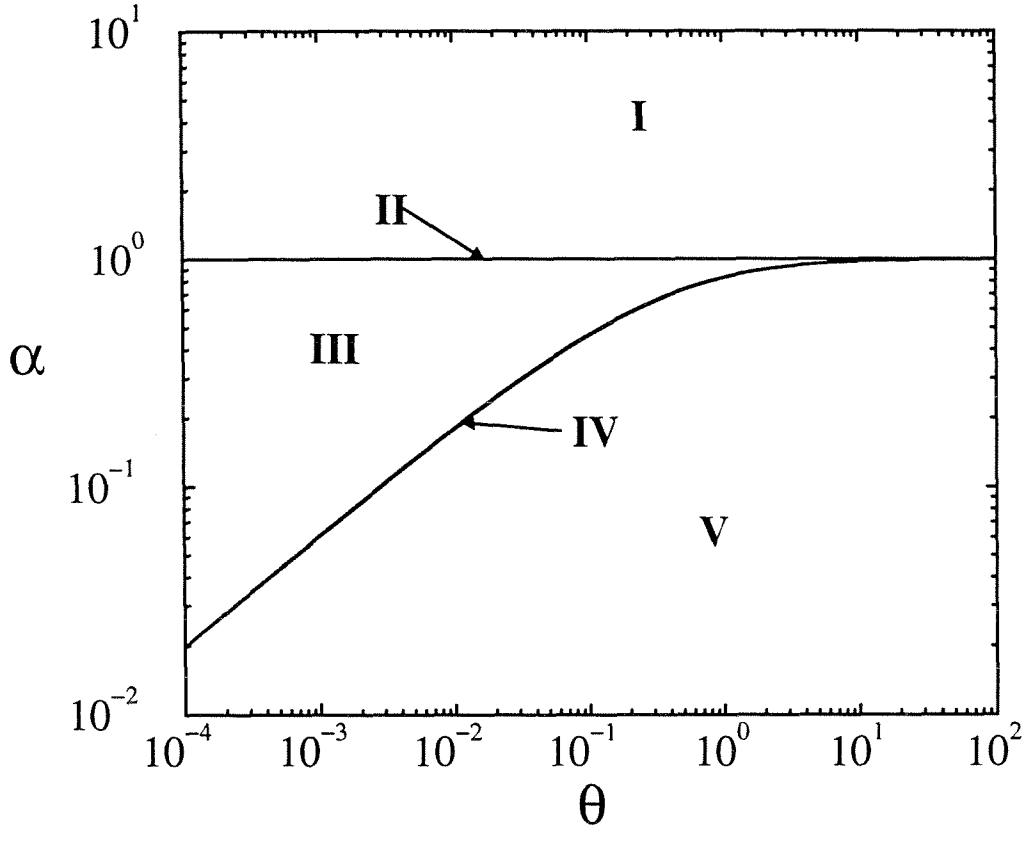

Figure 6.4: The multiplicity regions in the $\alpha-\theta$ parameter space. The latin numbers correspond to the scenarios shown in Figure 3. In region I two steady states are possible, in region III three steady states are possible, and in region V only steady state is possible.

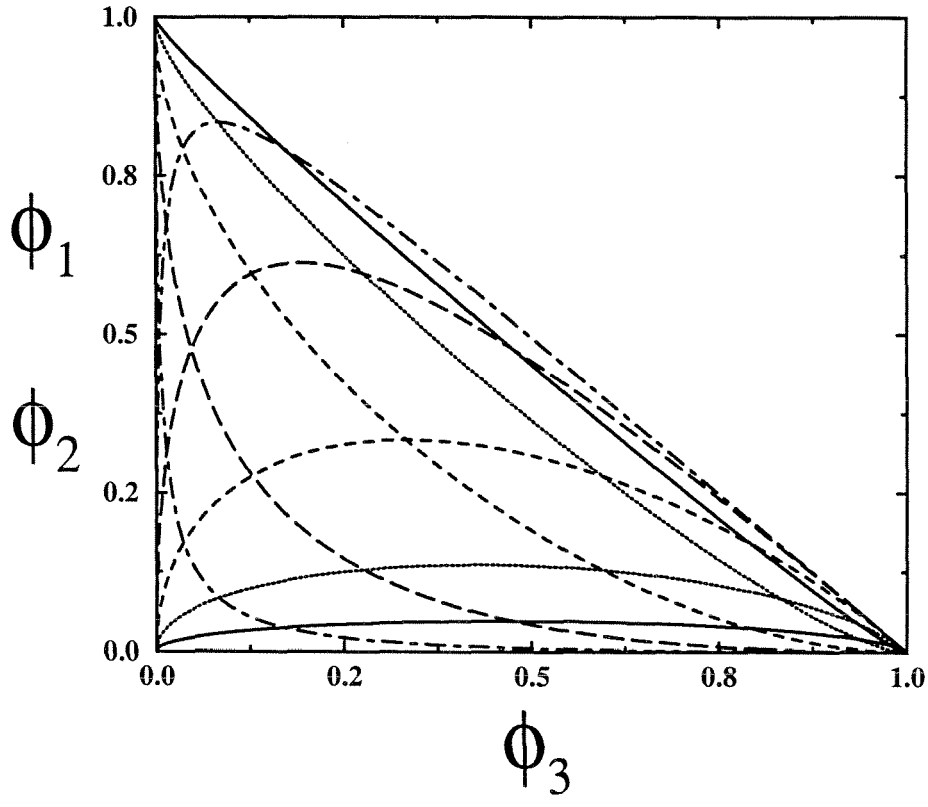

Figure 6.5: The dependency of ϕ_1 (convex lines) and ϕ_2 (concave lines) on ϕ_3 for different values of q_a : 0.01 (solid line), 0.1 (dotted line), 1 (dashed lines), 10 (long-dashed lines), and 100 (dashed-dotted line).

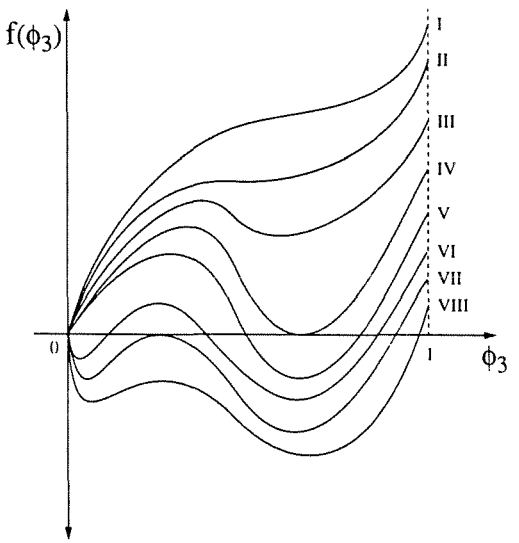

Figure 6.6: Qualitative diagram of the dependency of $f(\phi_3)$ on ϕ_3 for different values of the kinetic parameters ψ, ω, λ, and ξ.

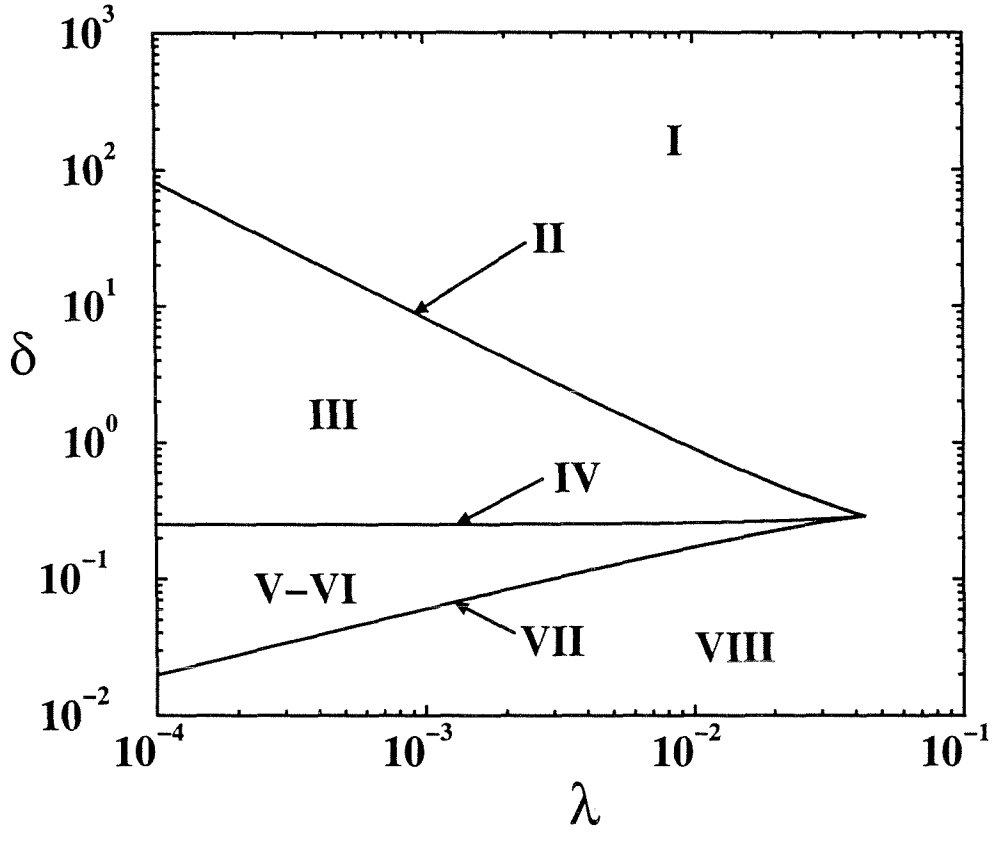

Figure 6.7: The multiplicity regions in the δ–λ parameter space for $q_a = 1$ and $\xi = 1$. The latin numbers correspond to the scenarios shown in Figure 6. In the parameter-value regions I, III, V-VI, and VIII, three, seven, nine or ten, and four steady states possibly exist for the system.

