

Quantitative Dynamic Modeling of Transcriptional Networks of Embryonic Stem Cells Using Integrated Framework of Pareto Optimality and Energy Balance

by Marco A. Avila

B.S. Chemical Engineering
Universidad de Concepcion, Concepcion, Chile, 2002

M.S. Chemical Engineering
Universidad de Concepcion, Concepcion, Chile, 2004

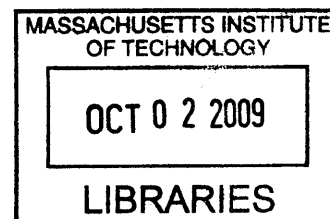
SUBMITTED TO THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MEDICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2009

© Marco A. Avila, 2009. All rights reserved.

ARCHIVES



The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature of Author: _____
Harvard-MIT Division of Health Sciences and Technology
June 22nd, 2009

Certified by: _____
Martin L. Yarmush, M.D., Ph.D.
Helen Andrus Benedict Professor of Surgery and Bioengineering
Harvard Medical School, Massachusetts General Hospital
Thesis Supervisor

Accepted by: _____
Ram Sasisekharan, Ph.D./Director
Harvard-MIT Division of Health Sciences and Technology
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering



Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.5668 Fax: 617.253.1690
Email: docs@mit.edu
<http://libraries.mit.edu/docs>

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Page 95 does not exist.
It appears to be a mistake
in numbering by the author.

ABSTRACT

Embryonic Stem Cells (ESCs) are pluripotent and thus are considered the “cell type of choice”. ESCs exhibit several phenotypic traits (e.g., proliferation, differentiation, apoptosis, necrosis, etc.) and when differentiated into a particular lineage they can perform an array of functions (e.g., protein secretion, detoxification, energy production). Typically, these cellular objectives compete against each other because of thermodynamic, stoichiometric and mass balance constraints. Analysis of transcriptional regulatory networks and metabolic networks in ESCs thus requires both a nonequilibrium thermodynamic and mass balance framework for designing and understanding complex ESC network approach as well as an optimality approach which can take cellular objectives into account simultaneously. The primary goal of this thesis was to develop an integrated energy and mass balance-based multiobjective framework for a transcriptional regulatory network model for ESCs. The secondary goal was to utilize the developed framework for large-scale metabolic flux profiling of hepatic and ESC metabolic networks.

Towards the first aim we first developed a complete dynamic pluripotent network model for ESCs which integrates several different master regulators of pluripotency such as transcription factors Oct4, Sox2, Nanog, Klf4, Nac1, Rex1, Dax1, cMyc, and Zfp281, and obtained the dynamic connectivity matrix between various pluripotency related gene promoters and transcription factors. The developed model fully describes the self-renewal state of embryonic stem cells. Next, we developed a transcriptional network model framework for ESCs that incorporates multiobjective optimality-based energy balance analysis. This framework predicts cofactor occupancy, network architecture and feedback memory of ESCs based on energetic cost.

The integrated nonequilibrium thermodynamics and multiobjective-optimality network analysis-based approach was further utilized to explain the significance of transcriptional motifs defined as small regulatory interaction patterns that regulate biological functions in highly-interacting cellular networks. Our results yield evidence that dissipative energetics is the underlying criteria used during evolution for motif selection and that biological systems during transcription tend towards evolutionary selection of subgraphs which produces minimum specific heat dissipation, thereby explaining the frequency of some motifs. Significantly, the proposed energetic hypothesis uncovers a mechanism for environmental selection of motifs, provides explanation for topological generalization of subgraphs into complex networks and enables identification of new functionalities for rarely occurring motifs.

Towards the secondary goal, we have developed a multiobjective optimization-based approach that couples the normalized constraint with both energy and flux balance-based metabolic flux analysis to explain certain features of metabolic control of hepatocytes, which is relevant to the response of hepatocytes and liver to various physiological stimuli and disease states. We also utilized this approach to obtain an optimal regimen for ESC differentiation into hepatocytes.

The presented framework may establish multiobjective optimality-based thermodynamic analysis as a backbone in designing and understanding complex network systems, such as transcriptional, metabolic and protein interaction networks.

ACKNOWLEDGMENTS

It has been almost five years since I started this adventure alone from my hometown city, Concepcion, in the middle of the winter in Chile. I had never left my country or my city for so long and I was extremely scared but also excited about these coming challenges. The aspirations of my family, friends and professors were placed in two bags together with my clothes, a sleeping bag, a mug and a spoon. Monica, my girlfriend at that time, traveled to Santiago to say goodbye. The uncertainty of our future kept our hearts crying. The journey started as well as the metamorphosis associated to it. And now, after five years, I look back and see all the changes that this trip involved: I married Monica after my first year in the U.S., I have a two-year old daughter named Emilia and we are expecting our second child, my father passed away in Chile and left my mother, his daughters and me with a hole in our lives, I met wonderful people, and I am just about to graduate from MIT. And since all of this was not achieved on my own, I would like to thank all the people that helped me to be at this precious time of my life.

First and foremost, I would like to thank my research advisor, Dr. Martin L. Yarmush, for the opportunity of joining the Center for Engineering in Medicine (CEM). Dr. Yarmush opened the doors of his laboratory to me at the beginning of 2006 and since then he has believed in me and my work. His financial support, academic guidance and outstanding research experience have been key elements in completing this doctoral thesis. By shaping my way of thinking, Dr. Yarmush encouraged me to explore new directions and find answers where I was not able to see before.

I also would like to express my most sincere gratitude to my thesis committee members: Dr. Jagesh Shah, who agreed to become the chair of the committee, Dr. Radhika Nagpal, and Dr. Forest White. They all accepted the challenges of getting involved in this research, kept me focused on the big picture, yet encouraged me to pursue all of my ideas.

My eternal gratitude goes to my parents, who always pushed me to become a better person, and to be simple and grateful. They gave me all of my values and made me the man I am right now. My father, Marco P. Avila, a high school teacher, gave everything for his children. He used to teach day and night in order to provide for us and yet he spent his whole life with us. He is my example, my horizon as a husband, father and man. I miss him a lot and I wish he were here in this world to see this accomplishment. My mother, Eliana I. Elchiver, an elementary

school teacher, always believed in me. After my father died while I was doing this Ph.D., she continuously inspired me to keep working and make them proud. This thesis would not have existed without my parents' sacrifice and love. I also would like to thank my sisters Paola and Karina, who have helped me by taking care of my father when he was sick and now by taking care of my mother, illuminating her solitude.

My deepest appreciation goes to my friend, Dr. Deepak Nagrath who with his wife Sunitha became the two most wonderful persons that I was able to meet not only during my Ph.D. but in my whole life. From the time I joined the CEM, I have been working side by side with Deepak, who lectured me not only about Pareto frontiers, multiobjective optimization and energy balance, but he also opened my mind, my eyes and my soul. Together with Sunitha, Deepak taught me how to face the world with all its related problems, to always see the positive aspect of everything and to never quit. In the seasonal stages of our soul, the pear tree may look very different if the wind heavily blows or if the flowers are in bloom. But no matter what, the pear tree is still the same and when the spring comes we will have plenty of fruit again. Thank you so much Deepak and Sunitha, my friends, my mentors, my parents.

I would like to thank all the people at the CEM, especially Arno Tilles and Ilana Reis. Although Arno and I did not spend much time together, I feel Arno as a very close friend. He was always willing to selflessly help me with my English and passed me the light and happiness he carries with him. Ilana also helped me with all administrative issues and I always felt she treated me like a son. Although Don Poulsen is not longer part of the Center's community, I also would like to thank him for his assistance in designing many of this thesis's figures and for teaching me to use Illustrator.

A big part of my life is amazing friends. I would like to recognize the impact of them in my life and Ph.D. In particular, I would like to thank my friend Roberto Rondanelli and his wife for buoying me up when I felt down and when this walk through the desert was too heavy. Many thanks to Dolores Vidal and Barbara Starr for your unlimited friendship, comprehension, and patience. To all of my friends here in Boston: Saurabh, Yoni and Amanda, Maria Louisa, Mridula and Raj, Arman, Daniela and Felipe, Andrea and Rodrigo, Ricardo and Carmen, thank you so much.

I would like to express my gratitude to my professors and personnel in the Department of Chemical Engineering at the University of Concepcion, Chile, who played a very important role

in my forming education and application process. I would also like to acknowledge the Fulbright Commission, in particular Maritza Marquez, and the Chilean National Commission for Research in Science and Technology (Comision Nacional de Investigacion Cientifica y Tecnologica, CONICYT) for their support during my first four years as a graduate student.

Finally, I dedicate this thesis to my wife Monica, my eternal companion and love of my life. She has traveled this entire journey together with me. Monica has cried, laughed and bled with me. She completes me, inspires me, and supports me. As this Ph.D. is not only in Medical Engineering but in life, love, trust, partnership, and parenthood, this thesis is not only mine but hers. Monica unconditionally followed me into this adventure and left her family and friends in Chile to become the *significant other*. It was not easy for her and yet we were able to start our family and complete our goals. She gave me the most beautiful daughter in the world, Emilia Antonia, who in her short two years has also been part of this process. Although she does not understand it very well, I have sacrificed some of her time on behalf of this work. Monica is also about to give me a son, who is supposed to be born in a couple of days after my thesis defense. He will be named Marco Ignacio, after my father and me. From Monica's womb, little Marco also heartened me to complete this achievement.

TABLE OF CONTENTS

1	INTRODUCTION.....	10
2	SOFT CONSTRAINTS-BASED MULTIOBJECTIVE FRAMEWORK FOR METABOLIC FLUX ANALYSIS	13
2.1	OVERVIEW	13
2.2	INTRODUCTION.....	13
2.3	MATERIALS AND METHODS	16
2.3.1	Metabolic Flux Analysis	16
2.3.2	Hepatocyte Metabolic Network	17
2.3.3	Multiobjective Optimization.....	22
2.3.4	Pareto Concept.....	23
2.3.5	Linear Physical Programming Lexicon	24
2.3.6	Constrained Multiobjective Metabolic Flux Analysis Framework.....	27
2.3.7	Quantitative Aspects of Linear Physical Programming.....	30
2.3.8	Generation of Pareto Points Using LPPFBA	34
2.4	RESULTS AND DISCUSSION.....	38
2.4.1	Analysis of Bi-Objective Hepatic Metabolic Network.....	40
2.4.2	Analysis of Tri-Objective Hepatic Metabolic Network.....	44
2.4.3	Quad-Objective Hepatic Metabolic Network	51
2.5	CONCLUSIONS AND SUMMARY	61
3	INTEGRATED ENERGY AND FLUX BALANCE BASED MULTIOBJECTIVE FRAMEWORK FOR LARGE-SCALE METABOLIC NETWORKS.....	63
3.1	OVERVIEW	63
3.2	INTRODUCTION.....	63
3.3	THEORY	66
3.3.1	Metabolic Flux Analysis	66
3.3.2	Energy Balance Analysis	69
3.3.3	Pareto Optimality	71

3.3.4	Normal Constraint Method	73
3.3.5	Normal Constraint Energy and Flux Balance Analysis (NCEFBA).....	77
3.3.6	Hepatic Function Specific Fluxes for Pareto Optimization	79
3.4	RESULTS.....	80
3.4.1	Pareto Frontiers of Liver Specific Functions	80
3.4.2	Effect of FBA+EBA on Pareto Frontier Compared to FBA Alone.....	91
3.4.3	Effect of Measured Fluxes on Pareto Frontiers	92
3.5	DISCUSSION.....	98
4	REGULATORY TRANSCRIPTIONAL NETWORK IN EMBRYONIC STEM CELLS	101
4.1	BACKGROUND.....	101
4.1.1	Mathematical Modeling	104
4.1.2	Steady State Responses.....	106
4.2	APPLICATION OF PARETO OPTIMALITY FOR THE PREDICTION OF THE REGULATION OF TARGET GENES BOX ARCHITECTURE.	108
4.3	EXTENSION OF THE TRANSCRIPTION FACTORS TO FOXD3	110
4.3.1	Mass Balance Equations	112
4.3.2	Steady State Responses.....	114
4.3.3	Pareto Optimality for Prediction of Network Architecture.....	115
5	KINETIC MODELING OF TRANSCRIPTIONAL REGULATORY NETWORKS USING CYCLE CONCEPT.....	119
5.1	OVERVIEW	119
5.2	NONEQUILIBRIUM THERMODYNAMICS.....	119
5.3	ACTIVATION OF Z BY X.....	123
5.3.1	Mechanism.....	123
5.3.2	Mass Balance Equations	125
5.3.3	General Steady State Solution	126
5.3.4	Approximate Steady State Solution.....	128
5.3.5	Energy of Activation.....	133
5.4	REPRESSION OF Z BY X	134

5.4.1	Mechanism.....	134
5.4.2	Mass Balance Equations	136
5.4.3	General Steady State Solution	137
5.4.4	Approximate Steady State Solution.....	139
5.4.5	Energy of Repression.....	141
5.5	TRANSCRIPTIONAL REGULATION OF Z BY TWO TRANSCRIPTION FACTORS X AND Y.....	141
5.5.1	Mechanism.....	142
5.5.2	Mass Balance Equations	143
5.5.3	Energy of Transcription	145
6	GLOBAL THERMODYNAMICS CONTROL THE EVOLUTION OF TRANSCRIPTIONAL REGULATORY NETWORKS	148
6.1	OVERVIEW	148
6.2	INTRODUCTION.....	148
6.3	DESCRIPTION OF NONEQUILIBRIUM TRNs.	150
6.4	HYPOTHESIS: SPECIFIC DISSIPATION ENERGY SHOULD BE MINIMAL FOR BIOLOGICAL SYSTEMS UNDER OPTIMAL CONDITIONS.	152
6.5	<i>SDE</i> LINKS NETWORK TOPOLOGY WITH DYNAMIC FUNCTIONALITY AND EVOLUTIONARY ADAPTATIONS.	157
6.6	<i>SDE</i> PREDICTS THE SELECTION OF MOTIFS UNDER VARYING ENVIRONMENTS	160
6.7	<i>SDE</i> PREDICTS NEW FUNCTIONALITIES OF MOTIFS	162
6.8	<i>SDE</i> AS THE BASIS FOR TOPOLOGICAL GENERALIZATION OF MOTIFS	164
6.9	DISCUSSION.....	168
6.10	METHODS.....	169
6.10.1	Pareto Dominance Concept Correlates Dissipative Energetics for Metabolic Networks and TRN Motifs.....	169
6.10.2	Pareto Optimal <i>SDE</i> for FFL TRN Motifs.....	178
6.10.3	<i>SDE</i> Correlation With Frequency of Occurrence of TRN Motif.....	183
6.10.4	Correlating <i>SDE</i> Function with Dynamical Functions of Network Motifs	185
6.10.5	Dynamical Characterization of Higher Order FFLs Using Energetic Cost	189
6.10.6	Prediction of Environmental Selection of Motifs Based on Energetic Cost.....	194

6.10.7	Energetic Cost as the Underlying Basis for Generalization of Subgraphs to Form Complex Networks.....	198
6.10.8	Energetic Phase Diagram Generation for Directional SIM Based Generalizations	200
6.10.9	Pulsed Energetics Topological Generalization Prediction (PETGP) Strategy for Energetics Based Dynamical Analysis for Characterizing Subgraphs.....	205
7	ENERGETIC ANALYSIS OF AN EXTENDED PLURIPOTENCY MODEL OF THE TRANSCRIPTIONAL NETWORK OF EMBRYONIC STEM CELLS	211
7.1	INTRODUCTION.....	211
7.2	THEORY	213
7.2.1	Improved Kinetic Transcriptional Regulatory Network (TRN) Model.....	213
7.2.2	General Michaelis-Menten Formulation.....	221
7.2.3	Relationship Between the Kinetic TRN Model and the Michaelis-Menten Formulation.....	223
7.2.4	Total Transcription Rate	223
7.2.5	Robustness Index	224
7.3	RESULTS.....	225
7.3.1	Insights From A Basic TRN Model of ESCs or Level-0	225
7.3.2	The Extended Pluripotent TRN of ESCs. Level-1	228
7.3.3	Energetic Cost Explains Regulation of Target Gene Expression by Transcription Factor Occupancy in Level-2	237
7.3.4	Analysis of ESC TRN Architecture: Role of Autoregulation, Feedback Loops, and Other Motifs.	241
7.4	CONCLUSIONS	248
8	SUMMARY AND FUTURE DIRECTIONS.....	249
9	REFERENCES.....	252

1 INTRODUCTION

Microorganisms and mammalian cells perform different functions, such as protein secretion, detoxification, storage and release of sources of energy, etc. Among these metabolic tasks, transcription of mRNA and further translation into the final proteins play an important role in determining the cellular state and functionality, and feedback the subsequent metabolic function of the system. In other words, cells are what they are based on the different proteins that they can express. For instance, the differentiation process of embryonic stem (ES) cells into ectoderm, mesoderm or endoderm depends exclusively on the transcription factors that may be up or down regulated, and thus protein synthesis and their interaction at DNA level forming transcriptional networks are key mechanisms in determining the cellular fate. In addition to the multiple metabolic and transcriptional functions that these systems perform, evolution also requires certain conditions in the time response to these tasks such as stability, robustness, fast or slow responses, or simply no response to small inputs in order to avoid mutations.

In order to determine the optimal state at which the cellular system is accomplishing all of these metabolic, transcriptional and dynamic functions, different and multiple objectives need to be taken into account. Typically, several objectives compete against each other; therefore, only “Pareto-optimal” solutions can be achieved. A solution is said to be Pareto-optimal if there are no other solutions that can better satisfy all of the objectives simultaneously (Nagrath, 2005). Specifically, a Pareto solution is one where any improvement in one objective necessitates the worsening of at least one other objective. Mass balance constraints (stoichiometry and Flux Balance Analysis) reduce the feasible space and consequently the Pareto-optimal solution. Energy Balance Analysis imposes the thermodynamic constraints on reaction fluxes both explicitly and implicitly for metabolic networks. Essentially, the reaction potentials are computed based on the chemical potentials and then these are used to obtain thermodynamic constraints that are based on the first and second laws of thermodynamics. Thermodynamic constraints further reduce the feasible solution space based on stoichiometric constraints alone. Available measurements, which bring in environmental constraints such as certain cell culture conditions, medium supplements, induced stress and extracellular matrices, typically limit the feasible solution space even further. If a sufficient number of measurements are available, the

analysis may yield a unique solution. Application of these concepts in the hepatocyte metabolic network has been studied earlier (Beard and Qian, 2005; Nagrath et al., 2007) . However, transcriptional and dynamical optimality subject to energy constraints is still an unexplored field, mainly due to the lack of a transcriptional model from which energy can be computed. Current models (Bolouri and Davidson, 2002) are based on thermodynamic equilibrium and thus the energy dissipated is always zero.

As mentioned before, a very important example of increasing interest of transcriptional networks is its application in the differentiation of ES cells. ES cells, derived from the inner cell mass of mammalian blastocysts, have the ability to grow indefinitely while maintaining pluripotency (i.e., ability to differentiate into various cell types). These properties have led to expectations that human ES cells might be useful to understand disease mechanisms, to screen effective and safe drugs, and to treat patients of various diseases and injuries, such as juvenile diabetes and spinal cord injury. The therapeutic potential of ES cells is contingent upon generating functional cells which have identical physiological characteristics as the damaged/diseased cells sought to be replaced. Because ES cells are pluripotent, they are considered the “cell type of choice”. Recently, a transcriptional network model for differentiation of ES cells was presented (Chickarmane et al., 2006) in which two proposed networks between three transcription factors Oct4, Sox2 and Nanog lead to a switch like behavior of the differentiation and self-renewal of target genes at steady state. However, whether the proposed interactions occur in nature nobody knows. In addition to the Oct4, Sox2 and Nanog, a set of six additional transcription factors has been found to play an important role in the maintenance of ES cells (Kim et al., 2008). This core transcriptional network includes Klf4, Nac1, Rex1, Dax1, cMyc, and Zfp281 in addition to the Oct4, Sox2 and Nanog.

In order to find the optimal conditions and architectures at which protein transcription occurs in this work we developed a nonequilibrium transcriptional model that allowed us to apply Pareto-optimality and energy balance theory in the ES cell network: analysis of the proposed interactions by (Chickarmane et al., 2006), extension of the transcription factors to the six additional transcription factors and potentially many others, reprogramming of somatic cells into the pluripotent state and the differentiation of ES cells into hepatocytes. In addition, this energy based optimization theory can be applied to the general gene circuits such as the

repressillator system and to the study of any metabolic network, with the fatty liver analysis one of our goals.

2 SOFT CONSTRAINTS-BASED MULTIOBJECTIVE FRAMEWORK FOR METABOLIC FLUX ANALYSIS

2.1 OVERVIEW

The current state of the art for linear optimization in Flux Balance Analysis has been limited to single objective functions. Since mammalian systems perform various functions, a multiobjective approach is needed when seeking optimal flux distributions in these systems. In most of the available multiobjective optimization methods, there is a lack of understanding of when to use a particular objective, and how to combine and/or prioritize mutually competing objectives to achieve a truly optimal solution. To address these limitations we developed a soft constraints based linear physical programming-based flux balance analysis (LPPFBA) framework to obtain a multiobjective optimal solution. The developed framework was first applied to compute a set of multiobjective optimal solutions for various pairs of objectives relevant to hepatocyte function (urea secretion, albumin, NADPH, and glutathione syntheses) in bioartificial liver systems. Next, simultaneous analysis of the optimal solutions for three objectives was carried out. Further, this framework was utilized to obtain true optimal conditions to improve the hepatic functions in a simulated bioartificial liver system. The combined quantitative and visualization framework of LPPFBA is applicable to any large-scale metabolic network system, including those derived by genomic analyses.

2.2 INTRODUCTION

Metabolic flux analysis (MFA) provides a framework for the estimation of intracellular metabolic fluxes at steady-state based on stoichiometric constraints of a metabolic pathway network. This technique, which has been extensively used for studying the metabolism of microorganisms (Antoniewicz et al., 2007a; Antoniewicz et al., 2007b; Stafford et al., 2002; Wong et al., 2004; Young et al., 2008), has been recently applied to characterize and compare different physiological states in mammalian systems (Banta et al., 2004; Chan et al., 2003a; Chan et al., 2003b; Chan et al., 2002; Chan et al., 2003c; Lee et al., 2004; Vo et al., 2004). In order to obtain a unique solution for the flux distribution in a particular cell or tissue system, a minimum number of measurements of rates of uptake and release of extracellular metabolites by the system

is needed. In cases where insufficient measurements are available, pathway fluxes have been predicted using linear optimization for one objective function, such as growth rate for unicellular organisms (Savinell and Palsson, 1992b). Mammalian systems, such as hepatocytes, typically do not undergo cell proliferation, but rather perform an array of metabolic functions (protein secretion, detoxification, energy production), therefore different and multiple objectives need to be taken into account. Hepatocytes are the major cell type in the liver with multiple functions including efficient uptake and subsequent metabolic conversion of amino acids, carbohydrates, lipids, and vitamins. Subsequently, these nutrients are either stored or released after biochemical transformations. These biochemical processes make hepatocytes the epicenter of the metabolic modulation of intermediary metabolism in the body, and thus can play an important role in biotechnological applications that use liver cells, such as bioartificial liver (BAL) devices. A recent analysis concluded that several objectives were necessary to profile metabolic information from perfused livers (Lee et al., 2004). Another study (Nagrath et al., 2007), combined both energy and flux balance based nonlinear multiobjective framework for hepatic systems. Recently, Bayesian (Knorr et al., 2007) and optimization (Burgard and Maranas, 2003) based techniques have been developed for selecting objectives.

Multiobjective optimization strategies previously used for MFA, such as weighted optimization and goal programming, suffer from several limitations. For example, it is often unclear when to use a particular objective and how to combine and/or prioritize mutually competing objectives to achieve a true optimal solution. Furthermore, visualization of the results is not straightforward. Importantly, most of the existing MFA methods employ “*hard constraints*” for the estimation/distribution and optimization of intracellular fluxes in metabolic networks. Burgard and Maranas (Burgard and Maranas, 2003) had developed an optimization-based framework for testing whether maximization of a weighted combination of fluxes can explain a set of observed experimental data. Their approach is based on weighted-sum (WS) based optimization and utilizes weights defined as the coefficients of importance to quantify the additive contribution of a given flux to a objective function. The limitations of using WS based approach are: (1) weights are arbitrary and have no physical meaning; (2) the spacing of optimal solutions is largely dependent on relative scaling of weights thus often leads to ill-conditioned problems; (3) an even distribution of scalar weights in WS does not yield an even optimal flux solutions; and (4) WS fails to capture the Pareto optimal solutions where the Pareto frontier is

non-convex. Here we introduce a multiobjective optimization based metabolic flux analysis approach, LPPFBA (Linear Physical Programming-Based Flux Balance Analysis), that employs Linear Physical Programming (LPP) (Maria et al., 2003; Messac, 2000) (Figure 2.1A), which enables the formulation of the optimization problem in terms of physically meaningful terms and parameters, and addresses the problems that exist in the previously used strategies by employing “soft constraints”.

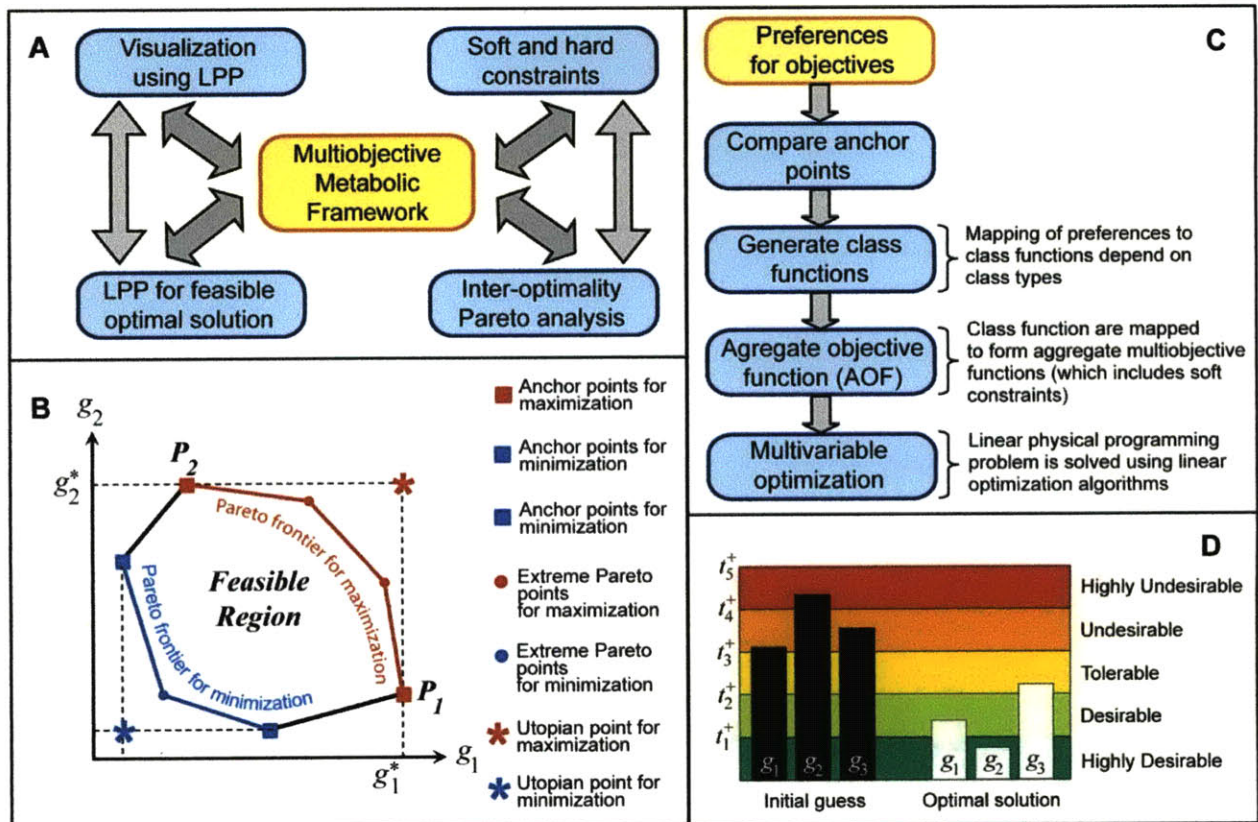


Figure 2.1: A). Characteristics of Linear Physical Programming based multiobjective metabolic network. B). Pareto frontiers and extreme Pareto points shown for a bi-objective maximization and minimization problems. C). Strategy for Linear Physical Programming based multiobjective optimization. D). LPPFBA visualization window. Multiobjective optimal solutions are displayed as bar graph in color coded regions, with highly desirable solution being green and highly undesirable being red.

In LPPFBA approach, first a set of so-called Pareto-optimal solutions is generated. A solution is said to be Pareto-optimal if there are no other solutions that can better satisfy all of the objectives simultaneously (Nagrath et al., 2005). In other words, any improvement in one objective necessitates the worsening of at least one other objective. In our specific application, we generate Pareto frontiers of optimal metabolic fluxes to identify potential solution regions

that provide a qualitative framework to assess the situation, and in particular, to determine the objectives that are conflicting. Second, we employ LPPFBA to prioritize the objectives and constraints. In LPPFBA, attributes of interest for each objective are defined to delineate degrees of desirability: *unacceptable*, *highly undesirable*, *undesirable*, *tolerable*, *desirable*, and *highly desirable*. Thus, LPPFBA completely eliminates the need for iterative weight setting, which is the object of the typical computational *bottleneck* in large optimization problems. Two key advantages of LPPFBA for metabolic flux optimization are: (1) once the preferences are articulated, obtaining the corresponding optimal fluxes is a non-iterative process – in stark contrast to conventional weight-based methods; and (2) it provides the means to reliably employ optimization with minimal prior knowledge thereof.

In this chapter, we present a LPPFBA approach for analyzing the multiobjective flux analysis of metabolic networks. The developed approach was then utilized for optimizing the metabolism of liver cells in the context of bioartificial liver (BAL) development. BALs are being developed to provide hepatic support to patients with fulminant hepatic failure (Chan et al., 2003a; Chan et al., 2003b; Chan et al., 2002; Chan et al., 2003c). One of the major design goals of BAL devices is to maintain viable hepatocytes that perform a high level of liver-specific functions (for example, albumin synthesis, urea secretion, cytochrome p450-mediated detoxification, etc.) (Berthiaume et al., 1996; Chan et al., 2002; Dunn et al., 1991). Obtained Pareto optimal metabolite fluxes were computed for various combinations of liver specific functions. Next, we obtained Pareto optimal solutions for tri-objective combinations of these hepatic functions. Lastly, we obtained the Pareto solutions for a simulated BAL system where the main goal was to operate the BAL at the highest possible level during human plasma exposure. This analysis exhibited that although lipid and carbohydrates fluxes may be similar but for hepatocytes amino acid synthesis, catabolism, is altered/rerouted in optimality analysis for maintaining hepatic functions.

2.3 MATERIALS AND METHODS

2.3.1 Metabolic Flux Analysis

The stoichiometric coefficients of the metabolic reactions are collected into a matrix S , where each element s_{ij} is the coefficient of metabolite i in reaction j . S has dimensions of $M \times N$,

where M is the number of metabolites and N is the number of reactions. In matrix form the mass balance is written as:

$$\frac{dx}{dt} = SJ \quad (2.1)$$

where each element x_i of x is the intracellular concentration of metabolite i and element J_i of J is the net rate of conversion in reaction j . External metabolite fluxes are generally measured (e.g., uptake of glucose, lactate, amino acids). Because of the very high turnover of the intracellular pools of most intracellular metabolites, the time scale of the intracellular metabolic reactions is short compared to other cellular reactions. Hence, the pseudo steady state assumption is generally applied to the metabolite mass balances, and thus

$$SJ = 0 \quad (2.2)$$

When the number of measured quantities is less than the number of measurements required for the system to be determined, the computation of unknown intracellular fluxes requires linear optimization (since infinite number of solutions exist) with linear bound constraints. Mathematically, this can be expressed as:

$$\max_{J_u} \quad c^T J_u \quad (2.3)$$

subject to

$$S_u J_u = -S_m J_m \quad (2.4)$$

$$J_{low} \leq J_u \leq J_{high} \quad (2.5)$$

where the vector c specifies which unknown flux vector elements are to be maximized (or minimized); vectors J_{low} and J_{high} provide the lower and upper bounds for the unknown fluxes; J_m and J_u indicate measured and unmeasured fluxes, respectively; and S_m and S_u contain the stoichiometric coefficients of measured and unknown reactions, respectively.

2.3.2 Hepatocyte Metabolic Network

A previously described hepatic metabolic network (Chan et al., 2003b; Chan et al., 2002) includes all of the major intracellular pathways to account for the majority of central carbon and nitrogen metabolism found in hepatocytes, namely the tricarboxylic acid (TCA) and urea cycles, the gluconeogenic and glycolysis pathways, the pentose phosphate shunt, pathways of entry,

transamination, and deamination of amino acids, protein synthesis, and the major components of lipid metabolism, including triglyceride synthesis and breakdown and β -oxidation of fatty acids, in addition to amino acid synthesis and apolipoprotein degradation. The current hepatic metabolic network model (Table 2.1) includes a few additional reactions, namely those of the 3-phosphoglycerate cycle as it is involved in glycerol production and glutathione synthesis, which results in a total of 81 reactions (as compared to 76 reactions in the previous model) and 47 metabolites (Table 2.2). Figure 2.2 presents the comprehensive hepatic metabolic network. The rationale for including glutathione synthesis is that glutathione is involved in several important detoxification functions of hepatocytes. The model assumes pseudo steady-state with no metabolic futile cycles. These assumptions are discussed in detail elsewhere (Chan et al., 2003a).

Table 2.1: Hepatic stoichiometric reactions

No	Stoichiometry
1	F6P \leftrightarrow G6P
2	F16P2 + H ₂ O \rightarrow F6P + P _i
3	2 G3P \leftrightarrow F16P2
4	3Pglyc + NADH + H ⁺ + ATP \leftrightarrow G3P + P _i + NAD ⁺ + ADP
5	PEP \leftrightarrow 3Pglyc
6	oac + GTP \rightarrow PEP + GDP + CO ₂
7	pyr + CO ₂ + ATP + H ₂ O \rightarrow oac + ADP + P _i + 2 H ⁺
8	oac + acCoA + H ₂ O \rightarrow ctt + CoASH
9	ctt + NAD ⁺ \leftrightarrow α kgl + CO ₂ + NADH + H ⁺
10	α kgl + NAD ⁺ + CoASH \rightarrow sucCoA + CO ₂ + NADH + H ⁺
11	sucCoA + P _i + GDP + FAD \leftrightarrow fum + GTP + FADH ₂ + CoASH
12	fum + H ₂ O \leftrightarrow mal
13	mal + NAD ⁺ \leftrightarrow oac + NADH + H ⁺
14	ctr + asp + ATP \rightarrow arg + fum + AMP + PP _i
15	orn + (CO ₂ + NH ₄ ⁺ + 2 ATP) + H ₂ O \rightarrow ctr + 2 ADP + 2 P _i + 3 H ⁺
16	arg + H ₂ O \rightarrow urea + orn
17	ala + 0.5 NAD ⁺ + 0.5 NADP ⁺ + H ₂ O \leftrightarrow pyr + NH ₃ + 0.5 NADH + 0.5 NADPH + H ⁺
18	ser \rightarrow pyr + NH ₃
19	cys + 0.5 NAD ⁺ + 0.5 NADP ⁺ + H ₂ O + SO ₃ ²⁻ \leftrightarrow pyr + thiosulfate + NH ₄ ⁺ + 0.5 NADPH + 0.5 NADH
20	thr + NAD ⁺ + ATP + CoASH \rightarrow gly + acCoA + NADH + H ⁺ + AMP + PP _i
21	thr + NAD ⁺ + CoASH \rightarrow propCoA + CO ₂ + NADH + H ⁺ + NH ₃ + H ₂
22	2 gly + NAD ⁺ + THF + H ₂ O \leftrightarrow NTHF + H ⁺ + CO ₂ + NH ₄ ⁺ + ser + NADH
23	3Pglyc + NAD ⁺ + glu + H ₂ O \rightarrow NADH + H ⁺ + α kgl + ser + P _i
24	trp + 3 O ₂ + 4 H ₂ O + 2 NAD ⁺ + FAD + CoASH \rightarrow Formate + ala + 2 CO ₂ + NH ₃ + 3 NADH + FADH ₂ + HCO ₃ ⁻ + acacCoA
25	propCoA + CO ₂ + ATP \rightarrow ADP + P _i + sucCoA
26	lys + 2 α kgl + 2 NAD ⁺ + CoASH + FAD + 2 H ₂ O + NADP ⁺ \rightarrow CO ₂ + NH ₃ + acacCoA + 5 NADH + FADH ₂
27	phe + O ₂ + H ₄ biopterin + H ⁺ \rightarrow tyr + H ₂ O + H ₂ biopterin
28	tyr + α kgl + 2 O ₂ + H ₂ O \rightarrow glu + CO ₂ + fum + acac
29	pro + 0.5 O ₂ + 0.5 NAD ⁺ + 0.5 NADP ⁺ \rightarrow glu + 0.5 NADH + 0.5 NADPH + H ⁺
30	his + H ₄ folate + 2 H ₂ O \rightarrow NH ₄ ⁺ + N ⁵ ,N ¹⁰ -CH ₂ -H ₄ folate + glu
31	met + ATP + ser + NAD ⁺ + H ₂ O + CoASH \rightarrow PP _i + P _i + adenosine + cys + NADH + H ⁺ + CO ₂ + NH ₄ ⁺ + propCoA

32	$\text{val} + \alpha\text{kgI} + 3 \text{NAD}^+ + 2 \text{H}_2\text{O} + \text{FAD} + \text{CoA} \rightarrow \text{glu} + 2 \text{CO}_2 + 3 \text{NADH} + 2 \text{H}^+ + \text{FADH}_2 + \text{CO}_2 + \text{propCoA}$
33	$\text{iso} + \alpha\text{kgI} + \text{H}_2\text{O} + 2 \text{NAD}^+ + \text{FAD} + 2 \text{CoASH} \rightarrow \text{glu} + \text{CO}_2 + 2 \text{NADH} + 2\text{H}^+ + \text{FADH}_2 + \text{acCoA} + \text{propCoA}$
34	$\text{leu} + \alpha\text{kgI} + \text{H}_2\text{O} + \text{NAD}^+ + \text{FAD} + \text{ATP} + \text{CoASH} + \text{HCO}_3^- \rightarrow \text{glu} + \text{CO}_2 + \text{NADH} + \text{H}^+ + \text{FADH}_2 + \text{acCoA} + \text{acac} + \text{ADP} + \text{P}_i$
35	$\text{oac} + \text{glu} \leftrightarrow \alpha\text{kgI} + \text{asp}$
36	$\text{asn} + \text{H}_2\text{O} \rightarrow \text{asp} + \text{NH}_3$
37	$\text{glu} + 0.5 \text{NAD}^+ + 0.5 \text{NADP}^+ + \text{H}_2\text{O} \leftrightarrow \text{NH}_4^+ + \alpha\text{kgI} + 0.5 \text{NADPH} + 0.5 \text{NADH} + \text{H}^+$
38	$\text{orn} + \text{NAD}^+ + \text{NADP}^+ + \text{H}_2\text{O} \rightarrow \text{glu} + \text{NH}_4^+ + \text{NADH} + \text{NADPH} + \text{H}^+$
39	$\text{gln} + \text{H}_2\text{O} \rightarrow \text{glu} + \text{NH}_4^+$
40	$\text{palm} + \text{ATP} + 7 \text{FAD} + 7 \text{NAD}^+ \rightarrow 8 \text{acCoA} + 7 \text{FADH}_2 + 7 \text{NADH} + \text{AMP} + \text{PP}_i$
41	$2 \text{acCoA} \leftrightarrow \text{acacCoA} + \text{CoA}$
42	$\text{acacCoA} + \text{H}_2\text{O} \rightarrow \text{acac} + \text{CoA}$
43	$\text{NADH} + \text{H}^+ + 0.5 \text{O}_2 + 3 \text{ADP} \rightarrow \text{NAD}^+ + \text{H}_2\text{O} + 3 \text{ATP}$
44	$\text{FADH}_2 + 0.5 \text{O}_2 + 2 \text{ADP} \rightarrow \text{FAD} + \text{H}_2\text{O} + 2 \text{ATP}$
45	$\text{gol} + \text{NAD}^+ + \text{ATP} \leftrightarrow \text{G3P} + \text{NADH} + \text{H}^+ + \text{ADP} + \text{P}_i$
46	$\text{G6P} + 12 \text{NADP}^+ + 7 \text{H}_2\text{O} \rightarrow 6 \text{CO}_2 + 12 \text{NADPH} + 12 \text{H}^+ + \text{P}_i$
47	$24 \text{arg} + 32 \text{asp} + 61 \text{ala} + 24 \text{ser} + 35 \text{cys} + 57 \text{glu} + 17 \text{gly} + 21 \text{tyr} + 33 \text{thr} + 53 \text{lys} + 26 \text{phe} + 25 \text{gln} + 30 \text{pro} + 15 \text{his} + 6 \text{met} + 20 \text{asn} + \text{trp} + 35 \text{val} + 13 \text{iso} + 56 \text{leu} + 2332 \text{ATP} \rightarrow \text{albumin} + 2332 \text{ADP} + 2332 \text{P}_i$
48	$\text{glu} + 2 \text{ATP} + \text{cys} + \text{gly} + \text{NADPH} \rightarrow \text{GSH} + 2 \text{ADP} + 2 \text{P}_i + \text{NADP}^+ + \text{H}^+$
49	$\text{gol} + 3 \text{acCoA} + \text{H}_2\text{O} + \text{ATP} \rightarrow 3 \text{CoASH} + \text{P}_i + \text{TG} + \text{ADP} + \text{P}_i$
50	$\text{lactate} + \text{NAD}^+ \leftrightarrow \text{pyr} + \text{NADH} + \text{H}^+$
51	$\text{acac} + \text{NADH} + \text{H}^+ \leftrightarrow \beta\text{-OH-butyrate} + \text{NAD}^+$
52	$\text{TG} + 3 \text{H}_2\text{O} \rightarrow \text{gol} + 3 \text{palm} + 3 \text{H}^+$
53	G6P release
54	gol uptake
55	palm release
56	cholesterol ester + H ₂ O → cholesterol + palm
57	TG stored
58	trp uptake
59	O ₂ uptake
60	pro uptake
61	glu secretion
62	asn uptake
63	orn secretion
64	arg uptake
65	NH ₄ ⁺ uptake
66	ala uptake
67	ser uptake
68	gly uptake
69	asp uptake
70	acac production
71	thr uptake
72	lys uptake
73	phe uptake
74	his uptake
75	met uptake
76	val uptake
77	iso uptake
78	leu uptake
79	gln uptake
80	cys uptake
81	tyr uptake

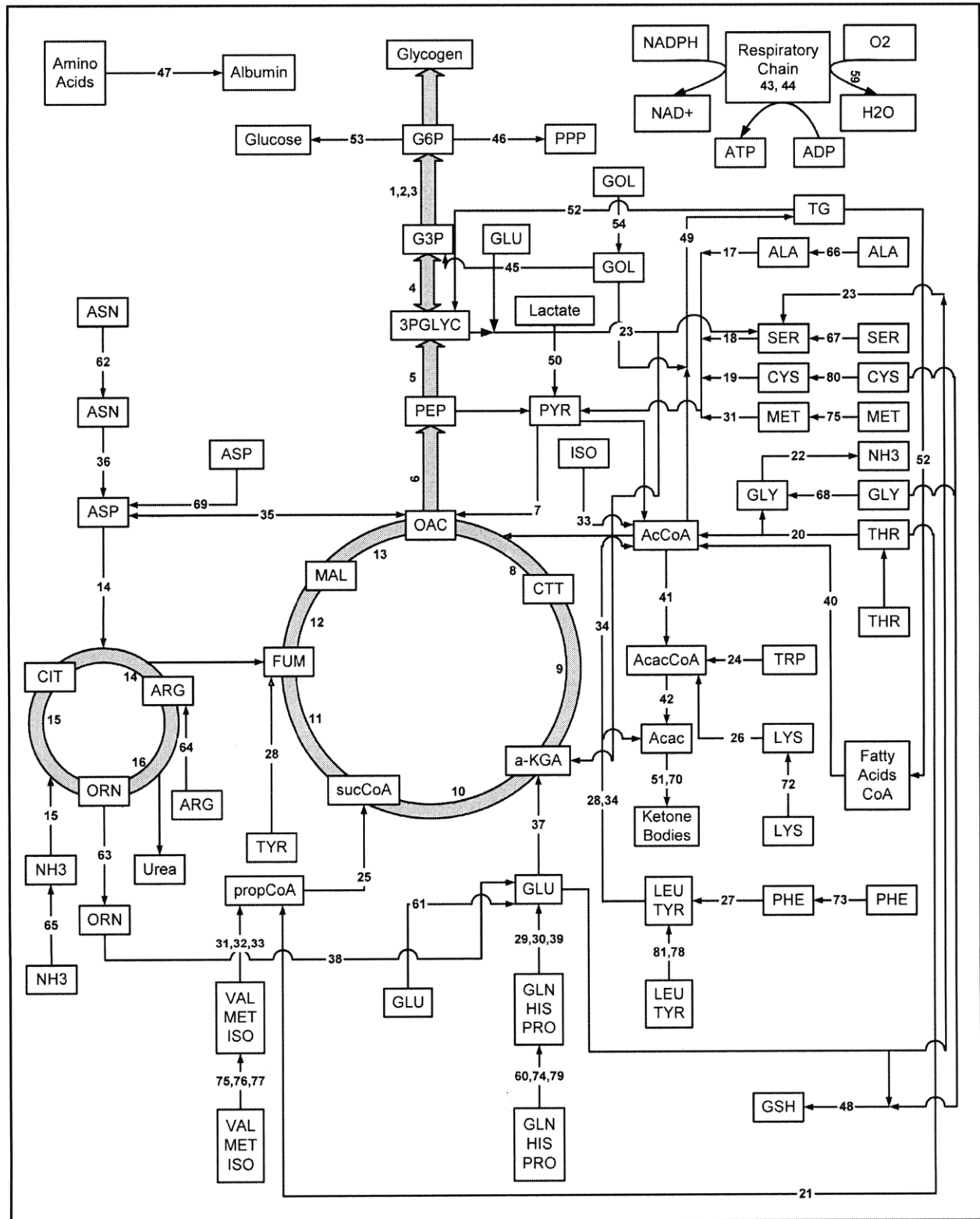


Figure 2.2: Hepatic metabolic network showing the linkage of various metabolites.