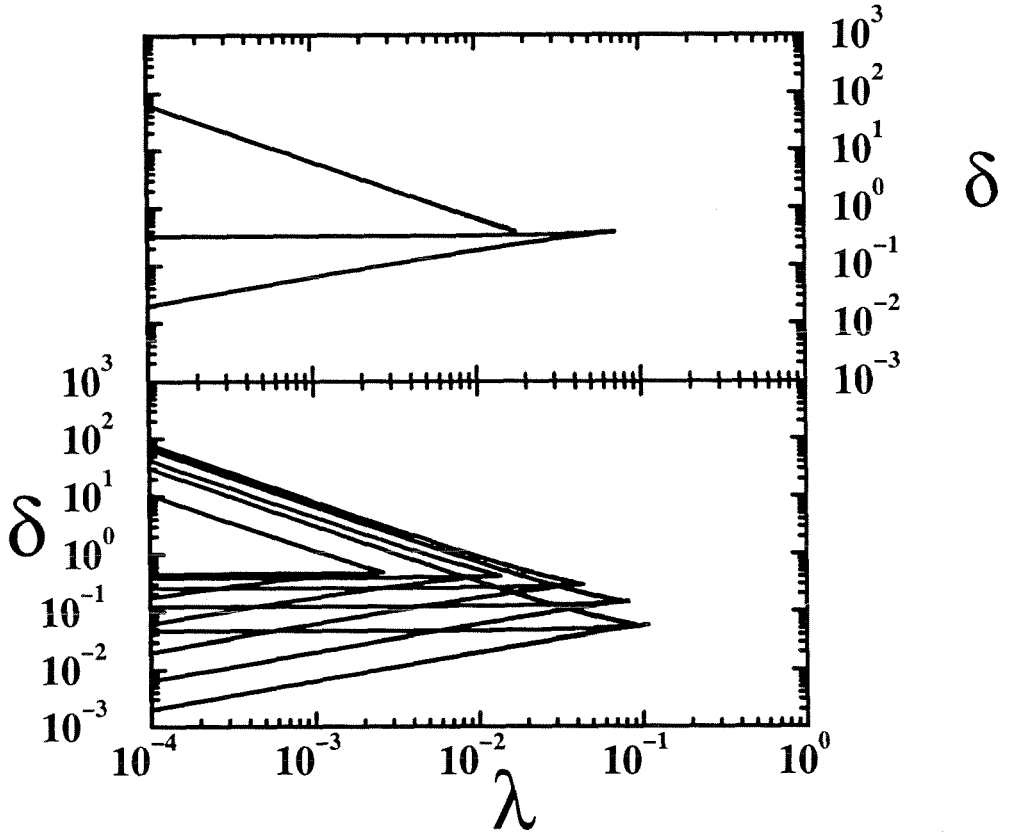

Figure 6.8: The multiplicity regions in the $\delta - \lambda$ parameter space for different values of q_a and ξ. Numbers on the arrows are the values considered for q_a; A: $\xi = 0.01$, B: $\xi = 0.1$, C: $\xi = 1$.

Chapter 7 Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Nonlinear Optimization

7.1 Introduction

The nonlinearity of the kinetics of the enzymes in metabolic pathways introduce complexities such as multiple steady-states, as was presented in the previous chapter. Therefore, the analysis and the design of the regulatory structures in a metabolic reaction network is even more difficult when a nonlinear description of the pathway is available. An optimization framework for nonlinear descriptions of metabolic systems is required.. In this chapter, such a framework will be presented.

The nonlinearity of the system and the discrete nature of the decisions concerning the regulatory structures leads to the formulation of this optimization problem as a mixed-integer nonlinear programming (MINLP) problem. These types of problems are very common in chemical engineering, and methods have been developed for their solution (Floudas, 1995). However, the success of these methods depends on the formulation of the problem, and this is the challenge in chemical engineering modeling and optimization. Similarly, the formulation of the optimization of the regulatory structure as an MINLP problem is a challenging, novel problem since the models for metabolic systems are not similar to any of the chemical process problems previously studied as MINLP problems.

7.2 Problem Statement and Formulation

7.2.1 Problem Statement

The general problem we address in this chapter is the same as in the previous chapter on optimization of regulatory structures:

A mathematical description of a metabolic pathway with a postulated number of regulatory loops is given. These regulatory loops are classified as either activation (increase the activity of the regulatory enzyme) or inhibition (decrease the activity of the regulatory enzyme) loops. The objective is to determine (i) which of the regulatory loops should be retained, and (ii) the number, type, and level of manipulation of amounts of enzymes, in order to optimize a certain function of the outputs of

the metabolic pathway (e.g., production of primary or secondary metabolites, growth, selectivity, etc.).

As discussed before, enzymes that catalyze the same reaction in different organisms are not necessarily the same in their catalytic and regulatory properties. Moreover, these enzymes very often follow different mechanisms. Therefore, in considering the introduction of different regulatory properties in a metabolic pathway, the different mechanisms should be taken into account. Introduction in an organism of a heterologous enzyme, which might follow a different mechanism, can change the regulatory structure of the metabolic network. This difference can result in changes in the performance of the overall pathway.

A *regulatory superstructure* can be again considered for the nonlinear description of metabolic pathways. The dimensionality of the large combinatorial problem that results from this consideration increases significantly if we take into account the existence of alternative mechanisms for the regulation of the enzymes. The importance of such considerations is becoming clear if we consider changes in other metabolic parameters such as enzyme and external substrate levels, since it should be expected that different regulatory mechanisms will result in different responses to changes in these parameters.

7.2.2 Mathematical Description of Metabolic Reaction Networks

As has been presented in the previous chapters, metabolic networks can be mathematically modeled using (log)linear models, S-system models or nonlinear models. (Log)linear and S-system models have been studied in previous chapters. In this chapter we will additionally consider nonlinear models. If we consider a metabolic system consisting of n metabolites and m enzymatically-catalyzed reactions, the mass balances on the metabolites of the system may be written:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{v}(\mathbf{x};\mathbf{p}),\mathbf{x};\mathbf{p}) \tag{7.1}$$

where $\mathbf{x}$ is the n-dimensional metabolite concentration vector, $\mathbf{f}$ is a function determined by the mass balances, $\mathbf{v}$ is the m-dimensional reaction rate vector, and $\mathbf{p}$ is the s-dimensional manipulated parameter vector (e.g., enzyme concentrations). Terms that account for other processes by which concentrations of metabolites change (such as the dilution brought about by increases in the biomass volume (Fredrickson, 1976) and transport through the cell wall envelope) are also included in the mass balance.

The kinetic expressions for the reaction rates, $\mathbf{v}$, depend on the molecular mechanism that they follow and are different between different reaction steps and even for the same reactions in different organisms. In Appendix A.I, some characterestic rate expressions are presented. One can observe the variety of alternative kinetic expressions and their nonlinearity.

The optimization problem can also be nonlinear even if the available mathematical description of the metabolic system is based on the (log)linear model or the S-system model. Nonlinearity of the problem can arise in the objective function and/or the constraints. For example, in optimizing the selectivity of the phenylalanine, as considered as an example in the MILP case, the objective function concerning selectivity can be described as the nonlinear ratio of the (log)linear rate expressions instead of the linearized expression used in that previous case.

7.2.3 Analysis Problem

If we consider the existence of a regulatory structure around a metabolic reaction network, the kinetics of the enzymes depend on the molecular mechanism by which regulatory metabolites act on the enzymes. In Appendix A.II some of the common regulatory mechanisms are considered and the corresponding rate expressions are presented.

The *analysis problem* is defined here as before:

Which of the existent regulatory loops should be inactivated, and what associated changes should be made in the manipulated variables (e.g., enzyme expression levels, environmental conditions, effectors external to the system), in order to optimize the

performance of the metabolic network?

In the formulation of the MILP problem, modifications in the regulatory structure were represented by changes in the elements of the regulatory elasticity matrix, $\mathcal{E}^r$, from non-zero values to zero and vice-versa. For nonlinear kinetic models, a differerent consideration is introduced and is illustrated next.

Illustration

We will consider first the case of the unireactant reaction:

$$S \rightarrow P$$

that is competitively inhibited by the metabolite I. The corresponding rate expression (Appendix A.II) can be written as:

$$v = v_m \frac{[S]}{K_s\left(1 + \frac{[I]}{K_i}\right) + [S]} \tag{7.2}$$

According to the molecular mechanism when the values of the K_S and K_i increase the affinity of the enzyme for the substrate, S, and the inhibitor, I, respectively, decreases. Therefore, elimination of the inhibitory action of I on the enzyme can be modeled by assuming:

$$K_i \rightarrow \infty$$

or, equivalently:

$$K_i^{-1} \rightarrow 0$$

and the rate expression (7.2) will be written as:

$$v = v_m \frac{[S]}{K_s + [S]} \tag{7.3}$$

Similarly, eliminating of the action of activators can be modeled by assuming that the dissociation constants corresponding to activators take infinitely high values. For

example, in the rate expression for the nonessential activation:

$$v = v_m \left(\frac{1 + \frac{\beta[A]}{\alpha K_A}}{1 + \frac{[A]}{\alpha K_A}} \right) \frac{[S]}{K_S \left(\frac{1 + \frac{[A]}{K_A}}{1 + \frac{[A]}{\alpha K_A}} \right) + [S]}$$

by setting

$$K_A^{-1} = 0$$

the enzyme becomes insensitive to activator, and the corresponding rate expression is the same as in equation (7.3).

In general, any regulatory loop can be inactivated by considering that the corresponding dissociation constants assume such values that, at the molecular level, the regulator cannot bind to enzyme, and, in the rate expression, the dependency on the concentration of the regulator is removed.

7.2.4 Synthesis Problem

The *synthesis problem* and the concept of the *regulatory superstructure* that have been introduced in previous chapters can be also considered in the nonlinear problem. The *synthesis problem* is defined again as:

What kind of regulation (i.e., activation or inhibition, by which metabolite and of what strength) should be assigned to each enzyme in the network, and what associated changes should be made in the manipulated parameters (e.g., enzyme expression levels, environmental conditions, effectors external to the system), in order to optimize the performance of the metabolic network?

The modeling of the regulatory superstructure is more complicated for nonlinear models. As mentioned above, an additional characteristic that should be taken into account is the possibility of different mechanisms for the same regulatory loop. The modeling of the regulatory superstructure will be illustrated next.

Illustration

We will consider first that the simple unireactant reaction initially is not subject to any regulation. The corresponding reaction rate will be described by equation (7.3). In postulating a regulatory superstructure around this enzyme, we can consider the existence of two potential inhibitors, I_1 and I_2, and two potential activators, A_1 and A_2. Moreover, for each of the regulators we will consider two alternative regulatory mechanisms: competitive and noncompetitive inhibition for the inhibitors, and ordered and random essential activation for the activators. These lead to the following general rate expression:

$$v = v_m \frac{[S]\left(\frac{1}{1+\frac{[I_1]}{K_{i,1}^{NC}}+\frac{[I_2]}{K_{i,2}^{NC}}}\right)\left(\frac{1}{1+\frac{K_{A,1}^{R}}{[A_1]}+\frac{K_{A,2}^{R}}{[A_2]}}\right)}{K_S\left(1+\frac{[I_1]}{K_{i,1}^{C}}+\frac{[I_2]}{K_{i,2}^{C}}\right)\left(1+\frac{K_{A,1}^{O}}{[A_1]}+\frac{K_{A,2}^{O}}{[A_2]}\right)+[S]} \tag{7.4}$$

where the superscrpipts "NC" and "C" denote the dissociation constant for noncompetitive and competitive inhibition, respectively, and the subscripts "R" and "O" denote random and ordered essential activation, respectively.

The superstructure postulated for this enzyme considers $2^8 = 256$ alternative regulatory structures if a minimum of zero loops is allowable and a maximum of eight (i.e., all the regulatory loops active). However, a regulator will not act on the same enzyme with two different mechanisms. Therefore, $2^5 = 32$ alternative regulatory structures should be considered. Even in the case that only one or none loop is allowed, the 9 alternative regulatory structures is a significant number for one enzyme, four regulators, and two alternative mechanisms for each loop, since, in a metabolic network, the same considerations for each enzyme and the possibility of each metabolite in the network to regulate any of the enzymes, by any type of action and mechanism, leads to a huge number of alternative regulatory structures.

If we consider here the simplest case, that allows only one or no regulatory loop, we will have to consider accordingly values for the dissociation constants of the regulatory

mechanisms. In the case that none of the loops is active we have:

$$\frac{1}{K_{i,1}^{C}} = \frac{1}{K_{i,2}^{C}} = \frac{1}{K_{i,1}^{UC}} = \frac{1}{K_{i,2}^{UC}} = 0$$

and

$$K_{A,1}^{R} = K_{A,2}^{R} = K_{A,1}^{O} = K_{A,2}^{O} = 0$$

In the case that we consider that I_1 is the only regulator, then we have:

$$\frac{1}{K_{i,1}^{C}} \neq 0$$

and

$$\frac{1}{K_{i,2}^{C}} = \frac{1}{K_{i,1}^{UC}} = \frac{1}{K_{i,2}^{UC}} = 0$$

and

$$K_{A,1}^{R} = K_{A,2}^{R} = K_{A,1}^{O} = K_{A,2}^{O} = 0$$

Similarly, the rest of the regulatory structures can be modeled based on the assumptions about the structure and equation (7.4).

The regulatory loops can also be different with respect to the value of the dissociation constants corresponding to each loop. In general, protein engineering of the binding sites for the regulators will result in values for the dissociation constants that will be orders of magnitude different between each other. Therefore, the same regulatory action, by the same regulator, following the same mechanism, can be considered with two, or more, different values for the dissociation constants. These considerations can be modeled as alternative regulatory structures, and, thus, the size of the regulatory superstructure can increase significantly.

7.3 Mathematical Formulation

In order to set up the mathematical formulation for the MINLP problem, the following index sets and variables are introduced to characterize the regulatory superstructure

(in general similar to those used in the MILP problem). The metabolites will be denoted by the index set $I = \{i\}$, the reaction rates by the index set $J = \{j\}$, the manipulated parameters by the index set $K = \{k\}$, the metabolic outputs by the index set $L = \{l\}$, and the following sets will be defined to establish the connections of the sets of metabolites with the reaction rates in the network:

$I_s^j = \{i_s \mid i_s \in I \text{ is a substrate for reaction } j, \ \ j \in J\}$

$I_r^j = \{i_r \mid i_r \in I \text{ is a regulator for reaction } j, \ \ j \in J\}$

As before the sets $M^+ = \{m^+\}$ and $M^- = \{m^-\}$ denote the indices for the activating and the inhibitory action, respectively, that can be applied to each enzyme by each metabolite, where as the set $M^0 = \{m^0\}$ denotes the indices for the catalytic action of each enzyme on its substrates. With each index m that belongs in the index set $M = M^+ \cup M^- \cup M^0$, there is a regulatory or catalytic mechanism associated. If the mathematical description of the metabolic system is a (log)linear or an S-system model, these sets denote the indices for the activation and the inhibition elasticities, respectively, as presented previously. Two index sets that are also needed are a set for the number of parameters associated with each regulatory mechanism and a set for the number of parameters associated with the substrate binding and transformation catalyzed by each enzyme:

$$C_r^{mji_r} = \{c_r \mid c_r \text{ is parameter for the regulatory mechanism } m \text{ by which reaction } j \text{ is regulated by metabolite } i_r\}$$

and

$$C_s^{mji_s} = \{c_s \mid c_s \text{ is a parameter for the catalytic mechanism } m \text{ by which reaction } j \text{ is transforming metabolite } i_s\}$$

In the case in which we consider alternative values for the parameters of a certain regulatory loop and mechanism or of a catalytic mechanism, we could treat the parameters with different values as independent parameters and we should introduce

the expanded index set:

$$G_r^{mji_rc_r} = \{g_r \mid g_r \text{ is an alternative value for the parameter } c_r \text{ of the regulatory mechanism } m \text{ by which reaction } j \text{ is regulated by metabolite } i_r\}$$

and

$$G_s^{mji_sc_s} = \{g_s \mid g_s \text{ is an alternative value for the parameter } c_s \text{ of the catalytic mechanism } m \text{ by which reaction } j \text{ is transforming metabolite } i_s\}$$

The parameters associated with the regulatory mechanisms of the postulated regulatory superstructure will be denoted as $K^r_{mji_rc_rg_r}$, where m belongs to the index set $M^r = M^+ \cup M^-$. The parameters associated with the catalytic mechanism of the enzyme (substate binding and transformation) will be denoted as $K^s_{mji_sc_sg_s}$, where m belongs to the set M^0.

The continuous variables of the nonlinear model are the metabolite concentrations, x_i, the manipulated variables, p_k, the metabolic outouts, h_l, and the reaction rates, v_j.

A binary variable, $y_{mji_rc_rg_r}$, is associated with each parameter, $K^r_{mji_rc_rg_r}$, of the regulatory mechanisms. If a regulatory loop, for which metabolite i_r is regulating reaction j following mechanism m that is described by c_r parameters with values g_r, is active in the network, $y_{mji_rc_rg_r}$ is set to 1; otherwise it is zero. In order to control the number of manipulated variables that are allowed to vary, similar to the MILP problem, a binary variable, d_k, is associated with each manipulated variable, p_k.

7.3.1 Reaction Rates and Metabolic Outputs

The postulation of the regulatory superstructure and the introduction of the parameters and variables presented above should be incorporated in the mathematical

description of the reaction rates. If we consider a unireactant reaction j in a metabolic network we can postulate a regulatory superstructure:

Every metabolite in the metabolic network can be a competitive inhibitor or an random essential activator of the enzyme catalyzing reaction j and the parameters for the catalytic and regulatory mechanism can have g_s and g_c, respectively, number of alternative values.

For this regulatory superstructure the rate expression for reaction j can be written in the following form:

$$v_j = \frac{(p_{k,o} + p_k d_k) x_{i_s} \left(\dfrac{1}{1 + \sum\limits_{i_r \in I_r^j} \sum\limits_{g_r \in G_r^{mji_r c_r}} y_{m^+ ji_r c_r g_r} \dfrac{K^r_{m^+ ji_r c_r g_r}}{x_{i_r}}} \right)}{\sum\limits_{g_s \in G_s^{mji_s c_s}} y_{m^0 ji_s c_s g_s} K^s_{m^0 ji_s c_s g_s} \left(1 + \sum\limits_{i_r \in I_r^j} \sum\limits_{g_r \in G_r^{mji_r c_r}} y_{m^- ji_r c_r g_r} \dfrac{x_{i_r}}{K^r_{m^- ji_r c_r g_r}} \right) + x_{i_s}} \tag{7.5}$$

The manipulated parameter, p_k, with a reference value, $p_{k,o}$, in this illustration corresponds to the v_m parameter and it is proportional to the amount of the enzyme that catalyzes reaction j. The binary variable, d_k, associated with it will determine if it can be changed from the reference value. The term in the parenthesis in the nominator corresponds to random essential activation, and the term in the parenthesis in the denominator corresponds to competitive inhibition. It should be mentioned that the regulatory superstructure, and the binary variables and the parameters associated with it, will appear only in the rate expressions.

The metabolic outputs can be written in a similar form since they will be, in general, functions of the metabolites of the system and of the reaction rates. Therefore, the mathematical description of the metabolic outputs as a function of the parameters and the variables associated with the regulatory structure does not require any special formulation.