Table 2.2: List of metabolites

No	Symbol	Metabolite
1	G6P	Glucose-6-phosphate
2	F6P	Fructose-6-phosphate
3	F16P2	Fructose-1,6-biphosphate
4	G3P	Glyceraldehyde-3-phosphate
5	PEP	Phosphoenolpyruvate
6	oac	Oxaloacetate
7	pyr	Pyruvate
8	ctt	Citrate
9	akgl	α -Ketoglutarate
10	sucCoA	Succinyl-CoA
11	fum	Fumarate
12	mal	Malate
13	arg	Arginine
14	orn	Ornithine
15	ctr	Citrulline
16	NH₄⁺	Ammonium
17	asp	Aspartate
18	his	Histidine
19	glu	Glutamate
20	gln	Glutamine
21	met	Methionine
22	thr	Threonine
23	val	Valine
24	iso	Isoleucine
25	phe	Phenylalanine
26	trp	Tryptophan
27	lys	Lysine
28	tyr	Tyrosine
29	ala	Alanine
30	asn	Asparagine
31	pro	Proline
32	cys	Cysteine
33	ser	Serine
34	gly	Glycine
35	propCoA	Propionyl-CoA
36	acCoA	Acetyl-CoA
37	palm	Palmitate
38	acacCoA	Acetoacetyl-Coa
39	acac	Acetoacetate
40	gol	Glycerol
41	NADH	Nicotinamide adenine dinucleotide, reduced form
42	NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
43	FADH₂	Flavin adenine dinucleotide, reduced form
44	O₂	Oxygen
45	leu	Leucine
46	3Pglyc	3-phosphoglycerate
47	TG	Triglyceride

2.3.3 Multiobjective Optimization

Definitions:

Multiobjective Optimization: A multiobjective optimization is a problem involving several competing objectives and constraints. The solution of this problem is considered the best solution that satisfies the conflicting objectives. Other commonly used terms in the literature for multiobjective optimization are multicriteria optimization, multidecision optimization, and vector optimization.

Pareto Solution: A Pareto solution is one where any improvement in one objective can only take place at the cost of another objective. A Pareto set is a set of Pareto-optimal solutions.

Design Parameters: A design parameter is a parameter over which the designer has direct control. Other terms used in the literature for design parameters include decision variables, design variables or decision parameters.

Design Metric: A design metric refers to an objective measure of a design attribute. Other commonly used terms are objective functions, design criterion, figure-of-merit, goal and performance metric. In the current work, the variable $g(x)$ denotes the vector of design metrics.

Design Constraint: A design constraint indicates the lower or upper bounds in the design metrics or design parameters.

Anchor value: The value obtained for a particular design objective if that design metric alone is optimized, given the bounds on the design parameters.

Mathematical Formulation of Multiobjective Problem P:

Mathematically, the multiobjective problem can be stated as follows:

$$\begin{array}{l} \text{minimize} \\ x \in D \end{array} g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \cdot \\ \cdot \\ \cdot \\ g_m(x) \end{pmatrix}, \text{ (P)} \quad (2.6)$$

where $D = \{x \in R^n \mid h(x) = 0, f(x) \leq 0, \alpha \leq x \leq \beta\}$, $h: R^n \rightarrow R^r$, $f: R^n \rightarrow R^s$, $\alpha \in (R \cup \{-\infty\})^n$, $\beta \in (R \cup \{+\infty\})^n$, m is the number of objectives, or criteria, $m \geq 2$, r and s the numbers of equality

and inequality constraints, respectively. For any design parameter $x = (x_1, \dots, x_n)$, a design metric vector $g = (g_1, \dots, g_m)$ is defined according to the function $g : R^n \rightarrow R^m$. $Z = \{z \in R^m \mid z = g(x), x \in D\}$ is the set of images of all points in D . D is called the feasible region in decision space and Z the feasible region in objective space; $(g_1(x), \dots, g_m(x))$ are the coordinates of the image of x in objective space.

2.3.4 Pareto Concept

For the multiobjective problem P , it is highly improbable to have a single optimal solution x^* which minimizes every g_i simultaneously; therefore, the solution is defined in terms of Pareto optimality in the following sense: a feasible solution for a multiobjective programming problem is Pareto optimal (noninferior, nondominated) if there exists no other feasible solution that will yield an improvement in one objective without causing a degradation in at least one other objective. So, $x \in D$ is Pareto optimal if there does not exist $y \in D$, whose criteria vector, $q = g(y)$, dominates the criteria vector of x , $p = g(x)$, i.e. $q \leq p$, $p \neq q$. (For any vectors v and w , $v \leq w$ implies that $v_i \leq w_i \forall i$).

Figure 2.1B presents a scheme of a Pareto set for a bi-objective linear maximization and minimization problem. If design metric g_1 alone is optimized (maximized), then the optimal value is g_1^* (shown as point P_1). Similarly, if design metric g_2 alone is optimized then the optimal value is g_2^* (shown as point P_2). Here g_1^* and g_2^* are the anchor values for design objectives g_1 and g_2 , respectively. The ideal or Utopian solution (g_1^*, g_2^*) obtained by the individual maximization of the objective functions is generally not a feasible solution of the multiobjective optimization problem. As seen in Figure 2.1B, lines joining points P_1 and P_2 defining the boundary of the feasible space are the efficient Pareto frontier. That is, for every point on arc P_1 - P_2 , it is not possible to improve both objectives simultaneously. If one objective is improved, it must be at the expense of the other. Points on the arc are often referred to as Extreme Pareto points. In view of their stated characteristics, Pareto points are usually the candidate of choice in the process of multi-objective optimization. Figure 2.1B shows Pareto frontiers for both maximization and minimization problems. It shows the shape of frontiers and

tradeoffs involved between two objectives for two separate parameters and conditions of maximization and minimization.

Often several Pareto optimal points are available in cellular systems, representing alternative designs, from which one can select the one that offers the best trade-off among multiple objectives. This optimization generally involves forming an Aggregate Objective Function (AOF) (or, some functional aggregation of the many conflicting criteria). Implicit in this process is the assumption that this AOF has the ability to indeed yield all the potentially useful/desirable optimal solutions. The most common AOF structure is the *weighted-sum* approach, which involves forming a linear combination of objectives -- minimized subject to the problem constraints. The algorithm for the Linear Physical Programming (LPP) based multiobjective FBA is shown in Figure 2.1C.

2.3.5 Linear Physical Programming Lexicon

The first step in the Linear Physical Programming (LPP) lexicon is to express the preferences with respect to each objective using four different *classes*. Each *class* comprises two cases, *soft* and *hard*, referring to the sharpness of the preference. All *soft* class functions are integrated in the AOF (that will be minimized). The desired behavior of an objective function, during optimization, is described by one of eight sub-classes, four *soft* (S) and four *hard* (H). These classes are defined as follows:

Soft:

- Class-1S Smaller-Is-Better, i.e. minimization.
- Class-2S Larger-Is-Better, i.e. maximization.
- Class-3S Value-Is-Better.
- Class-4S Range-Is-Better.

Hard:

- Class-1H Must be smaller, i.e., $g_i \leq t_{i,max}$
- Class-2H Must be larger, i.e., $g_i \geq t_{i,min}$
- Class-3H Must be equal, i.e., $g_i = t_{i,val}$
- Class-4H Must be in range, i.e., $t_{i,min} \leq g_i \leq t_{i,max}$

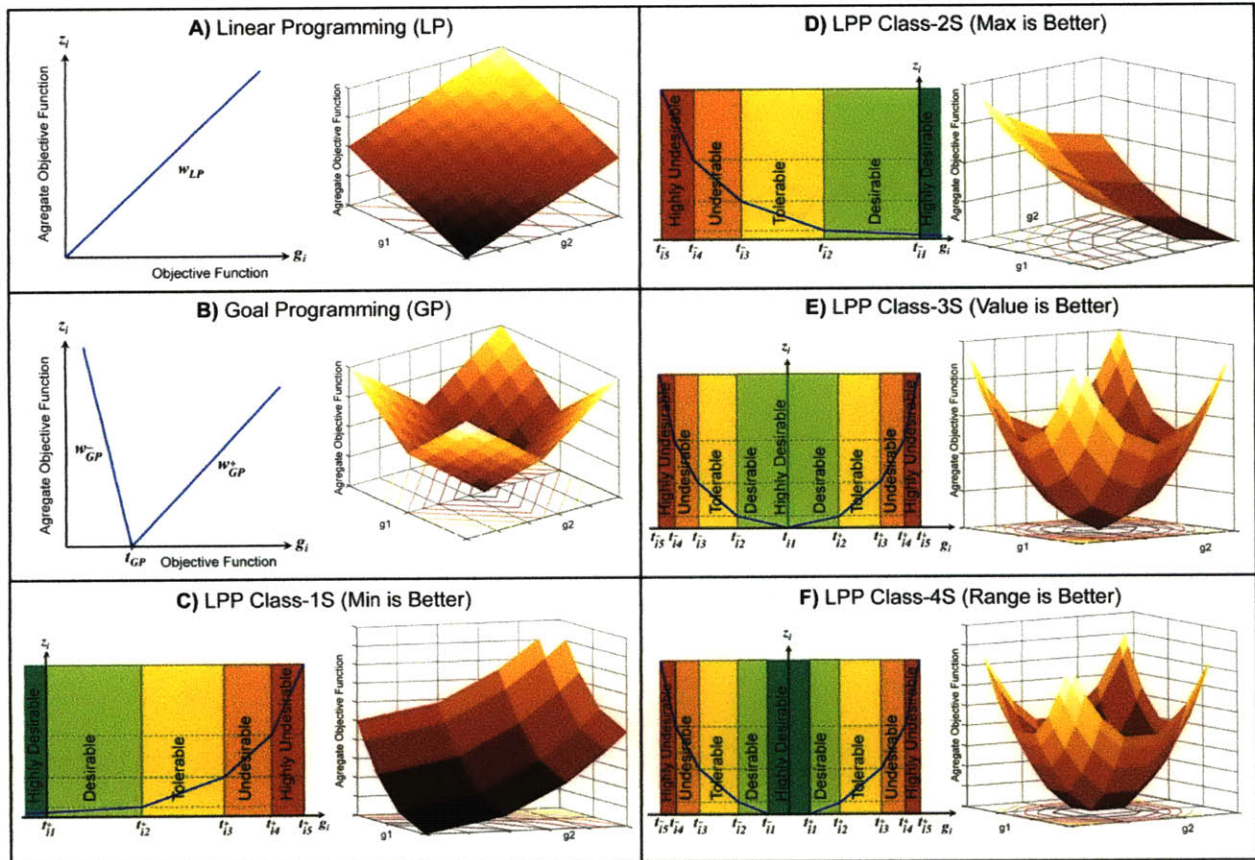


Figure 2.3: Relative capacity to express preferences using LP, GP and LPP for a given objective function, g_i . z_i is the actual quantity minimized in optimization. LP has a single weight w_{LP} and no target value, GP has 2 weights w_{GP}^+ and w_{GP}^- , and 1 target value t_{GP} . Soft classes-1S, 2S, 3S and 4S in LPP have 5, 5, 9, and 10 targets, respectively, and all LPP classes do not have any weights.

Figure 2.3 presents the relative capacity to express preferences using Linear Programming (LP), Goal Programming (GP), and LPP for a given objective function, g_i . The vertical axis, z_i , represents what is minimized in the optimization process. In the LP case, a single weight, w_{LP} , can be increased or decreased to express preference relative to other criteria. In the GP case, there are two weights, w_{GP}^- and w_{GP}^+ and a target value, t_{GP} . In the LPP case, there is flexibility to choose up to ten target values, and it eliminates the need to deal with weights entirely, as seen in Figures 2.3C (Class-1S), 2.3D (Class-2S), 2.3E (Class-3S) and 2.3F (Class-4S). The effectiveness of LPP comes from its ability to shape the class function (i.e. z_i) to suite the typically complex structure of the preference. Compared to choosing weights which can be difficult and undesirable because they are physically meaningless, choosing target values is preferable because they are physically meaningful. In both LP and GP, it is usually not clear

whether the weights should be increased by 10% or 100% in order to achieve the desired optimal result. This difficulty is compounded when there are several objectives involved. In this regard, LPP distinguishes itself by operating in a physically meaningful space.

The behavior of the AOF, in the cases of LP, GP, and LPP are shown as a collection of indifference curves and their three-dimensional view (Figures 2.3A-F). For the LP case, the weighted sum of two objectives leads to a simple plane. In two-sided-goal criteria for GP, there are four intersecting planes whose slopes depend on the weight-pair of each criterion. In contrast, the LPP case depicts a surface that comprises 81 intersecting planes (for class-4 criteria).

Soft Classes: The *soft* class functions allow ranges of differing levels of preferences for each objective to be expressed (Figures 2.3C-F). Based on their classes, the class functions are generated for all the objectives. These class functions are then minimized for each objective using a linear programming algorithm. The qualitative and quantitative depictions of each *class* are shown in Figures 2.3C-F. Where, the value of the objective i under consideration, g_i , is on the horizontal axis, and the corresponding *class function*, z_i , is on the vertical axis. A lower value of the class function is better (i.e., more valuable than) than a higher value, and a class function of zero is ideal. As would be done in conventional mathematical programming formalism, preferences for each criterion are required in LPP, compared to using the terms *minimize*, *maximize*, *greater than*, *less than*, or *equal to*. For the soft case, this lexicon comprises terms that characterize the degree of desirability of up to eleven ranges. The LPP lexicon comprises terms that characterize the degree of desirability of six ranges for each generic criterion for classes 1S and 2S, ten ranges for classes 3S, and eleven for class 4S. As an illustration of the LPP lexicon, consider the case of class 1S (Figure 2.3C), where the ranges are defined, in order of decreasing preference in Table 2.3. The parameters/targets t_{i1}^+ through t_{i5}^+ are physically meaningful values that are specified to quantify the preference associated with the i -th objective. These parameters delineate the desirability ranges within each objective. The shape of the class function depends on the numerical values of the range limits (targets). According to the definition of the *ideal* range, any two points of the ideal range are of *equal* value. The class function will be minimized only until the target value t_{i1}^+ is reached. Below that point, class 1S expresses explicit indifference. If a *smaller* value of the objective is *better*, the ideal range does

not apply. In this case, should be set to a value outside of the feasible space in order to exclude solutions in the *ideal* range. This will preclude the possibility of obtaining (incorrect) dominated solutions. A similar discussion would apply to the cases of class 2S and class 4S.

Table 2.3: Definition of ranges in their order of decreasing preferences for Class-1S

Range	Preference Values	Descriptions
Highly Desirable range	$g_i \leq t_{i1}^+$; Range-1	A range over which every value of the criterion is <i>ideal</i> (the most desirable possible). Any two points of that range are of equal value. (See following pertaining discussion.)
Desirable range	$t_{i1}^+ \leq g_i \leq t_{i2}^+$; Range-2	An acceptable range that is desirable
Tolerable range	$t_{i2}^+ \leq g_i \leq t_{i3}^+$; Range-3	An acceptable, tolerable range
Undesirable range	$t_{i3}^+ \leq g_i \leq t_{i4}^+$; Range-4	A range that, while acceptable, is undesirable
Highly Undesirable range	$t_{i4}^+ \leq g_i \leq t_{i5}^+$; Range-5	A range that, while still acceptable, is highly undesirable
Unacceptable range	$g_i \geq t_{i5}^+$; Range-6	The range of values that the generic metric may not take

Hard Classes: Constraints that are hard and have to be met are classified as *hard*. There are *hard* only two ranges defined for a hard objective, *acceptable* and *unacceptable*. All the *soft* class functions become part of the AOF to be minimized, and all of the *hard* class functions simply appear as constraints in the LPP model.

In the next section, the mathematical formulation of the class function for the soft objectives in a multiobjective setting is presented. The formulation includes implied intra-criteria and inter-criteria preferences.

2.3.6 Constrained Multiobjective Metabolic Flux Analysis Framework

This section first presents the procedure for forming the LPP problem model. The LPP application procedure entails four major steps:

- 1) Determining for each objective the appropriate class, i.e., which one of the four *soft* and *hard* classes applies.

- 2) Defining the limits of the ranges of differing degrees of desirability for each objective: the 'target'-values (see Figure 2.3). For classes 1S through 4S, there are respectively five, five, nine, and ten such values. For classes 1H through 4H, these values are respectively $t_{i,max}$, $t_{i,min}$, $t_{i,val}$, and $t_{i,min}$ and $t_{i,max}$.
- 3) Generate weights based on the optimality requirement of cellular or physiological systems. Input in the form of range boundaries (or targets) for each objective are used in the Linear Physical Programming Weights (LPPW) algorithm to generate the weights (Table 2.4).

Table 2.4: Linear Physical Programming Weights algorithm

STEP	ACTION
1	Initialize: $\beta=1.1$; $w_{il}^+ = 0$, $w_{il}^- = 0$; $\tilde{z}^2 =$ small positive number (say, 0.1) $i = 0$; $s = 1$, $n_{sc} = \#$ of soft criteria
2	Set $i = i + 1$
3	Evaluate, in sequence \tilde{z}^s , \tilde{t}_{is}^+ , \tilde{t}_{is}^- , w_{is}^+ , w_{is}^- , \tilde{w}_{is}^+ , \tilde{w}_{is}^- , \tilde{w}_{min}
4	Set $s = s + 1$
5	If \tilde{w}_{min} is less than some chosen small positive number (say, 0.01), then increase β , and go to step 2.
6	If $s \neq 5$, go to step 3.
7	If $i \neq n_{sc}$, go to step 2

('i' values correspond to *soft* criteria)

- 4) The following linear programming problem is then solved:

Piecewise Archimedean aggregate function:

$$\min_{d_{is}^-, d_{is}^+, x} J = \sum_{i=1}^{n_{sc}} \sum_{s=2}^5 (\tilde{w}_{is}^- d_{is}^- + \tilde{w}_{is}^+ d_{is}^+) \quad (2.7)$$

subject to

System Constraints:

$$x_{\min} \leq x \leq x_{\max} \quad (2.8)$$

$$g_i = g_i(x) \quad (2.9)$$

Goal Constraints:

$$g_i - d_{is}^+ \leq t_{i(s-1)}^+, d_{is}^+ \geq 0, g_i \leq t_{i5}^+ \text{ (for all } i \text{ in classes 1S, 3S, 4S, } i=1,2,\dots,n_{sc}, s=2,\dots,5) \quad (2.10)$$

$$g_i + d_{is}^- \geq t_{i(s-1)}^-, d_{is}^- \geq 0, g_i \geq t_{i5}^-, \text{ (for all } i \text{ in classes 2S, 3S, 4S, } i=1,2,\dots,n_{sc}, s=2,\dots,5) \quad (2.11)$$

and

$$g_j \leq t_{j,\max} \text{ (for all } j \text{ in class 1H, } j=1, 2,\dots, n_{hc}) \quad (2.12)$$

$$g_j \geq t_{j,\min} \text{ (for all } j \text{ in class 2H, } j=1, 2,\dots, n_{hc}) \quad (2.13)$$

$$g_j = t_{j,\text{val}} \text{ (for all } j \text{ in class 3H, } j=1, 2,\dots, n_{hc}) \quad (2.14)$$

$$t_{j,\min} \leq g_j \leq t_{j,\max} \text{ (for all } j \text{ in class 4H, } j=1, 2,\dots, n_{hc}) \quad (2.15)$$

In the above formulation, x is the decision variable vector of the objective function $g_i(x)$, and n_{sc} and n_{hc} denote the number of soft and hard criteria, respectively. Next, we outline a simple algorithm for evaluating the weights that are used in the LPP model of the class functions. It is important to note that these weights are related to the class function slopes. The next section and Table 2.5 show the quantitative aspects of LPP. In the LPP implementation of this section, the final value of the parameter β was kept constant for all criteria resulting in a more favorable numerical conditioning. The increase of β in the weight-algorithm loop above was set as 0.01. Then the weights obtained from the above LPPW algorithm are used to obtain expressions for the piecewise linear class function of each criterion. To maintain the linear programming framework, piecewise linear class functions were implemented using deviation variables (d_{is}^-, d_{is}^+). In the particular case of class-4S, for example, it can be shown that the LP model of the piecewise linear function is as follows:

Piecewise Archimedean aggregate function:

$$z_i = \min_{d_{is}^-, d_{is}^+} \sum_{s=2}^5 (\tilde{w}_{is}^- d_{is}^- + \tilde{w}_{is}^+ d_{is}^+) \quad (2.16)$$

subject to

System Constraints:

$$S_u J_u = -S_m J_m \quad (2.17)$$

$$J_{low} \leq J_u \leq J_{high} \quad (2.18)$$

$$x_{min} \leq x \leq x_{max} \quad (2.19)$$

Goal Constraints:

$$g_i - d_{is}^+ \leq t_{i(s-1)}^+, d_{is}^+ \geq 0, g_i \leq t_{i5}^+ (s = 2, \dots, 5) \quad (2.20)$$

$$g_i + d_{is}^- \geq t_{i(s-1)}^-, d_{is}^- \geq 0, g_i \geq t_{i5}^- (s = 2, \dots, 5) \quad (2.21)$$

$$g_j \leq t_{j,max} \text{ (for all } j \text{ in class 1H, } j=1, 2, \dots, n_{hc}) \quad (2.22)$$

$$g_j \geq t_{j,min} \text{ (for all } j \text{ in class 2H, } j=1, 2, \dots, n_{hc}) \quad (2.23)$$

$$g_j = t_{j,val} \text{ (for all } j \text{ in class 3H, } j=1, 2, \dots, n_{hc}) \quad (2.24)$$

$$t_{j,min} \leq g_j \leq t_{j,max} \text{ (for all } j \text{ in class 4H, } j=1, 2, \dots, n_{hc}) \quad (2.25)$$

where J_m and J_u indicate measured and unmeasured fluxes, respectively; and S_m and S_u contain the stoichiometric coefficients of measured and unknown reactions, respectively.

As discussed previously, the use of preferences for different degrees of desirability for each design metric also aids in simultaneous visualization of a large number of objectives. This facilitates the assessment of the effect of preference specifications on the objectives as well as the complex interplay of these objectives (Figure 2.1D). Each section is color coded according to the desirability level and labeled.

2.3.7 Quantitative Aspects of Linear Physical Programming

Intra-Criteria and Inter-Criteria Preference: For a given objective, once the design metric decides to which class the criterion belongs, and chooses the range targets (i.e. t_{is}^+ , t_{is}^-), the *intra-criterion* preference statement is complete. However, since decisions are made in a multi-objective environment, there must be an implicit or explicit *inter-criteria* preference. Linear physical programming operates within an implied inter-criteria heuristic rule that is hereby called

the *One vs. Others* criteria-rule (OVO rule). The OVO-rule entails the following inter-criteria preference for each soft criterion, g_i . If two options are considered, viz.,

Option 1: "Full improvement of g_i across a given range (say, range-3)"; and

Option 2: "Full reduction of all the other criteria across the next better range (i.e., range-2)"; then option 1 shall be preferred over option 2.

That is, the worst candidate is always helped first. Essentially, this philosophy has a built-in preemptive nature in which the minimization of the worst criterion is automatically implied. Say, for example that we are dealing with 15 criteria. The OVO rule states that: it is more beneficial for a single criterion to improve over the full *tolerable* range, than it is for the other 14 to improve over the full *desirable* range. The mathematical implication of the OVO rule is discussed later.

Development of Class Function: In this section we define and discuss the properties of class functions (Figures 2.3C-F), and develop the LP model representation of generic class functions. This model will be used in the statement of the linear physical programming mathematical model. We also discuss in some detail the structure of a generic class function. We first state the class function properties; which is followed by a discussion thereof. These properties are as follows:

1. A lower value of a class function is preferred over a higher value thereof.
2. A class function is strictly positive.
3. A class function is continuous, piecewise linear, and convex.
4. The value of a class function, z_i , at a given ranges-intersection (say, *desirable-tolerable*) is the same for any class-type.
5. The magnitude of the class-function's vertical excursion across any range must satisfy the OVO-rule.

Collectively, these properties provide the flexibility required of the linear physical programming methodology. We make the following observations:

- Since a lower value of the class function is better than a higher value thereof, *class functions* and *utility functions* have distinct structures (while fulfilling the same function).

- The positivity property of the class function allows the DM to define an *ideal* finite value, zero.
- The rationale for the convexity requirement is partially rooted in the axioms of utility theory. In our particular case, it means that the cost of traveling across, say, the *undesirable* range is always more than that of traveling across the *tolerable* range; regardless of the target values chosen by the design metric. (We note that the OVO rule, alone, does not guaranty convexity).
- The value of the class function at a given range limit (say, $z_i(t_{i2}^+)$) is always constant (see Figure 2.3). From criterion to criterion, only the *location* of the limits (say t_{i2}^+) change, not the corresponding z_i values. As a consequence, as one *travels* across a given range-type (say *tolerable*), the change in the class-function will always be of the same magnitude, \tilde{z}^3 , regardless of the criterion in question. This behavior of the class function values at the boundaries is the critical factor that makes each range-type have the same numerical *cost/value* for different criteria. This same behavior also has a *normalizing effect*, and results in favorable numerical conditioning properties.

Concept 1: Table 2.5 represents the mathematical representations of some of the Property (4) of the class function discussed above is expressed by the relation

$$z^s \equiv z_i(t_{is}^+) \equiv z_i(t_{is}^-) \quad \forall i \quad ; \quad (2 \leq s \leq 5) ; z^1 \equiv 0 \quad (2.26)$$

where ‘s’ and ‘i’ respectively denote a generic range-intersection and criterion number.

Concept 2: The change in z_i that takes place as one *travels* across the s -th range is always given by

$$\tilde{z}^s \equiv z^s - z^{s-1} \quad ; \quad (2 \leq s \leq 5) ; z^1 \equiv 0 \quad (2.27)$$

Concept 3: To enforce the OVO rule, we apply the relationship

$$\tilde{z}^s > (n_{sc} - 1)\tilde{z}^{s-1} \quad (3 \leq s \leq 5) ; (n_{sc} > 1) \quad (2.28)$$

or, equivalently

$$\tilde{z}^s = \beta (n_{sc} - 1)\tilde{z}^{s-1} \quad (3 \leq s \leq 5) ; n_{sc} > 1 ; \beta > 1 \quad (2.29)$$

where n_{sc} denotes the number of *soft* criteria, and β will be used as a *convexity parameter*. Equation (2.29) does not guaranty convexity of the class function, as the said convexity depends also on the targets chosen. To apply Equation (2.29), we need to be given a value for \tilde{z}^2 . In practice, a small positive number will be appropriate (say, 0.1).

Table 2.5: Quantitative aspects of Linear Physical Programming

CONCEPT	FORMULATION
1	$z^s \equiv z_i(t_{is}^+) \equiv z_i(t_{is}^-) \quad \forall i \quad ; \quad (2 \leq s \leq 5) ; z^1 \equiv 0$
2	$\tilde{z}^s \equiv z^s - z^{s-1} \quad ; \quad (2 \leq s \leq 5) ; z^1 \equiv 0$
3 (OVO rule enforcement)	$\tilde{z}^s > (n_{sc} - 1)\tilde{z}^{s-1} \quad (3 \leq s \leq 5) ; (n_{sc} > 1)$ or, equivalently, $\tilde{z}^s = \beta (n_{sc} - 1)\tilde{z}^{s-1} \quad (3 \leq s \leq 5) ; n_{sc} > 1 ; \beta > 1$ where n_{sc} denotes the number of <i>soft</i> criteria, and β will be used as a <i>convexity parameter</i> .
4 (Convexity requirement)	Define $\tilde{t}_{is}^+ = t_{is}^+ - t_{i(s-1)}^+ ; \tilde{t}_{is}^- = t_{is}^- - t_{i(s-1)}^- ; (2 \leq s \leq 5)$ The <i>magnitude</i> of the slopes of the class function of the i -th criterion is: $w_{is}^+ = \tilde{z}^s / \tilde{t}_{is}^+ ; w_{is}^- = \tilde{z}^s / \tilde{t}_{is}^- ; (2 \leq s \leq 5)$ $\tilde{w}_{min} = \min_{i,s} \{ \tilde{w}_{is}^+, \tilde{w}_{is}^- \} > 0 \quad ; \quad \left\{ \begin{array}{l} (2 \leq s \leq 5) \\ i : \text{soft criteria} \end{array} \right.$ where $\tilde{w}_{is}^+ = w_{is}^+ - w_{i(s-1)}^+ ; \tilde{w}_{is}^- = w_{is}^- - w_{i(s-1)}^- ; w_{i1}^+ = w_{i1}^- = 0 \quad \left\{ \begin{array}{l} (2 \leq s \leq 5) \\ i : \text{soft criteria} \end{array} \right.$

Concept 4: Turning our attention to the convexity requirement, we define

$$\tilde{t}_{is}^+ = t_{is}^+ - t_{i(s-1)}^+ ; \tilde{t}_{is}^- = t_{is}^- - t_{i(s-1)}^- ; (2 \leq s \leq 5) \quad (2.30)$$

which is the length of the s -th range of the i -th criterion. With this quantity, the *magnitude* of the slopes of the class function of the generic i -th criterion takes the form

$$w_{is}^+ = \tilde{z}^s / \tilde{t}_{is}^+ ; w_{is}^- = \tilde{z}^s / \tilde{t}_{is}^- ; (2 \leq s \leq 5) \quad (2.31)$$

We emphasize here that the slopes change from range to range and from criterion to criterion. Once the slopes are known, the convexity requirement can be verified by the relationship

$$\tilde{w}_{min} = \min_{i,s} \{ \tilde{w}_{is}^+, \tilde{w}_{is}^- \} > 0 \quad ; \quad \left\{ \begin{array}{l} (2 \leq s \leq 5) \\ i : \text{soft criteria} \end{array} \right. \quad (2.32)$$

where

$$\tilde{w}_{is}^+ = w_{is}^+ - w_{i(s-1)}^+ \quad ; \quad \tilde{w}_{is}^- = w_{is}^- - w_{i(s-1)}^- \quad ; \quad w_{i1}^- = w_{i1}^+ = 0 \quad \left\{ \begin{array}{l} (2 \leq s \leq 5) \\ i : \text{soft criteria} \end{array} \right. \quad (2.33)$$

We note here that the quantities \tilde{w}_{is}^+ and \tilde{w}_{is}^- are exactly the weights that will be used in the LP model of the class functions. In effect, equation (2.32) states that so long as all these weights are positive, the class function will be piecewise linear and convex. The important point here is to observe that convexity can always be satisfied by simply increasing the magnitude of the *convexity parameter* β .

2.3.8 Generation of Pareto Points Using LPPFBA

This section develops a simple Linear Physical Programming based Pareto frontier generation method. In LPPFBA approach, preferences regarding each cellular objective can be expressed by providing numerical values that are associated with ranges of differing desirability. In order to obtain the Pareto optimal solutions, preference values have to span the objective space, to generate all the combinations of actual preferences that can result in corresponding Pareto points. We generate Pareto points in a region where a particular objective function is small by simply expressing small pseudo-preference values for that objective function. In a similar fashion, we can generate Pareto points in a region where a certain objective function is large by simply choosing large pseudo-preference values for that design metric. The steps for Pareto generation are presented below:

Step 1-Definition of the objective space of interest: First a hypercube that defines the region of interest in the objective space is generated. In particular, the minimum and maximum values associated with each design metric are defined. For the i -th objective function, $g_{i,\min}$ and $g_{i,\max}$ respectively denote the associated minimum and maximum objective function values in the Pareto frontier, or in the desired region of investigation; where it is to be noted that the region of

investigation must be a subset of the Pareto frontier. For the i th objective function, we also define

$$d_i = g_{i,\max} - g_{i,\min} \quad (2.34)$$

Step 2-Tradeoff matrix construction: Define the Tradeoff matrix G , as follows

$$G = \begin{bmatrix} g_{1,\min} & g_{12} & \cdots & g_{1N} \\ g_{21} & g_{2,\min} & \cdots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N1} & g_{N2} & \cdots & g_{N,\min} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \quad (2.35)$$

where N is equal to the number of objective functions, n_{sc} . In the tradeoff matrix G above, the i th row, r_i , represents the set of objective function values that are obtained when $g_i = g_{i,\min}$.

Step 3-Diagonal translation of AOF surface: The generation of the Pareto points involves translating the AOF surface across the objective space. A pseudo-preference translation vector is formed as

$$S'_i = \gamma_i g_{i,\min} + (1 - \gamma_i) g_{i,\max} \quad (2.36)$$

$$\text{where } \sum_{i=1}^N \gamma_i = 1, \quad 0 \leq \gamma_i \leq 1 \quad (2.37)$$

The parameter γ_i varies between 0 and 1, and the i -th component of the translation vector varies between the minimum and maximum values of the i -th objective function. The number of values of γ_i is dictated by the resolution with which we wish to generate the Pareto frontier. We let n , the *point-density parameter* (which represents the resolution of the Pareto frontier), denote the number of γ_i values for each objective function. Accordingly, the γ_i increments are $1/(n - 1)$. Figure 2.4 depicts the scenarios for two and three soft objective function cases, with $n = 6$. In the case of two design metrics, we have a total of 6 translation vectors, which will yield 6 Pareto solutions. In the case of three objective functions, we have 21 translation vectors yielding 21 Pareto solutions. The AOF surface is shifted through the objective space. With each shift, an optimization is performed resulting in a Pareto point. In essence, translating the AOF surface in the objective space generates the entire Pareto frontier.

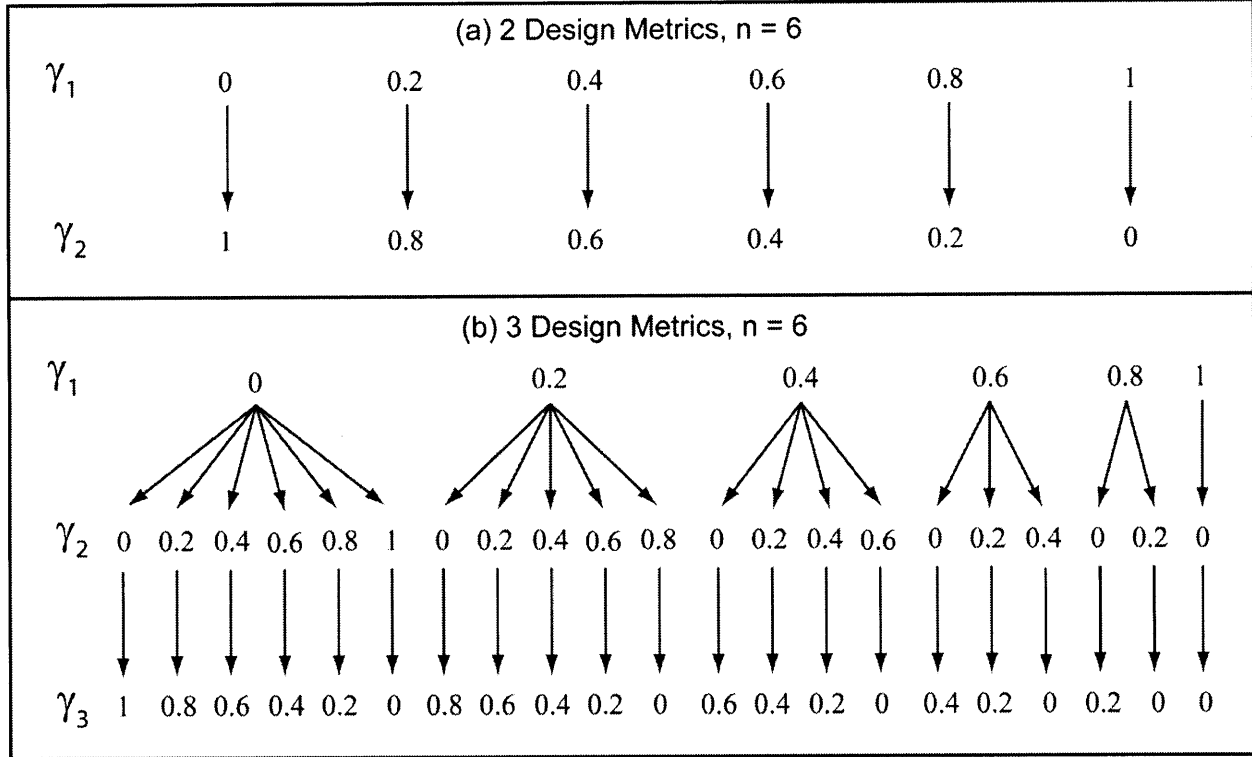


Figure 2.4: Non-dimensional parameters used in the generation of the Pareto optimal solutions for translating the AOF surface across the objective space.

Step 4-Offset in diagonal translation of AOF surface, and its magnification: In Step 3, we showed how the AOF surface could be translated, with the intent of generating the Pareto frontier. However, to avoid missing any Pareto solution if hypercube is too small we need following adjustments. First, we can offset the translation trajectory by replacing it with another that is parallel to the original. Alternatively, we can magnify the box size so as to overlap all regions of the objective space. First, we define the box size as

$$a_i = d_i / n_d \quad (2.38)$$

where n_d is a real positive number that defines the relative size of the hypercube of interest and the AOF. By letting n_d be a number less than or equal to two when there is no offset, we can ensure that all Pareto points can be generated. Second, we define the offset vector as

$$S_i^f = \alpha_i d_i, \quad -1 \leq \alpha_i \leq 1 \quad (2.39)$$

which is used to offset the translation trajectory, as discussed above.

Step 5-Formation of pseudo-preference vector: The pseudo-preference vector as follows. As discussed above, we directly use these values of preference input in the LPP algorithm to generate Pareto points. We define

$$S_i^p = -(n_p - 1)\frac{\alpha_i}{4}, \quad n_p = 1, \dots, 5 \quad (2.40)$$

and

$$P_i^0 = \left\{ 0 \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{3}{4} \quad 1 \right\} a_i \quad (2.41)$$

which yield the pseudo-preference vector as

$$P_i = \begin{Bmatrix} g_{i1} \\ g_{i2} \\ g_{i3} \\ g_{i4} \\ g_{i5} \end{Bmatrix}^T = (S_i^t + S_i^f + S_i^p)E + P_i^0 \quad (2.42)$$

where

$$E = \{1 \quad 1 \quad 1 \quad 1 \quad 1\} \quad (2.43)$$

The variable S_i^p is utilized to implement an additional offset of the translation trajectory. Also following constraints have to be satisfied in order for the pseudo-preferences to effectively impact the solution:

$$|S_i^f + S_i^p| \leq \frac{d_i}{2} \quad (2.44)$$

which leads to the requirement

$$-\frac{1}{2} + \frac{n_p - 1}{4n_d} \leq \alpha_i \leq \frac{1}{2} + \frac{n_p - 1}{4n_d} \quad (2.45)$$

2.4 RESULTS AND DISCUSSION

Multiobjective optimization is a useful tool with applications to numerous disciplines and more recently for cellular systems (Nagrath et al., 2007; Vo et al., 2004) where simultaneous targeting of several objectives is vital. Therefore, we developed this approach to optimize hepatocellular function in the context of a BAL device, in which case the main goal is for the hepatocytes to function at the highest possible level. Here we focused on a limited set of critical representative hepatocellular metabolic processes: urea secretion, albumin synthesis, NADPH synthesis, and glutathione synthesis. Urea secretion (flux 16 in Table 2.1) is a critical detoxification reaction, and is primarily derived from ammonia and aspartate generated through transamination reactions. Albumin synthesis (flux 47) was used as a marker of liver specific protein secretion. NADPH is produced by the pentose phosphate pathway (PPP), and is primarily used in nonproliferating hepatocytes as a co-factor for cytochrome p450 dependent oxidation reactions, de novo synthesis of glutathione, as well as reduction of oxidized glutathione. To increase NADPH production, the NADPH-generating oxidative branch of the PPP represented in a lumped fashion as flux 46 was increased. The tripeptide glutathione (GSH, γ -Glu-Cys-Gly) is a free radical scavenger and is involved in many detoxification reactions. The synthesis of glutathione is represented by flux 48.

Using LPPFBA, we assessed the sensitivity and geometry of the optimal region and determined the optimal results using various preferences and/or prioritization of the four objectives (fluxes 16, 46, 47, 48) mentioned above. The constraints for the hepatic metabolic network are listed in Table 2.6. As described in the Methods, LPPFBA requires characterization of design metrics into different classes. Here, all four objective functions are maximized and hence fall in “Class 2S”. As part of this analysis, we first obtained Pareto frontiers between various bi-objective combinations of liver specific functions (albumin synthesis, urea secretion, NADPH synthesis, and GSH synthesis). Next, we obtained Pareto optimal solutions for tri-objective combinations of these hepatic functions. Lastly, we obtained the Pareto solutions for a simulated BAL system where the main goal was to operate the BAL at the highest possible level during human plasma exposure. The experimentally measured flux data for hepatocytes during plasma exposure were obtained from Chan et al. (Chan et al., 2002).

Table 2.6: Lower and upper bounds for the metabolic network (Table 2.1)

Flux	lower bound	upper bound
1	0.1	30
2	0.00001	30
3	0.00001	30
4	0.00001	30
5	0.00001	30
6	0.00001	30
7	0.00001	30
8	5	30
9	5	30
10	5	30
11	5	30
12	5	30
13	5	30
14	0.00001	50
15	0.00001	50
16	0.00001	100
17	-10	50
18	0.1	50
19	0.1	50
20	0.01	50
21	0.001	50
22	-1	50
23	0.001	50
24	0.00001	50
25	0.00001	50
26	0.001	50
27	0.1	50
28	0.1	50
29	0.1	50
30	0.00001	50
31	0.01	50
32	0.01	50
33	0.01	50
34	0.01	50
35	-10	50
36	0.00001	50
37	-10	-0.01
38	0.00001	50
39	0.1	50
40	0.01	50
41	-10	50
42	0.1	50
43	5	100
44	5	100
45	-10	-0.001
46	0.01	10
47	0.00001	5
48	0.001	50
49	0.001	50
50	10	50
51	0.15	20
52	0.1	50
53	1	10
54	0.01	5
55	0.01	5
56	0.01	5
57	0.01	2
58	0.5	2
59	5	100
60	0.01	5
61	-1	5
62	0.0001	10
63	-1	5
64	0.00001	10
65	0.1	20
66	-1	5
67	-1	5
68	-1	5
69	-1	10
70	0.001	10
71	0.01	10
72	0.01	10
73	0.01	5
74	0.01	5
75	0.01	5
76	0.01	5
77	0.01	5
78	0.00001	10
79	0.5	10
80	0.00001	10
81	0.00001	5

Table 2.7: Optimal fluxes obtained for the corresponding Pareto optimal points A, B, C, D, E, F, G, H as for Pareto frontiers shown in Figure 2.5.

Flux	Albumin/Urea		NADPH/Albumin		NADPH/Urea		GSH/Albumin	
	A	B	C	D	E	F	G	H
1	7.183	10.709	3.983	3.793	3.946	7.543	2.769	2.400
2	7.183	10.709	3.983	3.793	3.946	7.543	2.769	2.400
3	7.183	10.709	3.983	3.793	3.946	7.543	2.769	2.400
4	14.391	21.446	7.968	7.586	7.894	15.161	5.540	4.803
5	14.392	21.476	30.000	30.000	30.000	30.000	12.515	7.888
6	14.392	21.476	30.000	30.000	30.000	30.000	12.515	7.888
7	1.302	9.031	29.506	30.000	25.000	25.000	13.463	14.214
8	5.145	7.555	15.000	20.022	15.000	15.000	10.441	10.950
9	5.145	7.555	15.000	20.022	15.000	15.000	10.441	10.950
10	5.000	5.000	5.000	10.036	5.000	5.000	5.628	5.001
11	5.182	5.031	11.815	17.638	9.967	5.387	10.583	9.586
12	28.235	30.000	25.494	29.987	30.000	30.000	10.724	9.700
13	28.235	30.000	25.494	29.987	30.000	30.000	10.724	9.700
14	22.953	24.869	13.579	12.231	19.933	24.513	0.030	0.004
15	22.953	24.869	13.579	12.231	19.933	24.513	0.030	0.004
16	29.700	34.400	13.115	11.232	25.100	34.400	4.290	3.350
17	-8.903	-1.330	-1.740	-8.712	-1.054	-0.792	-1.013	-3.254
18	0.104	0.167	20.675	19.002	15.954	15.692	0.111	0.100
19	0.100	0.192	0.100	0.100	0.100	0.100	0.100	0.100
20	0.032	0.014	7.425	4.684	9.704	9.844	7.636	5.548
21	0.001	0.001	1.915	0.894	0.001	0.001	0.001	0.014
22	-1.000	-0.910	-0.997	-1.000	0.009	2.331	-1.000	-0.572
23	0.001	0.030	22.032	22.414	22.106	14.839	6.975	3.084
24	0.364	0.510	0.480	0.462	0.491	0.495	0.489	0.375
25	0.182	0.031	6.815	7.602	4.967	0.387	4.955	4.586
26	0.022	0.004	0.001	0.027	0.848	2.927	0.001	1.259
27	1.411	0.102	0.426	0.100	0.248	0.100	0.100	0.103
28	0.100	0.100	0.100	0.118	0.100	0.100	0.111	0.109
29	0.109	0.104	0.100	0.100	0.100	0.100	0.165	0.221
30	2.930	4.696	0.357	2.990	4.792	0.000	4.831	2.990
31	0.161	0.010	4.880	4.196	4.946	0.366	4.934	4.196
32	0.010	0.010	0.010	0.277	0.010	0.010	0.010	0.291
33	0.010	0.010	0.010	2.235	0.010	0.010	0.010	0.085
34	2.347	0.018	0.010	0.010	0.010	0.010	0.010	0.021
35	-10.000	-10.000	-10.000	-9.965	-10.000	-10.000	-1.231	-5.076
36	7.289	9.609	0.310	7.320	9.821	9.906	0.019	0.083
37	-0.145	-2.555	-10.000	-9.985	-10.000	-10.000	-4.813	-5.949
38	2.978	4.713	0.000	0.000	0.167	4.887	0.075	0.256
39	6.562	0.307	9.500	6.650	4.984	0.382	4.227	6.628
40	0.329	0.959	0.970	1.604	0.408	0.320	0.328	0.322
41	-0.265	-0.413	-0.378	-0.334	-1.223	-1.564	-0.285	-1.532
42	0.121	0.101	0.103	0.155	0.116	1.858	0.205	0.102
43	26.053	28.997	33.527	47.726	52.950	55.606	43.906	49.837
44	5.000	5.000	5.005	5.035	13.197	5.689	13.394	13.866
45	-0.025	-0.027	-0.001	-0.001	-0.001	-0.074	-0.002	-0.003
46	0.439	0.710	2.983	2.793	2.946	1.783	1.766	1.399
47	0.136	0.020	0.020	0.134	0.009	0.005	0.011	0.134
48	0.071	6.497	14.080	9.406	14.533	10.102	14.449	9.406
49	0.135	0.329	0.319	0.135	0.145	0.184	0.136	0.114
50	10.000	10.001	10.471	19.610	10.000	10.000	14.265	17.268
51	2.561	0.211	0.154	0.215	0.204	1.967	0.268	0.180
52	0.100	0.291	0.307	0.110	0.123	0.100	0.100	0.100
53	6.744	9.999	1.000	1.000	1.000	5.760	1.003	1.001
54	0.010	0.011	0.012	0.024	0.021	0.010	0.034	0.011
55	0.014	0.043	0.025	0.637	0.019	0.010	0.014	0.011
56	0.014	0.043	0.025	0.637	0.019	0.010	0.014	0.011
57	0.035	0.038	0.013	0.025	0.022	0.084	0.036	0.014
58	0.500	0.530	0.500	0.596	0.500	0.500	0.500	0.509
59	23.524	26.182	35.490	54.993	36.042	37.873	30.526	33.412
60	4.175	0.691	0.700	4.120	0.368	0.241	0.495	4.241
61	4.928	4.765	4.737	2.681	5.000	5.000	-0.965	-1.000
62	10.000	10.000	0.710	10.000	10.000	10.000	0.239	2.763
63	3.769	4.818	-0.464	-0.999	5.000	5.000	4.185	3.089
64	10.000	10.000	0.016	2.217	5.382	10.000	4.525	6.561
65	20.000	19.342	20.000	20.000	20.000	20.000	0.132	0.174
66	-1.000	-0.648	-1.000	-1.000	-1.000	-1.000	-0.831	4.545
67	4.517	1.526	5.000	5.000	-1.000	-1.000	-0.666	5.000
68	0.343	4.995	5.000	5.000	5.000	5.000	5.000	4.992
69	10.000	5.885	3.909	-0.766	0.398	4.757	-0.868	-0.867
70	0.007	0.008	0.058	0.068	0.023	0.001	0.059	0.053
71	4.505	0.660	10.000	10.000	10.000	10.000	8.000	9.984
72	7.205	1.040	1.061	7.129	1.322	3.176	0.584	8.361
73	4.935	0.611	0.946	3.584	0.480	0.222	0.386	3.587
74	4.963	4.989	0.657	5.000	4.926	0.071	4.996	5.000
75	0.975	0.127	5.000	5.000	5.000	0.394	5.000	5.000
76	4.753	0.694	0.710	4.967	0.323	0.175	0.395	4.981
77	1.772	0.264	0.270	3.977	0.126	0.071	0.153	1.827
78	9.937	1.112	1.130	7.514	0.511	0.273	0.626	7.525
79	9.950	0.795	10.000	10.000	5.208	0.500	4.502	9.978
80	4.753	7.363	10.000	10.000	10.000	10.000	10.000	10.000
81	1.535	0.408	0.094	2.832	0.040	0.099	0.242	2.820

2.4.1 Analysis of Bi-Objective Hepatic Metabolic Network

Pareto frontiers for various sets of bi-objectives were generated in this section to identify potential optimal solution regions. These optimal solutions can provide a qualitative framework to assess the tradeoffs and robustness of the hepatic metabolic network for a quad-objective scenario (albumin synthesis, NADPH synthesis, urea secretion and GSH synthesis). The representative results are shown in Figure 2.5. These Pareto optimal solutions were obtained by changing the preferences from higher desirable values to highly undesirable values.

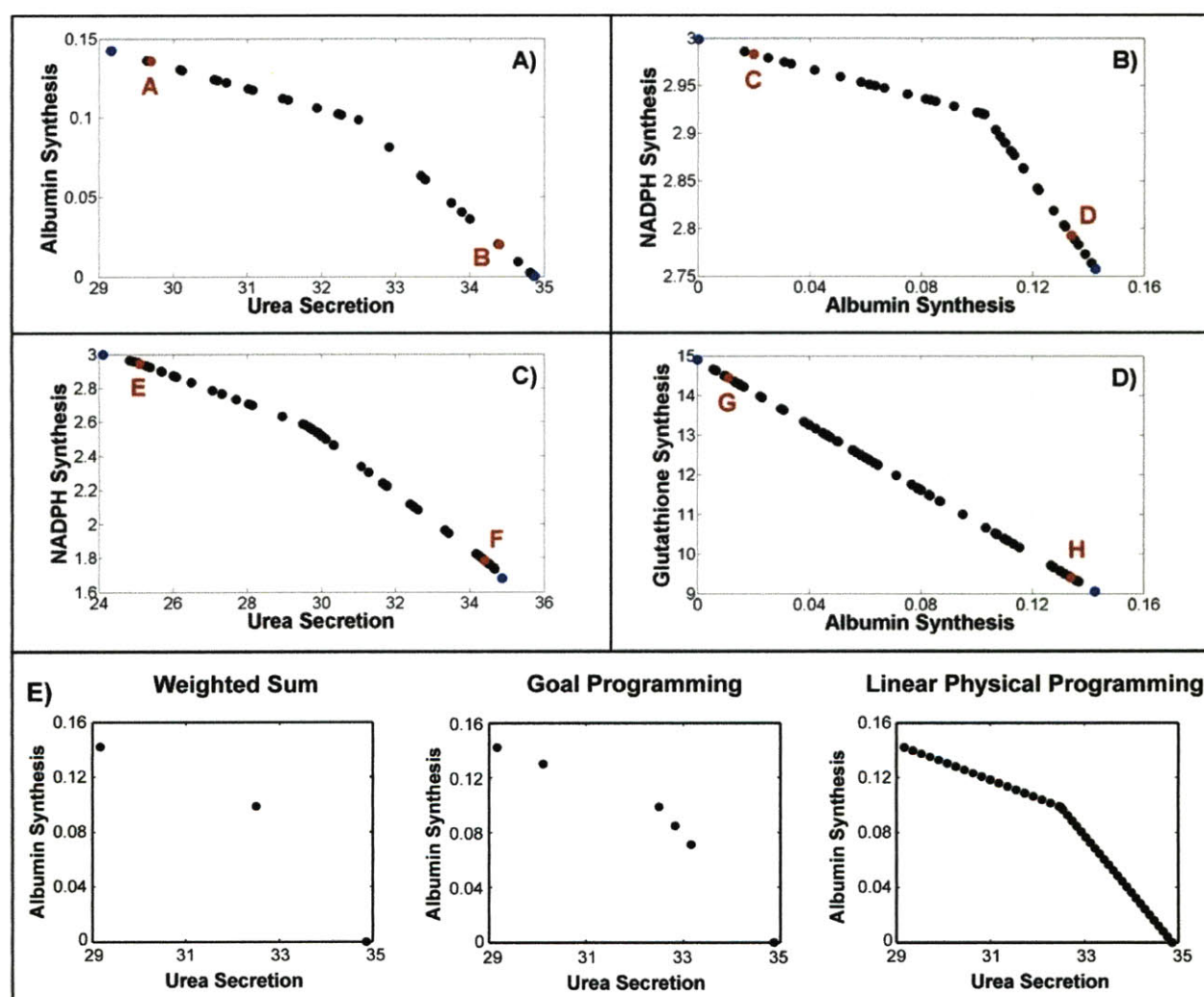


Figure 2.5: Pareto frontiers for bi-objective hepatic systems. Four major hepatic functions of albumin synthesis, urea secretion, NADPH and glutathione synthesis were used for bi-objective optimality in the combinations shown here: **A)** Pareto frontier between albumin and urea synthesis. **B)** Pareto frontier between NADPH and albumin synthesis. **C)** Pareto frontier between NADPH synthesis and urea secretion. **D)** Pareto frontier between glutathione synthesis and albumin synthesis. The blue circles are the Anchor points, black circles are Pareto optimal solutions for optimization and red circles are selected Pareto solutions for which complete set of optimal fluxes are shown in Table 2.7. A, B indicates Pareto points for albumin and urea bi-objective system; C, D indicates Pareto points for

NADPH and albumin system; E, F indicates Pareto points for NADPH and urea bi-objective system.; and G, H indicates Pareto points for glutathione and albumin bi-objective system. E) Comparison of Pareto optimal solutions obtained by LPPFBA weighted-sum and goal programming for a bi-objective problem of maximization of albumin synthesis and urea secretion.

As seen in Figure 2.5, all of these objectives exhibited a tradeoff with each other; for example, albumin and urea synthesis could not be at their maximal values at the same time (Figure 2.5A). Similarly, there was a tradeoff between NADPH and albumin synthesis, NADPH synthesis and urea secretion, glutathione and albumin synthesis (Figures 2.5B, 2.5C and 2.5D, respectively). In addition, the impact of changing preferences (for example favoring albumin synthesis over urea secretion, and vice-versa) varied depending on the objective. In particular, the tradeoff region or range of Pareto optimal solutions (i.e., how far the optimal value is from the “anchor value”) for albumin synthesis was very high compared to NADPH synthesis and urea secretion. Several other combinations were also tested and all of them indicated Pareto optimality between various objectives (data not shown). Figure 2.5E compares the Pareto optimal solutions obtained between albumin synthesis and urea secretion using weighted-sum, goal programming and LPPFBA. For all three equal number of simulations were used. As seen in these figures LPPFBA has significant advantage over both weighted sum and goal programming. This is because mapping of preferences to form an AOF in LPPFBA results in piecewise smooth hyper surfaces which leads to an even spread of Pareto optimal solutions for a given even spread of input preferences without missing any Pareto optimal solution. This behavior of optimal solution with respect to change in preference is highly desirable in large-scale mammalian metabolic network analysis (where tradeoffs between objectives are ubiquitous). In conventional methods, the spacing of points is largely dependent on relative scaling thus may lead to ill-conditioned problems. Importantly, these methods fail to capture significant number of optimal solutions resulting in an uneven distribution for even distribution of weights.

Figure 2.6 presents the distribution of Pareto optimal fluxes throughout the tradeoff region, which shows the changes required in flux values and direction (i.e. increasing or decreasing) as the objective preference is changed from one objective to another. The corresponding flux values are presented in Table 2.7.

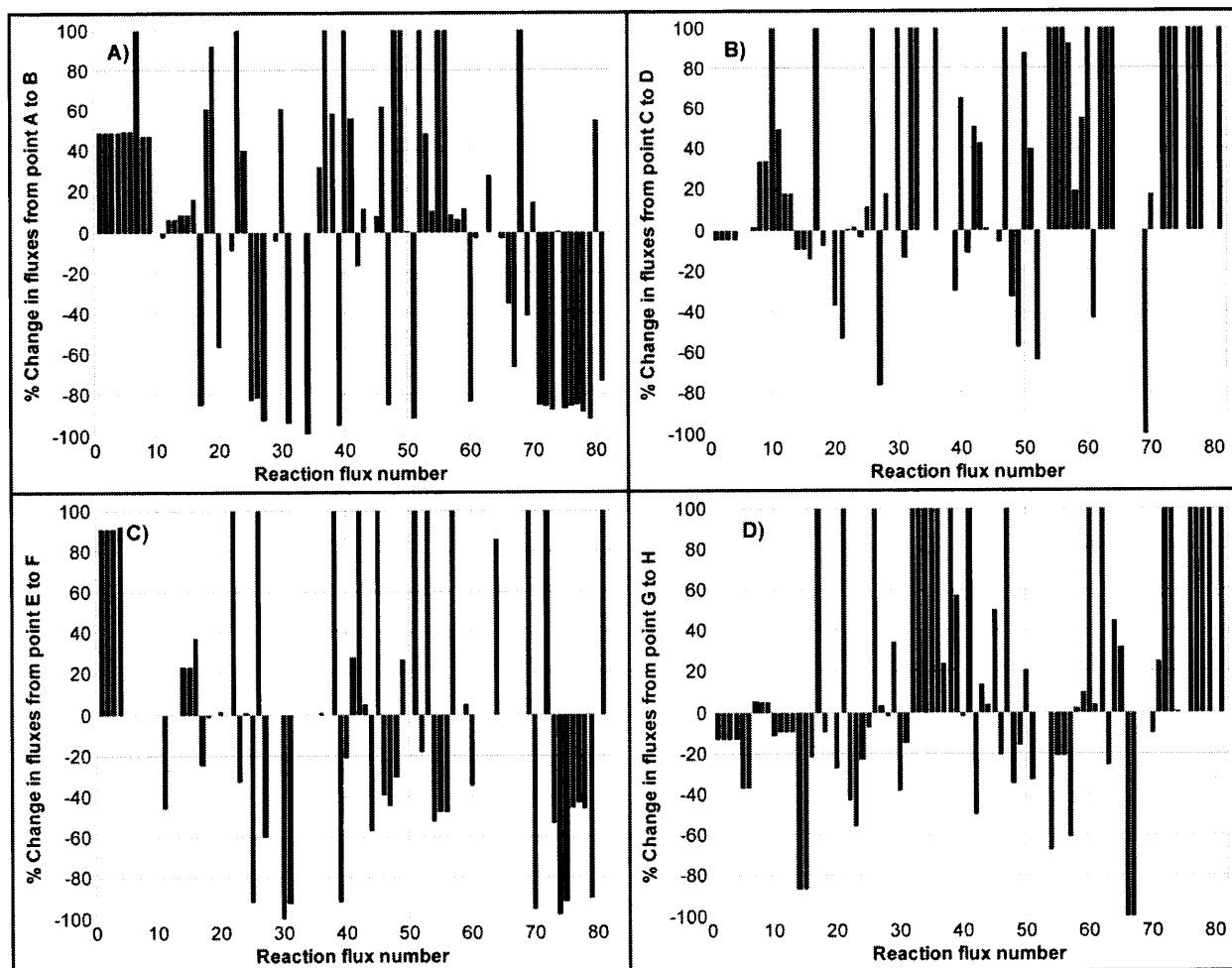


Figure 2.6: Metabolic profiling of percentage change in Pareto optimal fluxes for various solutions of Figure 2.5. These results are for various combinations of objectives and the corresponding flux values are in Table 2.7. Only changes up to $\pm 100\%$ are shown.

Figure 2.6A indicates the necessary change in fluxes when going from Pareto optimal solutions “A” to “B” in Figure 2.5A, in other words, when going from higher albumin synthesis/lower urea secretion rates to lower albumin synthesis/higher urea secretion rates. This change requires increasing gluconeogenic fluxes (1-9), formation of pyruvate from amino acids (fluxes 17-19, 23-24), aspartate synthesis (36), formation of glutamic acid (37, 39), and increasing oxidation of triglycerides (52). Noticeably, higher urea secretion/lower albumin synthesis necessitates decreased uptake of both glucogenic (proline, 60; serine, 67; aspartate, 69; threonine, 71; phenylalanine, 73; methionine, 75; valine, 76; isoleucine, 77; glutamine, 79; tyrosine, 81) and ketogenic (lysine, 72; leucine, 78) amino acids over the lower urea secretion/higher albumin synthesis case. On the contrary, glycine (68) and cysteine (80) uptake are increased when increasing urea secretion. Asparagine (62) and arginine (64) uptake rates

were at maximum for both urea and albumin maximizations. Histidine (74) uptake rate was also increased when increasing urea synthesis since it results in an increase of α -ketoglutarate. Essentially, the uptake of pyruvate forming amino acids (alanine, 66; serine, 67; and threonine, 71), fumarate forming amino acids (phenylalanine, 73; and tyrosine, 81), and succinyl CoA forming amino acids (threonine, 71; methionine, 75; and valine, 76) was decreased in order to increase urea secretion since these amino acids play a major role in increasing albumin synthesis. Also, increasing urea synthesis resulted in increased gluconeogenesis, which was associated with an increased rate of glucose clearance and an increase in glycogen synthesis.

Figure 2.6B indicates the change in optimal fluxes from C to D Pareto solutions for the case when NADPH production and albumin synthesis are considered to be the main objectives and whose maximization was studied. Noticeably, TCA cycle fluxes (8, 9, 10 and 11) were higher for the case when albumin was maximized. Additionally, oxygen uptake and electron transport system flux (59 and 43) were significantly lower at higher NADPH production and lower albumin synthesis. Moreover, NADPH use for alanine synthesis in reaction 17 was significantly reduced for the case of NADPH maximization. NADPH maximization also required increased aspartate uptake (69). Since the tradeoff region for NADPH synthesis is not large (2.793 to 2.983) there was not much change in gluconeogenic fluxes in this scenario. However, since the tradeoff region for albumin synthesis flux was high (from 0.02 to 0.134), increasing albumin synthesis required increasing the uptake rates of fatty acids (54-57), oxygen (59), proline (60), asparagine (62), arginine (64), lysine (72), phenylalanine (73) and histidine (74). Interestingly, histidine (30), which produces glutamate, was decreased for increasing NADPH production because reaction 37 is proceeding in reverse direction, which utilizes NADPH.

Figure 2.6C presents the flux profiles for Pareto optimal solutions E and F for the NADPH synthesis/urea synthesis bi-objective scenario. There was an increase in the tradeoff region for NADPH flux (1.783 to 2.946) when compared to the previous bi-objective case of NADPH/albumin synthesis. As seen in Figure 2.6C and Table 2.7, higher urea flux necessitated upregulation of gluconeogenic fluxes (1-4), ketogenic amino acid uptake (lysine: 26, 72), nonessential amino acid uptake (glutamate, 38), ketone body production (acetoacetate, 42; β -hydroxybutyrate, 51), glucose release (53), arginine uptake (64), aspartate uptake (69), and tyrosine uptake (81). On the contrary, increasing NADPH production required upregulation of the uptake of amino acids that produce succinyl CoA (methionine: 31, 75) to increase electron

transport chain fluxes (44), and uptake of amino acids that produce glutamate (histidine, 74, glutamine, 79). The increased glutamate resulted in higher α -ketoglutarate, which compensates for an increased fumarate production in TCA cycle through succinyl CoA (11).

Figure 2.6D presents the flux profiles for Pareto optimal solutions G and H for the GSH synthesis/albumin synthesis bi-objective scenario. Higher glutathione production flux upregulated the urea cycle fluxes (14, 15 and 16), increased serine production (22, 23), increased glutamate synthesis (23, 30), increased gluconeogenesis fluxes (5-6). However, the increased 3-phosphoglycerate flux (4) does not increase glyceraldehyde 3-phosphate flux (3) but instead result in the increased production of serine (23), which ultimately is used to produce glycine (22) for reduced glutathione. Also, there is a significantly reduced synthesis of alanine by alanine aminotransferase through pyruvate (17). Noticeably, higher albumin requires increased catabolism of lysine (26), valine (32), isoleucine (33), leucine (34), and ornithine (38) with an increased production of glutamate (38, 39), aspartate (36), pyruvate (50), and decreased ketone bodies (β -hydroxy butyrate, 51). Higher albumin flux also necessitates the increased uptake of amino acids (proline, 60; asparagine, 62; arginine, 64; alanine, 66; serine, 67; lysine, 72; phenylalanine, 73; valine, 76; isoleucine, 77; leucine, 78; glutamine, 79; tyrosine, 81).

As can be seen from the above results there is a significant re-routing of flux directions and cycle fluxes when switching from one objective to another within system constraints. In general, up regulation of gluconeogenesis was associated with higher urea secretion, which, in turn, was associated with higher arginine and aspartate fluxes. Increasing albumin synthesis required a significant increase in the uptake of various amino acids and the synthesis of some of the gluconeogenic amino acids. Interestingly, higher glutathione synthesis required an up regulation in glycine synthesis. It is important to note that the bi-objective cases analyses discussed in this section had preferences close to (although not exactly at) the anchor points.

2.4.2 Analysis of Tri-Objective Hepatic Metabolic Network

Next, as part of this analysis, we demonstrate the use of LPPFBA to perform tri-objective optimization. We studied various preferences for two different tri-objective combinations: NADPH production, albumin synthesis, and GSH synthesis and urea secretion, NADPH production, and albumin synthesis. The optimal results are presented in Tables 2.8 and 2.9,

respectively, and the corresponding fluxes are in Table 2.10. For each combination, we examined three cases, each case favoring one of the three functions. Figures 2.7A-B show the metabolic profiling for change in fluxes for NADPH synthesis/albumin synthesis/GSH synthesis scenario and Figures 2.7C-D show the metabolic profiling for change in fluxes for urea secretion/NADPH synthesis/albumin synthesis when preferences are changed from one objective to another. The anchor points which are obtained by individual optimization of urea, albumin, glutathione, and NADPH are 34.869, 0.14257, 14.9 and 2.9986, respectively.

Table 2.8: Linear Physical Programming optimization results for a tri-objective system (NADPH synthesis, albumin synthesis and glutathione synthesis) for hepatic metabolic network. HUD is highly undesirable, UD is undesirable, T is tolerable, D is desirable and HD is highly desirable preference values of objective functions.

Case #	Priority	Flux	t ₁ (HD)	t ₂ (D)	t ₃ (T)	t ₄ (UD)	t ₅ (HUD)	Optimal
1	High	NADPH	3.0000	2.7635	2.5270	2.2905	2.0539	2.8700
	High	albumin	0.1426	0.1313	0.1200	0.1087	0.0975	0.1149
	High	GSH	14.9000	13.7222	12.5444	11.3665	10.1887	10.1887
2	High	NADPH	3.0000	2.9439	2.8877	2.8316	2.7755	2.7795
	High	albumin	0.1426	0.1399	0.1372	0.1345	0.1319	0.1372
	Low	GSH	1.1280	0.8485	0.5690	0.2895	0.0100	9.2741
3	Low	NADPH	0.0130	0.0123	0.0115	0.0108	0.0100	2.6729
	High	albumin	0.14257	0.14253	0.14250	0.14246	0.14243	0.14257
	Low	GSH	0.0249	0.0212	0.0175	0.0137	0.0100	9.0546
4	High	NADPH	3.0000	2.9822	2.9645	2.9467	2.9290	2.9962
	Low	albumin	0.0034	0.0026	0.0017	0.0009	0.0000	0.0034
	Low	GSH	0.3637	0.2753	0.1869	0.0984	0.0100	14.7607
5	Low	NADPH	0.0226	0.0195	0.0163	0.0132	0.0100	2.8640
	Low	albumin	0.0006	0.0005	0.0003	0.0002	0.0000	0.0006
	High	GSH	14.9000	14.8843	14.8685	14.8528	14.8370	14.8749

Table 2.8 presents the multiobjective optimal solutions for 5 scenarios for the NADPH/albumin/GSH tri-objective case. Case 1 indicates the base case where preferences for all three hepatic objectives were given based on their anchor points, however, none of the objectives were given any specific priority, i.e., priority for each was set as equal to a high value. As seen from the Highly Desirable (HD) values for objective functions in Case 1, all objectives' highly desirable values are close to their anchor points. The preference ranges in the LPP optimization for this case were selected as 3-2.05 for NADPH production, 0.143-0.098 for albumin synthesis, and 14.9-10.19 for glutathione synthesis. The optimal values of objective

functions obtained for this case (Case 1) are 2.87 for NADPH synthesis (between highly desirable and desirable), 0.11 for albumin synthesis (between tolerable and undesirable) and 10.19 for glutathione synthesis (between undesirable and highly undesirable). Next, we present Cases 2-3, where higher priority is desired for albumin synthesis over objectives of glutathione and NADPH synthesis. In Case 2, preferences for albumin synthesis were increased as compared to Case 1 and preferences for glutathione synthesis were decreased significantly. This provides a higher priority for albumin synthesis and a lower priority for glutathione synthesis. Note that in Case 2, both higher albumin synthesis and NADPH synthesis were desired since the highly undesirable values of albumin and NADPH synthesis were increased closer to the highly desirable values. In this case, we see that the optimal value of albumin (0.137) lies between desirable and tolerable as compared to Case 1 where albumin was between tolerable and undesirable values. Inter-optimality or tradeoff between various objectives is clearly evident in Case 2 since albumin synthesis increased with a corresponding decrease in NADPH and GSH synthesis. To further increase the albumin synthesis, in Case 3 preferences for both NADPH and GSH were decreased and of albumin increased. This sets the priority of albumin high compared to NADPH and GSH. The obtained Pareto optimal values of albumin synthesis from LPPFBA are now similar to the highly desirable (HD) preference. As can be seen from the optimal values of NADPH (2.67) and GSH (9.05), the obtained values of NADPH and GSH decreased when compared to Cases 1 and 2 at the cost of an increase in optimal value of albumin. In Case 4, higher NADPH is desired and the priority for albumin and GSH synthesis is very low. To obtain multiobjective optimal solutions for this scenario, preference ranges were increased significantly for NADPH as compared to Case 1 and a low preference range was assigned to both albumin and GSH synthesis. The preference ranges for the undesired objectives need to be lowered in order to achieve a higher value for the desired objective. This is because highly undesirable values act as a hard constraint for the objective. Interestingly, optimal values obtained for all three objectives were close to their highly desirable (HD) preference values; however, when compared to Case 1, optimal albumin synthesis was significantly decreased at the cost of a marginal increase in NADPH synthesis and significant increase in GSH synthesis. This shows that albumin synthesis is highly sensitive as compared to NADPH and requires a significant decrease in its synthesis for other hepatic functions to increase. In Case 5, higher glutathione synthesis was desired; hence, preference ranges for GSH were increased significantly towards HD values. As seen in the

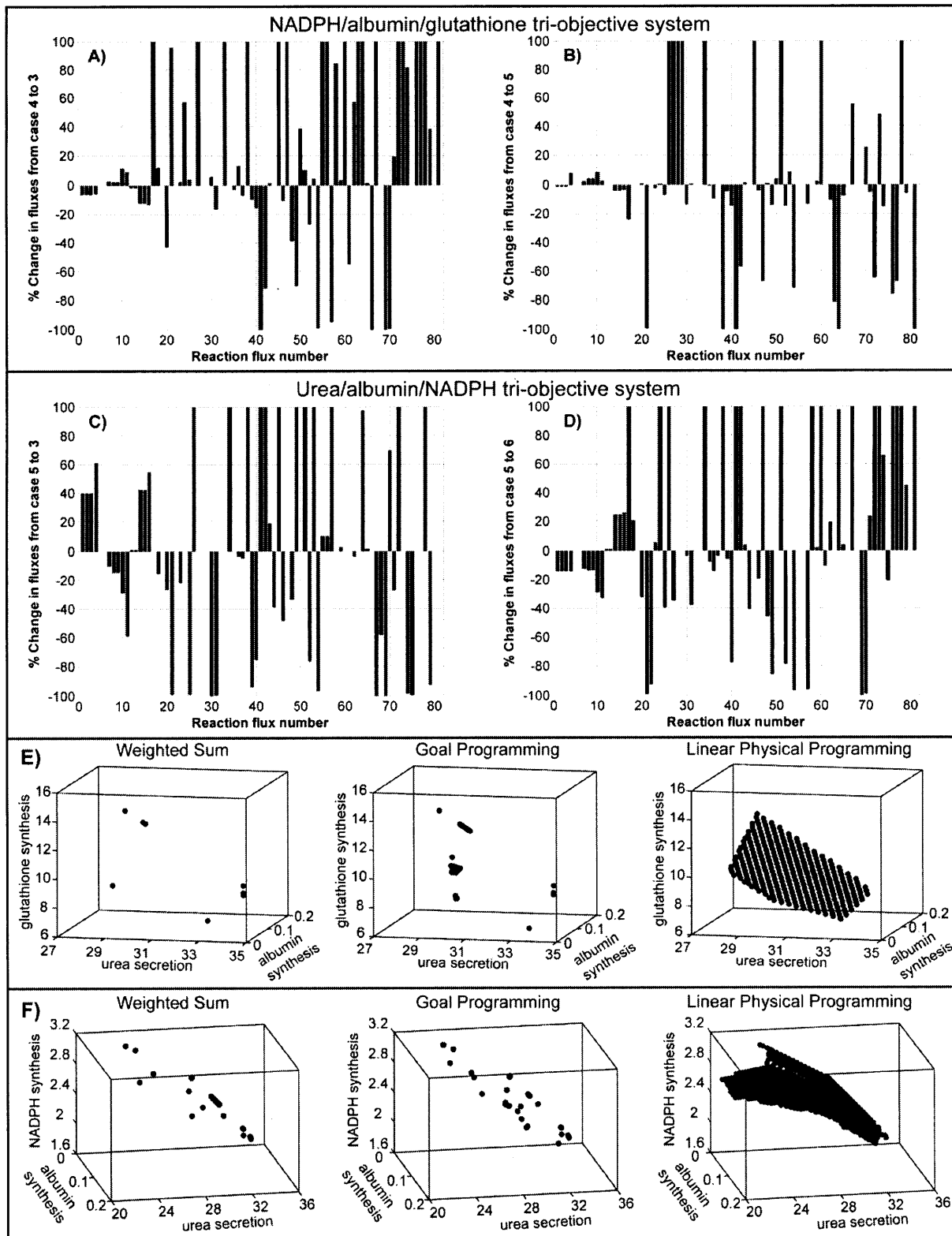


Figure 2.7: A-D) Metabolic profiling of percentage change in Pareto optimal fluxes for various solutions of Tables 2.8 and 2.9. These results are for various combinations of objectives and the corresponding flux values are in Table 2.10. Only changes up to $\pm 100\%$ are shown. E-F) Comparison of Pareto optimal solutions obtained by LPPFBA, weighted-sum and goal programming for two tri-objective cases for hepatic metabolism.

obtained optimal values from Table 2.8, there was a significant decrease in optimal albumin synthesis (0.0006) to increase GSH synthesis to the optimal value of 14.87. Table 2.10 presents the flux values for Cases 3, 4 and 5 with their corresponding profiling in Figures 2.7A-B. Cases 3, 4, and 5 present the scenarios where higher albumin is desired with lower preference for NADPH and GSH synthesis, higher NADPH is desired with lower preference for albumin and GSH synthesis, and higher GSH synthesis is desired with lower preference for NADPH and albumin synthesis, respectively. As seen in Figure 2.7A, going from Case 4 (higher NADPH) to Case 3 (higher albumin) requires up-regulation in the uptake of amino acid fluxes (58, 60, 62, 64, 67, 72, 73, 74, 77-79, 81), and catabolism of threonine (21), phenylalanine (27), and isoleucine (33). Figure 2.7B presents the re-routing of fluxes from Case 4 to Case 5 (higher GSH). This requires up-regulation of catabolism of lysine (26), tyrosine (28), proline (29), leucine (34), production of ketone bodies (51), and uptake of amino acid fluxes (59, 60, 67, 70, 73, 78).

Table 2.9: Linear Physical Programming optimization results for a tri-objective system (NADPH synthesis, albumin synthesis and urea synthesis) for hepatic metabolic network. HUD is highly undesirable, UD is undesirable, T is tolerable, D is desirable and HD is highly desirable preference values of objective functions.

Case #	Priority	Flux	t ₁ (HD)	t ₂ (D)	t ₃ (T)	t ₄ (UD)	t ₅ (HUD)	Optimal
1	High	urea	34.8690	33.3175	31.7660	30.2146	28.6631	28.6631
	High	NADPH	3.0000	2.8670	2.7339	2.6009	2.4678	2.4785
	High	albumin	0.1426	0.1362	0.1299	0.1235	0.1172	0.1299
2	High	urea	34.8690	33.7054	32.5418	31.3782	30.2146	30.2146
	Low	NADPH	0.4091	0.3093	0.2096	0.1098	0.0100	2.1567
	High	albumin	0.1426	0.1378	0.1331	0.1283	0.1235	0.1288
3	High	urea	34.8690	34.7137	34.5584	34.4030	34.2477	34.8078
	Low	NADPH	0.0633	0.0500	0.0366	0.0233	0.0100	1.5477
	Low	albumin	0.0026	0.0019	0.0013	0.0006	0.0000	0.0026
4	Low	urea	2.6181	1.9636	1.3091	0.6545	0.0000	19.8383
	High	NADPH	3.0000	2.9439	2.8877	2.8316	2.7755	2.7795
	High	albumin	0.1426	0.1399	0.1372	0.1345	0.1319	0.1372
5	Low	urea	0.6213	0.4660	0.3106	0.1553	0.0000	22.4754
	High	NADPH	3.0000	2.9867	2.9734	2.9600	2.9467	2.9968
	Low	albumin	0.0026	0.0019	0.0013	0.0006	0.0000	0.0026
6	Low	urea	0.6213	0.4660	0.3106	0.1553	0.0000	28.3484
	Low	NADPH	0.0633	0.0500	0.0366	0.0233	0.0100	2.4141
	High	albumin	0.1426	0.1419	0.1413	0.1407	0.1400	0.1426

Table 2.9 presents the multiobjective optimal solutions for 6 scenarios for urea/NADPH/albumin tri-objective case. Case 1 indicates the base case where preferences for all three were given based on their anchor points, however, none of the objectives were given any specific priority. Hence, higher preferences for all three were given in the optimization, and as seen from the Highly Desirable value for objective functions, all objectives' higher desirability is close to their anchor points. The preference ranges in the LPPFBA optimization for this case were selected as 34.87-28.66 for urea secretion, 3-2.46 for NADPH synthesis and 0.143-0.117 for albumin synthesis. The optimal values of objective functions obtained for this case (Case 1) are 28.66 for urea secretion (between undesirable and highly undesirable), 2.48 for NADPH synthesis (between undesirable and highly undesirable), and 0.13 for albumin synthesis (between tolerable and undesirable). Next, we present Cases 2-3, where higher priority is desired for urea secretion over objectives of albumin and NADPH synthesis. In Case 2, the desirable preferences for NADPH synthesis were decreased and urea and albumin synthesis were kept similar to Case 1. The obtained multiobjective optimality results for Case 2 clearly indicate the inter-optimality or tradeoff and inter-play between objectives, since the optimal value of urea secretion (30.21) increased from Case 1 solution with a concomitant decrease in NADPH and albumin synthesis to 2.15 and 0.128, respectively. Interestingly, Case 3 shows the case where preference for urea secretion is much higher and closer to anchor points than both of the other objectives of NADPH and albumin synthesis. In this case, urea secretion optimal values (between highly desirable and desirable) are close to the anchor point of urea secretion with a significant decrease in both NADPH and albumin synthesis. In Cases 4-5, the priority to achieve higher NADPH synthesis is desired over other objectives. As seen in the table, in Case 4 the preferences of urea were decreased (2.6-0.0) and both the NADPH and albumin synthesis preference ranges were increased to 3-2.78 and 0.14-0.13, respectively, resulting in higher optimal values for NADPH (2.78) and albumin (0.13) synthesis. In Case 5, still higher NADPH synthesis was desired with low priority for other objectives. All the objectives were similar to their highly desirable preference values. Case 6 presents the case where higher albumin synthesis was desired with low preferences for NADPH synthesis and urea secretion. Again, the higher optimal albumin synthesis was obtained at the cost of both urea and NADPH secretion. Table 2.10 presents the flux values for Case 3, 5 and 6 with their corresponding profiling in Figures 2.7C and 2.7D.

Table 2.10: Optimal fluxes obtained for tri-objective system for the corresponding Pareto optimal cases 3, 4 and 5 from Table 2.8, and cases 3, 5 and 6 from Table 2.9.