The bilinear products of continuous and binary variables introduced in equation (7.5) require, as in the MILP problem, the introduction of new continuous variables

for each bilinear product:

$$u^y_{m^-ji_rc_rg_r} = y_{m^-ji_rc_rg_r}x_{i_r} \quad \forall(m^-, j, i_r, c_r, g_r) \tag{7.6}$$

$$u^y_{m^+ji_rc_rg_r} = \frac{y_{m^+ji_rc_rg_r}}{x_{i_r}} \quad \forall(m^+, j, i_r, c_r, g_r) \tag{7.7}$$

and

$$u^d_k = p_{k,o} + d_k p_k \quad \forall k \tag{7.8}$$

and the introduction of the corresponding constraints that will be described later. The kinetic expression (7.5) can now be written in terms of the new continuous variables:

$$v_j = \frac{u^d_{k,o}x_{i_s}\left(\dfrac{1}{1+\sum\limits_{i_r\in I^j_r}\sum\limits_{g_r\in G^{mji_rc_r}_r} u^y_{m^+ji_rc_rg_r}K^r_{m^+ji_rc_rg_r}}\right)}{\sum\limits_{g_s\in G^{mji_sc_s}_s} y_{m^0ji_sc_sg_s}K^s_{m^0ji_sc_sg_s}\left(1+\sum\limits_{i_r\in I^j_r}\sum\limits_{g_r\in G^{mji_rc_r}_r}\dfrac{u^y_{m^-ji_rc_rg_r}}{K^r_{m^-ji_rc_rg_r}}\right)+x_{i_s}} \tag{7.9}$$

However, there is one more bilinear product of continuous and binary variables appearing in the denominator of the above kinetic expression:

$$y_{m^0ji_sc_sg_s}u^y_{m^-ji_rc_rg_r} \tag{7.10}$$

We can define two more new continuous variables:

$$u'_{m^0ji_sc_sg_s} = \left(1+\sum_{i_r\in I^j_r}\sum_{g_r\in G^{mji_rc_r}_r}\frac{u^y_{m^-ji_rc_rg_r}}{K^r_{m^-ji_rc_rg_r}}\right) \tag{7.11}$$

and

$$u^y_{m^0ji_sc_sg_s} = y_{m^0ji_sc_sg_s}u'_{m^0ji_sc_sg_s} \tag{7.12}$$

and then the kinetic expression (7.5) can be written in terms of continuous variables

only:

$$v_j = \frac{u^d_{k,o} x_{i_s} \left(\dfrac{1}{1 + \sum\limits_{i_r \in I^j_r} \sum\limits_{g_r \in G^{mji_rc_r}_r} u^y_{m^+ji_rc_rg_r} K^r_{m^+ji_rc_rg_r}} \right)}{\sum\limits_{g_s \in G^{mji_sc_s}_s} K^s_{m^0ji_sc_sg_s} u^y_{m^0ji_sc_sg_s} + x_{i_s}} \tag{7.13}$$

In general, when regulation acts by modification of the $K^s_{m^0ji_sc_sg_s}$ parameters, bilinear products of the form (7.10) will appear in the kinetic expressions, and introduction of the continuous variables $u'_{m^0ji_sc_sg_s}$ and $u^y_{m^0ji_sc_sg_s}$ is required.

7.3.2 The Objective Function

The objective function can be formulated based on the process to be optimized. The objective function can be any metabolic output or combination of the reaction rates and the metabolite concentrations. Since the mathematical description of the system is nonlinear, the objective function will, in general, be a nonlinear function.

However, as discussed before, when the available mathematical description is a (log)linear or an S-system model, the objective function can be any nonlinear function of the reaction rates and the metabolites, as estimated using either models. The mathematical formulation will be the same as for the MILP problem. However, the algorithmic procedure that will be followed will be the same as for the MINLP problem. An example of this type of study is the adjustment of the control coefficients, as discussed earlier.

7.3.3 Constraints

The metabolic optimization problem is also subject to the same type of constraints as in the MILP case:

(i) Mass balance for each metabolite i

At steady state, the left-hand side of the mass balance equation (7.1) is set to

zero, and, for each metabolite x_i, we have a nonlinear constraint of the form:

$$\mathbf{f}(\mathbf{v}(\mathbf{x}, \mathbf{u}^y, \mathbf{u}^d; \mathbf{K}^r, \mathbf{K}^s), \mathbf{x}, \mathbf{u}^y, \mathbf{u}^d; \mathbf{K}^r, \mathbf{K}^s) = 0 \tag{7.14}$$

(ii) Bounds on metabolites, manipulated variables, rates, metabolic outputs, and continuous variables

The significance of the postulation of bounds on metabolites, manipulated variables, rates, and metabolic outputs, has been discussed earlier. The four types of variables of the MINLP problem should again be constrained within bounds determined by their physiological ranges for the pathway of interest and by the available biological knowledge. Therefore, we will have bounds for the metabolite concentrations:

$$x_i^L \leq x_i \leq x_i^U \quad i \in I \tag{7.15}$$

for the manipulated variables:

$$p_k^L \leq p_k \leq p_k^U \quad k \in K \tag{7.16}$$

for the reaction rates:

$$v_j^L \leq v_j \leq v_j^U \quad j \in J \tag{7.17}$$

and the for metabolic outputs:

$$h_l^L \leq h_l \leq h_l^U \quad l \in L \tag{7.18}$$

In general, manipulated variables, reaction rates, and metabolic outputs can take negative values. However, the metabolite concentrations should always be constrained to be non-negative.

The new continuous variables that were introduced for the bilinear products of continuous and binary variables should also be bounded. The proper definition of bounds for every continuous variable is crucial for the performance of the algorithmic procedure that will be used to solve the MINLP problem. Examination of equations

(7.6), (7.7), (7.8), (7.11), and (7.12), that define the continuous variables allows the definition of the bounds for these variables:

$$0 \leq u^{y}_{m^- j i_r c_r g_r} \leq x^{U}_{i_r} \tag{7.19}$$

$$m^- \in M^-; j \in J,\ i_r \in I^j_r,\ c_r \in C^{m^- jir}_r,\ g_r \in G^{m^- ji_r c_r}_r$$

$$0 \leq u^{y}_{m^+ j i_r c_r g_r} \leq \frac{1}{x^{L}_{i_r}} \tag{7.20}$$

$$m^+ \in M^+; j \in J,\ i_r \in I^j_r,\ c_r \in C^{m^- jir}_r,\ g_r \in G^{m^- ji_r c_r}_r$$

$$p_{k,o} + p^L_k \leq u^d_k \leq p_{k,o} + p^U_k \tag{7.21}$$

$$k \in K$$

$$1 \leq u'_{m^0 j i_s c_s g_s} \leq \left(1 + \sum_{i_r \in I^j_r} \sum_{g_r \in G^{mjir c_r}_r} \frac{x^U_{i_r}}{K^r_{m^- j i_r c_r g_r}}\right) \tag{7.22}$$

$$m^0 \in M^0; m^- \in M^-; j \in J,\ i_r \in I^j_r,\ c_r \in C^{m^- jir}_r,\ g_r \in G^{m^- ji_r c_r}_r$$

$$0 \leq u^{y}_{m^0 j i_s c_s g_s} \leq \left(1 + \sum_{i_r \in I^j_r} \sum_{g_r \in G^{mjir c_r}_r} \frac{x^U_{i_r}}{K^r_{m^- j i_r c_r g_r}}\right) \tag{7.23}$$

$$m^0 \in M^0; m^- \in M^-; j \in J,\ i_r \in I^j_r,\ c_r \in C^{m^- jir}_r,\ g_r \in G^{m^- ji_r c_r}_r$$

The last two equations for the bounds, (7.22) and (7.23), are specific to the illustration presented above (equations (7.5)) in which a competitive inhibition was considered. In general, different mechanisms might require introduction of different continuous variables and, consequently, different bounds.

(iii) Constraints for the **u** *variables*

The variables that were introduced for the bilinear products of continuous and binary variables are connected with the continuous and binary variables via the following constraints:

a. Variables $u^{y}_{m^- j i_r c_r g_r}$

$$x_{i_r} - x^U_{i_r}(1 - y_{m^- j i_r c_r g_r}) \leq u^{y}_{m^- j i_r c_r g_r} \leq x_{i_r} - x^L_{i_r}(1 - y_{m^- j i_r c_r g_r}) \tag{7.24}$$

$$m^- \in M^-; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^- j i r},\ g_r \in G_r^{m^- j i_r c_r}$$

$$x_{i_r}^L y_{m^- j i_r c_r g_r} \leq u_{m^- j i_r c_r g_r}^y \leq x_{i_r}^U y_{m^- j i_r c_r g_r} \tag{7.25}$$

$$m^- \in M^-; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^- j i r},\ g_r \in G_r^{m^- j i_r c_r}$$

b. Variables $u_{m^+ j i_r c_r g_r}^y$

$$\frac{1}{x_{i_r}} - \frac{1}{x_{i_r}^L}(1 - y_{m^+ j i_r c_r g_r}) \leq u_{m^- j i_r c_r g_r}^y \leq \frac{1}{x_{i_r}} - \frac{1}{x_{i_r}^U}(1 - y_{m^+ j i_r c_r g_r}) \tag{7.26}$$

$$m^+ \in M^+; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^+ j i r},\ g_r \in G_r^{m^+ j i_r c_r}$$

$$\frac{1}{x_{i_r}^U} y_{m^+ j i_r c_r g_r} \leq u_{m^+ j i_r c_r g_r}^y \leq \frac{1}{x_{i_r}^L} y_{m^+ j i_r c_r g_r} \tag{7.27}$$

$$m^+ \in M^+; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^+ j i r},\ g_r \in G_r^{m^+ j i_r c_r}$$

c. Variables u_k^d

$$p_k - p_k^U + (p_k^U + p_{k,o}) d_k \leq u_k^d \leq p_k - p_k^L + (p_k^L + p_{k,o}) d_k \tag{7.28}$$

$$k \in K$$

$$d_k (p_{k,o} + p_k^L) \leq u_k^d \leq d_k (p_{k,o} + p_k^U) \tag{7.29}$$

$$k \in K$$

d. Variables $u_{m^0 j i_s c_s g_s}^y$

As mentioned before, these variables are specific to the illustration presented in which competitive inhibition was considered.