Flux	NADPH/Albumin/GSH			Urea/NADPH/Albumin		
	Case 3	Case 4	Case 5	Case 3	Case 5	Case 6
1	3.719	3.996	3.947	5.600	3.997	3.414
2	3.719	3.996	3.947	5.600	3.997	3.414
3	3.719	3.996	3.947	5.600	3.997	3.414
4	7.486	7.993	8.616	12.800	7.995	6.829
5	30.000	30.000	30.000	30.000	30.000	30.000
6	30.000	30.000	30.000	30.000	30.000	30.000
7	30.000	29.286	29.963	24.491	27.267	23.827
8	19.652	19.286	20.078	14.491	17.016	14.643
9	19.652	19.286	20.078	14.491	17.016	14.643
10	10.363	9.286	10.078	5.000	7.016	5.000
11	15.996	14.704	15.095	5.031	12.174	8.130
12	29.341	30.000	30.000	30.000	29.749	30.000
13	29.341	30.000	30.000	30.000	29.749	30.000
14	13.245	15.196	14.548	24.869	17.475	21.770
15	13.245	15.196	14.548	24.869	17.475	21.770
16	12.994	15.114	14.533	34.808	22.475	28.348
17	-7.032	-0.711	-0.538	-0.658	-0.658	-7.839
18	18.948	16.952	16.935	15.049	17.825	21.566
19	0.100	0.100	0.100	0.100	0.100	0.100
20	4.478	7.818	7.885	5.772	7.839	5.294
21	0.817	0.418	0.001	0.001	0.153	0.001
22	-1.000	-1.000	-1.000	-1.000	-1.000	-0.074
23	22.514	22.007	21.384	17.120	22.005	23.171
24	0.779	0.497	0.499	0.497	0.497	1.857
25	5.634	5.418	5.017	0.031	5.158	3.130
26	0.001	0.001	0.032	4.813	0.001	2.444
27	1.293	0.171	0.370	0.154	0.154	0.100
28	0.100	0.100	0.357	0.100	0.100	0.100
29	0.100	0.100	0.797	0.100	0.100	0.100
30	2.861	2.703	2.325	0.000	2.982	2.861
31	4.145	4.980	4.996	0.010	4.985	3.109
32	0.010	0.010	0.010	0.010	0.010	0.010
33	0.662	0.010	0.010	0.010	0.010	0.010
34	0.010	0.010	1.515	4.755	0.010	2.016
35	-9.689	-10.000	-9.885	-10.000	-10.000	-9.184
36	7.149	6.304	5.683	8.031	8.325	7.149
37	-9.289	-10.000	-10.000	-9.491	-10.000	-9.643
38	0.750	0.000	0.000	4.939	0.000	1.578
39	6.436	7.151	6.788	0.436	6.859	6.436
40	1.911	2.285	1.942	0.354	1.417	0.320
41	-0.520	0.416	-0.137	-3.260	-0.077	-2.548
42	0.260	0.914	0.394	2.050	0.422	1.753
43	46.102	45.534	46.316	52.766	44.351	46.037
44	5.000	5.000	5.000	5.721	9.346	5.544
45	-0.049	-0.001	-0.722	-1.680	-0.001	-0.001
46	2.673	2.996	2.864	1.548	2.997	2.414
47	0.143	0.003	0.001	0.003	0.003	0.143
48	9.055	14.761	14.875	9.821	14.795	8.019
49	0.608	2.000	1.715	1.800	0.779	0.111
50	17.984	12.945	13.466	10.000	10.000	10.000
51	0.369	0.335	1.403	6.758	0.444	3.868
52	0.549	0.755	0.641	0.111	0.466	0.100
53	1.046	1.000	1.083	4.052	1.000	1.000
54	0.010	1.244	0.352	0.010	0.312	0.010
55	0.131	0.010	0.010	0.011	0.010	0.010
56	0.131	0.010	0.010	0.011	0.010	0.010
57	0.059	1.245	1.074	1.690	0.313	0.011
58	0.922	0.500	0.500	0.500	0.500	2.000
59	55.265	53.406	54.394	43.016	42.023	42.876
60	4.377	0.202	0.815	0.177	0.177	4.377
61	2.255	5.000	5.000	5.000	5.000	4.473
62	10.000	6.372	5.695	8.082	8.376	10.000
63	-1.000	-0.082	-0.015	5.000	5.000	5.000
64	3.172	0.000	0.000	10.000	5.061	10.000
65	20.000	19.676	18.072	19.574	19.281	20.000
66	0.886	-1.000	-1.000	-1.000	-1.000	-1.000
67	5.000	1.006	1.562	-1.000	1.865	5.000
68	5.000	5.000	5.000	2.092	5.000	5.000
69	0.969	-1.000	-1.000	6.920	-0.768	10.000
70	0.001	0.688	0.863	0.148	0.087	0.001
71	10.000	8.349	7.907	5.858	8.076	10.000
72	7.557	0.181	0.064	4.948	0.136	10.000
73	5.000	0.260	0.385	0.220	0.220	3.807
74	5.000	2.754	2.335	0.038	3.020	5.000
75	5.000	5.000	5.000	0.025	5.000	3.964
76	5.000	0.129	0.031	0.099	0.099	5.000
77	2.516	0.054	0.018	0.043	0.043	1.863
78	7.994	0.200	1.549	4.898	0.153	10.000
79	10.000	7.236	6.803	0.500	6.923	10.000
80	10.000	10.000	10.000	10.000	10.000	10.000
81	1.801	0.000	0.000	0.000	0.000	2.994

As seen in Figure 2.7C, going from Case 5 (higher NADPH synthesis) to Case 3 (higher urea secretion) requires up-regulation of gluconeogenic fluxes (1-4), higher urea cycle fluxes (14-16), lower TCA cycle fluxes (10-11), and lower bypass reaction of 3-phosphoglycerate (23).

To further demonstrate the advantages of applying LPPFBA in metabolic systems, we compare the Pareto surface obtained using LPPFBA with weighted-sum (WS) and goal programming (GP) based MFA for two separate tri-objective systems: glutathione synthesis, urea secretion and albumin synthesis (Figure 2.7E); and NADPH synthesis, urea secretion and albumin synthesis (Figure 2.7F) of primary hepatocytes. We ran 50,000 simulations utilizing different sets of weights using WS and GP based MFA and only 1,000 simulations utilizing different set of preferences for LPPFBA. As seen in Figures 2.7E-F, even after using 50,000 set of weights using WS and GP based MFA very few Pareto optimal solutions could be obtained. This illustrates that the LPPFBA can predict all possible Pareto optimal solutions for large-scale metabolic network systems whereas existing methods can capture only limited optimal solutions on Pareto surface. This is a noteworthy advantage of LPPFBA.

2.4.3 Quad-Objective Hepatic Metabolic Network

In the previous sections we presented the application of LPPFBA for bi-objective and tri-objective systems. In this section, we present the application of LPPFBA for improving current hepatic cellular systems using quad-objective (albumin synthesis, glutathione synthesis, NADPH synthesis and urea secretion) optimization. In BAL systems, when hepatocytes are exposed to human plasma they become steatotic and exhibit severe loss of hepatic function (albumin and urea synthesis, (Chan et al., 2002)). The experimental metabolic fluxes for simulated BAL condition of hepatocytes exposed to human plasma were obtained from the literature (Chan et al., 2003a; Chan et al., 2002). The goal was to determine optimal fluxes for the hepatic metabolic network under the simulated BAL condition considering all the objective functions simultaneously, leading to a quad-objective scenario (albumin, urea, NADPH and GSH). Table 2.11 presents the preferences assigned to the four objectives to create the different scenarios. The experimentally measured fluxes with their corresponding intracellular fluxes were used in the “base cases” for all comparisons. Two separate cases were used as “base case” to compare the changes in current fluxes from these “base cases” to the optimized scenarios where a variety of

Table 2.11: Linear Physical Programming optimization results for a quad-objective system (NADPH synthesis, albumin synthesis, urea synthesis and glutathione synthesis) for hepatic metabolic network. HUD is highly undesirable, UD is undesirable, T is tolerable, D is desirable and HD is highly desirable preference values of design metrics. The base cases consider 10 measurements to each 5 cases: flux 54 = 0.5; flux 56 = 0.38; flux 57 = 0.83; flux 60 = 0.12; flux 66 = 0.24; flux 67 = -0.13; flux 69 = 0.006; flux 70 = 0.13; flux 75 = 0.016; and flux 81 = 0.022. Base case 1 is obtained by optimizing GSH (flux 48). In addition to the 10 measurements, another 3 values were imposed as constraints based on experimental data: flux 16 (urea) = 0.57, flux 46 (NADPH) = 0.012, and flux 47 (albumin) = $9.4 \cdot 10^{-5}$. Base case 2 is obtained by quad-objective optimization of urea, NADPH, albumin, and GSH using LPPFBA.

Case #	Priority	Flux	t ₁ (HD)	t ₂ (D)	t ₃ (T)	t ₄ (UD)	t ₅ (HUD)	Optimal
1 (base)	Measured	urea	—	—	—	—	—	0.57
	Measured	NADPH	—	—	—	—	—	0.012
	Measured	albumin	—	—	—	—	—	0.000094
	Max	GSH	—	—	—	—	—	0.26689
2 (base)	Mid	urea	30	20	10	1	0.1	29.3090
	Mid	NADPH	2.5	1.5	0.5	0.1	0.01	1.9736
	Mid	albumin	0.1	0.05	0.005	0.0005	0.00005	0.0005
	Mid	GSH	10	5	1	0.5	0.1	9.8955
3	High	urea	34.8690	34.8414	34.8137	34.7861	34.7585	34.8579
	Low	NADPH	0.0195	0.0171	0.0147	0.0124	0.0100	1.5195
	Low	albumin	0.0005	0.0003	0.0002	0.0001	0.0000	0.0005
	Low	GSH	0.0572	0.0454	0.0336	0.0218	0.0100	9.8938
4	Low	urea	0.1205	0.0929	0.0653	0.0376	0.0100	21.6799
	High	NADPH	3.0000	2.9976	2.9953	2.9929	2.9905	2.9984
	Low	albumin	0.0005	0.0003	0.0002	0.0001	0.0000	0.0005
	Low	GSH	0.0572	0.0454	0.0336	0.0218	0.0100	14.8811
5	Low	urea	0.0297	0.0248	0.0198	0.0149	0.0100	26.5255
	Low	NADPH	0.0117	0.0113	0.0108	0.0104	0.0100	2.3694
	High	albumin	0.1426	0.1425	0.1425	0.1425	0.1425	0.1426
	Low	GSH	0.0184	0.0163	0.0142	0.0121	0.0100	9.0546
6	Low	urea	0.0929	0.0722	0.0515	0.0307	0.0100	26.3734
	Low	NADPH	0.0171	0.0153	0.0136	0.0118	0.0100	2.6592
	Low	albumin	0.00035	0.00026	0.00018	0.00009	0.00001	0.00026
	High	GSH	14.9000	14.8911	14.8823	14.8734	14.8646	14.8892
7	High	urea	34.8690	32.1116	29.3542	26.5968	23.8394	27.9217
	High	NADPH	3.0000	2.7635	2.5270	2.2905	2.0539	2.6106
	High	albumin	0.1426	0.1313	0.1200	0.1087	0.0975	0.1087
	High	GSH	14.9000	13.7222	12.5444	11.3665	10.1887	10.4417

hepatic objectives were optimized. In one of the “base cases”, intracellular fluxes were obtained after optimizing for glutathione synthesis (in this case the other three hepatic objectives were used as measured fluxes obtained from (Chan et al., 2003a)). In the second scenario, all four

hepatic objectives were simultaneously optimized, hence, all four objectives were treated as unmeasured fluxes and other fluxes were used for optimization to compute the intracellular fluxes. The corresponding fluxes for these four cases are presented in Table 2.12. These cases were a subset of 16 different scenarios presented in Table 2.13. Various sets of preferences were changed for their corresponding objectives and the effect of changing preferences on the fluxes was investigated in the form of a “heat map”. Figures 2.8A-B show the distribution of flux changes for the cases shown in Table 2.11 in the form of a “heat map”. In Table 2.11, the “base case” is indicated as Case 1. The experimental values of albumin, urea and NADPH secretion are indicated with GSH synthesis being computed from the available measured data. Case 2 of the quad-objective system includes all the measured fluxes of Case 1; however, all four hepatic objectives are unmeasured in this case. Hence, Case 2 is a scenario where we optimized the current measured hepatic flux data using a quad-objective system. Both Case 1 and Case 2 serve as base cases or present the fluxes of the current BAL systems. Again, two heat maps are shown to compare the changes in fluxes from the two “base cases” to the optimized scenarios for hepatic objectives.

In Case 3 of the quad-objective system, we examined the impact of choosing urea synthesis as a priority over the other liver specific functions of albumin, GSH and NADPH synthesis. For this purpose, the preference for urea secretion was increased close to a highly desirable value, while the preferences for albumin synthesis, GSH synthesis and NADPH synthesis were kept far lower than their corresponding anchor points. We found that the optimal value for urea secretion (34.86) increased from Case 1 (0.57) and was close to its anchor point. There was a concomitant increase in NADPH, albumin and GSH synthesis from Case 1. However, compared to Case 2 there was an increase in urea secretion (29.3) in Case 3, however, with a moderate decrease in NADPH synthesis. In Case 4, we prioritized NADPH synthesis. Hence, the preference for NADPH synthesis was increased closer to the anchor point and all other preferences were decreased. We found that NADPH synthesis optimal values were close to the anchor point for NADPH flux (between highly desirable and desirable) concomitant with a significant decrease in urea synthesis when compared to Case 2.

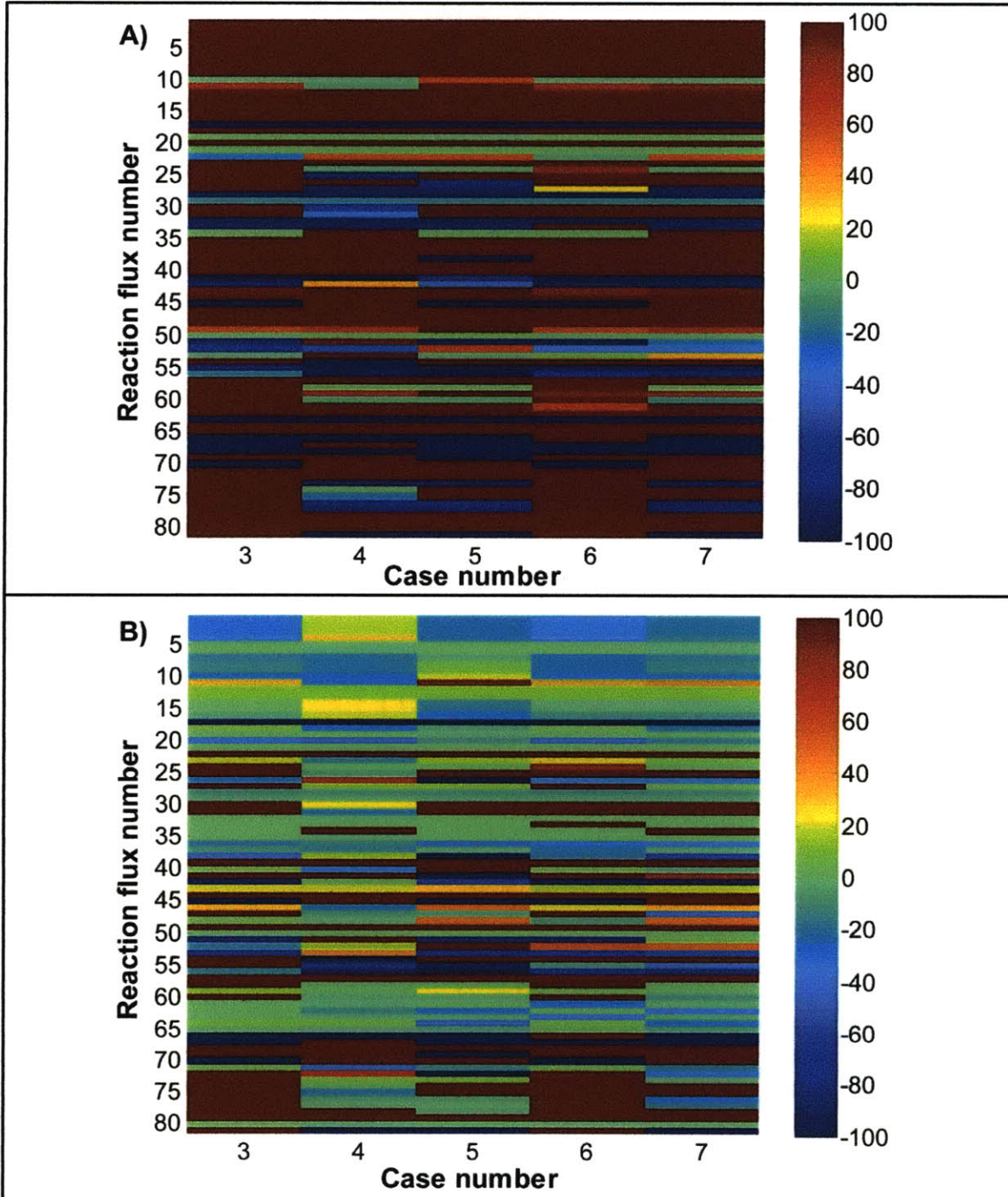


Figure 2.8: Metabolic profiling of percentage change in Pareto optimal fluxes for 16 solutions of the quad-objective optimization. The % change is taken from a referent optimal solution that consider 10 measurements to each 16 cases: flux 54 = 0.5; flux 56 = 0.38; flux 57 = 0.83; flux 60 = 0.12; flux 66 = 0.24; flux 67 = -0.13; flux 69 = 0.006; flux 70 = 0.13; flux 75 = 0.016; and flux 81 = 0.022. **A)** The reference is obtained by optimizing GSH (flux 48). In addition to the 10 measurements, another 3 values were imposed as constraints based on experimental data: flux 16 (urea) = 0.57, flux 46 (NADPH) = 0.012, and flux 47 (albumin) = $9.4 \cdot 10^{-5}$. **B)** The flux of reference is obtained by quad-objective optimization of urea, NADPH, albumin, and GSH using LPP. The preferences for each variable $[t_5^- \ t_4^- \ t_3^- \ t_2^- \ t_1^-]$ are [30 20 10 1 0.1] for urea, [2.5 1.5 0.5 0.1 0.01] for NADPH, [0.1 0.05 0.005 0.0005 0.00005] for albumin, and [10 5 1 0.5 0.1] for GSH. Only changes up to 100% are shown.

In Case 5, we prioritized albumin synthesis over other objectives. As seen in the table, in Case 5 the preferences of urea, NADPH and GSH synthesis were decreased and albumin synthesis preference range was increased to 0.1426, resulting in a higher optimal value of albumin (0.1426). In Case 6, we favored GSH synthesis over the other objectives. As seen in the table, preference for albumin synthesis has to be significantly decreased to achieve higher glutathione synthesis optimal values, whereas decreasing urea and NADPH synthesis is not necessary since it does not lead to any significant decrease in their optimal values. This is also confirmed in Figures 2.5D and 2.6D where going from Pareto optimal solution of higher glutathione synthesis (G) to higher albumin synthesis (H) does not require a significant change in urea and NADPH synthesis.

In Case 7, preferences for all four objectives were the same (highly desirable) and therefore none of them was given any specific priority. In this case, all objectives were close to their anchor points. The preference ranges in the LPP optimization for this case, were selected as 34.87-23.84 for urea secretion, 3-2.05 for NADPH, 0.143-0.09 for albumin, and 14.9-10.19 for glutathione syntheses. The optimal values of objective functions i.e., desired fluxes obtained for this case (Case 7) are 27.92 for urea synthesis (between tolerable and undesirable), 2.61 for NADPH synthesis (between desirable and tolerable), 0.11 for albumin synthesis (between tolerable and undesirable) and 10.44 for glutathione synthesis (between undesirable and highly undesirable). All objectives optimal values were less than their individual priority case, thus clearly illustrating the inter-dependence and tradeoff among the various objectives.

Now, we quantitatively profile the fluxes of various metabolites under various optimal conditions in several cases based on the relative changes indicated in heat maps. Table 2.12 presents the flux values for Cases 1-7 with their corresponding profiling presented as a heat map in Figure 2.8A. The plasma exposed hepatic fluxes for the current BAL system are shown in Column 1 of Table 2.12. If high urea synthesis is desired (Case 3) in the BAL then this necessitates a significant increase in gluconeogenic fluxes (2-4), with a concomitant increase in TCA cycle fluxes (8-9, 12), increase in urea cycle fluxes (14-16), and an increase in catabolism of glucogenic amino acids (serine, 18; threonine, 20; glutamine, 39). Another key observation is that higher urea secretion in the simulated BAL requires a significant increase in catabolism of ketogenic amino acids (lysine, 26; leucine, 34). As seen in the table, the increase in fluxes is much higher when priority is more for urea secretion than the other objectives. Notably, there is

Table 2.12: Optimal fluxes obtained for quad-objective system (Table 2.11). These solutions are compared in Figure 2.8 with two different base cases which consider 10 measured fluxes (pink cells): the first one is obtained by one-variable optimization of flux 48 given measured values for fluxes 16, 46 and 47 (yellow cells). The second one is obtained by quad-optimization of the target fluxes using LPP.

Flux #	Base Cases		Cases obtained by different quad-optimization using LPP				
	1	2	3	4	5	6	7
1	0.988	5.141	6.084	3.998	3.449	4.024	3.611
2	0.988	5.141	6.084	3.998	3.449	4.024	3.611
3	0.988	5.141	6.084	3.998	3.449	4.024	3.611
4	1.646	10.612	14.158	7.998	6.898	8.934	7.222
5	0.769	30.000	30.000	30.000	30.000	30.000	30.000
6	0.769	30.000	30.000	30.000	30.000	30.000	30.000
7	11.703	29.875	23.392	28.202	23.231	24.972	24.313
8	6.409	16.858	13.392	18.202	12.917	14.700	14.313
9	6.409	16.858	13.392	18.202	12.917	14.700	14.313
10	5.086	6.858	5.000	8.202	5.000	5.000	5.000
11	5.332	6.892	5.031	13.220	9.639	10.019	9.369
12	6.924	26.983	30.000	30.000	29.686	29.728	30.000
13	6.924	26.983	30.000	30.000	29.686	29.728	30.000
14	0.524	19.980	24.869	16.680	19.947	19.609	20.531
15	0.524	19.980	24.869	16.680	19.947	19.609	20.531
16	0.570	29.309	34.858	21.680	26.526	26.373	27.922
17	0.734	0.709	-0.529	-0.529	-6.812	-0.516	-6.394
18	0.100	19.066	13.821	18.203	19.944	15.388	20.381
19	0.100	0.100	0.100	0.100	0.100	0.100	0.100
20	0.010	9.429	5.253	8.301	5.294	7.894	6.411
21	0.001	0.001	0.001	0.001	0.001	0.001	0.001
22	-0.629	-0.167	-1.000	-1.000	-0.592	-1.000	-0.440
23	0.877	19.388	15.842	22.002	23.102	21.066	22.778
24	0.500	0.499	0.500	0.500	0.921	0.500	1.239
25	0.246	0.034	0.031	5.018	4.639	5.019	4.369
26	0.005	2.098	3.685	0.001	1.384	1.394	1.603
27	1.048	0.100	0.110	0.110	1.293	0.106	2.122
28	1.068	0.111	0.100	0.100	0.100	0.100	0.100
29	0.117	0.105	0.100	0.100	0.100	0.100	0.100
30	0.009	0.003	0.003	2.251	2.861	1.921	2.980
31	0.015	0.013	0.010	4.997	4.145	4.998	4.348
32	0.125	0.010	0.010	0.010	0.010	0.010	0.010
33	0.105	0.010	0.010	0.010	0.483	0.010	0.010
34	0.010	0.010	4.099	0.010	0.010	1.998	0.010
35	-0.515	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000
36	0.006	9.990	8.481	7.366	7.149	6.822	7.825
37	-1.323	-10.000	-8.392	-10.000	-7.917	-9.700	-9.313
38	0.489	4.329	4.989	0.000	3.444	1.765	2.390
39	0.498	0.487	1.436	7.556	6.436	6.419	6.857
40	0.010	1.068	0.707	2.310	1.119	1.011	1.101
41	0.724	-0.970	-2.509	0.153	-2.205	-1.794	-2.678
42	1.229	1.628	1.676	0.654	0.100	0.100	0.164
43	30.324	50.106	60.627	67.267	57.791	58.887	59.588
44	6.077	5.000	17.797	29.645	20.277	20.206	19.360
45	-0.330	-0.330	-1.990	-0.001	-0.001	-0.886	-0.001
46	0.012	1.974	1.520	2.998	2.369	2.659	2.611
4.70E+01	9.40E-05	5.00E-04	4.62E-04	4.62E-04	1.43E-01	2.64E-04	1.09E-01
48	0.267	9.896	9.894	14.881	9.055	14.889	10.442
49	1.297	1.019	2.214	2.763	2.076	2.293	2.093
50	10.000	10.000	10.000	10.427	10.000	10.000	10.227
51	2.177	1.620	4.234	0.222	0.209	1.688	0.273
52	0.467	0.189	0.214	0.763	0.299	0.293	0.153
53	1.000	3.167	4.564	1.000	1.079	1.365	1.000
54	0.500	0.500	0.010	1.999	1.776	1.114	1.939
55	1.791	0.120	0.033	0.010	0.111	0.066	0.320
56	0.380	0.380	0.033	0.010	0.111	0.066	0.320
57	0.830	0.830	2.000	2.000	1.777	2.000	1.940
58	0.500	0.500	0.500	0.500	1.064	0.500	1.348
59	22.942	41.405	41.556	50.587	43.341	42.206	46.148
60	0.120	0.120	0.114	0.114	4.377	0.108	3.362
61	2.163	5.000	5.000	5.000	3.577	5.000	5.000
62	0.008	10.000	8.490	7.375	10.000	6.827	10.000
63	-0.443	5.000	5.000	5.000	3.134	5.000	5.000
64	0.048	9.341	10.000	5.011	10.000	6.771	10.000
65	0.100	20.000	19.702	19.104	20.000	19.072	20.000
66	0.240	0.240	-1.000	-1.000	0.964	-1.000	-1.000
67	-0.130	-0.130	-1.000	2.210	5.000	0.327	5.000
68	-1.000	0.140	2.649	4.588	5.000	5.000	5.000
69	0.008	0.006	6.403	-0.671	7.361	2.796	6.186
70	0.130	0.130	1.642	0.542	0.001	0.510	0.001
71	0.014	9.447	5.269	8.317	10.000	7.903	10.000
72	0.010	2.125	3.710	0.025	8.940	1.408	7.366
73	1.050	0.113	0.122	0.122	5.000	0.112	4.950
74	0.010	0.010	0.010	2.258	5.000	1.925	4.611
75	0.016	0.016	0.013	5.000	5.000	5.000	5.000
76	0.128	0.027	0.026	0.026	5.000	0.019	3.816
77	0.106	0.016	0.016	0.016	2.336	0.013	1.424
78	0.015	0.038	4.125	0.036	7.994	2.013	6.099
79	0.500	0.500	1.448	7.568	10.000	6.425	9.575
80	0.355	10.000	10.000	10.000	10.000	10.000	10.000
81	0.022	0.022	0.000	0.000	1.801	0.000	0.261

a significant decrease in glycerol uptake (54) when higher priority is for urea synthesis. The other evident factor is that uptake of glucogenic amino acids is lower (serine, 67; glycine, 68; threonine, 71; histidine, 74; methionine, 75; glutamine, 79) when higher priority is to achieve higher urea synthesis over other objectives. In Case 4 we give higher preference to NADPH synthesis over other objectives. This is achieved using an increase in fatty acid oxidation, an increase in glycerol uptake (54), and a decrease in uptake of ketogenic amino acids (lysine, 72; leucine, 78). The trends or changes in fluxes in moving from “Case 1 to Case 3” and “Case 1 to Case 4” in this quad-objective case are similar to the trends when moving from Case F to E in the bi-objective scenario (Figure 2.5C). However, there are some major differences because of constraints in other objectives (albumin synthesis and glutathione synthesis). In quad-objective Case 3 and Case 4 the preference of albumin synthesis is low and this results in differences in actual flux values. The other important differences are the increase in oxygen uptake, lipid uptake (52), lipid stored (57) and fatty acid oxidation observed in Case 4 which were not significant in Case E of the bi-objective.

To move from “Case 1” to “Case 5”, i.e., to high albumin synthesis, requires glucogenic fluxes (2-4) and urea cycle fluxes (14-16) to be increased significantly with a concomitant decrease in ketone body production (51, 70) and triglyceride lipolysis (52). There is also an increased catabolism of essential amino acids which are both glucogenic and ketogenic (tryptophan, 24; phenylalanine, 27; and isoleucine, 33) and an increased uptake of glucogenic (60, 68, 71-77) and ketogenic (lysine, 72; and leucine 78) amino acids. Notably, in Case 5, i.e. of high albumin synthesis, there is an increased uptake of pyruvate forming amino acid (alanine, 66) and α -ketoglutarate forming amino acids (histidine, 30; proline, 60; arginine, 64) which is only evident under this condition of high albumin synthesis. In comparing Case 5 with 1-A of the bi-objective scenario (Figure 2.5A), again there is a decrease in urea secretion in Case 4, because of higher preferences for NADPH and glutathione synthesis over urea secretion. In Case 1-A, the values of two optimized objectives, urea secretion and albumin synthesis, were 29.7 and 0.136 respectively, and for the two non-optimized objectives, NADPH and glutathione synthesis, were 0.44 and 0.07, respectively. However, in Case 5, the values of optimized objectives for urea secretion, NADPH, albumin, and glutathione synthesis were 26.5, 2.37, 0.14 and 9.05, respectively. This results in decreased urea cycle fluxes, higher fatty acid oxidation (40), and decreased glucogenic fluxes in Case 5.

Table 2.13: Optimal fluxes obtained for quad-objective system. These solutions are compared in Figure 2.8 with two different base cases which consider 10 measured fluxes (pink cells): the first one is obtained by one-variable optimization of flux 48 given measured values for fluxes 16, 46 and 47 (yellow cells). The second one is obtained by quad-optimization of the target fluxes using LPP.

Flux #	Cases obtained by different quad-optimization using LPP																References	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Opt. of GSH	Quad-opt LPP
1	3.611	3.478	3.584	3.594	3.612	3.962	4.171	6.084	3.895	3.780	3.998	3.998	3.592	3.449	4.024	3.612	0.988	5.141
2	3.611	3.478	3.584	3.594	3.612	3.962	4.171	6.084	3.895	3.780	3.998	3.998	3.592	3.449	4.024	3.612	0.988	5.141
3	3.611	3.478	3.584	3.594	3.612	3.962	4.171	6.084	3.895	3.780	3.998	3.998	3.592	3.449	4.024	3.612	0.988	5.141
4	7.222	6.957	7.169	7.189	7.350	9.123	9.705	14.158	7.791	7.560	7.998	7.998	7.229	6.898	8.934	7.226	1.646	10.612
5	30.000	30.000	30.000	30.000	30.000	29.954	30.000	30.000	30.000	30.000	30.000	30.000	29.969	30.000	30.000	30.000	0.769	30.000
6	30.000	30.000	30.000	30.000	30.000	29.954	30.000	30.000	30.000	30.000	30.000	30.000	29.969	30.000	30.000	30.000	0.769	30.000
7	24.313	24.503	25.000	25.000	23.307	22.835	23.948	23.392	30.000	30.000	27.072	28.202	23.435	23.231	24.972	23.388	11.703	29.875
8	14.313	14.503	15.000	15.000	13.307	12.776	13.948	13.392	19.780	15.176	16.428	18.202	13.354	12.917	14.700	13.388	6.409	16.858
9	14.313	14.503	15.000	15.000	13.307	12.776	13.948	13.392	19.780	15.176	16.428	18.202	13.354	12.917	14.700	13.388	6.409	16.858
10	5.000	5.000	5.000	5.000	5.000	5.051	5.000	5.000	9.780	5.176	6.428	8.202	5.048	5.000	5.000	5.000	5.086	6.858
11	9.369	8.118	9.907	9.818	9.346	6.466	8.864	5.031	15.269	10.237	11.446	13.220	9.423	9.639	10.019	9.379	5.332	6.892
12	30.000	30.000	30.000	30.000	30.000	29.895	30.000	30.000	29.780	25.176	29.357	30.000	29.883	29.686	29.728	30.000	6.924	26.983
13	30.000	30.000	30.000	30.000	30.000	29.895	30.000	30.000	29.780	25.176	29.357	30.000	29.883	29.686	29.728	30.000	6.924	26.983
14	20.531	21.782	19.993	20.082	20.554	23.329	21.036	24.869	14.411	14.838	17.811	16.680	20.352	19.947	19.609	20.521	0.524	19.980
15	20.531	21.782	19.993	20.082	20.554	23.329	21.036	24.869	14.411	14.838	17.811	16.680	20.352	19.947	19.609	20.521	0.524	19.980
16	27.922	28.665	29.536	29.270	27.854	30.249	30.808	34.858	19.411	19.838	22.811	21.680	27.609	26.526	26.373	27.954	0.570	29.309
17	-6.394	-7.239	-1.680	-2.598	-6.865	-7.700	-1.091	-0.529	-6.959	-7.817	-0.538	-0.529	-6.842	-6.812	-0.516	-7.130	0.734	0.709
18	20.381	21.642	16.580	17.498	19.966	20.277	14.939	13.821	19.252	18.970	17.510	18.203	20.079	19.944	15.388	20.418	0.100	19.066
19	0.100	0.100	0.100	0.100	0.100	0.142	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
20	6.411	5.713	9.371	8.882	6.186	5.071	7.116	5.253	5.290	4.607	8.489	8.301	6.235	5.294	7.894	6.470	0.010	9.429
21	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	1.121	0.865	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
22	-0.440	0.027	-0.036	-0.103	-0.018	0.495	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.561	-0.592	-1.000	-0.432	-0.629	-0.167
23	22.778	23.043	22.831	22.811	22.650	20.831	20.295	15.842	22.209	22.440	22.002	22.002	22.740	23.102	21.066	22.774	0.877	19.388
24	1.239	1.684	0.481	0.466	0.542	0.594	0.490	0.500	0.674	1.554	0.499	0.500	0.486	0.921	0.500	0.393	0.500	0.499
25	4.369	3.118	4.907	4.818	4.346	1.415	3.864	0.031	5.489	5.062	5.017	5.018	4.375	4.639	5.019	4.379	0.246	0.034
26	1.603	3.116	1.319	1.468	1.493	3.054	1.492	3.685	0.239	1.766	0.001	0.001	1.461	1.384	1.394	1.430	0.005	2.098
27	2.122	1.623	0.500	0.811	1.989	1.400	0.300	0.110	2.128	1.432	0.113	0.110	2.012	1.293	0.106	2.220	1.048	0.100
28	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	1.068	0.111
29	0.100	0.100	0.100	0.100	0.100	0.101	0.100	0.100	0.100	0.100	0.100	0.148	0.100	0.100	0.100	0.100	0.117	0.105
30	2.980	2.838	0.579	0.912	3.250	2.172	0.341	0.003	3.369	2.942	2.657	2.251	3.248	2.861	1.921	3.362	0.009	0.003
31	4.348	3.097	4.886	4.797	4.325	1.351	3.843	0.010	4.348	4.177	4.996	4.997	4.348	4.145	4.998	4.358	0.015	0.013
32	0.010	0.010	0.010	0.010	0.010	0.013	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.125	0.010
33	0.010	0.010	0.010	0.010	0.010	0.050	0.010	0.010	0.010	0.010	0.010	0.010	0.015	0.483	0.010	0.010	0.105	0.010
34	0.010	0.010	0.010	0.010	0.092	1.935	2.236	4.099	0.010	0.010	0.010	0.010	0.155	0.010	1.998	0.010	0.010	0.010
35	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-10.000	-9.995	-10.000	-10.000	-10.000	-0.515	-10.000
36	7.825	7.402	9.619	9.323	7.667	7.434	7.752	8.481	7.825	7.256	8.375	7.366	7.642	7.149	6.822	7.861	0.006	9.990
37	-9.313	-9.503	-10.000	-10.000	-8.307	-7.725	-8.948	-8.392	-10.000	-10.000	-10.000	-10.000	-8.306	-7.917	-9.700	-8.388	-1.323	-10.000
38	2.390	1.883	4.543	4.188	2.737	3.091	4.771	4.989	0.000	0.000	0.000	0.000	2.591	3.444	1.765	2.434	0.489	4.329
39	6.857	6.530	4.983	5.242	7.108	5.818	4.792	1.436	7.282	6.570	7.153	7.056	7.125	6.436	6.419	7.327	0.498	0.487
40	1.101	0.531	1.081	1.131	1.279	0.969	1.032	0.707	1.809	1.441	1.596	2.310	1.289	1.119	1.011	1.347	0.010	1.066
41	-2.678	-3.250	-1.509	-1.677	-1.873	-2.385	-1.671	-2.509	-0.781	-3.220	-0.259	0.153	-1.844	-2.205	-1.794	-1.723	0.724	-0.970
42	0.164	1.550	0.291	0.258	0.162	1.264	0.312	1.676	0.133	0.100	0.241	0.654	0.103	0.100	0.100	0.100	1.229	1.628
43	59.588	61.588	62.843	63.356	57.795	59.212	59.339	60.627	62.751	63.635	57.033	67.267	57.712	57.791	58.887	57.624	30.324	50.106
44	19.360	16.469	17.376	18.669	20.439	18.891	19.638	17.797	21.339	23.422	22.201	29.645	20.544	20.277	20.206	20.664	6.077	5.000
45	-0.001	-0.001	-0.001	-0.001	-0.125	-1.199	-1.364	-1.990	-0.001	-0.001	-0.001	-0.001	-0.044	-0.001	-0.886	-0.001	-0.330	-0.330
46	2.611	2.478	2.584	2.594	2.542	2.905	2.232	1.520	2.895	2.780	2.998	2.998	2.566	2.369	2.659	2.612	0.012	1.974
47	0.109	0.130	0.019	0.034	0.112	0.128	0.010	0.000	0.109	0.137	0.001	0.000	0.109	0.143	0.000	0.107	0.000	0.001
48	10.442	8.451	14.119	13.513	10.288	6.716	13.409	9.894	10.442	9.274	14.875	14.881	10.442	9.055	14.889	10.516	0.267	9.896
49	2.093	0.869	2.020	2.100	2.321	2.266	2.337	2.214	0.521	2.474	1.788	2.763	2.351	2.076	2.293	2.442	1.297	1.019
50	10.227	10.000	10.000	10.000	10.106	10.115	10.000	10.000	17.607	18.747	10.000	10.427	10.098	10.000	10.000	10.000	10.000	10.000
51	0.273	1.165	0.400	0.367	0.347	2.262	2.137	4.234	0.150	0.150	0.350	0.222	0.313	0.209	1.688	0.209	2.177	1.620
52	0.153	0.168	0.100	0.100	0.378	0.309	0.337	0.214	0.473	0.474	0.316	0.763	0.399	0.299	0.293	0.442	0.467	0.189
53	1.000	1.000	1.000	1.000	1.070	1.958	1.939	4.564	1.000	1.000	1.000	1.000	1.026	1.079	1.365	1.000	1.000	3.167
54	1.939	0.490	1.919	1.999	1.818	0.759	0.636	0.010	0.047	1.999	1.471	1.999	1.908	1.776	1.114	1.999	0.500	0.500
55	0.320	0.014	0.391	0.415	0.073	0.021	0.010	0.033	0.195	0.010	0.324	0.010	0.046	0.111	0.066	0.010	1.791	0.120
56	0.320	0.014	0.391	0.415	0.073	0.021	0.010	0.033	0.195	0.010	0.324	0.010	0.046	0.111	0.066	0.010	0.380	0.380
57	1.940	0.491	1.920	2.000	1.943	1.957	2.000	2.000	0.048	2.000	1.472	2.000	1.952	1.777	2.000	2.000	0.830	0.830
58	1.348	1.814	0.500	0.500	0.654	0.723	0.500	0.500	0.783	1.691	0.500	0.500	0.595	1.064	0.500	0.500	0.500	0.500
59	46.148	46.149	44.232	44.499	42.992	42.487	42.199	41.556	53.983	50.126	42.423	50.587	42.921	43.341	42.206	42.793	22.942	41.405
60	3.362	3.997	0.671	1.115	3.475	3.950	0.386	0.114	3.362	4.217	0.118	0.114	3.410	4.377	0.108	3.308	0.120	0.120
61	5.000	5.000	5.000	5.000	4.802	4.879	5.000	5.000	4.111	2.516	5.000	5.000	4.777	3.577	5.000	5.000	2.163	5.000

To move from “Case 1” to “Case 6”, i.e. to high glutathione synthesis, necessitates higher glucogenic amino acids (67, 71, 74, 75 and 79) and ketogenic amino acids (72, 78). Also, there is an increased uptake of amino acids relevant to urea cycle fluxes (arginine, 64; aspartate, 69; and asparagine, 62) compared to Case 1 but decreased when compared to Case 5 of high albumin synthesis; but, there is a significant increase in amino acids involved in the synthesis of glutathione (glycine, 68; and glutamine, 79).

Case 7 presents the results where equal priority was assigned to all four objectives. As can be seen from the heat map, moving from Case 1 to Case 7 necessitates most of the fluxes to be increased. Significant increase in fluxes of amino acids forming pyruvate (serine, 18), acetoacetylCoA (tryptophan, 24, 58), fumarate (phenylalanine, 27), and α -ketoglutarate (glutamine, 39) is required to simultaneously increase the fluxes of various objectives. In summary, to increase the hepatic function in BAL from its current state of Case 1 to the Case where all four major hepatocyte functions are increased necessitates increases in glucogenic fluxes, TCA cycle fluxes, and increased uptake of both glucogenic and ketogenic amino acid fluxes. Interestingly, the increase in hepatic function also necessitates decreased lipid synthesis and storage fluxes. This is in concurrence with the experimental results of stimulated BAL where lipid storage and synthesis decreased hepatic function. Table 2.14 provides a qualitative overall summary of the changes in fluxes for various important cases relevant for BAL systems when compared with the “base cases”.

Figure 2.8B presents the profiling if Case 2 is used as the base case and metabolite changes are computed based on this base case. The difference between Base Cases 1 and 2 is essentially between the exclusion of measurements for urea, albumin and NADPH syntheses. Base Case 2 optimizes the hepatic function using the existing measurements of amino acids and lipid uptake. The respective values of equally prioritized optimal hepatic objectives for Case 2 using LPPFBA are urea synthesis (29.31), NADPH synthesis (1.97), albumin synthesis (0.0005) and GSH synthesis (9.9).

Table 2.14: Summary of results. Fluxes for Cases 1-A, 2-B, 3-C, 4-D, 5-E, 6-F, 7-G and 8-H are presented in Table 2.7, and for Cases 9-2, 9-3, 9-4 and 9-5 are presented in Table 2.12. Low-r indicates reverse direction.

Case	Albumin synthesis (47)	GSH synthesis (48)	NADPH synthesis (46)	Urea secretion (16)	β oxidation (40)	Electron-transport (43,44)	Lipid uptake (52)	Glucogenesis (2,6)	Urea cycle (14,15)	TCA cycle (8)	TCA cycle (13)	Lipid stored (57)	Asp uptake (69)	Glucogenic amino acids (60, 67, 69, 71, 73, 75, 81, 76, 77, 79)	Ketogenic amino acids (72,78)
BASE CASE															
1-A	high	low	low	med	med	low	med	high	high	low	high	low	med	med	low
2-B	med	med	low	high	med	low	med	med	med	high	high	low	low	high	low
3-C	med	high	high	low	med	low	med	med	med	high	high	low	low	high	low
4-D	high	med	high	low	high	med	low	med	med	high	high	low	low-r	high	high
5-E	low	high	high	med	low	high	low	med	med	high	high	low	low	low	low
6-F	low	high	med	high	low	med	low	high	high	high	high	med	med	med	low
7-G	med	high	med	low	low	med	low	low	low	med	low	low	med	med	low
8-H	high	med	med	low	low	med	low	low	low	med	low	low	med	high	high
9-2	low	med	med	high	med	high	low	high	high	high	high	high	med	low	med
9-3	low	high	high	med-low	high	high	high	med	med	high	high	high	low-r	low	low
9-4	high	med	med	med	med	high	low	med	med	high	high	high	high	high	med
9-5	low	low	high	med	med	high	low	med	med	high	high	high	low	high	low

As seen in the heat map in Figure 2.8B, the flux changes to achieve various objectives when Base Case 2 is used as reference case both increased and decreased significantly from their current state. This is in contrast to Figure 2.8A where most of the changes were in the direction of increasing uptakes. Interestingly, to move from Case 2 to Case 7, i.e. to increase the hepatic functions in simulated BAL we need to decrease glucogenic fluxes (1-4) and increase the TCA cycle fluxes (11-14). Importantly, pyruvate is required to increase the synthesis of alanine (17) rather than using it directly as a substrate in the TCA cycle to compensate for the increased demand of alanine by other hepatic functions. Notably, TCA cycle flux from succinyl CoA to oxaloacetate is increased by increasing the synthesis of succinylCoA from methionine (net succinyl CoA flux is 4.37 in Case 7 compared to 0.034 in Case 2). There is net increase in uptake of various amino acids (proline, 60; serine, 67; glycine, 68; aspartate, 69; 72-79) required in Case 7 to increase the hepatic function in simulate BAL system.

2.5 CONCLUSIONS AND SUMMARY

Mammalian systems perform several different functions in nature and hence, optimization of such systems may involve more than one objective as the goal. For example, hepatocytes perform several different functions as the key component of BAL systems, and these functional objectives are potentially conflicting. As seen above, higher albumin synthesis changes the uptake of various metabolites in such a manner that necessarily decreases urea secretion. In order to investigate the trade-offs between these conflicting objectives and to explore available design options, one needs to formulate the optimization problem with multiple objectives (vector optimization). Vector optimization obtains a Pareto optimal solution that satisfies the strict constraints imposed by multiple objectives. However, most of the current algorithms suffer from several disadvantages, such as: requiring a priori selection of weights or targets for each of the objective functions which are inadequate in capturing desired preferences; providing a single Pareto solution; inability to generate proper Pareto points for non-convex problems (e.g., the weights method); inability to generate sensitivity information for trade-off and decision making; and no inherent capabilities for design exploration. LPPFBA captures the designer's preferences a priori in a mathematically consistent manner using preference functions. The application of LPPFBA does not require specifying weights for each objective function. Rather, the ranges of differing degrees of desirability for each objective function are specified. A clear advantage of the LPPFBA approach is that it is a strategy that allows one to obtain conditions where tradeoff of all the desired objectives could be observed in their physical space.

Another advantage of LPPFBA is its ability to deal with multiple objectives with ease. As seen in this work, BAL systems have many objectives and as can be seen from the quad-objective scenario, working in physical space allowed us to analyze optimal conditions easily and obtain various desired optimal solutions. As seen in Figure 2.3, using linear programming for this quad-objective problem would have necessitated specifying four weights and no target values, using goal programming would have necessitated 8 weights and four target values; however, LPPFBA requires no specification of weights and only requires target values (20). Specifying the target values is much easier since these are specified in physical space which is always a known space for the designer. Another significant advantage of LPPFBA is that it facilitates optimization of poorly scaled problems. An example of such problems in metabolic

networks is the maximization/minimization of two fluxes of different magnitudes, such as the minimization of albumin (on the order of 10^{-5}) and NADPH synthesis (on the order of 10^1).

In summary, in this section, a constrained multiobjective formulation LPPFBA to analyze large scale linear metabolic networks is presented. The LPPFBA approach provides a new effective tool to obtain Pareto optimal solutions. The incorporation of LPP into the standard Metabolic Flux Analysis method enables an unambiguous formulation of an aggregate objective function that facilitates effective multiobjective flux balance analysis for large-scale problems. The presented LPPFBA approach initiates a meaningful step towards infusion of genomic and proteomics data into metabolite perturbations. Importantly, the presented methodology could be employed in various metabolic networks that invariably have multiple objectives (ranging from physiological to design objectives) to be optimized. The combined quantitative and visualization framework presented in this work sets the stage for the development of true optimal solutions for large scale genomics based metabolic network systems. In the context of BAL, the results presented in this section illustrate that BAL design using constraints based multiobjective optimization can result in an increase in overall hepatic functions by modifying various metabolite fluxes from its current simulated state during BAL operation. The results presented in this work have the potential to improve the hepatic function by using optimal pre-conditioning medium in BAL devices.

3 INTEGRATED ENERGY AND FLUX BALANCE BASED MULTIOBJECTIVE FRAMEWORK FOR LARGE-SCALE METABOLIC NETWORKS

3.1 OVERVIEW

Flux Balance Analysis (FBA) provides a framework for the estimation of intracellular fluxes and Energy Balance Analysis (EBA) ensures the thermodynamic feasibility of the computed optimal fluxes. Previously, these techniques have been used to obtain optimal fluxes that maximize a single objective. Because mammalian systems perform various functions, a multi-objective approach is needed when seeking optimal flux distributions in such systems. For example, hepatocytes perform several metabolic functions at various levels depending on environmental conditions; furthermore, there is a potential benefit to enhance some of these functions for applications such as bioartificial liver (BAL) support devices. Herein we developed a multi-objective optimization approach that couples the normalized Normal Constraint (NC) with both FBA and EBA to obtain multiobjective Pareto optimal solutions. We investigated the Pareto frontiers in gluconeogenic and glycolytic hepatocytes for various combinations of liver specific objectives (albumin synthesis, glutathione synthesis, NADPH synthesis, ATP generation, and urea secretion). Next, we evaluated the impact of experimental flux measurements on the Pareto frontiers. We found that measurements induce dramatic changes in Pareto frontiers and further constrain the network fluxes. This multi-objective optimality analysis may help explain certain features of the metabolic control of hepatocytes, which is relevant to the response to hepatocytes and liver to various physiological stimuli and disease states.

3.2 INTRODUCTION

The quantification of intracellular metabolic fluxes is widely used for investigation of the metabolism in microorganisms (Edwards and Palsson, 2000a; Edwards and Palsson, 2000b; Papin et al., 2002; Savinell and Palsson, 1992a; Schilling et al., 2002; Segre et al., 2005; Segre et al., 2002) and mammalian systems (Banta et al., 2004; Chan et al., 2003a; Chan et al., 2003b; Chan et al., 2002; Chan et al., 2003c; Lee et al., 2000; Lee et al., 2003a; Lee et al., 2003b; Lee et al., 2004; Yarmush and Banta, 2003; Yokoyama et al., 2005). Flux Balance Analysis (FBA) uses stoichiometric and mass balance constraints to compute the intracellular fluxes.

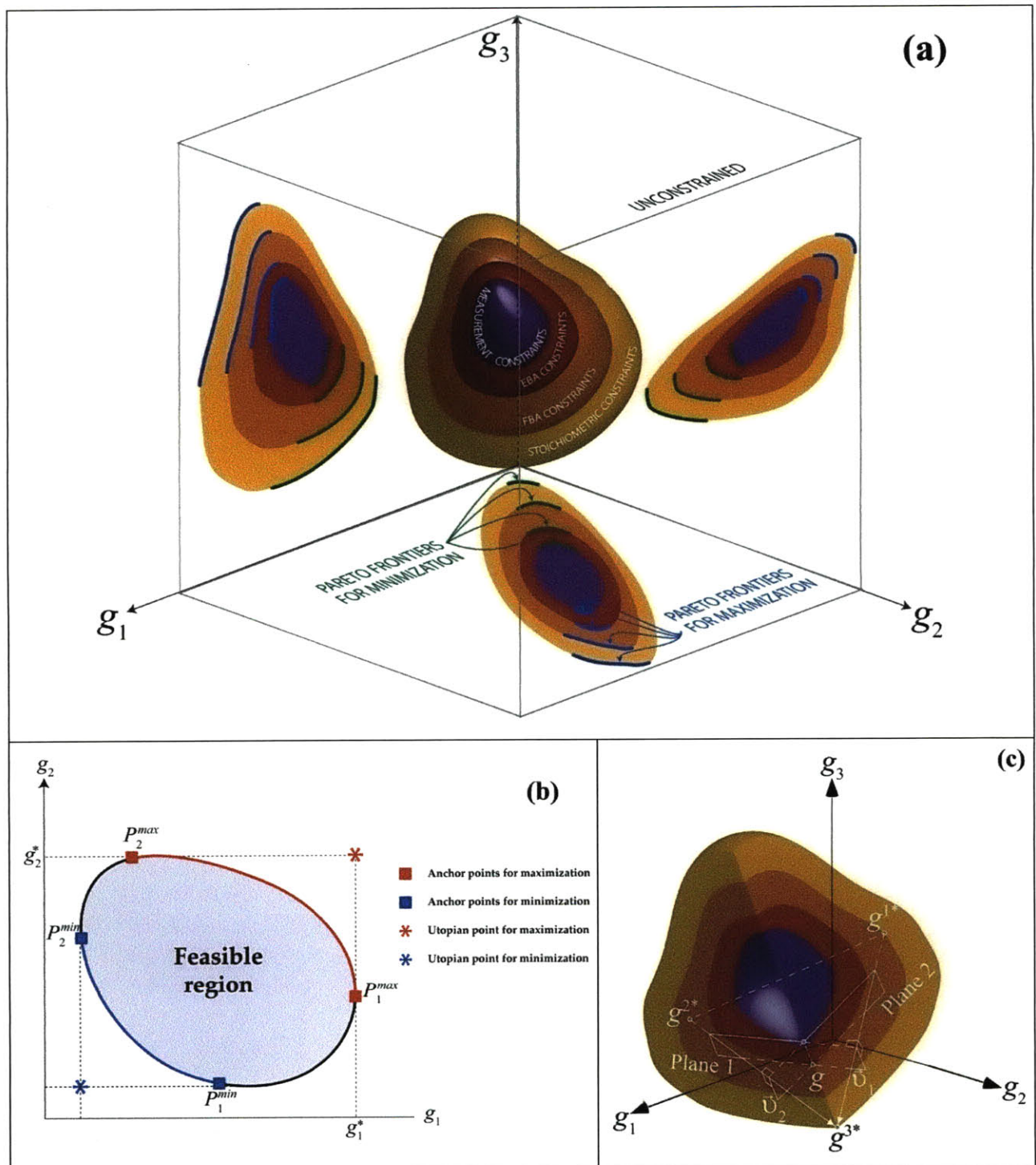


Figure 1: a) Feasible space reduction due to of the imposed stoichiometric, flux balance, energy balance, and measurement constraints. The Pareto surface of the feasible space is projected onto the g_1g_2 , g_2g_3 , and g_1g_3 planes and their corresponding Pareto frontiers are shown. The mutually orthogonal axes g_1 , g_2 , and g_3 , represent the individual design objectives. b) Pareto frontiers and Pareto optimal solutions shown are for bi-objective maximization and minimization problems. c) Normal Constraint method is based on sequence of systematic objective space reductions. In this case we utilize normal constraints to reduce the objective space as shown (i.e., obtain the dot product of a plane with the normal to ensure orthogonality) and then solve the nonlinear optimization in this reduced space.

Recently, Energy Balance Analysis (EBA) was developed to ensure the thermodynamic feasibility of the computed fluxes (Beard et al., 2004; Beard et al., 2002; Beard and Qian, 2005; Yang et al., 2005). EBA imposes the thermodynamic constraints on reaction fluxes both explicitly and implicitly. Essentially, the reaction potentials are computed based on the chemical potentials and then these are used to obtain thermodynamic constraints that are based on the first and second laws of thermodynamics. Thermodynamic constraints further reduce the feasible solution space based on stoichiometric constraints alone (Figure 3.1a). Available measurements, which bring in environmental constraints such as certain cell culture conditions, medium supplements, induced stress and extracellular matrices, typically limit the feasible solution space even further. If a sufficient number of measurements is available, the analysis may yield a unique solution.

Since mammalian systems perform an array of metabolic functions (protein secretion, detoxification, energy production), several different objectives need to be taken into account simultaneously when seeking optimal fluxes. Typically, several objectives compete against each other; therefore, only “Pareto-optimal” solutions can be achieved. A solution is said to be Pareto-optimal if there are no other solutions that can better satisfy all of the objectives simultaneously (Nagrath, 2005). Specifically, a Pareto solution is one where any improvement in one objective necessitates the worsening of at least one other objective. Non-Pareto optimal solutions are sub-optimal and their performance is inferior to systems operating and designed based on Pareto optimality of objectives. The class of Generate First-Choose Later (GCFL) multiobjective optimization approaches entails first generating a representative set of Pareto solutions, and then choosing the most suitable and appropriate solution within this set. The Normal Constraint (NC) method (Messac et al., 2003), unlike other popular methods such as the Normal Boundary Intersection (NBI) method, can generate complete Pareto frontiers for multi-objective problems from the full range of Pareto bi-objective solutions. The NC method essentially generates an even distribution of Pareto points throughout the complete Pareto frontier; and it is guaranteed to yield any Pareto point in the feasible design space. Further, it is insensitive to objective function scaling, and is valid for any arbitrary number of design objectives.

In the current work, we develop a Normalized Constraint Energy and Flux Balance Analysis (NCEFBA) based multiobjective framework for characterizing the intermediary metabolism of large-scale metabolic networks. The implementation is general and could be

easily modified for other metabolic networks but here it is presented in the context of hepatic metabolism. In the context of a bioartificial liver (BAL) device, this multiobjective optimal flux analysis could play an important role in: a) understanding the underlying mechanisms of perturbing a sub-optimal hepatic cellular system towards an optimal state, b) optimizing hepatocyte functions in an extracorporeal BAL device, c) studying the intracellular activity of liver under various physiological and disease states, and d) the preconditioning and preservation of donor livers. The presented multiobjective optimization platform NCEFBA couples the normalized NC method with both FBA and EBA to obtain multiobjective Pareto optimal solutions.

Here the NCEFBA method was implemented to investigate Pareto optimal solutions for the hepatic metabolic network under both gluconeogenic and glycolytic conditions. We analyzed various combinations of liver-specific objectives (albumin synthesis, glutathione synthesis, NADPH synthesis, ATP generation, and urea secretion). Next, the sensitivity to available measurements of these Pareto frontiers and changes in objective inter-optimality is presented. Noticeably, measurements induced dramatic changes in Pareto frontiers and further constrained the network fluxes.

3.3 THEORY

3.3.1 Metabolic Flux Analysis

The stoichiometric coefficients of the metabolic reactions are collected into a matrix S , where each element s_{ij} is the coefficient of metabolite i in reaction j . S has dimensions of $M \times N$, where M is the number of metabolites and N is the number of reactions. In matrix form the mass balance is written as:

$$\frac{dx}{dt} = SJ \quad (3.1)$$

where each element x_i of x is the intracellular concentration of metabolite i and element J_j of J is the net rate of conversion in reaction j . External metabolite fluxes are generally measured (e.g., uptake of glucose, lactate, amino acids). Because of the very high turnover of the intracellular pools of most intracellular metabolites, the time scale of the intracellular metabolic

reactions is short compared to other cellular reactions. Hence, the pseudo steady state assumption is generally applied to the metabolite mass balances and thus

Stoichiometric Equality Constraints for Unmeasured Fluxes

$$S_j = 0 \quad (\text{Mass Balance Constraint}) \quad (3.2)$$

Stoichiometric Equality Constraints for Measured Fluxes

$$S_u J_u = -S_m J_m \quad (\text{Mass Balance Constraint}) \quad (3.3)$$

where J_m and J_u indicates measured and unmeasured fluxes, respectively, and S_m and S_u contain the stoichiometric coefficients of measured and unknown reactions, respectively.

A previously described hepatic metabolic network (Chan et al., 2003a; Chan et al., 2003b; Chan et al., 2002; Chan et al., 2003c) includes all of the major intracellular pathways that account for the majority of central carbon and nitrogen metabolism found in hepatocytes, namely the tricarboxylic acid (TCA) and urea cycles, the gluconeogenic and glycolytic pathways, the pentose phosphate shunt, pathways of entry, transamination, and deamination of amino acids, protein synthesis, and the major components of lipid metabolism, including triglyceride synthesis and breakdown and β -oxidation of fatty acids, in addition to amino acid synthesis and apolipoprotein degradation. The current hepatic metabolic network model (Table 3.1) includes a few additional reactions, namely those of the 3-phosphoglycerate cycle as it is involved in glycerol production and glutathione synthesis reaction, which results in a total of 81 reactions (as compared to 76 reactions in the previous model) and 47 metabolites (Table 2.2). Glutathione is an important mediator in detoxification reactions of hepatocytes. The model assumes pseudo steady-state with no metabolic futile cycles. These assumptions are discussed in detail elsewhere (Chan et al., 2003a; Chan et al., 2003b; Chan et al., 2002; Chan et al., 2003c).

Table 3.1: Hepatic stoichiometric reactions

N°	Stoichiometry		
1	F6P → G6P	Gluconeogenesis	
2	F16P2 + H ₂ O → F6P + P _i		
3	2 G3P → F16P2		
4	3Pglyc + NADH + H ⁺ + ATP ↔ G3P + P _i + NAD ⁺ + ADP		
5	PEP ↔ 3Pglyc		
6	oac + GTP → PEP + GDP + CO ₂		
7	pyr + CO ₂ + ATP + H ₂ O → oac + ADP + P _i + 2 H ⁺		
1	G6P → F6P	Glycolysis	
2	F6P + P _i → F16P2 + H ₂ O		
3	F16P2 → 2 G3P		
4	G3P + P _i + NAD ⁺ + ADP → 3Pgyc + NADH + H ⁺ + ATP		
5	3Pgyc → PEP		
6	PEP + ADP → pyr + ATP		
7	pyr + CoA + NAD ⁺ → acCoA + CO ₂ + NADH		
8	oac + acCoA + H ₂ O → ctt + CoASH		
9	ctt + NAD ⁺ ↔ αkgl + CO ₂ + NADH + H ⁺		
10	αkgl + NAD ⁺ + CoASH → sucCoA + CO ₂ + NADH + H ⁺		
11	sucCoA + P _i + GDP + FAD ↔ fum + GTP + FADH ₂ + CoASH		
12	fum + H ₂ O ↔ mal		
13	mal + NAD ⁺ ↔ oac + NADH + H ⁺		
14	ctr + asp + ATP → arg + fum + AMP + PP _i		
15	orn + (CO ₂ + NH ₄ ⁺ + 2 ATP) + H ₂ O → ctr + 2 ADP + 2 P _i + 3 H ⁺		
16	arg + H ₂ O → urea + orn		
17	ala + 0.5 NAD ⁺ + 0.5 NADP ⁺ + H ₂ O ↔ pyr + NH ₃ + 0.5 NADH + 0.5 NADPH + H ⁺		
18	ser → pyr + NH ₃		
19	cys + 0.5 NAD ⁺ + 0.5 NADP ⁺ + H ₂ O + SO ₃ ²⁻ ↔ pyr + thiosulfate + NH ₄ ⁺ + 0.5 NADPH + 0.5 NADH		
20	thr + NAD ⁺ + ATP + CoASH → gly + acCoA + NADH + H ⁺ + AMP + PP _i		
21	thr + NAD ⁺ + CoASH → propCoA + CO ₂ + NADH + H ⁺ + NH ₃ + H ₂		
22	2 gly + NAD ⁺ + THF + H ₂ O ↔ NTHF + H ⁺ + CO ₂ + NH ₄ ⁺ + ser + NADH		
23	3Pglyc + NAD ⁺ + glu + H ₂ O → NADH + H ⁺ + αkgl + ser + P _i		
24	trp + 3 O ₂ + 4 H ₂ O + 2 NAD ⁺ + FAD + CoASH → Formate + ala + 2 CO ₂ + NH ₃ + 3 NADH + FADH ₂ + HCO ₃ ⁻ + acacCoA		
25	propCoA + CO ₂ + ATP → ADP + P _i + sucCoA		
26	lys + 2 αkgl + 2 NAD ⁺ + CoASH + FAD + 2 H ₂ O + NADP ⁺ → CO ₂ + NH ₃ + acacCoA + 5 NADH + FADH ₂		
27	phe + O ₂ + H ₄ biopterin + H ⁺ → tyr + H ₂ O + H ₂ biopterin		
28	tyr + αkgl + 2 O ₂ + H ₂ O → glu + CO ₂ + fum + acac		
29	pro + 0.5 O ₂ + 0.5 NAD ⁺ + 0.5 NADP ⁺ → glu + 0.5 NADH + 0.5 NADPH + H ⁺		
30	his + H ₄ folate + 2 H ₂ O → NH ₄ ⁺ + N ⁵ ,N ¹⁰ -CH ₂ -H ₄ folate + glu		
31	met + ATP + ser + NAD ⁺ + H ₂ O + CoASH → PP _i + P _i + adenosine + cys + NADH + H ⁺ + CO ₂ + NH ₄ ⁺ + propCoA		
32	val + αkgl + 3 NAD ⁺ + 2 H ₂ O + FAD + CoA → glu + 2 CO ₂ + 3 NADH + 2 H ⁺ + FADH ₂ + CO ₂ + propCoA		
33	iso + αkgl + H ₂ O + 2 NAD ⁺ + FAD + 2 CoASH → glu + CO ₂ + 2 NADH + 2H ⁺ + FADH ₂ + acCoA + propCoA		
34	leu + αkgl + H ₂ O + NAD ⁺ + FAD + ATP + CoASH + HCO ₃ ⁻ → glu + CO ₂ + NADH + H ⁺ + FADH ₂ + acCoA + acac + ADP + P _i		
35	oac + glu ↔ αkgl + asp		
36	asn + H ₂ O → asp + NH ₃		
37	glu + 0.5 NAD ⁺ + 0.5 NADP ⁺ + H ₂ O ↔ NH ₄ ⁺ + αkgl + 0.5 NADPH + 0.5 NADH + H ⁺		
38	orn + NAD ⁺ + NADP ⁺ + H ₂ O → glu + NH ₄ ⁺ + NADH + NADPH + H ⁺		
39	gln + H ₂ O → glu + NH ₄ ⁺		
40	palm + ATP + 7 FAD + 7 NAD ⁺ → 8 acCoA + 7 FADH ₂ + 7 NADH + AMP + PP _i		Gluconeogenesis
40	8 acCoA + 7 ATP + 14 NADPH + 14 H ⁺ → palm + 8 CoA + 6 H ₂ O + 7 ADP + 7P _i + 14 NADP ⁺		Glycolysis

41	$2 \text{ acCoA} \leftrightarrow \text{acacCoA} + \text{CoA}$
42	$\text{acacCoA} + \text{H}_2\text{O} \rightarrow \text{acac} + \text{CoA}$
43	$\text{NADH} + \text{H}^+ + 0.5 \text{ O}_2 + 3 \text{ ADP} \rightarrow \text{NAD}^+ + \text{H}_2\text{O} + 3 \text{ ATP}$
44	$\text{FADH}_2 + 0.5 \text{ O}_2 + 2 \text{ ADP} \rightarrow \text{FAD} + \text{H}_2\text{O} + 2 \text{ ATP}$
45	$\text{gol} + \text{NAD}^+ + \text{ATP} \leftrightarrow \text{G3P} + \text{NADH} + \text{H}^+ + \text{ADP} + \text{P}_i$
46	$\text{G6P} + 12 \text{ NADP}^+ + 7 \text{ H}_2\text{O} \rightarrow 6 \text{ CO}_2 + 12 \text{ NADPH} + 12 \text{ H}^+ + \text{P}_i$
47	$24 \text{ arg} + 32 \text{ asp} + 61 \text{ ala} + 24 \text{ ser} + 35 \text{ cys} + 57 \text{ glu} + 17 \text{ gly} + 21 \text{ tyr} + 33 \text{ thr} + 53 \text{ lys} + 26 \text{ phe} + 25 \text{ gln} + 30 \text{ pro} + 15 \text{ his} + 6 \text{ met} + 20 \text{ asn} + \text{trp} + 35 \text{ val} + 13 \text{ iso} + 56 \text{ leu} + 2332 \text{ ATP} \rightarrow \text{albumin} + 2332 \text{ ADP} + 2332 \text{ P}_i$
48	$\text{glu} + 2 \text{ ATP} + \text{cys} + \text{gly} + \text{NADPH} \rightarrow \text{GSH} + 2 \text{ ADP} + 2 \text{ P}_i + \text{NADP}^+ + \text{H}^+$
49	$\text{gol} + 3 \text{ acCoA} + \text{H}_2\text{O} + \text{ATP} \rightarrow 3 \text{ CoASH} + \text{P}_i + \text{TG} + \text{ADP} + \text{P}_i$
50	$\text{lactate} + \text{NAD}^+ \leftrightarrow \text{pyr} + \text{NADH} + \text{H}^+$
51	$\text{acac} + \text{NADH} + \text{H}^+ \leftrightarrow \beta\text{-OH-butyrate} + \text{NAD}^+$
52	$\text{TG} + 3 \text{ H}_2\text{O} \rightarrow \text{gol} + 3 \text{ palm} + 3 \text{ H}^+$
53	G6P release
54	gol uptake
55	palm release
56	cholesterol ester + H ₂ O → cholesterol + palm
57	TG stored
58	trp uptake
59	O ₂ uptake
60	pro uptake
61	glu secretion
62	asn uptake
63	orn secretion
64	arg uptake
65	NH ₄ ⁺ uptake
66	ala uptake
67	ser uptake
68	gly uptake
69	asp uptake
70	acac production
71	thr uptake
72	lys uptake
73	phe uptake
74	his uptake
75	met uptake
76	val uptake
77	iso uptake
78	leu uptake
79	gln uptake
80	cys uptake
81	tyr uptake

3.3.2 Energy Balance Analysis

Energy Balance Analysis imposes constraints based on law of thermodynamics on the cellular reaction networks (Beard et al., 2004; Beard et al., 2002). For any reaction set, if stoichiometry is represented by matrix S , μ denotes an M -dimensional vector of chemical potentials, $\Delta\mu$ denotes the N -dimensional vector of reaction potentials, then these potentials can

be computed as $\Delta\mu = S^T \mu$. The null space matrix of S (for r linearly independent rows, with $r \leq N$) is denoted by K and forms a basis for the null space of S , so that $SK = 0$. The product of the null space K of the stoichiometric S with the chemical potential difference gives the energy balance equation as $K\Delta\mu = 0$. This balances the global potential energy of the network. The first law of thermodynamics necessitates energy conservation, which then leads to an equality constraint as

$$K^T \Delta\mu = K^T S^T \mu = 0 \quad (\text{First Law of Thermodynamics-Based Energy Equality Constraint}) \quad (3.4)$$

This constraint requires that the sum of reaction potentials around any cycle of reactions equals zero, which is similar to Kirchoff's voltage or loop law of electrical circuit theory, and is known as the energy balance constraint of EBA (Beard et al., 2004; Beard et al., 2002; Beard and Qian, 2005). The second law of thermodynamics takes the form of an inequality constraint for each flux as $-J_i \Delta\mu_i \geq 0$. However, this equation is written in terms of net fluxes. Beard et al. (2002) compute the net flux distribution through the reaction network by introducing the concept of reversibility of each reaction which entails defining the nonnegative forward and reverse reaction fluxes, J_+ and J_- respectively, with j th entries representing the one-way fluxes through the j th reaction. The vector of net flux distribution through the reaction network can be then computed as $J = J_+ - J_-$, which is then used to compute the j th reaction potential as

$$\Delta\mu^j = RT \ln \left(\frac{J_-^j}{J_+^j} \right) \quad (3.5)$$

where R is the ideal gas constant and T is the temperature.

This relationship leads directly to the second law of thermodynamics, i.e.,

$$-J^j \Delta\mu^j = -RT (J_+^j - J_-^j) \ln \left(\frac{J_-^j}{J_+^j} \right) \geq 0 \quad (\text{Second Law of Thermodynamics Based Energy Equality Constraint}) \quad (3.6)$$

which says that the system must dissipate heat, and entropy must increase as a result of the work being done on the system through the external fluxes. For equilibrium systems, this is an equality since for these systems $J^j = \Delta\mu^j = 0$.

The other inequality constraint is obtained for energy balance by ensuring that the total heat dissipation rate of the living system is always positive as indicated by

$$hdr = -J^T \Delta\mu > 0 \quad (\text{Inequality Heat Dissipation Constraint}) \quad (3.7)$$

Since, $hdr \rightarrow 0$ in the limit as $J_- / J_+ \rightarrow 1$ component-wise while maintaining $J = J_+ - J_-$ so to prevent this physically unrealistic possibility, an additional inequality constraint

$$(hdr)_{lb} \leq hdr \leq (hdr)_{ub} \quad (\text{Inequality Heat Dissipation Constraint}) \quad (3.8)$$

is also imposed as part of Energy Balance Analysis

3.3.3 Pareto Optimality

Table 3.2 shows some of the definitions and mathematical formulation of the generic terms involved in multiobjective optimization. The mathematical representation of the generic multiobjective optimization problem is as follows

Problem P1

$$\min_x \{g_1(x), g_2(x), \dots, g_n(x)\} \quad (n \geq 2) \quad (3.9a)$$

subject to:

$$f_j(x) \leq 0 \quad (1 \leq j \leq r) \quad (3.9b)$$

$$h_k(x) = 0 \quad (1 \leq k \leq s) \quad (3.9c)$$

$$x_l \leq x_i \leq x_u \quad (1 \leq i \leq n_x) \quad (3.9d)$$

where the vector x denotes the design variables and g_i denotes the i -th objective function. Equations (3.9b), (3.9c) and (3.9d) denote the inequality, equality and side constraints, respectively. Problem P1 does not yield a unique solution on its own, as it requires a preference or prioritization of objectives to obtain a single optimum solution. The NC method requires anchor points, g_i^* , or optimum vertices to obtain the desired optimal solutions. The i -th anchor point (or end point) is obtained when the generic i -th objective is minimized independently.

Table 3.2: Definitions and mathematical formulations of some of the relevant multiobjective optimization keywords used in this section.

TERMS	MATHEMATICAL FORMULATION	DEFINITION
<i>Multiobjective optimization</i>	<p>Problem P1</p> $\min_x \{g_1(x), g_2(x), \dots, g_n(x)\}, (n \geq 2)$ <p>subject to</p> $f_j(x) \leq 0 \quad (1 \leq j \leq r)$ $h_k(x) = 0 \quad (1 \leq k \leq s)$ $x_l \leq x_i \leq x_u \quad (1 \leq i \leq n_x)$ <p>The vector x denotes the design variables and g_i denotes the ith objective function.</p>	A multiobjective optimization is a problem involving several competing objectives and constraints. The solution of this problem is considered the best solution that satisfies the conflicting objectives.
<i>Pareto solution</i>	Solutions joining the anchor points and part of the feasible space.	A Pareto solution is one where any improvement in one objective can only take place at the cost of another objective. A Pareto set is a set of Pareto-optimal solutions.
<i>Design parameters</i>	The vector x denotes the design parameters	A design parameter is a parameter over which the designer has direct control. Other terms used in the literature for design parameters include decision variables, design variables or decision parameters.
<i>Design metric</i>	The variable $g(x)$ denotes the vector of design metrics.	A design metric refers to an objective measure of a design attribute. Other commonly used terms are objective functions, design criterion, figure-of-merit, goal and performance metric.
<i>Design constraint</i>	$f_j(x) \leq 0 \quad (1 \leq j \leq r)$ $h_k(x) = 0 \quad (1 \leq k \leq s)$ $x_l \leq x_i \leq x_u \quad (1 \leq i \leq n_x)$	A design constraint indicates the lower or upper bounds in the design metrics or design parameters
<i>Anchor value</i>	<p>Problem PU$_i$</p> $\min_x \{g_i(x)\} \quad (1 \leq i \leq n)$ <p>subject to</p> $f_j(x) \leq 0 \quad (1 \leq j \leq r)$ $h_k(x) = 0 \quad (1 \leq k \leq s)$ $x_l \leq x_i \leq x_u \quad (1 \leq i \leq n_x)$ <p>or</p> $g^{i*} = [g_1(x^{i*}) \quad g_2(x^{i*}) \dots g_n(x^{i*})]^T$ <p>where $x^{i*} = \arg\{\min g_i(x)\}$</p>	The value obtained for a particular design objective if that design metric alone is optimized, given the bounds on the design parameters
<i>Nadir point</i>	$g^N = [g_1^N \quad g_2^N \dots g_n^N]^T$ <p>where g_i^N is defined as</p> $g_i^N = \max_x g_i(x)$	A point in the design space where all the objectives are simultaneously at their worst values

Figure 3.1b presents a schematic of a Pareto set for a bi-objective problem. If the design metric g_1 alone is optimized (maximized), then the optimal value is $g_1^*(P_1)$. Similarly, if the design metric g_2 alone is optimized, then the optimal value is $g_2^*(P_2)$. Here g_1^* and g_2^* are the anchor values for design metrics g_1 and g_2 , respectively. The ideal or Utopian solution (g_1^*, g_2^*) obtained by the individual maximization of the objective functions is generally not a feasible solution of the multiobjective optimization problem. The arc joining points P_1 and P_2 is the Pareto frontier that represents the optimal tradeoff solutions. Generally, the desired solution can be chosen from the Pareto set; the line joining two anchor points in bi-objective cases, the *utopia line*, and the plane that comprises all anchor points in the multiobjective case, the *utopia hyper plane*. The anchor points are obtained by solving Problem PUi, defined as follows

Problem PUi

$$\min_x \{g_i(x)\}, \quad (1 \leq i \leq n) \tag{3.10a}$$

subject to:

$$f_j(x) \leq 0 \quad (1 \leq j \leq r) \tag{3.10b}$$

$$h_k(x) = 0 \quad (1 \leq k \leq s) \tag{3.10c}$$

$$x_l \leq x_i \leq x_u \quad (1 \leq i \leq n_x) \tag{3.10d}$$

3.3.4 Normal Constraint Method

As seen in Figure 3.1c, the NC method is based on the design space reductions using reduction constraints. The reduction constraint is constructed by ensuring the orthogonality by constructing the dot product between the normal \vec{w} and r_0 an arbitrary point on a plane. The vector equation of a plane is expressed as

$$\vec{w} \cdot (r - r_0) = 0 \tag{3.11}$$

To solve for multiobjective solutions, a reduced feasible space is constructed using the above equation as

$$\vec{w} \cdot (r - g) \leq 0 \tag{3.12}$$

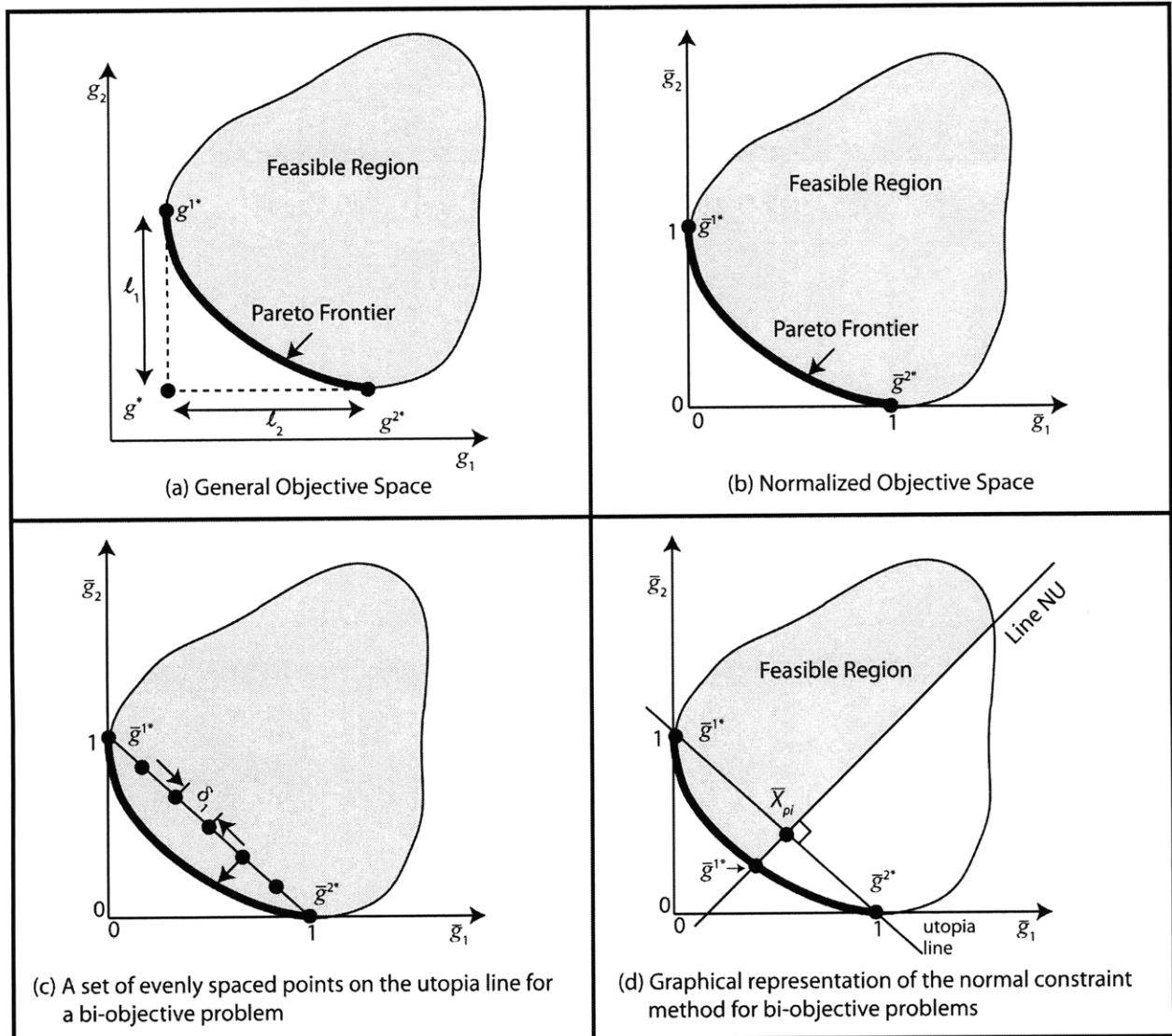


Figure 3.2: Steps involved for obtaining bi-objective Pareto optimal solutions using the Normalized Normal Constraint method for minimization. The mutually orthogonal axes g_1 , and g_2 , represent the individual design objectives. **a)** Pareto frontier for a minimization problem and the anchor points obtained using nonlinear optimization. **b)** The usage of anchor points to work in a normalized objective space. **c)** Drawing the utopia line and constructing evenly spaced points on the utopia line. **d)** Constructing the normal on the utopia line and reducing the feasible space.

where g is any point in the feasible space. Figure 3.2a shows the non-normalized design space and the Pareto frontier of a bi-objective problem. Figure 3.2b represents the normalized Pareto frontier in the normalized design space. In the normalized objective space, all anchor points are one unit away from the utopia point, and the utopia point is at the origin. A bar over a variable implies that it is normalized. The two anchor points denoted by g_1^* and g_2^* , are obtained by

successively minimizing the first and second design metrics (Problem PU_i) by solving Problem PU1 and PU2 respectively. The line joining these two points is the *utopia line*. The actual optimization takes place in the normalized design metric space. Let \bar{g} be the normalized form of g and g^u , the utopia point defined as

$$g^u = [g_1(x^{1*}) \quad g_2(x^{2*})]^T \quad (3.13)$$

and ℓ_1 and ℓ_2 be the distances between g^{2*} and g^{1*} , and the Utopia point, g^u , respectively (Figure 3.2a). Then

$$\ell_1 = g_1(x^{2*}) - g_1(x^{1*}) \quad (3.14)$$

$$\ell_2 = g_2(x^{1*}) - g_2(x^{2*}) \quad (3.15)$$

The normalized design objectives can then be evaluated as

$$\bar{g}^T = \left[\frac{g_1(x) - g_1(x^{1*})}{\ell_1} \quad \frac{g_2(x) - g_2(x^{2*})}{\ell_2} \right] \quad (3.16)$$

\bar{N}_1 is defined as the direction from \bar{g}^{1*} to \bar{g}^{2*} , yielding

$$\bar{N}_1 = \bar{g}^{2*} - \bar{g}^{1*} \quad (3.17)$$

Next, the utopia line is divided into $m_1 - 1$ segments, resulting in m_1 points. A normalized increment, δ_1 along the direction \bar{N}_1 for a prescribed number of solutions, m_1 , is obtained as

$$\delta_1 = \frac{1}{m_1 - 1} \quad (3.18)$$

As seen in Figure 3.2b, the next step involves generating a set of evenly distributed points on the utopia line as

$$\bar{X}_{pj} = \alpha_{1j} \bar{g}^{1*} + \alpha_{2j} \bar{g}^{2*} \quad (3.19a)$$

where

$$0 \leq \alpha_{1j} \leq 1 \quad (3.19b)$$

$$\sum_{k=1}^2 \alpha_{kj} = 1 \quad (3.19c)$$

and α_{1j} is incremented by δ_1 between 0 and 1 (Figure 3.2c), with values of j as $j \in \{1, 2, \dots, m_1\}$.

Figure 3.2c shows one of the generic points intersecting the segments used to define a normal to the utopia line. This normal line is used to reduce the feasible space as indicated in Figure 3.2d. As can be seen, if we minimize \bar{g}_2 the resulting optimum point is \bar{g}^{2*} . By translating the normal line, we can see that a corresponding set of solutions will be generated. This is essentially done by generating a corresponding set of Pareto points by solving a succession of optimization runs of Problem P2. Each optimization run corresponds to a point on the utopia line. Specifically, for each generated point on the utopia line, solve for the j th point.

Problem P2 (For j th point)

$$\min_x \{\bar{g}_2(x)\} \quad (3.20a)$$

subject to:

$$f_j(x) \leq 0 \quad (1 \leq j \leq r) \quad (3.20b)$$

$$h_k(x) = 0 \quad (1 \leq k \leq s) \quad (3.20c)$$

$$x_i \leq x_i \leq x_u \quad (1 \leq i \leq n_x) \quad (3.20d)$$

$$\bar{N}_1(\bar{g} - \bar{X}_{p_j})^T \leq 0 \quad (3.20e)$$

This results in a set of vectors for the design parameters, one vector x for each Pareto point. Then, design objectives are computed by evaluating the non-normalized design metrics that correspond to each Pareto point. The non-normalized design objectives can be obtained through an inverse mapping of Equation (3.16) by using the relation

$$g = [\bar{g}_1 \ell_1 + g_1(x^{1*}) \quad \bar{g}_2 \ell_2 + g_2(x^{2*})]^T \quad (3.21)$$

Importantly, we note that the generation of the set of Pareto points is performed in the normalized objective space, which results in critically beneficial scaling properties. Since some of the points generated in some pathological cases will be dominated by other points in the set, we use a *Pareto filter* (Table 3.3) to finally compute the true Pareto optimal solutions. This filter compares a point generated on the Pareto frontier with every other generated point. If a point is not globally Pareto, it is discarded. The steps involved and the essential mathematical formulation for the NC method for a n -objective case are presented in Table 3.2.

Table 3.3: Pareto filtering algorithm.

STEPS	ACTION
<i>Step-1:</i>	Initialize Initialize the algorithm indices and variables: $i = 0, j = 0, k = 1$ and $m =$ number of generated solutions; $m = f(m_k)$
<i>Step-2:</i>	Set $i = i + 1; j = 0$
<i>Step-3</i>	Eliminate non-global Pareto points $j = j + 1$ If $i = j$ go to the beginning of Step 3. Else continue. If $\mu^i \neq \mu^j$ and $(\mu^i - \mu^j)_s \geq 0, \forall s$ then μ^i is not a global Pareto point. Go to Step 4. Else if $j = m$ Then μ^j is a global Pareto point. $p^k = \mu^j$ $k = k + 1$ Go to Step 4. Else go to the beginning of Step 3.
<i>Step-4</i>	If $i \neq m$, go to Step 2, else end.