$$u'_{m^0 j i_s c_s g_s} - \left(1 + \sum_{i_r \in I_r^j} \sum_{g_r \in G_r^{m j i_r c_r}} \frac{x_{i_r}^U}{K_{m^- j i_r c_r g_r}^r}\right)(1 - y_{m^0 j i_s c_s g_s}) \leq u_{m^0 j i_r c_r g_r}^y \leq u'_{m^0 j i_s c_s g_s} \tag{7.30}$$

$$m^0 \in M^0; m^- \in M^-; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^- j i r},\ g_r \in G_r^{m^- j i_r c_r}$$

$$0 \le u^y_{m^- j i_r c_r g_r} \le \left(1 + \sum_{i_r \in I_r^j} \sum_{g_r \in G_r^{m j i_r c_r}} \frac{x^U_{i_r}}{K^r_{m^- j i_r c_r g_r}}\right) y_{m^0 j i_s c_s g_s} \tag{7.31}$$

$$m^- \in M^-; j \in J,\ i_r \in I_r^j,\ c_r \in C_r^{m^- j i r},\ g_r \in G_r^{m^- j i_r c_r}$$

The bounds for the variables $u^y_{m^0 j i_s c_s g_s}$ (equation (7.23)) were used in the above constraints.

(iv) Logical constraints

Introduction of constraints on the binary variables are needed in order to model the existence or nonexistence of various regulatory loops and the activation or deactivation of different continuously adjustable manipulated variables.

An important logical constraint that almost always should be included is one that does not allow activation and inhibition of an enzyme by the same metabolite. Moreover, for each parameter of a postulated regulatory mechanism, only one value should be considered. In general, each metabolite, when it acts as regulator on an enzyme, should be considered to do so by only one mechanism. Therefore, the number of regulatory mechanisms for each regulator should also be constrained to one. These constraints appear in the formulation for each (j, i_r):

$$\sum_{m^+ \in M^+} \sum_{c_r \in C_r^{m^+ j i_r c_r}} \sum_{g_r \in G_r^{m^+ j i_r c_r}} y_{m^+ j i_r c_r g_r} + \sum_{m^- \in M^-} \sum_{c_r \in C_r^{m^- j i_r c_r}} \sum_{g_r \in G_r^{m^- j i_r c_r}} y_{m^- j i_r c_r g_r} \le 1$$
$$\forall (j, i_r, c_s) \tag{7.32}$$

$$m^+ \in M^+; m^- \in M^-; j \in J,\ i_r \in I_r^j,$$
$$c_r \in \{C_r^{m^- j i r} \cup C_r^{m^+ j i r}\},\ g_r \in \{G_r^{m^- j i_r c_r} \cup G_r^{m^+ j i_r c_r}\}$$

The parameters associated with the catalytic mechanism can take different values but only one value for each parameter must be allowed. The corresponding constraint can be formulated for each (m^0, j, i_s, c_s):

$$\sum_{g_s \in G_s^{m j i_s c_s}} y_{m^0 j i_s c_s g_s} = 1 \quad (m^0, j, i_s, c_s) \tag{7.33}$$

$$m^0 \in M^0; j \in J,\ i_s \in I_s^j,\ c_s \in C_s^{m^0 jis}$$

Two additional constraints will constrain the maximum number of regulatory actions for each enzyme, $\mid I_{r,max}^j \mid$, and the maximum number of enzymes that a metabolite regulates, $\mid J_{max}^{i_r} \mid$, and can be formulated, respectively, as:

$$\sum_{m \in \{M^+ \cup M^-\}} \sum_{i_r \in I_{i_r}^j} y_{mji_r c_r g_r} \leq \mid I_{r,max}^j \mid \tag{7.34}$$

$$j \in J,\ c_s \in \{C_r^{m^- jir} \cup C_r^{m^+ jir}\},\ g_r \in \{G_r^{m^- ji_r c_r} \cup G_r^{m^+ ji_r c_r}\}$$

and as:

$$\sum_{m \in \{M^+ \cup M^-\}} \sum_{j \in J} y_{mji_r c_r g_r} \leq \mid J_{max}^{i_r} \mid \tag{7.35}$$

$$i_r \in I_r^j,\ c_s \in \{C_r^{m^- jir} \cup C_r^{m^+ jir}\},\ g_r \in \{G_r^{m^- ji_r c_r} \cup G_r^{m^+ ji_r c_r}\}$$

Similar to the MILP problem, the simultaneous manipulation of the variables q_k will be subject to the following constraint:

$$\sum_{k \in K} d_k \leq \mid K_{max} \mid \tag{7.36}$$

where $\mid K_{max} \mid$ is the maximum number of the manipulated variables that can be modified simultaneously. This constraint arises from practical and physiological limitations, and $\mid K_{max} \mid$ varies from system to system.

Finally, integer cuts should be also introduced here so that we can exclude all the previous optimal solutions found for the system. Because of the nonlinearity of the model, the previously found solutions are not necessarily better than the one that could be identified next. Therefore, these integer cuts can help appreciably in finding a better solution. In this case, when we solve for the n-th best solution we have to include $n-1$ constraints of the form:

$$\sum_{i \in B_i} y_i - \sum_{i \in NB_i} y_i \leq \mid B_i \mid -1 \tag{7.37}$$

$$B_i = \{i \mid y_i = 1\}$$

$$NB_i = \{i \mid y_i = 0\}$$

where $\mid B_i \mid$ is the cardinality of the set B_i (i.e. the number of the elements in the set).

The above mathematical formulation of the problem of optimal regulatory structure for a metabolic network involves continuous and binary variables and features nonlinearities in the mathematical description of the kinetic expressions and the metabolic outputs. Therefore, it is an MINLP formulation, and the algorithmic procedure that can be followed for its solution will be presented next.

7.4 Algorithmic Procedure

Many problems in chemical process engineering involve integer or discrete variables in addition to the continuous variables. The mathematical models used to describe and study these problems are also nonlinear problems. These classes of optimization problems are Mixed-Integer Nonlinear Programming (MINLP) problems. The difficulties encountered in solving these problems are associated with the combinatorial domain of the discrete variables and the nonlinearity of the continuous domain.

As the number of integer variables increase, the problem becomes a large combinatorial problem and computational problems arise (Nemhauser and Wolsey, 1988). At the same time, the nonlinear problem is in general nonconvex, which implies the potential existence of multiple local solutions, while the determination of the global optimal solution is a computationally hard problem and there is no theoretical development that will identify a solution as global.

The numerous significant problems that can be studied within the MINLP framework has led to intensive research. Several algorithms have been proposed, their convergence properties have been investigated, and they have been successfully used in many applications. Some of the most commonly used, in chronological order of development, are the following:

1. Generalized Benders Decomposition, **GBD**, (Geoffrion, 1972; Paules and Floudas,

1989; Floudas *et al.*, 1989);

2. Branch and Bound, **BB**, (Beale, 1977; Gupta, 1980; Ostrovsky *et al.*, 1990; Borchers and Mitchell, 1991);

3. Outer Approximation, **OA**, (Duran and Grossmann, 1986);

4. Feasibility Approach, **FA**, (Mawenngkang and Murtagh, 1986);

5. Outer Approximation with Equality Relaxation, **OA/ER**, (Kocis and Grossmann, 1987);

6. Outer Approximation with Equality Relaxation and Augmented Penalty, **OA/ER/AP**, (Viswanathan and Grossmann, 1990);

7. Generalized Cross Decomposition, **GCD**, (Holmberg, 1990)

8. Generalized Outer Approximation, **GOA**, (Fletcher and Leyffer, 1994);

In our studies we will apply the Generalized Benders Decomposition, **GBD**, algorithm which will be presented in the following section.

7.4.1 Generalized Benders Decomposition, GBD

In this section the **GBD** algorithm will be presented. In order to simplify the notation, the variables, the functions, and the indices that will be used in this section will be independent from the previously used notation and will be defined in the text below.

The general MINLP formulation can be stated as:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x},\mathbf{y}) \\ s.t. \quad & \mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathbf{X} \subseteq \mathcal{R}^n \\ & \mathbf{y} \in \{0,1\} \end{aligned} \tag{7.38}$$

where $\mathbf{x}$ is a vector of n continuous variables (e.g., metabolite concentrations, reaction rates), and $\mathbf{y}$ is a vector of q binary (0-1) variables (e.g., existence of a regulatory loop ($y_i = 1$) or non-existence ($y_i = 0$); $\mathbf{h(x,y)=0}$ denote the m equality constraints (e.g., mass balance equations); $\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}$ are the p inequality constraints (e.g., bounds for the metabolite concentrations and the reaction rates, logical constraints); $f(\mathbf{x}, \mathbf{y})$ is the objective function (e.g., growth rate, production rate of a product).

The basic idea in **GBD** is the generation, at each iteration, of an upper bound and a lower bound on the sought solution of the MINLP model. The upper bound results from the *primal* problem, while the lower bound results from the *master* problem. The primal problem corresponds to problem (7.38) with fixed $\mathbf{y}$-variables (i.e. it is in the $\mathbf{x}$-space only), and its solution provides information about the upper bound and the Langrange multipliers associated with the equality and the inequality constraints. The master problem is derived via nonlinear duality theory, makes use of the Langange multipliers obtained in the primal problem, and its solution provides information about the lower bound, as well as the next set of fixed $\mathbf{y}$-variables to be used subsequently in the primal problem. As the iterations proceed, it is shown that the sequence of updated upper bounds is non-increasing, the sequence of the lower bounds is non-decreasing, and that the sequences converge in a finite number of iterations. The steps followed in the **GBD** algorithm will be presented next in detail.