3.3.5 Normal Constraint Energy and Flux Balance Analysis (NCEFBA)

This section combines FBA, EBA and NC constraints. In the combined EBA and FBA, nonlinear thermodynamic constraints analogous to electrical circuit system constraints are included with the linear FBA constraints. The addition of nonlinear thermodynamic constraints leads to a nonlinear optimization problem. To avoid repetition in the presented NC method we will show here only the fluxes that are changed in the previously presented NC. There are changes in anchor points which lead to a different utopian hyperplane. Further, both FBA and EBA constraints are added to the NC formulation with the optimized quantity being the desired fluxes as objectives.

Computation of the Utopia hyperplane: The anchor points for NCEFBA are obtained by solving the Problem PUi, which is now defined as follows:

Problem PUi

$$\min_x \{g_i(x)\} \quad (1 \leq i \leq n) \tag{3.22a}$$

subject to:

$$S_u J_u = -S_m J_m \quad (3.22b)$$

$$K^T \Delta\mu = K^T S^T \mu = 0 \quad (3.22c)$$

$$J = J_+ - J_- \quad (3.22d)$$

$$-J^j \Delta\mu^j = -RT(J_+^j - J_-^j) \ln\left(\frac{J_-^j}{J_+^j}\right) \geq 0 \quad (3.22e)$$

$$hdr = -J^T \Delta\mu > 0 \quad (3.22f)$$

$$(hdr)_{lb} \leq hdr \leq (hdr)_{ub} \quad (3.22g)$$

$$J_{lb} \leq J \leq J_{ub} \quad (3.22h)$$

$$0 \leq J_+ \leq \infty \quad (3.22i)$$

$$0 \leq J_- \leq \infty \quad (3.22j)$$

$$J_{lb}^{ext} \leq J^{ext} \leq J_{ub}^{ext} \quad (3.22k)$$

$$\Delta\mu_{lb} \leq \Delta\mu \leq \Delta\mu_{ub} \quad (3.22l)$$

where vector x is defined as

$$x^T = [J^T \quad \Delta\mu^T \quad J_+^T \quad J_-^T] \quad (3.22n)$$

and the boundary constraints are meant to be satisfied component-wise.

Computation of Pareto points: Once anchor points are obtained, a set of well-distributed Pareto solutions are generated in the normalized objective space, by solving Problem Pn:

Problem Pn (for j th point)

$$\min_x \{\bar{g}_n(x)\} \quad (3.23a)$$

subject to:

$$S_u J_u = -S_m J_m \quad (3.23b)$$

$$K^T \Delta\mu = K^T S^T \mu = 0 \quad (3.23c)$$

$$J = J_+ - J_- \quad (3.23d)$$

$$-J^j \Delta\mu^j = -RT(J_+^j - J_-^j) \ln\left(\frac{J_-^j}{J_+^j}\right) \geq 0 \quad (3.23e)$$

$$hdr = -J^T \Delta\mu > 0 \quad (3.23f)$$

$$(hdr)_{lb} \leq hdr \leq (hdr)_{ub} \quad (3.23g)$$

$$J_{lb} \leq J \leq J_{ub} \quad (3.23h)$$

$$0 \leq J_+ \leq \infty \quad (3.23i)$$

$$0 \leq J_- \leq \infty \quad (3.23j)$$

$$J_{lb}^{ext} \leq J^{ext} \leq J_{ub}^{ext} \quad (3.23k)$$

$$\Delta\mu_{lb} \leq \Delta\mu \leq \Delta\mu_{ub} \quad (3.23l)$$

$$\bar{N}_k (\bar{g} - \bar{X}_{p_j})^T \leq 0 \quad (1 \leq k \leq n-1) \quad (3.23n)$$

where vector x is defined as

$$x^T = [J^T \quad \Delta\mu^T \quad J_+^T \quad J_-^T] \quad (3.23o)$$

and the boundary constraints are meant to be satisfied component-wise.

3.3.6 Hepatic Function Specific Fluxes for Pareto Optimization

The main goal that we wish to achieve in the BAL device is for hepatocytes to perform at the highest level of liver specific functions. Therefore, for hepatic metabolic optimization, the set of objective functions maximizing urea, albumin, NADPH and glutathione synthesis, ATP generation are chosen. NADPH, which is produced in the pentose phosphate pathway (PPP), is primarily used in nonproliferating hepatocytes for cytochrome p450 dependent oxidation reactions (detoxification reactions) and glutathione synthesis. Hence, to increase the NADPH flux, the NADPH-generating oxidative branch of the PPP represented in a lumped fashion as reaction 46 (Table 3.1) is maximized. As a marker of secretory liver specific function, we use albumin synthesis, which is maximized by modulating flux 47. Urea synthesis is primarily derived from ammonia and aspartate generated through transamination reactions and is

maximized by modulating reaction 16. The tripeptide glutathione (GSH, γ -Glu-Cys-Gly) is an important reductant and has many detoxifying and cytoprotective effects. The synthesis of glutathione is increased by maximizing reaction 48. The ATP generation is maximized by increasing the TCA cycle fluxes (11, 43, 44). Figure 2.2 presents the comprehensive hepatic metabolic network with all the cycles shown with each reaction and the constraints for this metabolic network are listed in Table 2.6.

3.4 RESULTS

Pareto optimal solutions are accepted solutions of multi-objective optimization problems, and can serve as a useful tool to understand the underlying tradeoffs between conflicting design objectives and cellular phenotypes. Pareto optimality analysis has been applied to numerous disciplines and more recently to cellular systems (Vo et al., 2004). As mentioned earlier, usage of FBA alone can lead to thermodynamically infeasible fluxes. Consequently, we chose to combine both FBA and EBA constraints with Pareto optimality to optimize hepatocellular function in the context of a BAL device. As part of this analysis, we first obtained Pareto frontiers between various bi-objective combinations of liver specific functions (albumin synthesis, urea secretion, NADPH synthesis, GSH synthesis, and ATP generation). This was done for hepatocytes in a gluconeogenic state and in a glycolytic state. Next, for a representative case, i.e. ATP generation vs. urea secretion, we compared the Pareto frontier using NCEBFBA (i.e. both FBA and EBA) with FBA alone. Lastly, we obtained the Pareto solutions in the presence of measurement constraints. The experimentally measured flux data for gluconeogenic and glycolytic state were taken from Chan et al., 2003a and Chan et al., 2003b.

3.4.1 Pareto Frontiers of Liver Specific Functions

Pareto optimality analysis here is carried out first in gluconeogenic hepatocytes (Figure 3.3) and then for glycolytic hepatocytes (Figure 3.5). This distinction was necessary because the hepatic metabolic network used in each case is different. Note that in both figures, the same panels analyze the same objectives to facilitate the comparison of results obtained in the gluconeogenesis and glycolysis modes. For each Pareto curve shown in Figure 3.3 for

gluconeogenic hepatocytes, Figure 3.4 shows the relative flux changes that are necessary when switching objective priority. This information is summarized in Table 3.4, where important groups are clustered together. Similarly, for each Pareto curve shown in Figure 3.5 for glycolytic hepatocytes, Figure 3.6 and Table 3.5 summarize the flux changes that are necessary when switching objective priority. Tables 3.6 and 3.7 provide the comprehensive set of flux data that are summarized in Figures 3.3 and 3.5, respectively.

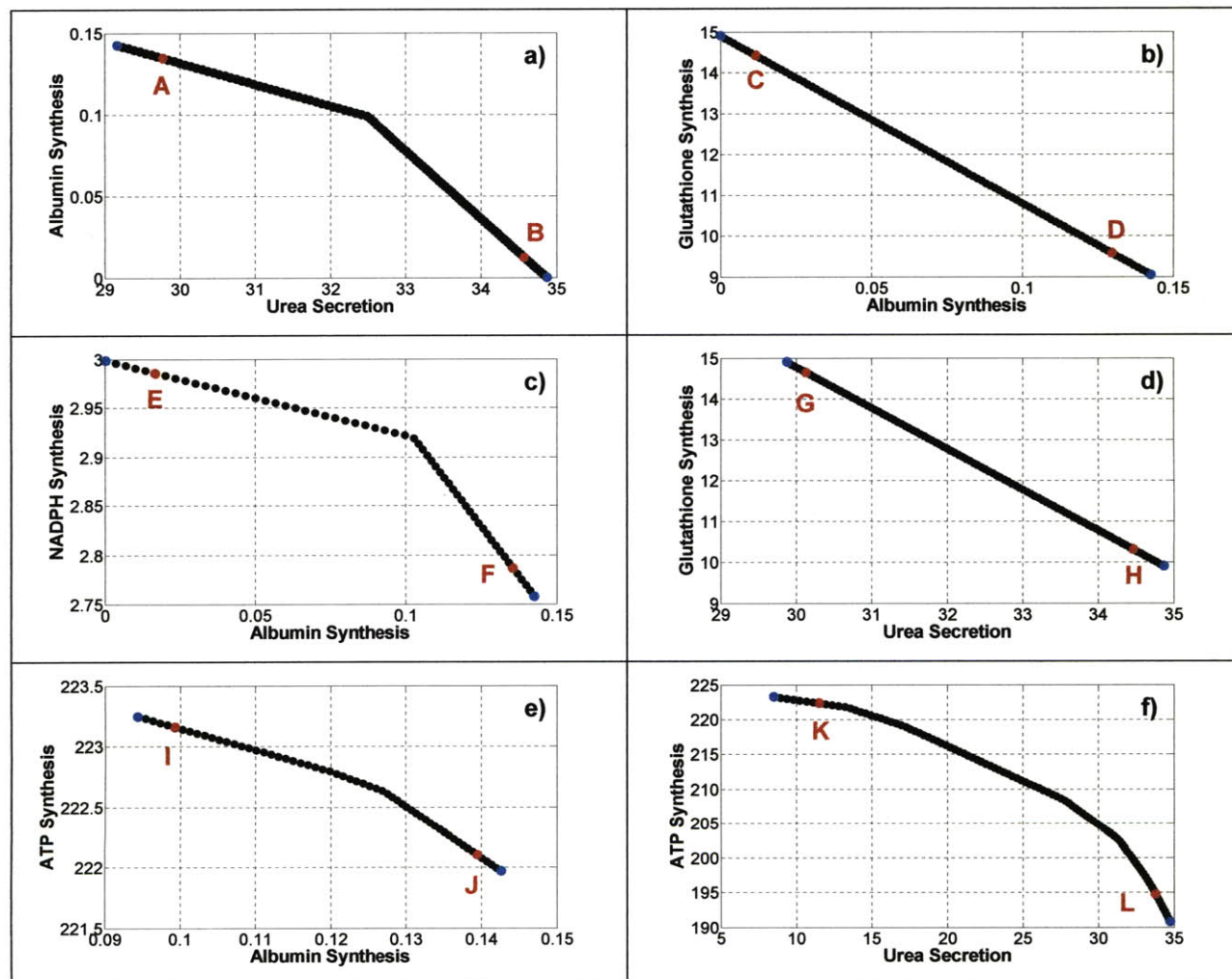


Figure 3.3: Pareto frontiers for bi-objective problems in hepatocytes operating in a gluconeogenic mode. Five major hepatic functions were considered: albumin, urea, ATP, NADPH and glutathione synthesis. **a)** Albumin vs. urea synthesis. **b)** Glutathione vs. albumin synthesis. **c)** NADPH vs. albumin synthesis. **d)** Glutathione vs. urea synthesis. **e)** ATP vs. albumin synthesis. **f)** ATP vs. urea synthesis. The blue circles are the anchor points, black circles are Pareto optimal solutions, and red circles labeled A through L are selected Pareto solutions for which flux distributions are shown in Table 3.6.

The bi-objective Pareto optimal solutions were first obtained using the NCEFBA approach for various binary combinations of liver-specific objectives in gluconeogenic

hepatocytes (Figure 3.3). The Pareto frontiers for albumin synthesis vs. urea secretion, glutathione synthesis vs. albumin synthesis, NADPH synthesis vs. albumin synthesis, glutathione synthesis vs. urea secretion, ATP generation vs. albumin synthesis, and ATP generation vs. urea secretion are shown in Figures 3.3a, 3.3b, 3.3c, 3.3d, 3.3e, and 3.3f, respectively.

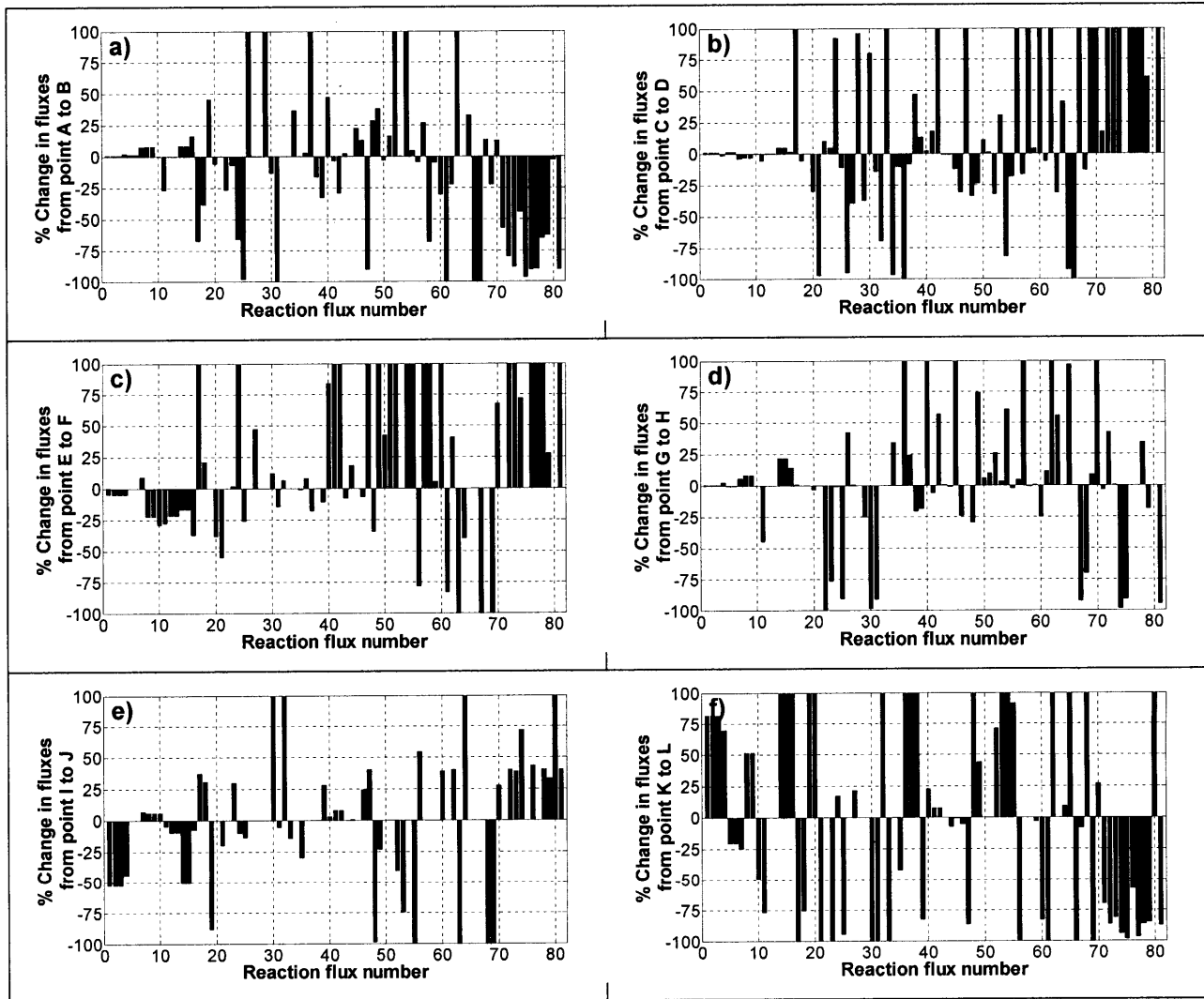


Figure 3.4: Distribution of flux changes when moving along the Pareto surface in Figure 3.3. a) % Flux changes from point A to point B in Figure 3.3a. b) % Flux changes from point C to point D in Figure 3.3b. c) % Flux changes from point E to point F in Figure 3.3c. d) % Flux changes from point G to point H in Figure 3.3d. e) % Flux changes from point I to point J in Figure 3.3e. f) % Flux changes from point K to point L in Figure 3.3f. The corresponding flux values are in Table 3.6. Note that the % flux changes for all figures are on y-axis and the corresponding reaction flux number is shown on the horizontal axis. Only changes up to 100% are shown in the figure.

Table 3.4: Summary of results indicating the changes in hepatic gluconeogenic fluxes in Figures 3.3 and 3.4 when going from selected Pareto optimal solutions “A” to “B”, “C” to “D”, “E” to “F”, “G” to “H”, “I” to “J” and “K” to “L” for which flux distributions are shown in Table 3.6. **a)** “A” to “B” indicates from high albumin/low urea synthesis to low albumin/high urea synthesis. **b)** “C” to “D” indicates from high glutathione/low albumin synthesis to low glutathione/high albumin synthesis. **c)** “E” to “F” indicates from high NADPH/low albumin synthesis to low NADPH/high albumin synthesis. **d)** “G” to “H” indicates from high glutathione/low urea synthesis to low glutathione/high urea synthesis. **e)** “I” to “J” indicates from high ATP/low albumin synthesis to low ATP/high albumin synthesis. **f)** “K” to “L” indicates from high ATP/low urea synthesis to low ATP/high urea synthesis. Here, s-dec indicates significantly decreased, dec indicates decreased, m-dec moderately decreased. Similarly, s-inc indicates significantly increased, inc indicates increased, m-inc moderately increased. Inc-r, indicates increased in reverse direction.

Case	Albumin synthesis (47)	GSH synthesis (48)	NADPH synthesis (46)	Urea secretion (16)	β oxidation (40)	Electron transport (43,44)	Lipid uptake (52)	Gluconeogenesis (2,6)	Urea cycle (14,15)	TCA cycle (8)	TCA cycle (13)	Lipid stored (57)	Asp uptake (69)	Gluconeogenic amino acids (60, 67, 69, 71, 73, 75, 81, 76, 77,79)	Ketogenic amino acids (72,78)
A to B	s-dec	inc	inc	inc	inc	m-inc	s-inc	m-inc	inc	m-inc	same	inc	dec	dec	s-dec
C to D	inc	dec	dec	inc	same	m-dec	dec	m-dec	m-inc	m-dec	m-dec	m-dec	s-inc	s-inc	s-inc
E to F	s-inc	dec	dec	dec	s-inc	dec	s-inc	m-dec	dec	dec	dec	s-inc	same	s-inc	s-inc
G to H	same	dec	dec	inc	s-inc	m-inc	inc	m-inc	inc	inc	same	s-inc	inc	dec	inc
I to J	s-inc	s-dec	inc	m-dec	m-inc	m-inc	dec	s-dec	s-dec	inc	dec	same	inc-r	inc	s-inc
K to L	s-dec	s-inc	m-dec	s-inc	inc	m-dec	s-inc	s-inc	s-inc	s-inc	same	same	inc-r	s-dec	s-dec

As seen in Figure 3.3, all of these objectives have a tradeoff region with each other. For example, we cannot have both albumin and urea synthesis at their maximal values. Additionally, there is a tradeoff between other liver specific functions such as GSH and albumin synthesis, NADPH and albumin synthesis, GSH synthesis and urea secretion, ATP synthesis and albumin synthesis, and ATP synthesis and urea secretion. As seen in these figures, the tradeoff region or range of Pareto optimal solutions (how far the optimal value is from the “anchor value”) for albumin synthesis is very high compared to NADPH, GSH and ATP synthesis and urea secretion. Several other combinations were also tested and all of them indicated Pareto optimality between various objectives (data not shown). Figure 3.4 presents the metabolic flux profiling of Pareto optimal fluxes throughout the tradeoff region, which shows the changes required in flux values and direction (i.e. increasing or decreasing) as the objective preference is changed from one objective to another. The corresponding flux values for these cases are presented in Table 3.6. The Pareto frontiers for various binary combinations of objectives were also obtained for glycolytic hepatocytes (Figure 3.5). As in the case of gluconeogenic

hepatocytes, these objectives have a tradeoff region with each other, and some objectives change over a wide range (e.g., albumin, urea and GSH) while some change only little (NADPH and ATP). Figure 3.6 presents the metabolic flux profiling of Pareto optimal fluxes throughout the tradeoff region, which shows the changes required in flux values and direction (i.e. increasing or decreasing) as the objective preference is changed from one objective to another. The corresponding flux values for these cases are presented in Table 3.7.

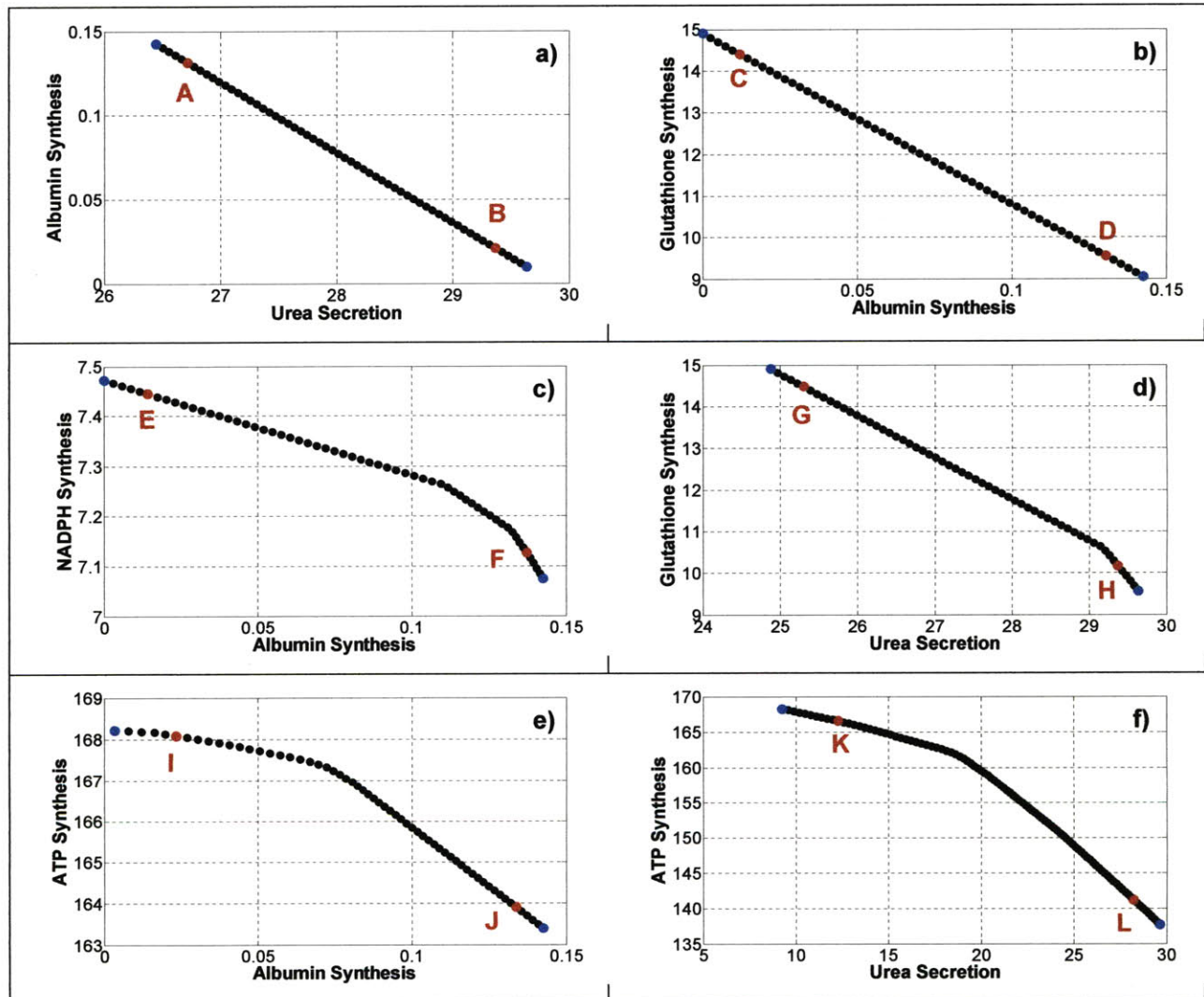


Figure 3.5: Pareto frontiers for bi-objective problems in hepatocytes operating in a glycolysis mode. Five major hepatic functions were considered: albumin, urea, ATP, NADPH and glutathione synthesis. **a)** Albumin vs. urea synthesis. **b)** Glutathione vs. albumin synthesis. **c)** NADPH vs. albumin synthesis. **d)** Glutathione vs. urea synthesis. **e)** ATP vs. albumin synthesis. **f)** ATP vs. urea synthesis. The blue circles are the anchor points, black circles are Pareto optimal solutions, and red circles labeled A through L are selected Pareto solutions for which flux distributions are shown in Table 3.7.

Figure 3.3a examines the tradeoff between albumin and urea secretion in gluconeogenic hepatocytes. Many flux changes were required to go from Pareto optimal solutions “A” to “B” in Figure 3.3a, in other words, when going from a state of high albumin/low urea secretion rate to a low albumin/high urea secretion rate.

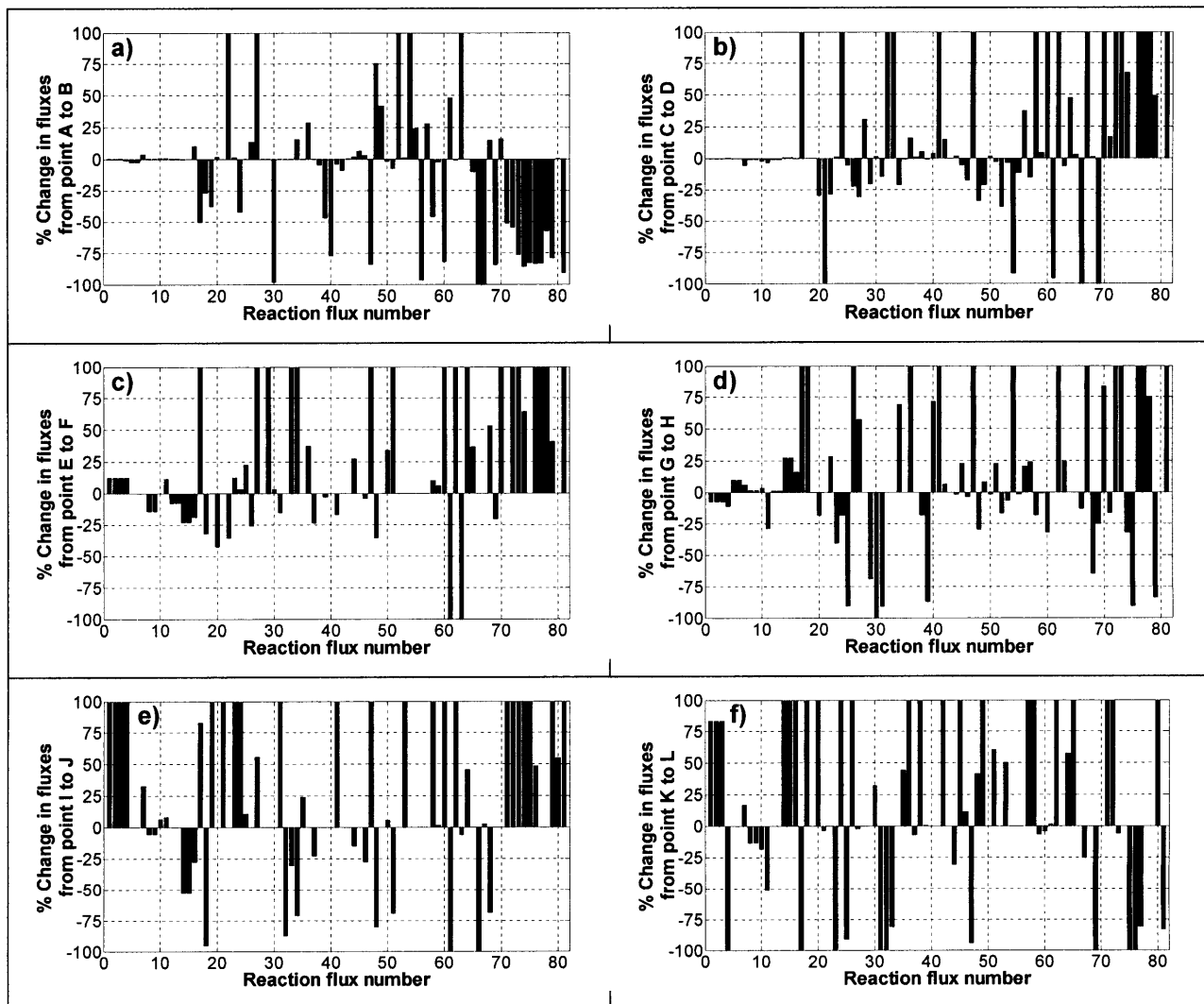


Figure 3.6: Distribution of flux changes when moving along the Pareto surface in Figure 3.5. **a)** % Flux changes from point A to point B in Figure 3.5a. **b)** % Flux changes from point C to point D in Figure 3.5b. **c)** % Flux changes from point E to point F in Figure 3.5c. **d)** % Flux changes from point G to point H in Figure 3.5d. **e)** % Flux changes from point I to point J in Figure 3.5e. **f)** % Flux changes from point K to point L in Figure 3.5f. The corresponding flux values are in Table 3.7. Note that the % flux changes for all figures are on y-axis and the corresponding reaction flux number is shown on the horizontal axis. Only changes up to 100% are shown in the figure.

Table 3.5: Summary of results indicating the changes in hepatic glycolytic fluxes in Figures 3.5 and 3.6 when going from selected Pareto optimal solutions “A” to “B”, “C” to “D”, “E” to “F”, “G” to “H”, “I” to “J” and “K” to “L” for which flux distributions are shown in Table 3.7. **a)** “A” to “B” indicates from high albumin/low urea synthesis to low albumin/high urea synthesis. **b)** “C” to “D” indicates from high glutathione/low albumin synthesis to low glutathione/high albumin synthesis. **c)** “E” to “F” indicates from high NADPH/low albumin synthesis to low NADPH/high albumin synthesis. **d)** “G” to “H” indicates from high glutathione/low urea synthesis to low glutathione/high urea synthesis. **e)** “I” to “J” indicates from high ATP/low albumin synthesis to low ATP/high albumin synthesis. **f)** “K” to “L” indicates from high ATP/low urea synthesis to low ATP/high urea synthesis. Here, s-dec indicates significantly decreased, dec indicates moderately decreased. Similarly, s-inc indicates significantly increased, inc indicates increased, m-inc moderately increased. Inc-r, indicates increased in reverse direction.

Case	Albumin synthesis (47)	GSH synthesis (48)	NADPH synthesis (46)	Urea secretion (16)	β oxidation (40)	Electron transport (43,44)	Lipid uptake (52)	Gluconeogenesis (2,6)	Urea cycle (14,15)	TCA cycle (8)	TCA cycle (13)	Lipid stored (57)	Asp uptake (69)	Glucogenic amino acids (60, 67, 69, 71, 73, 75, 81, 76, 77,79)	Ketogenic amino acids (72,78)
A to B	s-dec	s-inc	m-inc	m-inc	dec	m-inc	s-inc	m-dec	same	m-inc	m-inc	inc	dec	dec	s-dec
C to D	s-inc	dec	dec	m-dec	m-inc	m-dec	dec	m-dec	m-inc	m-dec	m-dec	dec	s-inc-r	s-inc	s-inc
E to F	s-inc	s-dec	m-dec	dec	same	inc	same	m-inc	dec	dec	m-dec	same	dec	s-inc	s-inc
G to H	s-inc	dec	m-dec	inc	s-inc	m-dec	m-dec	m-dec	inc	m-inc	m-inc	inc	dec	dec	s-inc
I to J	s-inc	s-dec	dec	dec	same	m-dec	same	s-inc	s-dec	dec	same	same	same	inc	s-inc
K to L	s-dec	s-inc	m-inc	s-inc	same	s-dec	same	inc	s-inc	dec	same	s-inc	s-inc	inc	inc

As seen in Figure 3.4a and summarized in Table 3.4, this change required increasing marginally gluconeogenic fluxes (1-9), increasing moderately urea cycle fluxes (14-15), decreasing formation of glutamate (38-39), increasing oxidation of triglycerides (52), decreasing uptake of both glucogenic (proline, 60; serine, 67; aspartate, 69; threonine, 71; phenylalanine, 73; methionine, 75; ; valine, 76; isoleucine, 77; glutamine, 79, tyrosine, 81) and ketogenic (lysine, 72; leucine, 78) amino acids, with the exception of glycine (34) uptake, which was increased. Arginine (64) uptake rate did not change because it was at its maximum at both optimal points. Histidine (18) uptake increased because it results in an increase of α -ketoglutarate. The uptake of pyruvate-forming amino acids (alanine, 66; serine, 67; and threonine, 71), fumarate-forming amino acids (phenylalanine, 73; and tyrosine, 81), and succinyl CoA-forming amino acids (methionine, 75; valine, 76; and threonine, 71), which play a major role in albumin synthesis, were decreased. When considering the tradeoff between albumin and urea secretion in glycolytic hepatocytes (Figures 3.5a and 3.6a), the main difference with the case of gluconeogenic hepatocytes was in the β -oxidation flux which was higher in gluco-

neogenesis and decreased in glycolysis. This is expected because glycolysis is dominant in the fed state and gluconeogenesis in the fasted state. Further, the production of ketone bodies through β -oxidation occurs mostly in the fasted state.

Next, we investigated the tradeoff between glutathione and albumin synthesis in gluconeogenic hepatocytes. The Pareto curve is shown in Figure 3.3b. Going from Pareto optimal solutions “C” to “D” also required many flux changes, which are reported in Figure 3.4b and Table 3.4. There was a marginal increase in urea cycle fluxes (14-15), a decrease in lipid uptake (52) and lipid stored (57), and a significant increase in aspartate uptake (69). Additionally, the uptake of both gluconeogenic amino acids (60, 67, 69, 71, 73, 76, 77, 79, and 81,) and ketogenic amino acids (72, 78) increased. The corresponding flux values for NADPH synthesis decreased. There were no significant differences in the results of this analysis when considering glycolytic hepatocytes (Figures 3.5b and 3.6b, and Table 3.5).

Considering the tradeoff between NADPH synthesis and albumin synthesis (Figures 3.3c and 3.5c), flux changes required to move from points E to F along the Pareto frontier were generally similar in both gluconeogenic and glycolytic hepatocytes (Figures 3.4c and 3.6c), with the exception of β -oxidation, electron transport (43, 44), lipid uptake and lipid storage fluxes. This is because fatty acid synthesis significantly consumes NADPH (14 molecules of NADPH per molecule of palmitate).

Considering the tradeoff between glutathione synthesis and urea secretion (Figures 3.3d and 3.5d), the changes in flux required to move from points G to H along the Pareto frontier were also generally similar in both gluconeogenic and glycolytic hepatocytes (Figures 3.4d and 3.6d), with the exception of aspartate uptake (69).

Considering the tradeoff between ATP synthesis and albumin synthesis (Figures 3.3e and 3.5e), the changes in flux required to move from points I to J along the Pareto frontier significantly differed between gluconeogenic and glycolytic hepatocytes (Figures 3.4e and 3.6e), mainly with respect to gluconeogenesis fluxes (2-6), and TCA cycle fluxes (8-13). In the gluconeogenesis mode, TCA cycle fluxes are higher because of increased demand to produce ATP (gluconeogenesis consumes ATP as well), since glycolysis itself produces ATP (2 molecules of ATP for 1 molecule of glucose consumed).

Table 3.6: Flux values for the selected Pareto solutions shown in Figure 3.3 in gluconeogenesis mode.

Flux #	J (at A)	J (at B)	J (at C)	J (at D)	J (at E)	J (at F)	J (at G)	J (at H)	J (at I)	J (at J)	J (at K)	J (at L)
1	10.45616	10.53164	2.931751	2.962597	3.985876	3.786608	10.68169	10.75281	5.084417	2.411236	6.023587	10.94565
2	10.45616	10.53164	2.931751	2.962597	3.985876	3.786608	10.68169	10.75281	5.084417	2.411236	6.023587	10.94565
3	10.45616	10.53164	2.931751	2.962597	3.985876	3.786608	10.68169	10.75281	5.084417	2.411236	6.023587	10.94565
4	22.34042	22.81248	7.25294	7.144599	7.972752	7.574215	21.42532	21.97203	12.15883	6.812471	14.03717	23.86665
5	24.49237	24.81512	12.83025	13.00202	30	30	22.22823	22.16054	30	30	30	23.86765
6	24.49237	24.81512	12.83025	13.00202	30	30	22.22823	22.16054	30	30	30	23.86765
7	9.51886	10.23318	10.06462	9.666193	27.50332	30	9.749638	10.29762	19.2476	20.63132	17.90442	13.5001
8	5.026489	5.418401	7.949764	7.695878	17.05776	13.21547	7.525195	8.14122	11.43791	12.1179	9.897259	15
9	5.026489	5.418401	7.949764	7.695878	17.05776	13.21547	7.525195	8.14122	11.43791	12.1179	9.897259	15
10	5	5	5.005367	5	7.057761	5.000135	5	5	11.42791	12.1079	9.887259	5
11	6.900497	5.031	10.05995	9.483926	13.92636	10.08871	9.768923	5.434227	26.26654	24.96552	25.55738	5.858272
12	30	30	14.35929	14.28768	29.55444	23.1121	30	30	29.51789	26.62793	30	30
13	30	30	14.35929	14.28768	29.55444	23.1121	30	30	29.51789	26.62793	30	30
14	22.9995	24.869	3.98076	4.178844	15.52808	12.92339	20.13108	24.46577	3.151353	1.562414	4.342624	24.04173
15	22.9995	24.869	3.98076	4.178844	15.52808	12.92339	20.13108	24.46577	3.151353	1.562414	4.342624	24.04173
16	29.7685	34.56632	10.43101	10.59504	19.75209	12.45496	30.13084	34.46553	2.807173	2.601182	11.5165	33.76909
17	-1.631214	-0.536267	-0.528837	-2.006131	-1.539125	-7.981121	-0.497689	-0.50062	-6.653947	-9.146062	-5.662553	0
18	0.461622	0.284045	0.438175	0.412978	15.697	18.98731	0.1	0.1	15.05454	19.67738	13.46402	3.283547
19	0.330588	0.481115	0.1	0.1	0.1	0.1	0.1	0.1	0.847005	0.1	0.10295	0.216556
20	2.87565	2.707592	8.00082	5.613211	7.497758	4.648697	9.203206	8.907835	0.01	0.01	0.01	2.588962
21	0.001	0.001	0.038607	0.001001	1.949159	0.880762	0.001	0.001	6.714512	5.388064	7.242189	0.001
22	-0.971072	-0.714002	-0.906803	-1	-1	-1	-0.314121	-3.38E-11	-1	-1	-1	-1
23	2.151948	2.002633	5.577312	5.857419	22.02725	22.42578	0.802911	0.188519	17.84117	23.18753	15.96283	0.001
24	1.580908	0.533345	0.524345	1.007869	0.48324	1.282606	0.502921	0.49999	0.400743	0.360547	0.416733	0.48864
25	1.900497	0.031	0.5054581	4.483926	6.868599	5.088575	4.768923	0.434227	14.83863	12.85762	15.87012	0.858272
26	0.001	0.754159	0.030353	0.00154	0.001	0.001005	1.077168	1.534977	0.001	0.001	0.001	0.001
27	0.1	0.1	0.213869	0.129024	0.114423	0.168828	0.100009	0.1002	0.1	0.1	0.1	0.12156
28	0.1	0.1	0.318582	0.624913	0.1	0.100001	0.1	0.1	0.1	0.1	0.1	0.1
29	0.864605	3.048025	1.741664	1.088041	0.1	0.1	0.14072	0.105309	0.1	0.1	0.1	0.1
30	2.980626	2.58457	1.144032	2.063769	2.654201	2.967936	1.498649	0.020831	0.985504	2.16986	1.701053	0.00001
31	1.879497	0.01	4.930819	4.223332	4.899439	4.187174	4.747923	0.413227	4.404457	4.163284	4.500398	0.01
32	0.01	0.01	0.032753	0.01	0.01	0.010637	0.01	0.01	0.01	0.0119159	0.01	0.837272
33	0.01	0.01	0.052401	0.250593	0.01	0.010002	0.01	0.01	3.709856	3.187116	3.917529	0.01
34	1.870186	2.554259	0.329499	0.010005	0.01	0.01	0.849459	1.142287	0.01	0.01	0.01	0.01
35	-10	-9.999663	-3.64389	-3.25598	-10	-9.89663	-9.996213	-9.995851	-7.327574	-5.141357	-8.007158	-4.632452
36	7.307501	7.499843	0.017841	1E-05	6.765451	7.290581	0.934579	4.470369	0.00001	0.00001	0.00001	9.772798
37	-0.026489	-0.418401	-2.944397	-2.695878	-10	-8.215331	-2.525195	-3.14122	-0.01	-0.01	-0.01	-10
38	6.489144	5.411853	2.499289	3.68853	0.00001	0.00001	7.316898	5.816622	0.00001	0.00001	2.173876	4.727358
39	3.149794	2.111601	5.755599	6.499171	7.413951	6.613227	7.644235	6.203562	3.645606	4.669928	1.244402	0.215998
40	0.449495	0.661466	0.901383	0.920863	1.242776	2.287185	0.0502	0.542191	9.499658	9.784768	9.555281	11.74025
41	-0.658454	-0.632682	-0.335649	-0.396005	0.000194	0.321518	-0.707952	-0.665219	27.12534	29.2863	27.35839	29.40036
42	0.923454	0.654823	0.219049	0.613403	0.484434	1.605129	0.872136	1.369748	27.52708	29.64784	27.77612	29.89
43	40.07639	40.97647	45.94547	45.59571	60.43861	55.72856	52.62026	53.04138	100	100	100	100
44	13.51905	13.52302	17.33898	17.20997	23.14003	27.41325	12.56987	12.42682	96.89554	97.13672	96.7996	89.38694
45	-1.428096	-1.749204	-1.389438	-1.219405	-0.001	-0.001	-0.061944	-0.466409	-1.99	-1.99	-1.99	-1.975352
46	0.471627	0.531642	1.414381	0.980214	2.985876	2.786608	1.091501	0.825673	1.134247	1.411236	1.000814	0.94565
47	0.134625	0.012612	0.01153	0.129611	0.01676	0.135471	0.00001	0.00001	0.099257	0.139453	0.083267	0.01136
48	6.837036	8.791876	14.42727	9.585937	14.21284	9.345691	14.64757	10.31288	0.083459	0.001	1.483113	9.39584
49	1.547404	2.136846	2.771774	2.11282	0.133939	3.035892	1.451659	2.528957	4.679441	3.598256	5.255245	7.57675
50	10.35786	10.00425	10.05528	11.15935	13.24545	18.89381	10.04733	10.59824	10	10	10	10
51	1.3159	1.52952	0.851862	0.860876	0.445374	1.465321	1.259451	1.381499	20	20	20	20
52	0.100346	0.302264	1.296234	0.877215	0.103022	1.749043	0.491578	0.617581	2.679441	1.598256	3.255245	5.57675
53	9.984536	9.999997	1.517371	1.982383	1	1	9.590187	9.927135	3.95017	1	5.022774	10
54	0.018962	0.085378	0.086102	0.015661	0.029918	1.285849	0.898136	1.444967	0.01	0.01	0.01	0.024648
55	4.318612	4.515504	3.504287	2.866019	1.543921	3.49074	2.722895	2.665903	1.774313	0.01	2.610203	5
56	4.46707	4.270178	0.516968	1.155236	2.477633	0.530796	1.298361	1.355353	3.235647	5	2.399749	0.01
57	1.447058	1.834582	1.47554	1.235066	0.030918	1.296849	0.960081	1.911377	2	2	2	2
58	1.715533	0.545957	0.535875	1.13748	0.5	1.418077	0.502931	0.5	0.5	0.5	0.5	0.5
59	32.27275	30.6738	34.93713	36.34932	43.60347	45.83755	34.4742	34.58693	100	100	100	96.53095
60	4.903353	3.426371	2.087567	4.97638	0.602803	4.164128	0.14102	0.105609	3.077716	4.283578	2.59801	0.040803
61	-1	4.063714	-0.999501	-0.938391	5	0.828969	4.477554	4.974097	-1	-1	-1	5
62	10	7.752075	0.248443	2.592236	7.100653	10	0.934779	4.470569	1.985154	2.789062	1.66535	10
63	0.279857	4.28547	3.950962	2.727669	4.224003	-0.468432	2.682862	4.183138	-0.34419	1.038758	5	5
64	10	10	6.726974	9.52687	4.626255	2.78288	10	10	2.037993	4.385631	9.172284	10
65	11.60736	15.43998	1.339349	0.1	20	20	8.297209	16.33414	0.1	0.1	0.1	20
66	5	-0.300308	-0.349845	4.89229	-1	-1	-1	-1	-1	-1	-1	0.204326
67	4.391241	-0.691909	0.975208	2.888563	-0.028565	5	4.359373	0.324949	5	5	5	4.565189
68	4.307866	4.870677	4.808851	4.176118	5	5	4.816294	1.405212	-0.239168	0.361694	0.888652	5
69	10	7.773064	0.687993	5.070415	-0.701049	0.071244	9.200605	9.999872	-1	0.88353	-1	10
70	1.577739	1.779562	0.015269	0.387445	0.14906	0.249809	0.562144	1.230536	7.637083	9.757845	7.886124	10
71	7.319273	3.124773	8.419921	9.891385	10	10	9.204536	8.909165	10	10	10	2.964845
72	7.136122	1.422572	0.641449	6.870939	0.889285	7.180965	1.077698	1.535507	5.261632	7.391988	4.41415	0.603085
73	3.600249	0.427901	0.513652	3.498918	0.550185	3.691073	0.100269	0.10046	2.680687	3.725768	2.264942	0.416922
74	5	2.773744	1.316984	4.007939	2.905602	5	1.498799	0.020981	2.474362	4.26165	2.950058	0.170411
75	2.687247	0.085669	5	5	5	5	4.747983	0.413287	5	5	5	0.078161
76	4.721873	0.451405	0.436307	4.546396	0.596603	4.75212	0.01035	0.01035	3.484002	5	2.924345	1.234876
77	1.760124	0.17395	0.202293	1.935539	0.227881	1.771124	0.01013	0.01013	5	5	5	0.157681
78	9.409182	3.260506	0.975186	7.268237	0.948565	7.596373	0.850019	1.142847	5.568404	7.819346	4.672951	0.646165
79	6.515418	2.42689	6.043852	9.739454	7.832953	10	7.644485	6.203812	6.127037	8.156243	3.326076	0.5
80	10	9.70443	10	10	10	10	10	10	0.00001	0.818557	0.00001	10
81	2.827124	0.264843	0.346846	3.217726	0.337539							

Table 3.7: Flux values for the selected Pareto solutions shown in Figure 3.5 in glycolysis mode.

Flux #	J (at A)	J (at B)	J (at C)	J (at D)	J (at E)	J (at F)	J (at G)	J (at H)	J (at I)	J (at J)	J (at K)	J (at L)
1	3.320282	3.33758	7.912406	7.88244	2.556222	2.873484	7.81942	7.224603	0.345462	1.526889	0.543029	0.995505
2	3.320282	3.33758	7.912406	7.88244	2.556222	2.873484	7.81942	7.224603	0.345462	1.526889	0.543029	0.995505
3	3.320282	3.33758	7.912406	7.88244	2.556222	2.873484	7.81942	7.224603	0.345462	1.526889	0.543029	0.995505
4	5.137323	5.075633	14.07575	14.1165	5.111443	5.745967	14.03731	12.48403	0.689924	3.052777	0.335526	0.00101
5	3.418129	3.330969	8.163969	8.130087	0.00001	0.00001	8.203415	9.000354	0.00001	0.00001	1E-05	0.00001
6	3.418129	3.330969	8.163969	8.130087	0.00001	0.00001	8.203415	9.000354	0.00001	0.00001	1E-05	0.00001
7	12.32408	12.77062	17.80928	16.82315	30	30	18.66582	19.77409	10.4446	13.88948	10.73724	12.52604
8	18.60514	18.67631	17.89311	17.69616	12.62882	10.82335	18.77853	19.08346	24.19015	22.79568	23.07392	20
9	18.60514	18.67631	17.89311	17.69616	12.62882	10.82335	18.77853	19.08346	24.19015	22.79568	23.07392	20
10	8.605136	8.676313	11.9678	11.70473	5	5	8.778528	9.083457	14.19015	15.07521	13.07392	10.69914
11	8.636136	8.707313	17.36348	16.79063	9.94211	11.07681	13.37344	9.518703	25.83536	27.97411	23.97935	11.66512
12	28.60514	28.67631	27.13863	26.86795	22.62882	20.82335	28.77853	29.08346	30	30	30	30
13	28.60514	28.67631	27.13863	26.86795	22.62882	20.82335	28.77853	29.08346	30	30	30	30
14	19.869	19.869	8.973567	9.028741	12.58671	9.64654	15.30508	19.46475	4.064643	1.925887	5.920655	18.23488
15	19.869	19.869	8.973567	9.028741	12.58671	9.64654	15.30508	19.46475	4.064643	1.925887	5.920655	18.23488
16	26.7175	29.36555	14.44119	14.38821	12.3232	9.978069	25.30484	29.36816	8.695292	6.284636	12.26472	28.23464
17	-1.617039	-0.800649	-0.558489	-1.660425	0	-7.501398	-5.24E-05	-0.299137	-1.943861	-3.553949	-0.510528	0
18	0.152018	0.111136	0.1	0.1	3.996811	2.712269	0.164013	0.927821	2.288453	0.105207	1.147754	2.426031
19	0.161049	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.772237	0.1
20	2.682588	2.735775	7.935394	5.584393	9.565107	5.474062	7.915699	6.468561	0.01	0.01	0.01	2.90982
21	0.001	0.001	0.133875	0.001	0.001	0.001	0.001	0.001	2.690735	5.577851	0.001038	0.001
22	-0.264356	-0.95412	-0.949376	-0.6777	-0.877953	-0.566872	-0.779019	-0.999998	-1	-1	-1	-1
23	1.719194	1.744664	5.911782	5.986414	5.111433	5.745957	5.833898	3.483673	0.689914	3.052767	0.335516	0.001
24	1.397487	0.806964	0.538067	1.374724	1.802044	1.862881	0.977117	0.795465	0.476712	1.866299	0.49983	1.00061
25	0.031	0.031	5.395675	5.085898	4.94211	6.076806	4.594917	0.435246	11.64521	12.8989	10.90542	0.965983
26	2.983685	3.389258	0.025628	0.019927	3.69123	2.732676	0.412554	1.401005	0.001	0.001	0.001	5.893824
27	0.117672	0.285134	0.616989	0.427887	0.1	1.149643	0.1002	0.158044	0.193928	0.301728	0.102338	0.1
28	0.1	0.1	0.80158	1.048576	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
29	0.100108	0.1	1.085767	0.864271	0.1	0.753995	0.325084	0.1	0.1	0.1	0.1	0.1
30	0.319822	0.004941	2.497591	2.533106	2.849161	2.94321	0.089	0.00001	0.00001	0.00001	0.007453	0.00985
31	0.01	0.01	4.927451	4.21706	4.922114	4.175953	4.573917	0.414246	1.708434	3.738724	2.784947	0.01
32	0.01	0.01	0.049263	0.149523	0.01	0.01	0.01	0.01	2.548784	0.320448	3.121647	0.01
33	0.01	0.01	0.285086	0.718315	0.01	1.889853	0.01	0.01	4.697255	3.261881	4.997792	0.944983
34	2.646489	3.066213	1.863949	1.467829	0.120164	2.321318	3.828026	6.496781	8.695867	2.512717	9.990491	9.99944
35	-10	-10	-9.245516	-9.171791	-10	-10	-10	-5.809852	-7.204324	-6.926079	-10	-10
36	7.373746	9.495871	0.723824	0.84152	2.837567	3.899447	2.180237	7.267313	0.00001	0.00001	0.00001	0.521579
37	-10	-10	-5.925306	-5.991425	-7.628821	-5.823346	-10	-10	-7.720467	-10	-9.30086	-10
38	5.018684	4.76798	2.021491	2.127802	0.00001	0.00001	5.995824	4.903404	0.00001	0.00001	1.344061	4.99976
39	0.47692	0.252761	5.601236	5.561892	6.783794	6.572017	3.064091	0.399379	0.1	0.1	0.495755	0.49975
40	0.280872	0.063599	0.261434	0.270928	4.69	4.69	0.494189	0.848234	0.01	0.01	0.01	0.01
41	-4.265566	-4.091052	-0.013714	-0.763224	-5.393275	-4.495557	-0.122731	-0.851096	-0.377712	-1.767299	0	0
42	0.115615	0.10517	0.54998	0.631427	0.1	0.1	1.26694	1.345374	0.1	0.1	0.50083	6.894434
43	91.8723	92.14348	93.69348	93.2521	100	100	99.99988	99.99388	100	100	100	100
44	15.6838	15.98975	20.12547	20.52095	15.57555	19.89353	18.61114	18.23195	42.25498	35.93646	42.59011	29.51398
45	-1.503241	-1.599526	-1.749061	-1.64838	-0.001	-0.001	-1.601527	-1.965178	-0.001	-0.001	-0.750533	-1.99
46	1.066859	1.098593	2.063146	1.702992	7.443778	7.126516	2.160044	2.077314	0.654538	0.472312	0.456971	0.510559
47	0.131313	0.020977	0.012091	0.13049	0.013148	0.137119	0.00001	0.004025	0.023288	0.133701	0.00017	0.00001
48	5.226328	9.175799	14.40425	9.549912	14.36092	9.276777	14.47357	10.17338	6.614103	1.29614	7.007113	9.90965
49	1.780721	2.526538	2.645517	2.085518	0.111	0.111	2.644322	2.860764	0.111	0.111	0.860533	2.100993
50	10.20993	10.02917	10.0038	10.15349	25.90318	34.68912	10.19844	10.19844	10	10	10.56598	10
51	0.208388	0.192798	3.214313	3.119414	0.319164	2.473694	2.638503	3.242341	8.894867	2.711717	10.59032	16.99287
52	0.263952	0.586841	0.681379	0.419993	0.1	0.1	1.032793	0.860764	0.1	0.1	0.1	0.100093
53	4.387141	4.436173	9.975552	9.585433	10	10	9.979465	9.301916	1	1.9992	1	1.506064
54	0.013528	0.340172	0.215076	0.017145	0.01	0.01	0.010002	0.034822	0.01	0.01	0.01	0.01
55	1.480581	1.838039	3.341685	2.954351	5	5	3.985182	3.904161	2.66	2.66	2.66	2.660309
56	0.407854	0.013918	1.036112	1.423446	0.01	0.01	0.392615	0.473636	2.35	2.35	2.35	2.350028
57	1.516769	1.939697	1.964137	1.665526	0.011	0.011	1.611529	2	0.011	0.011	0.760533	2
58	1.528799	0.827942	0.550158	1.505214	1.815193	2	0.977127	0.799489	0.5	2	0.5	1.00062
59	58.33824	57.02264	61.28671	63.96787	63.54391	67.26205	62.6996	61.90736	73.00155	74.11885	73.14688	68.10882
60	4.039489	0.729315	1.448512	4.77897	0.494448	4.867575	0.325384	0.220745	0.798642	4.111045	0.105094	0.1003
61	3.204382	4.754184	2.037928	0.090657	2.251413	-1	4.999861	5	2.258497	-0.996638	4.930476	5
62	10	9.915414	0.965654	3.451319	3.100532	6.641834	2.180437	7.347809	0.465772	2.67404	0.003406	0.521779
63	1.829812	4.728567	3.446137	3.231665	-0.263524	0.33152	4.003936	5	4.630639	4.358739	5	5
64	10	10	5.757823	8.491226	0.052044	3.622393	10	10	5.189563	7.567585	6.348136	10
65	19.825	17.85305	11.86937	12.23529	5.290796	7.230804	19.99972	19.97508	0.1	0.1	0.1	6.126085
66	4.995549	-0.328006	-0.358976	4.924739	-1	-1	-0.976559	-0.849086	-1	2.735543	-1	-1
67	1.858686	-0.165957	0.35524	2.140105	5	5	-0.316709	-0.999962	4.865887	5	4.601261	3.435271
68	4.247343	4.888396	4.775659	4.828447	3.263427	5	5	1.773243	5	1.559066	5	5
69	6.697261	1.044399	-0.608845	3.191109	0.169889	0.134911	3.125166	2.326236	-1	-1	-1	7.713618
70	2.653716	3.078586	0.001196	0.028419	0.001	0.047624	2.556462	4.699814	0.001	0.001	0.001	0.001
71	7.016907	3.429022	8.468287	9.891562	10	10	7.917029	6.602381	3.469242	10	0.016642	2.91115
72	9.943258	4.501048	0.666476	6.935894	4.388089	10	0.413084	1.614322	1.235268	7.087179	0.01	5.894354
73	3.531802	0.83054	0.931367	3.820625	0.441855	4.714746	0.10046	0.26269	0.799418	3.779767	0.106753	0.10026
74	2.289513	0.319598	2.678964	4.490455	3.046385	5	0.08915	0.060383	0.349331	2.005532	0.01	0.01
75	0.797876	0.135863	5	5	5	4.998669	4.573977	0.438395	1.848163	4.540933	2.785966	0.01006
76	4.605945	0.744201	0.472465	4.716672	0.470189	4.809176	0.01035	0.150869	3.363867	5	3.12759	0.01035
77	1.717065	0.282703	0.442275	2.414684	0.180927	3.672404	0.01013	0.062323	5	5	5	0.945113
78	10	4.240935	2.541072	8.775266	0.856467	10	3.828586	6.722172	10	10	10	10
79	3.759738	0.77719	5.903523	8.824141	7.112501	10	3.064341	0.5	0.682202	3.442537	0.5	0.5
80	9.973322	10	10	10	10	10	10	10	5.820751	9.009205	4.32811	10
81	2.739895	0.255387	0.43851									

Table 3.8: Optimal selected solutions of ATP-urea in glycolysis mode from Figure 3.7.

Flux #	Point A					Point B					Point C				
	FBA	FBA + EBA				FBA	FBA + EBA				FBA	FBA + EBA			
	J	J	$\Delta\mu$	J ₊	J ₋	J	J	$\Delta\mu$	J ₊	J ₋	J	J	$\Delta\mu$	J ₊	J ₋
1	0.496845	0.543029	-3.648108	369.0639	368.5208	0.70133	0.995505	-6.703557	368.4273	367.4318	0.501749	0.995505	-6.698737	368.692	367.6965
2	0.496845	0.543029	-3.670871	366.777	366.234	0.70133	0.995505	-6.745699	366.1288	365.1333	0.501749	0.995505	-6.740603	366.4052	365.4097
3	0.496845	0.543029	-3.623763	371.5415	370.9984	0.70133	0.995505	-6.658487	370.9178	369.9223	0.501749	0.995505	-6.653956	371.17	370.1745
4	0.992689	0.335526	-1.287794	645.6821	645.3466	0.00101	0.00101	-0.00388	644.9403	644.9393	1.002497	0.00101	-0.003881	644.7304	644.7294
5	1E-05	1E-05	-72.08323	0.000349	0.000339	0.00001	0.00001	-66.793	0.000376	0.000366	0.00001	1E-05	-69.75755	0.00036	0.00035
6	1E-05	1E-05	-25.32626	0.000983	0.000973	0.00001	0.00001	-19.55836	0.001272	0.001262	0.00001	1E-05	-22.94519	0.001085	0.001075
7	13.71808	10.73724	-15.41062	1731.605	1720.867	15.09015	12.52604	-17.9772	1732.577	1720.051	14.28603	12.64926	-18.1732	1730.82	1718.171
8	22.29936	23.07392	-84.2784	689.9173	666.8434	20	20	-73.67926	682.5786	662.5786	20	20	-73.43776	684.79	664.79
9	22.29936	23.07392	-83.02457	700.1602	677.0863	20	20	-72.55268	693.0206	673.0206	20	20	-72.32988	695.1243	675.1243
10	12.29936	13.07392	-49.42679	661.9033	648.8293	10.69914	10.69914	-40.63622	657.686	646.9868	10	11.13376	-42.21718	658.9821	647.8484
11	23.20479	23.97935	-112.0429	542.3283	518.3489	11.66512	11.66512	-56.50075	517.3733	505.7081	19.92117	19.92117	-94.37706	532.9912	513.07
12	30	30	-92.03133	822.7217	792.7217	30	30	-92.14675	821.7103	791.7103	30	30	-92.1654	821.5471	791.5471
13	30	30	-93.12278	813.257	783.257	30	30	-93.2443	812.2169	782.2169	30	30	-93.26031	812.0801	782.0801
14	6.695214	5.920655	-24.14858	610.4066	604.4859	18.23488	18.23488	-73.66843	622.427	604.1921	9.978834	9.978834	-40.48745	615.6435	605.6647
15	6.695214	5.920655	-24.64234	598.2355	592.3148	18.23488	18.23488	-75.12945	610.5019	592.267	9.978834	9.978834	-41.3084	603.5081	593.5293
16	12.26472	12.26472	-33.82397	904.5245	892.2598	28.23464	28.23464	-77.48377	917.0037	888.7691	19.97859	19.97859	-54.88287	911.9178	891.9392
17	-0.510528	-0.510528	1611.801	0.556976	1.067504	0.99938	0	1631.194	0	0	0.99938	0	1618.894	0	0
18	4.128594	1.147754	-1.762845	1613.856	1612.708	3.99076	2.426031	-3.725793	1614.472	1612.046	3.186638	2.549252	-3.910883	1616.244	1613.695
19	0.1	0.1	-1315.077	0.2428	0.1428	0.1	0.1	-1329.488	0.240806	0.140806	0.1	0.1	-1318.454	0.242329	0.142329
20	0.01	0.01	-1297.755	0.024526	0.014526	2.90982	2.90982	-1221.85	7.474318	4.564498	3.60896	2.475204	-1268.189	6.178357	3.703153
21	0.903575	0.001038	-689.2836	0.004273	0.003235	0.001	0.001	-755.9063	0.003803	0.002803	0.001	0.001	-716.3788	0.003983	0.002983
22	-1	-1	804.2452	2.607621	3.607621	-1	-1	831.8429	2.50634	3.50634	-1	-1	381.3141	6.010278	7.010278
23	0.992679	0.335516	-2.268137	366.6646	366.329	0.001	0.001	-0.006832	362.6184	362.6174	1.002487	0.001	-0.006776	365.6321	365.6311
24	0.49983	0.49983	-1045.02	1.452448	0.952618	1.99999	1.00061	-1044.71	2.908348	1.907738	1.99999	1.00061	-1043.818	2.910347	1.909737
25	10.90542	10.90542	-484.5683	61.3892	50.48378	0.965983	0.965983	-513.2221	5.162922	4.196939	9.921166	8.787411	-499.6327	48.11618	39.32876
26	0.001	0.001	-682.9912	0.00415	0.00315	5.736849	5.893824	-666.9043	24.97472	19.08089	0.032669	0.00947	-676.2073	0.039647	0.030177
27	0.103556	0.102338	-861.5948	0.34841	0.246071	0.1002	0.1	-853.4371	0.343171	0.24317	0.1002	0.100083	-857.7435	0.342012	0.241928
28	0.1	0.1	-718.8723	0.397062	0.297062	0.1	0.1	-704.4679	0.40406	0.30406	0.1	0.1	-714.1738	0.399313	0.299313
29	0.1	0.1	-836.0113	0.349163	0.249163	0.1	0.1	-826.9843	0.352368	0.252368	0.1	0.1	-833.3708	0.350093	0.250093
30	0.007453	0.007453	-1032.111	0.021875	0.014422	0.00985	0.00985	-1027.566	0.029014	0.019164	0.00985	0.00985	-1913.028	0.018309	0.008459
31	0.01	2.784947	-1077.01	7.899583	5.114636	0.01	0.01	-1075.093	0.028406	0.018406	0.70914	0.01	-1076.191	0.028383	0.018383
32	4.994057	3.121647	-736.3901	12.14076	9.01911	0.01	0.01	-725.2552	0.039405	0.029405	4.211156	3.776541	-731.6427	14.76962	10.99308
33	4.997792	4.997792	-839.9307	17.38198	12.38419	0.944983	0.944983	-835.4658	3.301341	2.356358	4.99987	4.99987	-837.5958	17.42993	12.43006
34	9.990491	9.990491	-621.4147	45.03579	35.0453	9.99944	9.99944	-604.3868	46.19365	36.19421	9.99944	9.99944	-614.4784	45.52384	35.5244
35	-7.700638	-6.926079	692.956	21.46145	28.38753	-10	-10	638.5789	34.01276	44.01276	-10	-10	393.5788	58.08216	68.08216
36	1E-05	0.00001	-734.1562	3.9E-05	2.9E-05	0.521868	0.521579	-728.1611	2.048222	1.526643	0.00001	0.068314	-731.1543	0.267322	0.199008
37	-10	-10	966.5681	20.95695	30.95695	-9.30086	-9.30086	985.4419	19.04102	28.34188	-10	-8.866244	975.4458	18.37674	27.24298
38	0.569502	1.344061	-972.4947	4.140073	2.796012	4.99976	4.99976	-965.2205	15.4954	10.49564	4.99976	4.99976	-968.8515	15.44791	10.44815
39	0.495755	0.495755	-1089.716	1.393134	0.89738	0.49975	0.49975	-1082.046	1.412288	0.912538	0.49975	0.49975	-1088	1.406126	0.906376
40	0.01	0.01	-353.2098	0.075263	0.065263	0.01	0.01	-331.022	0.079961	0.06996	0.01	0.010436	-342.1699	0.080901	0.070465
41	3.001998	0	569.3062	0	0	2.164721	0	590.1179	0	0	6.240649	0	568.8523	0	0
42	3.502828	0.50083	-239.9423	5.425878	4.925048	9.90156	6.894434	-213.1516	83.63425	76.73982	8.273307	1.01008	-234.7942	11.17148	10.1614
43	100	100	-514.0888	533.6625	433.6625	100	100	-525.6537	523.0983	423.0983	100	100	-517.1978	530.776	430.776
44	43.68796	42.59011	-620.5908	192.2147	149.6246	30.35638	29.51398	-632.174	131.053	101.539	41.16429	39.7071	-614.1645	175.8284	136.1213
45	-0.001	-0.750533	1020.01	1.473341	2.223964	-1.40165	-1.99	1020.015	3.906706	5.896706	-0.001	-1.99	1020.863	3.902746	5.892746
46	0.503155	0.456971	-1.666165	679.7398	679.2827	0.475458	0.510559	-1.861376	679.831	679.3204	0.498251	0.376028	-0.672528	1385.464	1385.088
47	0.00017	0.00017	-1012.053	0.000506	0.000337	0.00001	0.00001	-1012.354	2.98E-05	1.98E-05	0.00001	0.00001	-1012.164	2.98E-05	1.98E-05
48	7.007113	7.007113	-481.211	39.69386	32.68674	9.90965	9.90965	-457.5804	58.76312	48.85347	10.60879	9.475034	-475.6835	54.23923	44.7642
49	0.111	0.860533	-1379.718	2.015263	1.15473	1.51165	2.100093	-1386.98	4.898923	2.79883	0.111	3.346763	-1382.067	7.827759	4.480995
50	10	10	-170.0975	150.7132	140.7132	10	10	-157.4185	162.4405	152.4405	10	10	-164.2995	155.8513	145.8513
51	13.59232	10.59032	-104.3687	256.7323	246.142	20	16.99287	-154.8004	280.5549	263.562	18.37175	11.10852	-107.8014	260.8989	249.7904
52	0.1	0.1	-1298.127	0.245204	0.145204	0.1	0.100093	-1305.328	0.244403	0.14431	0.1	1.346763	-1300.559	3.297618	1.950855
53	1	1				1.176788	1.506064					1.371533			
54	0.01	0.01				0.01	0.01					0.01	0.01		
55	0.32	2.66				4.860498	2.660309					4.885641	4.532825		
56	0.01	2.35				4.550498	2.350028					4.575841	0.482099		
57	0.011	0.760533				1.41165	2					0.011	2		
58	0.5	0.5				2	1.00062					2	1.00062		
59	73.69702	73.14688				71.52836	68.10882					76.93232	73.20546		
60	0.105094	0.105094				0.1003	0.1003					0.1003	0.1003		
61	4.155917	4.930476				5	5					5	5		
62	0.003406	0.003406				0.522068	0.521779					0.00021	0.068514		
63	5	5				5	5					5	5		
64	5.573578	6.348136				10	10					10	10		
65	0.1	0.1				2.872659	6.126085					0.1	1.712752		
66	-1	-1				-1	-1					-1	-1		
67	4.14999	4.601261				5	3.435271					3.89353	3.558492		
68	5	5				5	5					5	5		
69	-1	-1				7.713329	7.713618					-0.020856	-0.08916		
70	0.001	0.001				0.001	0.001					0.001	0.001		
71	0.919179	0.016642				2.91115	2.91115								

Considering the tradeoff between ATP synthesis and urea synthesis (Figures 3.3f and 3.5f), the changes in flux required to move from points K to L along the Pareto frontier in both gluconeogenic and glycolytic hepatocytes were mainly lipid uptake (52), TCA cycle (8), aspartate uptake (69) and the uptake of gluconeogenic and ketogenic amino acids. This is expected because higher urea secretion could be achieved with increased uptake of arginine or aspartate under gluconeogenic conditions. Higher urea secretion has been seen to require an increase in gluconeogenic fluxes and this is coupled with an increase in TCA cycle fluxes, which necessitates an increase in aspartate uptake.

3.4.2 Effect of FBA+EBA on Pareto Frontier Compared to FBA Alone

We compared Pareto frontiers for the representative case of ATP synthesis vs. urea secretion considering FBA (i.e., mass balance) constraints only and then both FBA and EBA (i.e., both mass balance and thermodynamic) constraints. Figure 3.7a shows the Pareto frontiers when hepatocytes are in a glycolysis mode. The addition of EBA constraints generally reduced the feasible space of the flux distribution, and changed the Pareto frontiers accordingly. For example, for the representative case of ATP synthesis vs. urea secretion in glycolytic hepatocytes (shown in Figure 3.7a), the Pareto frontier obtained using both FBA & EBA constraints was below that obtained using FBA alone. Furthermore, the fluxes obtained using both approaches were vastly different throughout the Pareto frontier. Figures 3.7b, 3.7c, and 3.7d show the effect of adding EBA constraints on the Pareto optimal solutions A, B, and C, respectively, and the corresponding flux values are presented in Table 3.8. In all cases, EBA reduced the feasible space. It is to be noted that urea secretion (flux 16 on the abscissa) was kept constant to analyze these differences. As seen in Figures 3.7b, 3.7c and 3.7d, several glycolytic fluxes (2-6) and catabolic fluxes that produce pyruvate (17, 18 and 19) were changed at points A, B, and C on the Pareto frontier when adding EBA constraints. On the other hand, there was a marginal difference in TCA cycle flux (8) at Pareto solution A, and no difference at points B or C. Similarly, when going from A to C, there was a decreased difference in the uptake of succinyl CoA forming amino acids (threonine, 71, methionine, 75, and valine, 76). Notably, the difference at point A in the aspartate production through asparagine (36) first increases with the increased urea secretion, then the difference decreases significantly at Pareto solution B. Throughout the Pareto frontier

there was a decreased ketone body production (41) when adding EBA constraints. The change in lipid uptake and lipid storage fluxes when adding EBA constraints became more prominent when going from point A to point C.

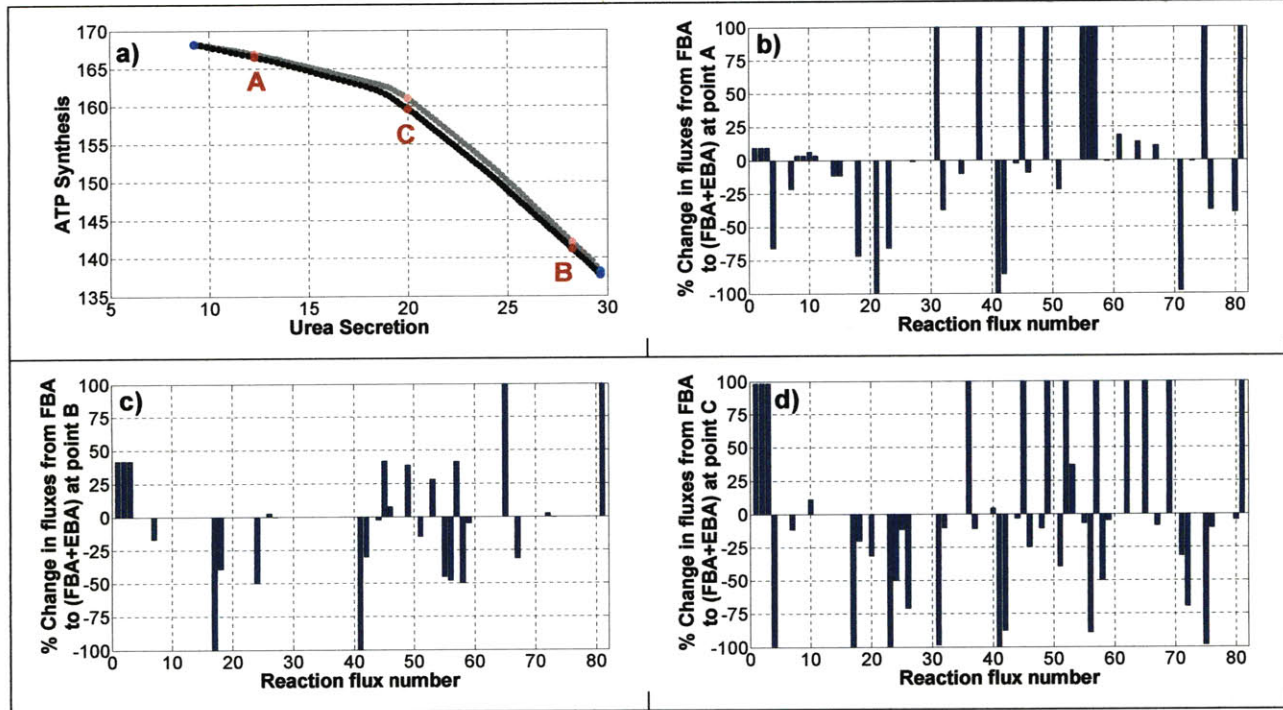


Figure 3.7: Effect of adding EBA constraints on optimal fluxes for the representative case of ATP synthesis vs. urea secretion bi-objective problem in glycolytic hepatocytes. **a)** Pareto frontiers using FBA constraints alone (gray line) and FBA+EBA constraints (black line). The blue circles are the anchor points and the red circles are selected Pareto solutions A, B, C for which the complete set of fluxes is provided in Table 3.8. **b)** Distribution of flux changes when adding EBA constraints at point A. **c)** Distribution of flux changes when adding EBA constraints at point B. **d)** Distribution of flux changes when adding EBA constraints at point C. In panels b), c) and d), the data are expressed as % flux change and the corresponding reaction flux number is shown on the horizontal axis. Note that urea secretion was kept constant when analyzing these differences.

3.4.3 Effect of Measured Fluxes on Pareto Frontiers

The incorporation of experimental measurements also reduced the feasible space for the fluxes and the Pareto optimal solutions (Figure 3.8). The control case has no measurements and is labeled as M0 (shown as black dots in Figure 3.8). Experimental measurements were included as equality constraints in the stoichiometric matrix according to Equation (3.3). For this analysis, four different bi-objective sets were examined as examples. Figure 3.8 shows the effect of adding four different measurement sets on the Pareto frontiers. Measurements were included in a sequential fashion. The Pareto curve M1 (shown as red diamonds) was obtained after adding

lactate (50), glucose (53), and glutamate (61) measurements. The Pareto curve M2 (shown as blue squares) was obtained after adding to M1 measurements, glutamine (79) and tyrosine (81). The Pareto curve M3 (shown as yellow triangles) is obtained after adding to M2 measured fluxes, alanine (66), serine (67), and glycine (68) flux measurements. The Pareto curve M4 (shown as green stars) is obtained after adding to M3 measured fluxes, methionine flux measurement (75). Experimental data for gluconeogenesis and glycolysis were taken from (14) and (16), respectively.

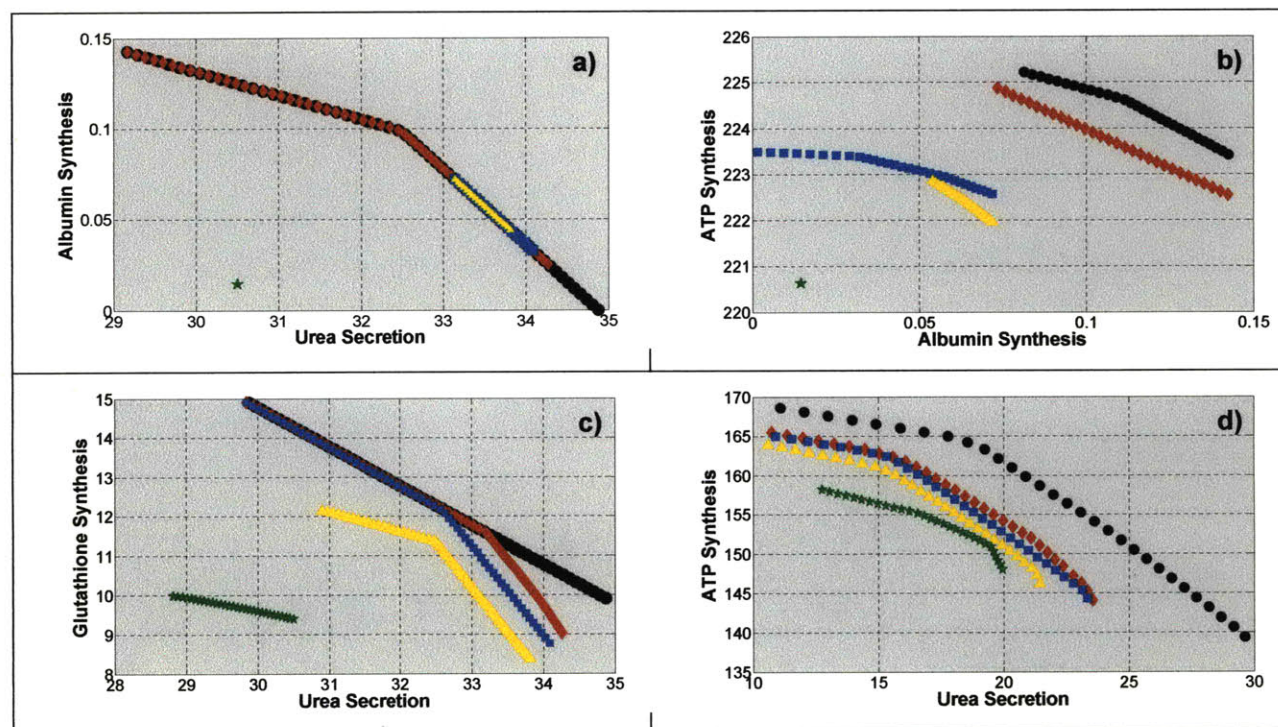


Figure 3.8: Effect of adding flux measurements to Pareto frontiers. Measurements were incorporated as equality constraints in the stoichiometric matrix. Four different bi-objective cases are shown in panels a-d, respectively: albumin vs. urea synthesis (gluconeogenesis mode), ATP vs. albumin synthesis (gluconeogenesis mode), glutathione vs. urea synthesis (gluconeogenesis mode), and ATP vs. urea synthesis (glycolysis mode). The control case has no measurements (M0 in black). The Pareto curve M1 (shown as red diamonds) is obtained after adding measured flux 50 (value of 1.0815 and 1.08 for gluconeogenesis and glycolysis, respectively) + flux 53 (value of 1.1472 and 0.15 for gluconeogenesis and glycolysis, respectively) + flux 61 (value of -0.3789 and -0.38 for gluconeogenesis and glycolysis, respectively). The Pareto curve M2 (shown as blue squares) is obtained after adding to M1 measured + flux 79 (value of 1.8962 and 1.9 for gluconeogenesis and glycolysis, respectively) + flux 81 (value of 0.0319 and 0.032 for gluconeogenesis and glycolysis, respectively). The Pareto curve M3 (shown as yellow triangles) is obtained after adding to M2 measured flux 66 (value of 0.0316 and 0.032 for gluconeogenesis and glycolysis, respectively) + flux 67 (value of -0.2292 and -0.23 for gluconeogenesis and glycolysis, respectively) + flux 68 (value of 0.1368 and 0.14 for gluconeogenesis and glycolysis, respectively). The Pareto curve M4 (shown as green stars) is obtained after adding to M3 measured flux 75 (value of 0.0978 and 0.098 for gluconeogenesis and glycolysis, respectively). Experimental data for gluconeogenesis and glycolysis were taken from; Chan et al., 2003b; Chan et al., 2002, respectively.

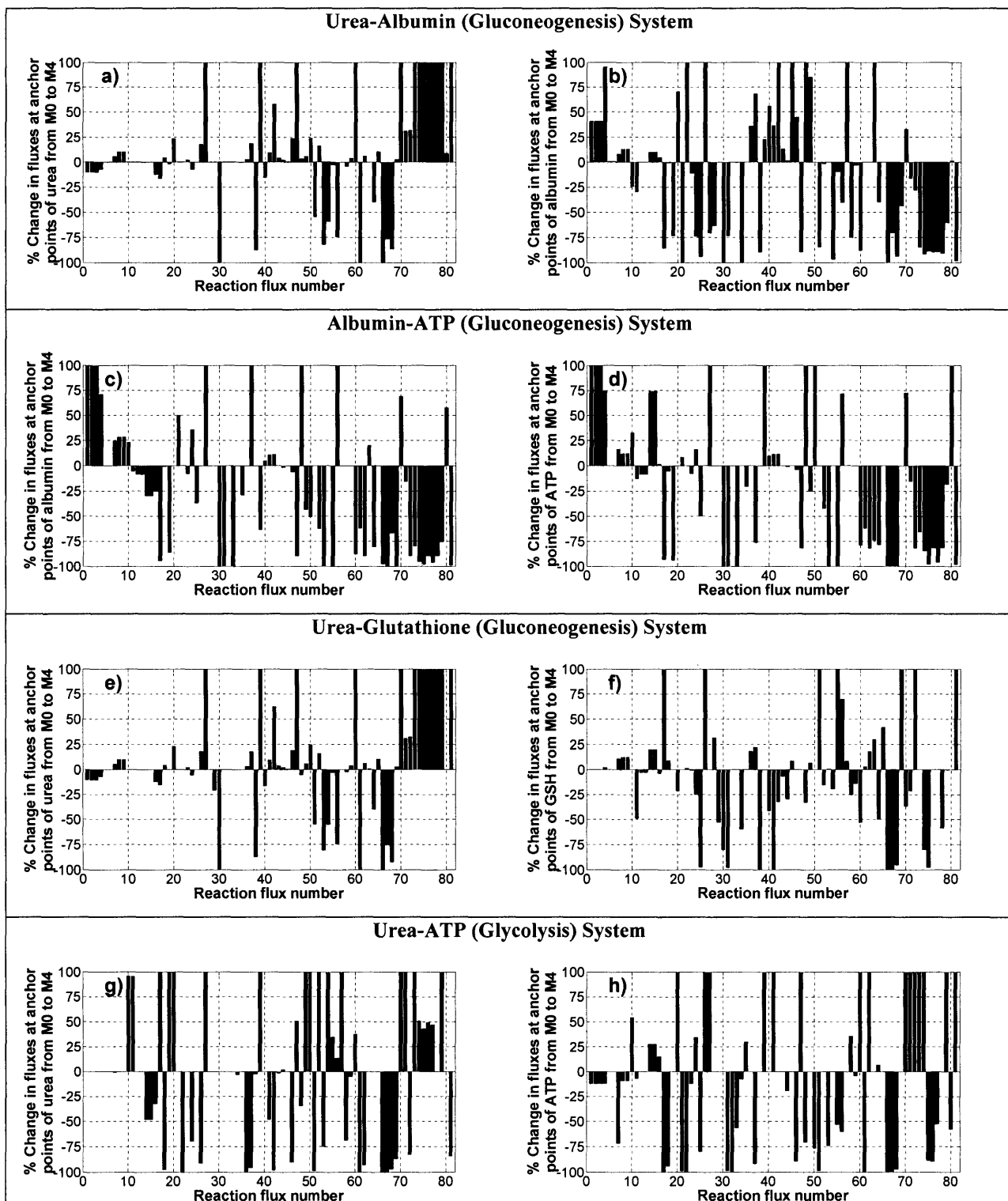


Figure 3.9: Distribution of optimal flux changes between the anchor points of the system solved without constraints (M0) and with the maximum number of constraints (M4) for the bi-objective system of Figure 3.8. **a)** and **b)**: albumin vs. urea (gluconeogenesis mode). **c)** and **d)**: albumin vs. ATP (gluconeogenesis mode). **e)** and **f)**: urea vs. glutathione (gluconeogenesis mode). **g)** and **h)**: urea vs. ATP (glycolysis mode). The absolute flux values are in Table 3.9. Note that the % flux changes for all figures are on y-axis and the corresponding reaction flux number is shown on the horizontal axis. Only changes up to 100% are shown in the figure.

PAGES (S) MISSING FROM ORIGINAL

Page 95 does not exist.
It appears to be a mistake
in numbering by the author.

We looked at four representative bi-objective sets (albumin vs. urea; ATP vs. albumin; glutathione vs. urea; and ATP vs. urea) to ascertain the changes in Pareto frontiers. The three first sets are in gluconeogenic mode and the last one is in glycolytic mode.