The GBD Algorithm: Feasible NLP Subproblems

The **GBD** algorithm under the assumption of feasible NLP subproblems can be stated as follows:

Step 1. Select an initial assignment for the projected binary variables to be used in the first NLP subproblem, $\mathbf{y}^1$. Set the iteration counter $k = 1$. Initialize the upper bound $Z_U = +\infty$.

Step 2. Solve the $\mathbf{y}^k$ parameterized NLP subproblem $\mathbf{S_{GBD}}^k$

$$Z(\mathbf{y}^k) = \min_{\mathbf{x}} C(\mathbf{x}, \mathbf{y}^k)$$

subject to

$$\mathbf{h}(\mathbf{x}, \mathbf{y}^k) = \mathbf{0}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}^k) \leq \mathbf{0}$$

$$\mathbf{x} \in \mathbf{X}$$

The solution of problem $\mathbf{S_{GBD}}^k$ gives $Z(\mathbf{y}^k), \mathbf{x}^k, \eta^k$ and λ^k. Here, η^k are the Langange multipliers for the constraints of the form $\mathbf{g}(\mathbf{x}, \mathbf{y}^k)$, and λ^k are the Langrange multipliers for constraints of the form $\mathbf{h}(\mathbf{x}, \mathbf{y}^k)$. Update the current upper bound

$$Z_U = \min\{Z_U, Z_U(\mathbf{y}^k)\}$$

If $Z_U = Z_U(\mathbf{y}^k$, set $\mathbf{y}^* = \mathbf{y}^k$ and $\mathbf{x}^* = \mathbf{x}^k)$.

Step 3. Formulate the pseudointeger master problem $\mathbf{M_{GBD}}^k$

$$Z_M^k = \min_{\mathbf{y}, \mu_{BD}} \mu_{BD}$$

subject to

$$\mu_{BD} \geq L(\mathbf{x}^k, \mathbf{y}, \eta^k, \lambda^k) \quad k = 1, \ldots, k$$

$$\mathcal{E}(\mathbf{y})$$

$$\mathbf{y} \in \mathbf{Y}$$

$$\mu_{BD} \in \mathcal{R}^1$$

where

$$L(\mathbf{x}^k, \mathbf{y}, \eta^k, \lambda^k) = C(\mathbf{x}^k, \mathbf{y}) + (\eta^k)^T \cdot \mathbf{g}(\mathbf{x}^k, \mathbf{y}) + (\lambda^k)^T \cdot \mathbf{h}(\mathbf{x}^k, \mathbf{y})$$

and T denotes the transpose of the corresponding vectors of the Langrange multipliers.

Step 4. Solve the master problem $\mathbf{M_G BD}^k$. The solution of this problem gives Z_M^k and $\mathbf{y}^{k+1}$. If $Z_M^k = Z_U$ then stop; the solution is Z_U, $\mathbf{y}^*$, $\mathbf{x}^*$. Otherwise, set $k = k + 1$ and return to Step 2.

The GBD Algorithm: Infeasible NLP Subproblems

If the initial integer guess or a proposal returned from the master problem generates an infeasible NLP subproblem, the procedure is as follows. Replace the infeasible NLP problem with the problem:

$$Z(\mathbf{y}^k) = \min_{\alpha} \alpha$$

subject to

$$\mathbf{h}(\mathbf{x}, \mathbf{y}^k) = \mathbf{0}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}^k) \leq \mathbf{0}$$

$$\mathbf{x} \in \mathbf{X}$$

$$\alpha \in \mathcal{R}^1$$

where α is a scalar variable that will minimize the infeasibilities. Define a new Lagrangian function at the optimal solution to this problem

$$L'(\mathbf{x}^k, \mathbf{y}, \eta^k, \lambda^k) = C(\mathbf{x}^k, \mathbf{y}) + (\eta^k)^T \cdot \mathbf{g}(\mathbf{x}^k, \mathbf{y}) + (\lambda^k)^T \cdot \mathbf{h}(\mathbf{x}^k, \mathbf{y})$$

Proceed from Step 3 and include the constraint

$$L'(\mathbf{x}^k, \mathbf{y}, \eta^k, \lambda^k) \leq 0$$

The GBD Algorithm: The Integer Cuts

The form of the integer cuts included in the MINLP master problems to prevent a binary combination that has already been tried, from being proposed again is:

$$\sum_{j \in Q} y_j - \sum_{j \in NQ} y_j \leq \mid Q \mid -1$$

$$Q \in \{y_j = 1\} \quad NQ \in \{y_j = 0\}$$

where $| Q |$ is the cardinality of the set of activated binary variables Q.

These integer cuts are different than the ones introduced as logical constraints. The integer cuts here are included and updated after each iteration of the algorithm. The previously introduced integer cuts are included after the algorithm has converged to an optimal solution and updated for every optimal solution that has been found in order to prevent a binary combination to be found again as optimal solution.

7.4.2 Initialization Schemes

Nonlinear programming (NLP) problems may involve nonconvexities which imply difficulties with determining the global optimal solution. Hence the performance of conventional algorithms for NLP problems is highly dependent on the starting point provided for the algorithm, and they often fail to determine even a feasible solution. Recent advances in applied mathematics and computer sciences allowed the development of approaches for global optimization for problems that feature certain types of nonlinearities, such as polynomial functions (Pardalos and Rosen, 1986; Zilinskas, 1986; Torn and Zilinskas, 1987; Floudas *et al.*, 1989; Floudas and Visweswaran, 1990), and there is no approach for global optimization of MINLP problems.

Metabolic network optimization problems, due to the nonlinearity of the kinetic models for metabolic kinetics and of the objective functions, cannot be solved with any of the proposed global optimization approaches. The mixed-integer nonlinear nature of the problems complicates further the search for optimal solutions. Because of the combinatorial character of the problem, as shown for the MILP problem, existence of multiple global optima is also possible. However, clever initialization schemes for the MINLP algorithm can significantly improve the performance of the NLP solver and aid in determining a large number of local optima.

For the metabolic reaction networks, as shown before, there are two alternative modeling frameworks that approximate the nonlinear models very accurately: the (log)linear model and the S-system representation. Therefore, having a nonlinear model for a metabolic network, we could construct the corresponding (log)linear and

S-system models. These models could be studied using the MILP framework and the suggested solutions could be used as starting points for the MINLP algorithm.

Consider that, in a given reaction network, the enzyme levels are the manipulated variables. Every reaction rate is proportional to enzyme levels and depends on the set of metabolites, $\mathbf{x}$, and the binary variables for the regulatory structure, $\mathbf{y}^r$:

$$v_j = v_{m,j}(d_j)\phi(\mathbf{x}, \mathbf{y}^r) \tag{7.39}$$

where $v_{m,j}(y_j^d)$ is the parameter corresponding to the enzyme level, and d_j is the binary variable which determines if changes in that enzyme level are allowed ($d_j = 1$) or not ($d_j = 0$). In order to ensure that the initial point will be a feasible point the following procedure is proposed:

Step 1. Construct the approximate model around the reference steady state and solve the MILP problem. The solution will provide a set of optimal regulatory structures, $\mathbf{y}_L^r$, and the corresponding manipulated variables that are allowed to be changed, $\mathbf{d}_L$, where the subscript L indicates that the vector is a solution of the MILP problem. A set of continuous variables will also correspond to each optimal regulatory structure. These continuous variables will include the metabolite concentrations, $\mathbf{x}_L$, and the reaction rates, $\mathbf{v}_L$.

Step 2. From the first regulatory structure in the solution set and for every reaction, v_j, in the nonlinear model, use as initial values for metabolite concentrations, reaction rates, and binary variables the corresponding values from the MILP solution:

$$\mathbf{x}_{NL}^1 = \mathbf{x}_L, \mathbf{v}_{NL}^1 = \mathbf{v}_L, (\mathbf{y}_{NL}^r)^1 = \mathbf{y}_L^r$$

Calculate the initial values of the manipulated variables that have been suggested from the MILP solution to change from Equation (7.39):

$$v_{m,j} = \frac{v_L}{\phi(\mathbf{x}_L, \mathbf{y}_L^r)} \tag{7.40}$$

For every manipulated variable that has not been proposed to change in the solution(s)

to the MILP problem, keep it fixed. This will probably result in infeasibilities for the mass balances.

Step 3. Solve the NLP problem minimizing the infeasibilities of the mass balances, similar to the infeasible NLP subproblems in the GBD algorithm. If solution results in a feasible NLP; i.e., the mass balances are exactly satisfied, then solve the MINLP using the **GBD** algorithm presented above. If the solution results in an infeasible NLP; i.e., the infeasibilities have been minimized but the mass balances are not exactly satisfied, discard the suggested MILP solution and proceed to the next one.

The above initialization scheme can significantly enhance the search for local optima in the MINLP problem. The possibility of exploiting both the (log)linear and the S-system modeling approaches can, in general, provide different suggestions, increases the number of initial binary variable combinations and continuous variables.

7.5 Computational Study

The synthesis problem will be illustrated using again the aromatic amino acid biosynthetic pathway in bacteria. The pathway has eight regulatory loops, all of which are feedback inhibitory loops. The nonlinear model is presented in the Appendix A.III.

The question addressed was the same as in the first of the MILP problems:

Which of the existing loops should be inactivated and what should be the associated changes in the enzyme expression levels to maximize the phenylalanine selectivity ?

The upper bounds considered for the concentrations of the intermediate metabolite DAHP, CHR, and PHP were 1000 mM, whereas the concentrations of the amino acids PHE, TYR, and TRP, were left unbounded. The specific growth rate, μ, was constrained within $\pm 10\%$ its reference value. This constraint allows solutions that will not have a strong effect on the rest of the metabolic activities. The levels of the six enzymes in the pathway were constrained between their reference value and twice the level of the reference value since great overexpression is practically feasible. The number of the simultaneously overexpressed enzymes was constrained to four.

The MINLP optimization problem was solved by using the GBD algorithm and

initial points proposed by the MILP problem. The regulatory structure of the best solution is presented in Figure 1 (case a) and the corresponding value of the phenylalanine selectivity is

$$S_{phe,opt_1} = 0.903621$$

which is 111% higher than the value for the reference state

$$S_{phe,o} = 0.427335$$

This solution features five regulatory loops and three overexpressed enzymes. The metabolite concentration levels and the overexpressed enzyme levels associated with the solution are presented in Table I. The specific growth rate was decreased by 8.94%, and none of the constrained intermediates reached the allowable upper bound.

A second solution with significantly improved selectivity

$$S_{phe,opt_2} = 0.897813$$

was also found. This value for the selectivity is 110% higher than the value for the reference state. The solution, presented in Figure 1 (case b) features four regulatory loops and four overexpressed enzymes. The metabolite concentration levels and the overexpressed enzyme levels associated with the solution are presented in Table I, and the corresponding specific growth rate is again decreased by 8.94%. Although the two solutions are very similar with respect to the phenylalanine selectivity and the effect on the specific growth rate, the first solution is more attractive since it suggests the minimum number of the regulatory loops that should be inactivated and the minimum number of the enzymes that should be overexpressed.

More solutions with similarly high improvements of the objective function were also found when the constraints for the specific growth rate and the concentration of DAHP were relaxed. The solutions found with a significant decrease of the specific growth rate contained fewer regulatory loops suggesting that the multiple regulatory loops present in the original system minimize the effects of the perturbations in the

pathway on the rest of the cellular metabolism.

The solutions found with significantly high DAHP concentration suggested inactivation of the feedback inhibitory loop from DAHP on the first reaction that is responsible for its formation. This pattern is a strong indication that the functionality of this loop is the regulation of the concentration levels of DAHP.

7.6 Linear vs. Nonlinear Framework

The solutions found using the nonlinear framework are different from the solutions proposed from the linear framework (Problem 1 in the MILP chapter). However, the regulatory structures in cases c and f from the linear framework (Problem 1) have many common characteristics with the MINLP solution and the enzymes that catalyze reaction steps 3 and 4 should be overexpressed, according to both frameworks. Moreover, the relative overexpression levels of the two enzymes are the same for the two frameworks; i.e., the enzyme that catalyze reaction step 4 should be overexpressed to the maximum allowable level whereas the enzyme that catalyzes reaction step 3 should be overexpressed only up to 20%, according to the nonlinear model, and up to 40%, according to the linear model.

The MINLP problem is computationally more complex than the MILP one. However, the solutions from the nonlinear model are more reliable since the description of the system is based on the molecular mechanisms of enzyme catalysis, and, although it is still an approximation of the physical system, it is a more complete, more globally accurate one. The coupling of the frameworks by using the linear framework for obtaining initial points for the nonlinear framework improves significantly the performance of the MINLP algorithmic procedure.

7.7 References

Floudas, C.A., "Nonlinear and Mixed Integer Optimization: Fundamentals and Applications," Oxford University Press, New York (1995).

Floudas, C. A., A. Aggarwal, and A. R. Ciric, "Global Optimum Search for Nonconvex NLP and MINLP Problems." *Computers chem. Engng* **13**, 1117 (1989).

Floudas, C. A. and V. Visweswaran, "A Global Optimization Algorithm (GOP) for Certain Classes of Nonconvex NLPs – I. Theory." *Computers chem. Engng* **14**, 1397 (1990).

Pardalos, P. and J. B. Rosen, Constrained Global Optimization: Algorithms and Applications. *SIAM Rev.* **28**, 367 (1986).

Savageau, M. A., "Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology," Reading, MA, Addison–Wesley (1976).

Schlosser, P. M., and J. E. Bailey, "An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways," *Math. Biosci.*, **100**, 87-114 (1990).

Segel, I. H., "Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems," John Wiley & Sons, New York (1975).

Torn, A. and A. Zilinskas, "Global Optimization." *Lecture Notes in Computer Science*, **350**. Springer-Verlag, New York (1987).

Zilinskas A., *Global Optimiation–Axiomatics of Statistical Models, Algorithms and Their Applications*, Moklas, Vilnius (1986)

APPENDIX A.I

The number of the reactants in a metabolic reaction varies from one to three, though most of the them are bireactant; i.e., the reaction involves two metabolites as substrates (reactants). Rate expressions are given here for some of the most common cases (Segel, 1975):

1. Unireactant Enzymes

   These enzymes catalyze the reaction:

$$S \rightarrow P$$

   The three most common rate expressions used, depending on the molecular mechanism, are:

   (a) Michaelis-Menten kinetics:

$$v = v_m \frac{[S]}{K_m + [S]}$$

   (b) Henri-Michaelis-Menten kinetics:

$$v = \frac{v_{m_f} \frac{[S]}{K_{m_S}} - v_{m_r} \frac{[P]}{K_{m_P}}}{1 + \frac{[S]}{K_{m_S}} + \frac{[P]}{K_{m_P}}}$$

   (c) Hill kinetics

$$v = v_m \frac{[S]^n}{K_m^n + [S]^n}$$

2. Bireactant Enzymes

   These enzymes catalyze the reaction:

$$A + B \rightarrow P + Q$$

The three most common rate expressions used, depending on the molecular mechanism, are:

(a) Random Bireactant System:

$$v = v_m \frac{[A][B]}{\alpha K_A K_B + \alpha K_A[B] + \alpha K_B[A] + [A][B]}$$

(b) Ordered Bireactant System:

$$v = v_m \frac{[A][B]}{K_A K_B + K_B[A] + [A][B]}$$

(c) Ordered Bi Bi System (Theorell-Chance):

$$v = \frac{v_f v_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{\Pi}$$

where

$$\begin{aligned} \Pi &= v_r K_{ia} K_{m_B} + v_r K_{m_B}[A] + v_r K_{m_A}[B] + \frac{v_f K_{m_Q}[P]}{K_{eq}} + \frac{v_f K_{m_P}[Q]}{K_{eq}} \\ &\quad + v_r[A][B] + \frac{v_f K_{m_Q}[A][P]}{K_{eq} K_{ia}} + \frac{v_r K_{m_A}[B][Q]}{K_{iq}} + \frac{v_f[P][Q]}{K_{eq}} \end{aligned}$$

3. Terreactant Enzymes These enzymes catalyze the reaction:

$$A + B + C \rightarrow P + Q + R$$

The most common rate expression used is of the form:

$$v = v_m \frac{[A][B][C]}{K_{m_A}[B][C] + K_{m_B}[C][A] + K_{m_C}[A][B] + [A][B][C]}$$

APPENDIX A.II

The kinetic expressions of enzymes that are subject to regulation depend on the molecular mechanism of the reaction and of the action of the regulator. The regulators are characterized, depending on the effect they have on the reaction rate, as *inhibitors*, if their concentration increase decreases the reaction rate, or as *activators*, if their concentration increase increases the reaction rate. The rate expression of some of the most common regulatory systems will be presented here (Segel, 1975):

1. Unireactant enzymes

   As mentioned before, these enzymes catalyze the reaction

$$S \rightarrow P$$

   and the most common types of regulation are:

   (a) Competitive Inhibition:

$$v = v_m \frac{[S]}{K_s\left(1 + \frac{[I]}{K_i}\right) + [S]}$$

   where I is the inhibitor.

   (b) Noncompetitive Inhibition:

$$v = v_m \frac{[S]}{K_s\left(1 + \frac{[I]}{K_i}\right) + [S]\left(1 + \frac{[I]}{K_i}\right)}$$

   where I is the inhibitor.

   (c) Uncompetitive Inhibition:

$$v = v_m \left(\frac{K_i}{K_i + [I]}\right) \frac{[S]}{K_s\left(\frac{K_i}{K_i + [I]}\right) + [S]}$$

   where I is the inhibitor.

(d) Nonessential Activation:

$$v = v_m \left(\frac{1 + \frac{\beta[A]}{\alpha K_A}}{1 + \frac{[A]}{\alpha K_A}} \right) \frac{[S]}{K_S \left(\frac{1 + \frac{[A]}{K_A}}{1 + \frac{[A]}{\alpha K_A}} \right) + [S]}$$

where A is the activator.

(e) Nonessential Activation in the presence of an inhibitor:

$$v = v_m \frac{[S]}{K_s \frac{\left(1 + \frac{[I]}{K_i}\right)}{\left[1 + \frac{[I]}{K_i\left(1 + \frac{K_A}{[A]}\right)}\right]} + [S] \frac{\left(1 + \frac{[I]}{K_i}\right)}{\left[1 + \frac{[I]}{K_i\left(1 + \frac{K_A}{[A]}\right)}\right]}}$$

where the binding of the activator, A, reverses the inhibitory action of a pure noncompetitive inhibitor, I.

(f) Essential Activation:

In essential activation, the activator site must be filled before any catalytic activity is possible. Depending on the order of binding the expressions for the kinetic rates have any of the following forms:

i. Ordered: a molecule of the activator A binds randomly before the substrate S:

$$v = v_m \frac{[S]}{K_S \left(1 + \frac{K_A}{[A]}\right) + [S]}$$

ii. Ordered: Substrate S binds first, then a molecule of the activator A binds randomly:

$$v = v_m \frac{[A]}{K_A} \frac{[S]}{(K_s + [S])\frac{K_A}{[A]} + [S]}$$

iii. Random binding of the substrate S and the activator A:

$$v = v_m \left(\frac{[A]}{K_A + [A]}\right)\left(\frac{[S]}{K_S + [S]}\right)$$

(g) Competitive Inhibition by two different nonexclusive inhibitors:

$$v = v_m \frac{[S]}{K_s\left(1 + \frac{[I]}{K_i} + \frac{[I][X]}{K_i K_X}\right) + [S]}$$

where I and X are the two competitive inhibitors.

(h) Inhibition by one competitive and one noncompetitive inhibitor:

$$v = v_m \frac{[S]}{K_s\left(1 + \frac{[I]}{K_i} + \frac{[X]}{K_X}\right) + [S]\left(1 + \frac{[X]}{\alpha K_X}\right)}$$

where I is the competive, and X is the noncompetitive, with respect to S, inhibitor.

In an analogous way similar rate expressions can be derived for any rate expression. There are two main ways a regulator can affect a reaction rate: either by modification of the dissociation constants, i.e. K_S's, or by modification of the catalytic activities, i.e. v_m's.

7.8 Appendices

APPENDIX A.III

The rate expressions for the aromatic amino acid pathway are taken from Schlosser and Bailey (1990). The parameter values used for the dissociation constants are the same as in Schlosser and Bailey (1990) where the references for the estimation of these parameters can be found. The values for the $v_{j,max}$ have been adjusted to give steady-state values similar to those found in bacterial cells for $[G6P] = 0.8\ mM$, $[PEP] = 0.1\ mM$, $[ATP] = 2.5\ mM$, $[ADP] = 0.4271\ mM$, and $[AMP] = 0.0729\ mM$. The rate expressions for the 15 reactions in the pathway are:

$$v_1 = v_{m,1}\frac{[G6P][PEP]\left(\frac{60}{60+[PHE]}\right)\left(\frac{406}{406+[TYR]}\right)\left(\frac{1512}{1512+[TRP]}\right)}{(0.1+[G6P])\left[0.006\left(1+\frac{[DAHP]}{0.02}\right)+[PEP]\right]}$$

$$v_2 = v_{m,2}\frac{[DAHP][PEP][ATP]}{(2+[DAHP])\,(0.00867+[PEP])\,(0.9281+[ATP])}$$

$$v_3 = v_{m,3}\frac{[CHR]}{(2+[CHR])\,(1+[PHE]/50)\,(1+[TYR]/40)}$$

$$v_4 = v_{m,4}\frac{[PHP]}{(1+[PHP])\,(1+[PHE]/50)}$$

$$v_5 = v_{m,5}\frac{[PHP]}{(1+[PHP])}$$

$$v_6 = v_{m,6}\frac{[G6P][CHR][ATP]}{(1.269+[G6P])(2+[CHP])(0.9821+[ATP])(1+[TRP]/16)}$$

$$v_7 = 54\mu, \quad v_8 = 131\mu, \quad v_9 = 176\mu$$

$$v_{10} = \mu[DAHP], \quad v_{11} = \mu[CHR], \quad v_{12} = \mu[PHP]$$

$v_{13} = \mu[PHE]$, $v_{14} = \mu[TYR]$, $v_{15} = \mu[TRP]$, where:

$$\mathbf{V}_m^T = [710,\ 22,\ 474,\ 64,\ 10.5,\ 28]$$

The growth function, μ, used is:

$$\mu = 0.014 \frac{Y[PHE][TYR][TRP][PEP]}{(0.25+Y)(18+[PHE])(13+[TRY])(5+[TRP])(0.005923+[PEP])}$$

where

$$Y = \frac{[ATP]+0.5[ADP]}{[ATP]+[ADP]+[AMP]}$$

The mass balance equations for each of the metabolites in the aromatic amino acid pathway is given by

$$\begin{aligned}
DAHP: \quad 0 &= v_1 - v_2 - v_{10} \\
CHR: \quad 0 &= v_2 - v_3 - v_6 - v_{11} \\
PHP: \quad 0 &= v_3 - v_4 - v_5 - v_{12} \\
PHE: \quad 0 &= v_4 - v_9 - v_{13} \\
TYR: \quad 0 &= v_5 - v_8 - v_{14} \\
TRP: \quad 0 &= v_6 - v_7 - v_{15}
\end{aligned}$$

For the stoichiometric matrix we have

$$N = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}$$

The stable steady-state at which the linear model was constructed is:

$$\mathbf{x}_o^T = [3.311539,\ 24.06868,\ 0.6025115,\ 262.6012,\ 318.8831,\ 81.84384]$$

and the values for the rates at these values are:

$$\mathbf{v}_o^T = [\ \begin{matrix} 9.340355 & 9.310362 & 7.843243 & 3.862561 & 3.975329 \\ 1.200839 & 0.474409 & 1.150880 & 1.546221 & 0.029994 \\ 0.266281 & 0.005353 & 2.316340 & 2.824449 & 0.726430] \end{matrix}$$

	case a		case b	
i	x_i	$V_{m,i}$	x_i	$V_{m,i}$
1	853.316	710.0	13.3259	1276.16
2	0.590432	44.0	0.200516	43.2071
3	0.262508	574.608	0.214807	948.0
4	3150.81	128.0	2653.19	128.0
5	141.903	10.5	101.082	10.5
6	27.9298	28.0	35.9307	28.0

Table 7.1: The metabolite concentration levels and the enzyme levels for the two optimal solutions of the MINLP problem. The subscript i denotes the corresponding reaction steps for the V_m variables, and for the metabolite concentrations: x_1 =[DAHP], x_2 =[CHR], x_3 =[PHP], x_4 =[PHE], x_5 =[TYR], x_6 =[TRP].

7.9 Figures

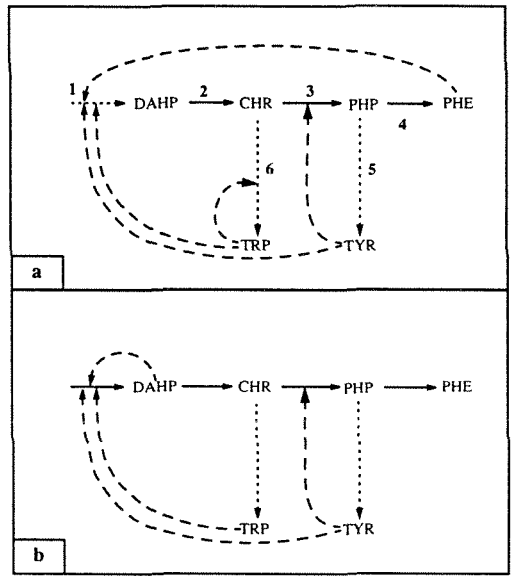

Figure 7.1: The two optimal solutions of the MINLP problem. Solid arrows indicate enzyme overexpression for the respective reaction, dotted arrows indicate reactions with enzyme levels at the reference state, and dashed arrows indicate inhibitory loops. In solution b the reaction numbering has been omitted for clarity.

Chapter 8 Epilogue

8.1 Future Prospects

Metabolism is an extremely complicated, multicomponent system. In addition to numerous different chemical reactions and intermolecular interactions, many coupled through common reactants, metabolism is complicated by an overlay of multiple control systems operating at both genetic and protein levels. There is an intercalated hierarchy of increasingly sophisticated mathematical tools and of increasingly detailed experimental bases for analyzing and manipulating metabolism. Three new developments in mathematical methods and applications were presented in this thesis: a framework for estimating intracellular metabolic reaction rates, a (log)linear kinetic model for approximating responses of metabolism to changes in their parameters, and an optimization framework for the analysis and design of regulatory structures in metabolic reaction networks. As with every novel method, these developments presented introduce numerous new topics for further investigation.

Experimental data used in metabolic flux analysis are subject to errors that propagate in the estimation of the reaction rates. Moreover, advanced experimental techniques using isotopically labelled substrates (such as ^{13}C-labelled substrates with products analyzed by NMR) are available for reducing the number of the unknown intracellular reaction rates. Further development of the metabolic flux analysis framework should include an error analysis and should consider the formulation of constraints for the additional experimental information. The construction of growth and product formation models for bioprocess control and optimization based on the results of this type of analysis should also be investigated. These models could be used for determining process operations to achieve an optimal pattern of intracellular metabolic reaction rates.

Detailed and reliable nonlinear mathematical models of single cells have been developed for various organisms. However, because of their size and complexity, they cannot be used for bioprocess control and optimization. The (log)linear modeling framework could be used in order to transform these models into simpler mathematical structures. Detailed analysis of the agreement between the nonlinear models and their

(log)linear representations can be performed in order to examine the performance of the (log)linear models. If the (log)linear representations are found to be good approximations of the nonlinear models, they would be excellent tools for studying the effects of genetic manipulations on bioprocess performance, since the information about molecular level regulatory interactions will be explicitly preserved in the values of the corresponding elasticities.

One of the most important problems in constructing any kind of kinetic models of metabolism is the uncertainty of the kinetic parameters. The experimental methods used to estimate their values is already subject to error. Moreover, the intracellular conditions will never be exactly known and will always be subject to fluctuations. Therefore, the optimization framework should be further developed in order to take into account these uncertainties for both (log)linear and nonlinear kinetic models. While the methods presented here provide with good qualitative guidance, consideration of the uncertainties could further limit the presently proposed approaches, since strategies that are optimal in the presence of uncertainties will be the primary candidates for experimental implementation.

One of the important issues that this thesis has not addressed is the stability of the metabolic steady states. Simulation studies of various metabolic networks has suggested that their steady-state stability characteristics is strongly influenced by the regulatory structures. Moreover, in metabolic networks with complex regulatory structures, I have observed that some of the regulatory loops regulate the dynamic characteristics of the network whereas some others have no effect on the dynamics of the network but regulate the level of various metabolic parameters, such as steady-state metabolite concentrations and reaction rates. The development of the (log)linear model allows the application of linear robust control methods for the analysis of the regulatory structures with respect to their effect on the stability of the metabolic network. Prior investigations have shown that tools such as μ-analysis can be successfully applied in metabolic systems.

Biology and biotechnology can only profit from integration of mathematical methods to advanced tools for experimental analysis. This is emphasized when one consid-

ers the methods presented in this thesis and the above suggestions for their development and applications. What is needed in order for mathematical approaches to the increasingly complex biotechnological problems to be more successful and instructive, is wise and imaginative perception and formulation of critical questions and problems, taking into account the highest level of mathematical and computational tools which exist (or can be developed).