Figure 3.8a show Pareto frontiers for albumin synthesis vs. urea secretion and Figure 3.8b shows the Pareto frontiers for ATP synthesis vs. albumin synthesis. In both cases, as more measured data are included, the anchor points of the Pareto frontiers move towards the center and eventually become a single point solution. Figure 3.8c shows that Pareto frontiers for glutathione synthesis vs. urea secretion, in the higher glutathione synthesis range did not change when including measurement sets M1 and M2, although they did when including measurement sets M3 and M4. Figure 3.8d shows the Pareto frontiers for ATP synthesis vs. urea secretion. Pareto frontiers were lowered when adding each measurement set. The corresponding fluxes for these four cases are presented in Table 3.9. Figures 3.9a-b, 3.9c-d, 3.9e-f, and 3.9g-h show the distribution of flux changes for the cases shown in Figures 3.8a, 3.8b, 3.8c, and 3.8d, respectively.

When considering the albumin vs. urea case (Figure 3.8a), the change in Pareto curve at high urea secretion was associated with many differences in fluxes (Figure 3.9a), including a moderate decrease in gluconeogenic fluxes (2-4), a moderate increase in TCA cycle flux (8), a decrease in urea secretion (16) and β -oxidation (40), an increase in electron transport (43, 44), lipid uptake (52), lipid stored (57), albumin (47), NADPH (46) and GSH (48) synthesis. The change in Pareto curve at high albumin synthesis also caused flux changes (Figure 3.9b), including a moderate increase in gluconeogenic fluxes (2-6), TCA cycle fluxes (8, 13), urea cycle fluxes (14, 15) and urea secretion (16), a significant increase in β -oxidation (40), electron transport (43, 44), lipid storage (57), NADPH (46) and GSH (48) synthesis, significant decrease in lipid uptake (52) and albumin synthesis (47). Additionally, there was a decrease in uptake of both gluconeogenic and ketogenic amino acids.

When considering the ATP vs. albumin case (Figure 3.8b), incorporation of measurements also changed the Pareto curve. At high albumin secretion, the associated flux differences (Figure 3.9c) included a significant increase in gluconeogenic fluxes (2-4), TCA cycle flux (8), and GSH (48) synthesis, a moderate increase in β -oxidation (40), a decrease in urea cycle fluxes (14, 15), urea secretion (16), lipid uptake (52) and albumin (47), a moderate decrease in electron transport (44) and NADPH (46) synthesis. At the anchor point of ATP

generation on the Pareto curve, flux changes caused by introduction of the measurements (Figure 3.9d) significantly increased gluconeogenesis fluxes (2-6), urea cycle fluxes (14, 15), and GSH (48) synthesis, moderately increased TCA cycle flux (8), urea secretion (16), and β -oxidation (40), decreased electron transport (44), lipid uptake (52), albumin synthesis (47), and NADPH (46). Additionally, there was decreased uptake of both gluconeogenic and ketogenic amino acids. The effect of measurements on the Pareto curve of glutathione vs. urea are shown in Figure 3.8c. The major differences in fluxes at the anchor point of high urea secretion (Figure 3.9e) included a significant increase in albumin (47), a moderate increase in TCA cycle flux (8), electron transport (43, 44), lipid uptake (52), and NADPH (46), a decrease in urea secretion (16), a moderate decrease in gluconeogenic fluxes (2-6), β -oxidation (40), and glutathione synthesis (48). As seen earlier, the increased albumin synthesis necessitates significant increase in the uptake of both gluconeogenic and ketogenic amino acids. The major differences in fluxes at the anchor point of high glutathione synthesis (Figure 10f) included a moderate increase in TCA cycle flux (8) and lipid storage, a significant increase in urea cycle fluxes (14, 15), a decrease in electron transport (43, 44), a moderate decrease in urea secretion (16) and lipid uptake (52), a significant decrease in β -oxidation (40) and glutathione synthesis (48).

The effect of measurements on the Pareto curve of urea secretion vs. ATP generation (in glycolysis mode) are shown in Figure 3.8d. The major differences in fluxes at the anchor points of high urea secretion and ATP generations are shown in Figures 3.9g and 3.9h, respectively. We found that addition of the measurements did not change glycolysis fluxes (1-5) at either anchor point. On the other hand, at the anchor point of high urea secretion (Figure 3.9g), there was a significant decrease in urea cycle fluxes (14, 15), urea secretion (16), NADPH (46) and GSH synthesis (48), a moderate decrease in electron transport (43), a significant increase in lipid uptake (52), lipid storage (57), and albumin (47). The increased albumin synthesis necessitates the increased uptake of gluconeogenic amino acids. The major differences in fluxes at the anchor point of high ATP generation (Figure 3.9h) included a moderate decrease in TCA cycle flux (8), a significant increase in urea cycle fluxes (14, 15) and albumin synthesis (47), a moderate increase in urea secretion (16), a decrease in electron transport (44), and a significant decrease in NADPH and GSH synthesis. Additionally, there was a significant increase in the uptake of both gluconeogenic and ketogenic amino acids.

3.5 DISCUSSION

Mammalian cells exhibit various phenotypic states including proliferation, differentiation, etc. Metabolic flux distributions in these various states must obey constraints imposed by the environment, reaction stoichiometry, thermodynamics, and laws of conservation. Mathematically these constraints translate into a reduction of the feasible space for the flux distribution. Most of the literature on constraints-based metabolic network optimality deals with unicellular organisms where the main objective is growth of biomass (Edwards and Palsson, 2000a; Edwards and Palsson, 2000b). In mammalian systems, various phenotypes are encountered, some of which exhibit proliferation, and others expression of organ-specific or “differentiated” functions. Several objectives should be considered before making any conclusions about the optimal states of such systems. Often times there is a competition between the various objectives because they are differentially altered by the constraints. This paradigm of conflicting objectives is addressed herein using a class of multi-objective optimality called Pareto optimality. Furthermore, we used the Normal Constraint method, which yields any Pareto point in the feasible objective space, guarantees an even distribution of the Pareto frontier, and is insensitive to design objective scaling. Combining these concepts with FBA and EBA, we developed a framework called NCEFBA, which we applied to the specific case of cultured hepatocytes.

Hepatocytes are the key cellular component in BAL devices. The ability to optimize hepatocellular metabolism is important to maximize the clinical efficacy of the BAL, and increasing the function per cell may help reduce the number of hepatocytes needed in the device. Hepatocytes express various liver-specific functions that require common substrates, such as glucose, amino acids, and so on. Thus, it is expected that increasing one function (for instance, albumin synthesis) will decrease another (for example, urea secretion). In order to systematically investigate the tradeoffs between the various hepatocellular functions, we used NCEFBA. More specifically, we investigated the interactions among five key hepatocyte metabolic functions, namely albumin synthesis, urea secretion, glutathione synthesis, NADPH synthesis, and ATP generation. These analyses were carried out first in gluconeogenic hepatocytes (Figure 3.3) and then glycolytic hepatocytes (Figure 3.5).

Using NCEFBA, we observed the Pareto optimality between various liver specific functions. Some of the representative bi-objective combinations were shown in this section. Here, the implementation was done for several biobjective combinations in order to develop a

suitable framework for designing a compartmental BAL device that can perform all essential liver-specific functions. This BAL device could have several individual bioreactor modules interconnected in series and each individual bioreactor could be designed based on the various combinations of liver specific Pareto optimal solutions. The idea here will be to obtain an optimal BAL system that can exhibit very high and stable levels of key liver-specific functions and thus translate into a proportional reduction of required cell mass and total perfusion volume of the bioreactor required for a given processing capacity. The six combinations of liver-specific functions for both gluconeogenic hepatocytes and glycolytic hepatocytes were analyzed to obtain a global Pareto optimal solution with respect to each liver specific function in BAL assembly. Table 3.10 shows the flux values for these solutions at few representative Pareto optimal solutions A, B, C, D, E and F for both gluconeogenic and glycolytic hepatocytes of Figure 3.3 and Figure 3.5. Based on the results obtained, to design a BAL assembly system that provides higher liver specific functions in gluconeogenic mode, one option could be to operate five different bioreactors in series at H, G, F, E and D points, respectively from Figure 3.3. If the five reactors are operated at these points then the total fluxes can be calculated by summing up the individual fluxes at these points. For albumin, urea, glutathione and NADPH synthesis these values are 0.281, 107.4, 58.31, and 8.67, respectively. On the other hand, if the reactor assembly is just operated at the equal preference point of E then the total fluxes of albumin, urea, glutathione and NADPH synthesis will be 0.08, 98.76, 72.1, and 14.93, respectively. This indicates that the variable operating condition BAL will produce overall higher albumin and urea synthesis compared to the case where assembly is just operated at point E condition. It is to be noted that glutathione and NADPH synthesis in variable operating condition reactor is lower than that if assembly is operated at point E alone. This could be tolerable because of higher priority to attain high albumin and urea synthesis. However, if there is a situation where there is a higher demand of ATP (because of stress and mitochondrial dysfunction) BAL system for gluconeogenic mode could be designed for H, G, J, K and L points, respectively from Figure 4. In glycolytic mode of BAL operation the preferred combination of reactor operations could be H, G, C, D, and F points, respectively from Figure 3.5. If the five bioreactors are operated at these points then the total fluxes of albumin, urea, glutathione and NADPH synthesis are 0.283, 93.48, 57.88, and 15.13, respectively. On the other hand, if the reactor assembly is just operated at the equal preference point of C then the total fluxes of albumin, urea, glutathione and NADPH

synthesis are 0.06, 72.21, 72.02, and 10.32, respectively. As seen earlier for gluconeogenic hepatocytes, we see also in glycolytic hepatocytes that the variable operating condition BAL will produce in overall higher albumin, urea and NADPH synthesis compared to the case where assembly is operated at point C condition. Again, if there is a situation demanding higher energy production BAL system for glycolytic mode could be H, G, I, K and L point, respectively from Figure 3.5.

Table 7: Objective function flux values for Pareto optimal solutions in Figure 3.3 (for gluconeogenic mode) and Figure 3.5 (for glycolysis mode). The detailed flux values are provided in Table 3.6 and Table 3.7

OBJECTIVES	A	B	C	D	E	F	G	H	I	J	K	L
GLUCONEOGENESIS												
UREA	29.7685	34.56	10.431	10.595	19.752	12.454	30.13	34.465	2.8071	2.601	11.516	33.769
NADPH	0.47	0.53	1.414	0.98	2.985	2.786	1.091	0.825	1.134	1.411	1.008	0.945
ALBUMIN	0.136	0.0126	0.011	0.1296	0.016	0.135	0.00001	0.00001	0.0992	0.139	0.0832	0.01136
GSH	6.837	8.79	14.427	9.585	14.4212	9.345	14.647	10.312	0.083	0.001	1.483	9.395
GLYCOLYSIS												
UREA	26.717	29.365	14.441	14.388	12.323	9.978	25.304	29.368	8.695	6.28	12.264	28.234
NADPH	1.066	1.098	2.063	1.702	7.443	7.126	2.16	2.077	0.654	0.472	0.456	0.51
ALBUMIN	0.131	0.02	0.012	0.1304	0.0131	0.137	0.00001	0.004	0.0232	0.133	0.0017	0.00001
GSH	5.226	9.175	14.404	9.549	14.36	9.276	14.473	10.173	6.614	1.296	7.007	9.909

In conclusion, the NCEFBA platform is a useful tool for optimality analysis of large scale metabolic networks that are bound to possess multi-objective Pareto optimal solutions. This technique enables the systematic identification of tradeoff situations between various metabolic objectives that characterize a particular cellular phenotype. The addition of FBA to EBA constraints ensures that thermodynamically feasible solutions are obtained. Furthermore, experimental flux data can be easily incorporated into the analysis, which further reduces the feasible space of fluxes. Although the NCEFBA approach described here was applied to the specific case of hepatocellular metabolism, it can be readily used on any large-scale metabolic network. This study highlights how Pareto optimal solutions may contribute to operating BAL devices, alter the metabolic states of hepatocytes, achieve the desired range of objectives and has relevance for understanding the impact of environmental stress, inducers, hormones and supplements on cellular metabolism.

4 REGULATORY TRANSCRIPTIONAL NETWORK IN EMBRYONIC STEM CELLS

4.1 BACKGROUND

Embryonic stem cells (ESCs) are pluripotent because they can give rise to cells derived from all three germ layers. ESCs are considered a potential source of cells for human disease therapies due to their limitless capacity for self-renewal and proliferation, and their ability to differentiate into all major cell lineages. Octamer-4 (Oct4), Sox2 and Nanog are important markers of pluripotency, expressed by primitive embryonic cells both *in vivo* and *in vitro*. Oct4 and Nanog expression is downregulated during early differentiation. In embryonic stem cells (ESCs), the stem cell-ness is determined by the expression of three major transcription factors: Oct4, Sox2 and Nanog (Chickarmane et al., 2006).

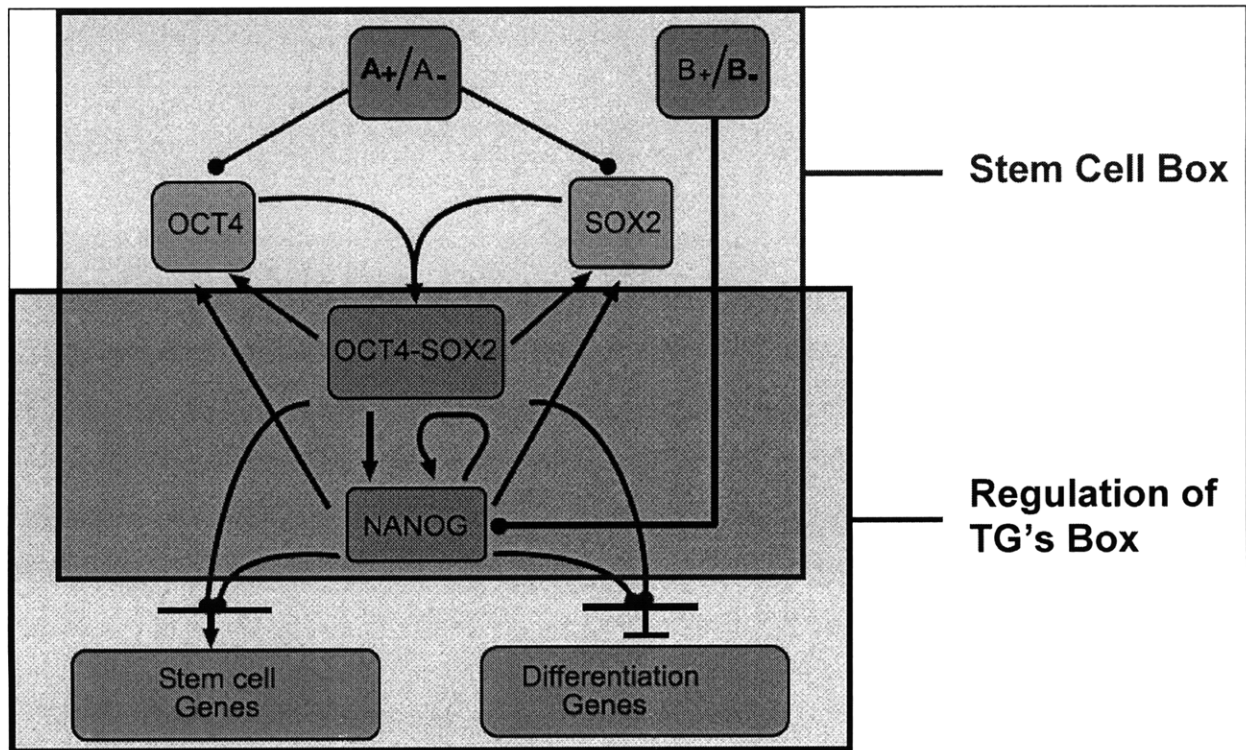


Figure 4.1: The core transcriptional network of ES cells (adapted from Chickarmane et al, 2006).

As seen in Figure 4.1, the signals A₊ and A₋ positively and negatively regulate Oct4 and Sox2, expression, respectively, whereas signals B₊ and B₋ positively and negatively regulate

Nanog expression. Oct4 and Sox2 are activated by an external signal A_+ which are different for murine (A_+ can be LIF or BMP4) and human (A_+ can be bFGF, Activin A and Noggin) systems. In both murine and human ESCs, the wnt pathway is important in upregulating Oct4 and Sox2 expression. As seen in Figure 4.1, the transcriptional network loops involved in pluripotent and differentiated state of ESCs is divided into two separate hierarchical type networks “stem cell box” and “target gene regulation box”. In stem cell box, because of an external signal A_+/A_- , there is a transcription of Oct4 and Sox2 transcription factors which further leads to the formation of Oct4-Sox2 complex that in turn binds on the promoter regions of both Oct4 and Sox2 and thus, lead to a positive feedback regulatory network (Figure 4.1). In principle, there is also a signal A_- which represses both Oct4 and Sox2 and the subsequent transcription of Nanog. The external signal B_+/B_- (B_- could be BMP4 in human ES cells) acts as an activator/repressor of Nanog. In addition, the Oct4-Sox2 (OS) complex promotes the transcription of Nanog which upregulates the transcription of Oct4, Sox2, and itself, thus forming a more complex self-regulatory and positive feedback network loop defined as the stem cell box. Furthermore, the activation of Oct4 and Sox2 by binding of Nanog on the promoter region of Oct4 and Sox2 necessitates the initial binding of Oct4-Sox2 on each of their promoters. As seen in the target gene regulation box, during differentiation, Nanog acts as the repressor on the target genes for differentiation. However, since the formation of the OS complex is required first, this action is less effective compared to the direct effect of B_- on Nanog. Additionally, there may also be a signal B_+ which externally upregulates the expression of Nanog. However, since the described mechanism requires Oct4 and Sox2 binding to recruit Nanog, this mechanism is not very effective. Then, signal A_+ will activate the system and B_- will deactivate it.

It is well known that Oct4, Sox2 and Nanog regulate the expression of many other genes, which in turn are transcription factors responsible for either maintaining the stem cell-ness i.e. pluripotent cellular state or the differentiation into all the three embryonic germ layers (endoderm, mesoderm and ectoderm). As seen in Figure 4.1, this network is referred as the regulation of target genes box. The individual effect of Oct4 and Sox2 on the target genes is here modeled by the single interaction of the complex Oct4-Sox2. By isolating this box from the external signals A_+ and B_- , it can be seen that Oct4-Sox2 regulates Nanog expression and both regulate the target genes responsible for pluripotency and differentiation. This three-node architecture is the well known Feed Forward Loop (FFL). The two possible interactions

(activation or repression) from the three nodes lead to 8 combinations that can be divided into Coherent and Incoherent. In the Coherent-Type FFL, the direct interaction from the master gene (in current case Oct4-Sox2) to the target gene is same as the indirect interaction through the intermediate transcription factor (in current case Nanog). The opposite is true for the Incoherent-Type FFL. The classification of the FFL is seen in Figure 4.2A (Mangan and Alon, 2003). One possible interaction between Oct4-Sox2, Nanog and the target genes in the regulation of target genes box is a Coherent-Type motif. As shown in Figure 4.2B, the stem cell target genes (SC) are being activated by both Oct4-Sox2 (OS) and Nanog (N) as a Coherent Type-1 FFL, and the differentiation target genes (D) are being repressed by both OS and N as a Coherent Type-3 FFL. Figure 4.2C shows an Incoherent Type-1 FFL for these transcription factors, Nanog being a weak repressor for the stem cell-ness target genes. It is not known, which mechanism is dominating for maintenance of pluripotency or differentiated state. In this study, we propose a Pareto-optimal transcriptional regulatory network framework to elucidate and predict the most likely interaction between these transcription factors for maintenance of stem cell-ness and differentiated state.

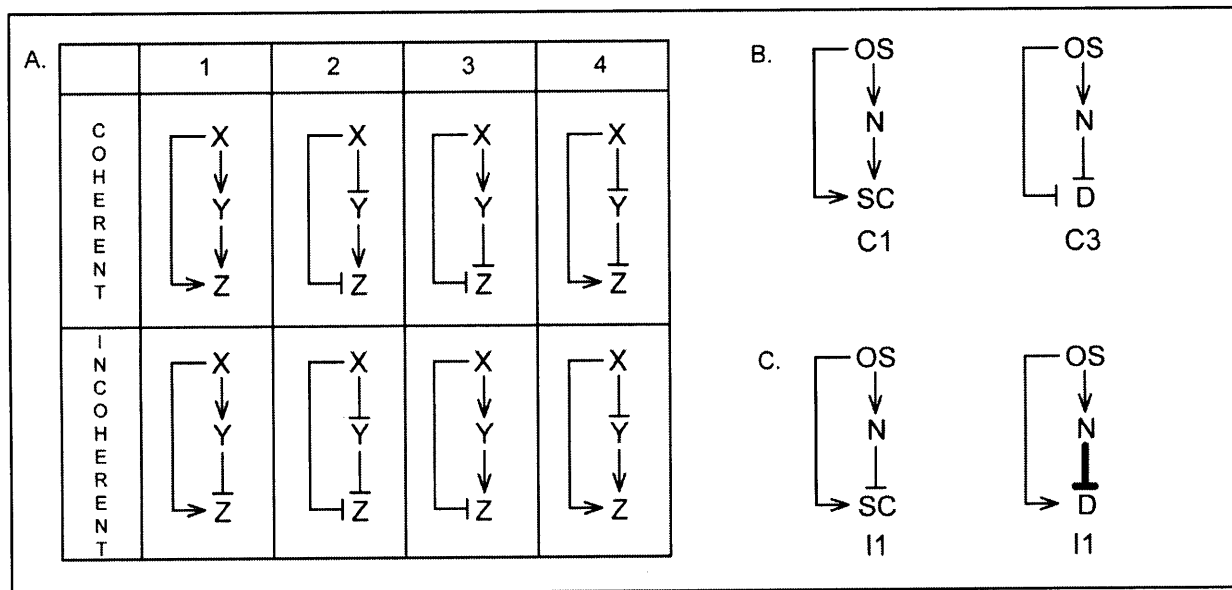


Figure 4.2: The Feed Forward Loop (FFL). **A)** Classification of the FFL in Coherent-Types 1-4 and Incoherent-Types 1-4. **B)** Coherent interaction between Oct4-Sox2 (OS), Nanog (N) and the target genes. For stem cellness (SC) and differentiation (D), Coherent Type-1 and Coherent Type-3 are assumed. **C)** Incoherent interaction between Oct4-Sox2 (OS), Nanog (N) and the target genes. For both stem cellness (SC) and differentiation (D), Incoherent Type-1 is assumed. However, weak repression of Nanog is assumed for the pluripotent state.

4.1.1 Mathematical Modeling

The presented model (Chickarmane et al., 2006), is based on the Shea-Ackers formalism (Shea and Ackers, 1985):

General Shea-Ackers Formulation

In general, the Michaelis-Menten approach assumes binding of the TFs in the promoter region of gene Z at thermodynamic equilibrium and that the total transcription rate is proportional to the concentration of DNA sites bound by RNA polymerase. Let us assume that the translation and transcription activity of protein Z is regulated by N_A activation TFs (X_j^A), N_R repressive TFs (X_j^R) and RNA polymerase (R), with $N = N_A + N_R$. In a competitive binding X_j^A , X_j^R , or R can bind a free DNA site in the promoter region D_Z for transcription of protein Z , forming a transcription factor-DNA site complex $X_j D_Z$: $X_j + D_Z \leftrightarrow X_j D_Z$, where X_j represents either X_j^A , X_j^R , or R . Let $K_1^{ZX_j}$ be the association equilibrium constant of this step. Then, at thermodynamic equilibrium:

$$K_1^{ZX_j} [X_j] [D_Z] = [X_j D_Z] \quad (4.1)$$

If X_j is an activator ($X_j = X_j^A$), then RNA polymerase (R) is recruited by the activator X_j^A bound to D_Z : $R + X_j^A D_Z \leftrightarrow R X_j^A D_Z$. If $K_2^{ZX_j}$ is the association equilibrium constant of this reaction, then at thermodynamic equilibrium:

$$K_2^{ZX_j} [R] [X_j^A D_Z] = [R X_j^A D_Z] \quad (4.2)$$

If X_j is a repressor for the expression of Z ($X_j = X_j^R$), then RNA polymerase is not bound and translation does not occur. The total number of sites in the promoter region for transcription of Z can be expressed as follows:

$$\begin{aligned} [D_Z]_{tot} = & \\ & [D_Z] + [R D_Z] + \\ & \sum_{j=1}^N [X_j D_Z] + \sum_{j=1}^{N_A} [R X_j^A D_Z] = [D_Z] \left(1 + K_1^{ZR} [R] + \sum_{j=1}^N K_1^{ZX_j} [X_j] + \sum_{j=1}^{N_A} K_2^{ZX_j} [R] [X_j^A] \right) \end{aligned} \quad (4.3)$$

As mentioned, the transcription rate (TR) of protein Z is proportional to the total number of sites bound by RNA polymerase $TR = k_z([X_0 D_Z] + \sum_{j=1}^{N_A} [R X_j^A D_Z])$, with k_z being the proportionality constant. Using Equations (4.1)-(4.3), it can be shown that:

$$TR = \frac{k_z [D_Z]_{tot} [R] \left(K_1^{ZR} + \sum_{j=1}^{N_A} K_1^{ZX_j} K_2^{ZX_j} [X_j^A] \right)}{1 + K_1^{ZR} [R] + \sum_{j=1}^{N_A} K_1^{ZX_j} [X_j^A] \left(1 + K_2^{ZX_j} [R] \right) + \sum_{j=1}^{N_R} K_1^{ZX_j} [X_j^R]} \quad (4.4)$$

By defining the following variables, $\eta^{ZR} = k_z [D_Z]_{tot} [R] K_1^{ZR}$, $a^{ZX_j^A} = k_z [D_Z]_{tot} [R] K_1^{ZX_j} K_2^{ZX_j}$, $\gamma^{ZR} = K_1^{ZR} [R]$, $b^{ZX_j^A} = K_1^{ZX_j} \left(1 + K_2^{ZX_j} [R] \right)$, and $b^{ZX_j^R} = K_1^{ZX_j}$, the transcription rate of protein of Z can be expressed as:

$$TR = \frac{\eta^{ZR} + \sum_{j=1}^{N_A} a^{ZX_j^A} [X_j^A]}{1 + \gamma^{ZR} + \sum_{j=1}^{N_A} b^{ZX_j^A} [X_j^A] + \sum_{j=1}^{N_R} b^{ZX_j^R} [X_j^R]} \quad (4.5)$$

The Stem Cell Box

The dynamic equations that model this box are:

$$\frac{d[O]}{dt} = \frac{\eta_1 + a_1 [A_+] + a_2 [OS] + a_3 [OS][N]}{1 + \gamma_1 + b_1 [A_+] + b_2 [OS] + b_3 [OS][N]} - \alpha_1 [O] - k_{1c} [O][S] + k_{2c} [OS] \quad (4.6)$$

$$\frac{d[S]}{dt} = \frac{\eta_2 + c_1 [A_+] + c_2 [OS] + c_3 [OS][N]}{1 + \gamma_2 + d_1 [A_+] + d_2 [OS] + d_3 [OS][N]} - \alpha_2 [S] - k_{1c} [O][S] + k_{2c} [OS] \quad (4.7)$$

$$\frac{d[OS]}{dt} = k_{1c} [O][S] - k_{2c} [OS] - k_{3c} [OS] \quad (4.8)$$

$$\frac{d[N]}{dt} = \frac{\eta_3 + e_1 [OS] + e_2 [OS][N]}{1 + \gamma_3 + f_1 [OS] + f_2 [OS][N] + f_3 [B_-]} - \alpha_3 [N] \quad (4.9)$$

where $[O]$: concentration of Oct4, $[S]$: concentration of Sox2, $[OS]$: concentration of Oct4-Sox2, $[N]$: concentration of Nanog. Here, a_i, b_i, c_i, d_i, e_i and f_i are constants related to the free energies of binding of transcription factors to the operator regions of Oct4, Sox2, and Nanog. Finally, η_i, γ_i , denote the basal transcription rates parameters, α_i degradation rates, and k_{1c} the kinetic constants (see Section 4.2.1)

The Regulation of Target Genes Box

For the Incoherent-Type FFL motif, the equations describing the concentrations of the target genes for both self-renewal (SC) and differentiation (D) are:

$$\frac{d[TG]}{dt} = \frac{\eta_4 + g_1[OS]}{1 + \gamma_4 + h_1[OS] + h_2[OS][N]} - \alpha_4[TG], \text{ with } TG = \{SC, D\} \quad (4.10)$$

where $[TG]$ is the concentration of the target genes for either self-renewal or differentiation, and g_i, h_i , are constants related to the free energies of binding of transcription factors to the operator regions of Oct4-Sox2, and Nanog.

For the Coherent-Type FFL motif, the equations describing the concentrations of the target genes for self-renewal (SC) and differentiation (D) are:

$$\frac{d[SC]}{dt} = \frac{\eta_5 + m_1[OS] + m_2[OS][N]}{1 + \gamma_5 + n_1[OS] + n_2[OS][N]} - \alpha_5[SC] \quad (4.11)$$

$$\frac{d[D]}{dt} = \frac{\eta_6}{1 + \gamma_6 + q_1[OS] + q_2[OS][N]} - \alpha_6[D] \quad (4.12)$$

where m_i, n_i, q_i , are constants related to the free energies of binding of transcription factors to the operator regions of Oct4-Sox2, and Nanog.

4.1.2 Steady State Responses

At steady state, Equations (4.6)-(4.12) describe a nonlinear algebraic system which requires numerical solutions with concentrations of A_+ and B as inputs. This concentration dependence is shown in Figure 4.3. As seen in the figure, the presented model presents a bistable switch behavior with respect to the input signal A_+ . With the increase in concentration of A_+ , there is an increase in the transcription rates of Oct4 and Sox2, with a subsequent increase in formation of Oct4-Sox2 complex. The increased transcription of OS complex further accelerates the transcription of Oct4 and Sox2 and resultantly the formation of Oct4-Sox2 complex. Concomitantly, the increased binding of Oct4-Sox2 on the promoter region of Oct4 and Sox2, results in the recruitment of Nanog for further transcription of both Oct4 and Sox2 and itself. This activation of the positive feedback loops suddenly activates the three transcription factors, turning the stem cell switch in the ON position at some concentration of A_+ . Noticeably, at the

latter concentration of A_+ , the target genes for differentiation in both Coherent and Incoherent-Type FFLs are completely OFF and the target genes for self-renewal are completely ON.

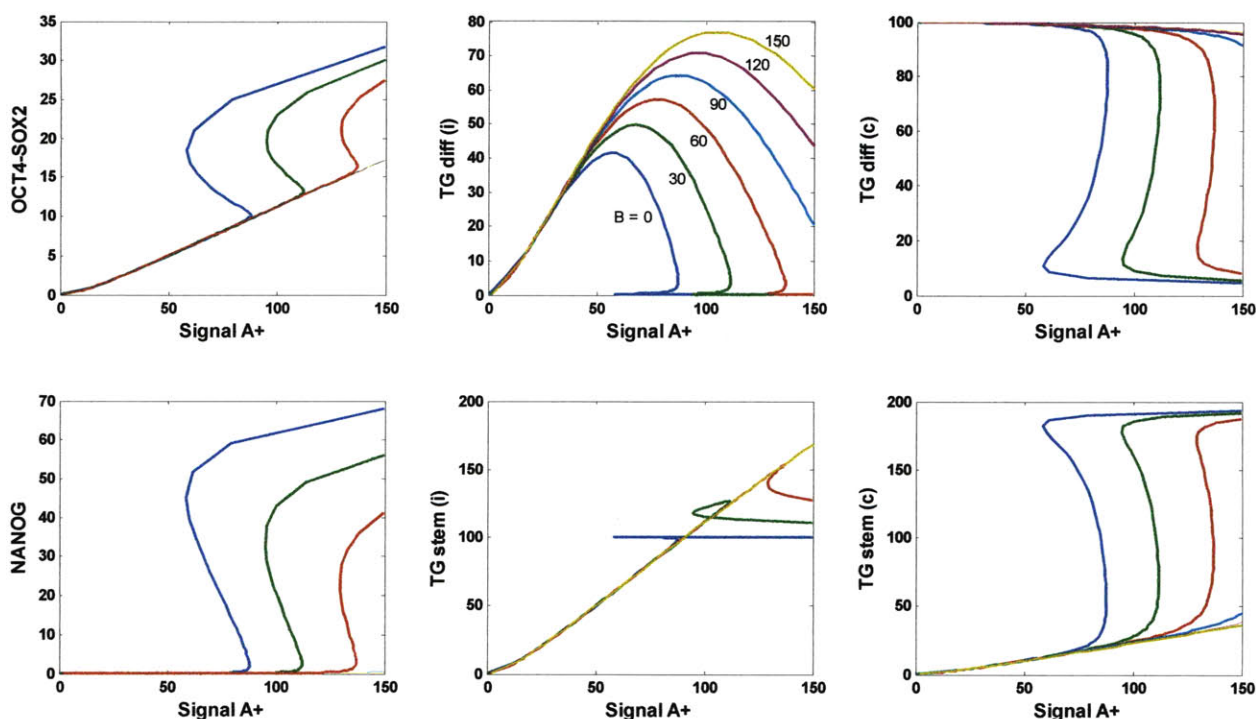


Figure 4.3: Dependence of the transcription factors and target genes on the concentration of A_+ and B . First column shows the concentration of Oct4-Sox2 and Nanog. Second column shows the target genes for differentiation and self-renewal given by the Incoherent-Type FFL, and the third column shows these target genes obtained by the Coherent-Type FFL.

In the Incoherent-Type FFL, both target genes for differentiation and pluripotency initially increase with the increase in concentration of A_+ . However, at the threshold concentration of A_+ at which the system turns ON, the differentiation target genes falls down to the OFF position of the switch and the pluripotent target genes remain steady at their maximum value. However, in the Coherent-Type FFL, the differentiation target genes start at their ON position and as the levels of the external signal A_+ increase, their concentration decreases and differentiation genes are suddenly switched to the OFF position. As seen in the figure, opposite behavior is observed for the stem cell target genes. This positive feedback also leads to hysteresis in the system. The threshold concentration of A_+ required to turn ON the system is higher than the one required turning the system OFF as seen in Figure 4.3. It is important to note that as the concentration of

external signal B . increases, there is a repression of Nanog resulting in its decreased transcription, which keeps the system in the OFF position. At higher concentrations of B . the threshold concentration of A_+ is higher for both turning ON and OFF the target genes. When the concentration of Nanog goes down, the transcription rate of Oct4 and Sox2 also decreases, leading to reduced formation of Oct4-Sox2 complex. This results in decreased binding of the OS complex on the promoter region of Oct4 and Sox2, and reduced Nanog recruitment on the promoter region of Oct4 and Sox2, decelerating further the transcription of all the three transcription factors. For a fixed concentration of A_+ , the dependence of the transcription factors and target genes follows a bistable switch behavior which tends to turn OFF the system with increase in B . concentration.

4.2 APPLICATION OF PARETO OPTIMALITY FOR THE PREDICTION OF THE REGULATION OF TARGET GENES BOX ARCHITECTURE.

As seen in Section 4.1, the external variables A_+ and B . regulate the transcription factors in the stem cell box which in turn modifies the expression of the target genes in the regulation of target genes box. Since there is a positive feedback, changes in the regulation of transcription genes box also modifies the expression in the stem cell box, as described by the non-linear algebraic system given by Equations (4.6)-(4.12). Based on the developmental cell biology literature and known concepts in embryonic cell stem cell cultures we know that irrespective of the nature of interaction between Oct4-Sox2, Nanog and the target genes, there is always a tradeoff between self-renewal and differentiated state, hence, there must exist a Pareto frontier between self-renewal and differentiation (Figure 4.4). This stems from the fact that in embryonic and mature cellular systems it has been noticed that there is trade-off between proliferation and differentiation. Hence, in a proliferative state cells may not be completely differentiated and likewise for the differentiation state. Consequently, there is a tradeoff between the external variables that allow the cell to make the decision to go from a pluripotent state to a fully specialized cell. As expected from the switch-like behavior, if the cell is in the stem cell mode, the target genes for self renewal must be completely ON and those for the differentiation must be fully OFF. The opposite is true for the differentiated state in cells. Here, we hypothesize that cells work in an optimal state which is thermodynamically favorable for the entire system.

Therefore, in addition to the expected switch ON-OFF behavior, most favorable network's Pareto frontier is expected to have a maximal range and should be closest to the utopian points. This maximal Pareto frontier between differentiation and stem cell genes may also represent the thermodynamically favorable energetic state of the cell. Hence, we postulate that the network which provides maximal Pareto frontiers (obtained by maximization of target genes) is energetically more favorable because of the efficient utilization of energetic resources.

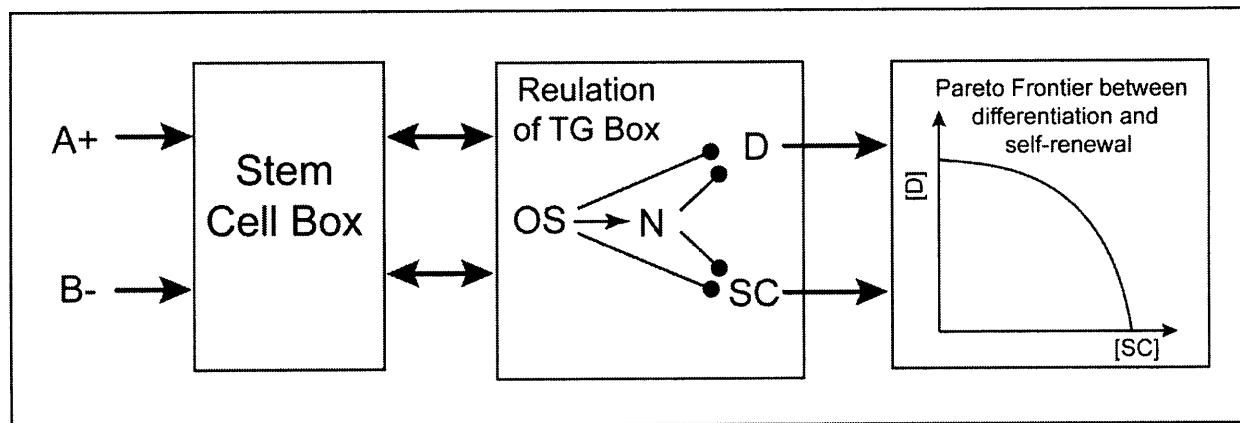


Figure 4.4: Strategy for determination of the Pareto frontier between differentiation and self-renewal target genes. Dotted-headed arrows in the regulation of target genes box indicate either activation or repression. OS: Oct4-Sox2, N: Nanog, D: target genes for differentiation, SC: target genes for self-renewal.

The Pareto problem shown in Figure 4.4 can be mathematically expressed as follows:

For Coherent and Incoherent Type FFLs:

$$\text{Maximize}_{\vec{x}} ([SC], [D])$$

subject to:

$$f(\vec{x}) = 0 \tag{4.13}$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

where $f(\vec{x})$ represents the nonlinear equations at steady state given by (3.24)-(3.30) and $\vec{x} = [[A_+] \ [B_-]]^T$.

Figure 4.5 shows the Pareto frontiers for both Incoherent (Figure 4.5A) and Coherent (Figure 4.5B) Type FFLs in the regulation of target genes box. As seen in the figure, the Pareto frontier obtained from the Incoherent case partially turns the switch ON and OFF when the cell is allowed to move from the pluripotent state ($[SC] \approx 170$ and $[D] \approx 60$ [nM]) to the differentiated

mode ($[SC] \approx 115$ and $[D] \approx 75$ [nM]). In addition, the expression of B. is constant along the Pareto frontier at its maximum value, thus constraining the tradeoff region. The opposite behavior is seen in the coherent case (Figure 4.5B). When the cell is in the pluripotent state ($[SC] \approx 195$ and $[D] \approx 5$ [nM]), the concentration of the stem cell genes are at higher value than in the incoherent case, and at the same time the differentiation genes are practically OFF. At the anchor point, the concentration of B. is nearly zero. As the cell differentiates, B. increases, and reaches its maximum value when the cell is fully differentiated. At the anchor point, $[SC] \approx 0$ and $[D] \approx 100$ [nM]. The target genes for differentiation for Coherent type-FFL relation are at higher value than that in the Incoherent-Type FFL and at the same time self-renewal genes are nearly zero. Importantly, through Pareto optimal solutions, we can see the switch-like behavior *only* in the Coherent case, and thus the most likely architecture of the Regulation of Target Genes Box is the Coherent-Type 1 FFL for pluripotency and Coherent-Type 3 for differentiation.

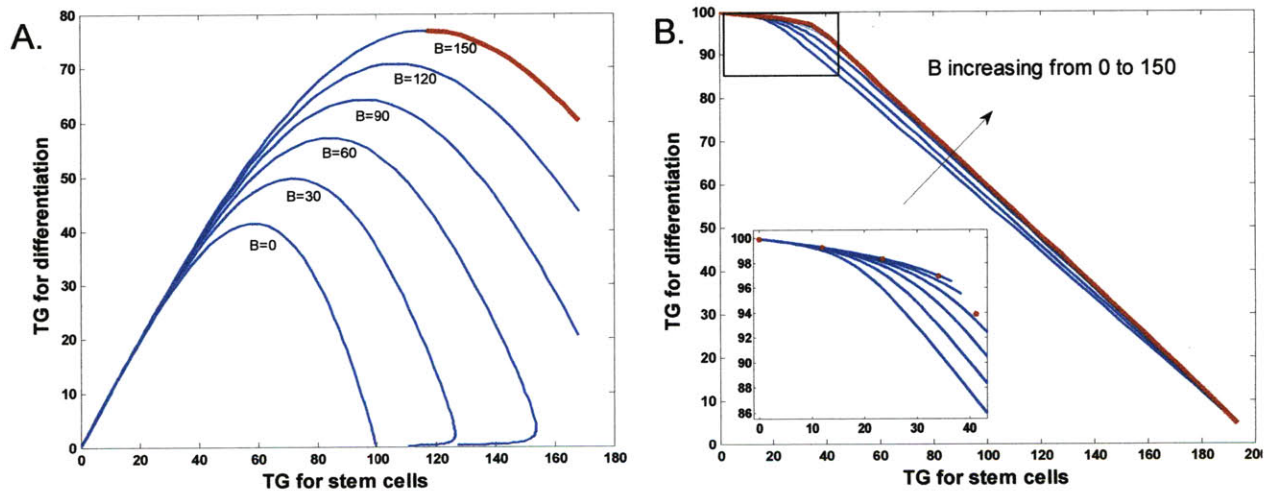


Figure 4.5: Feasible region (in blue) and Pareto Frontier (in red) between the target genes for differentiation and stem cell. **A)** Incoherent-Type FFL. **B)** Coherent-Type FFL.

4.3 EXTENSION OF THE TRANSCRIPTION FACTORS TO FOXD3

In addition to the upregulation of Oct4, Sox2 and Nanog during the embryonic cell state, it has been reported that FoxD3, a forkhead family member D3 transcription factor, is also highly

expressed in mouse pluripotent cells (Pan et al., 2006). FoxD3 activates Nanog by binding its promoter region *in vivo* and forms a negative feedback loop between Oct4 and Nanog. Oct4 has been recently shown to directly bind the promoter region of Nanog in a concentration dependent manner, thus modifying the *Stem Cell Box* shown in Figure 4.6.

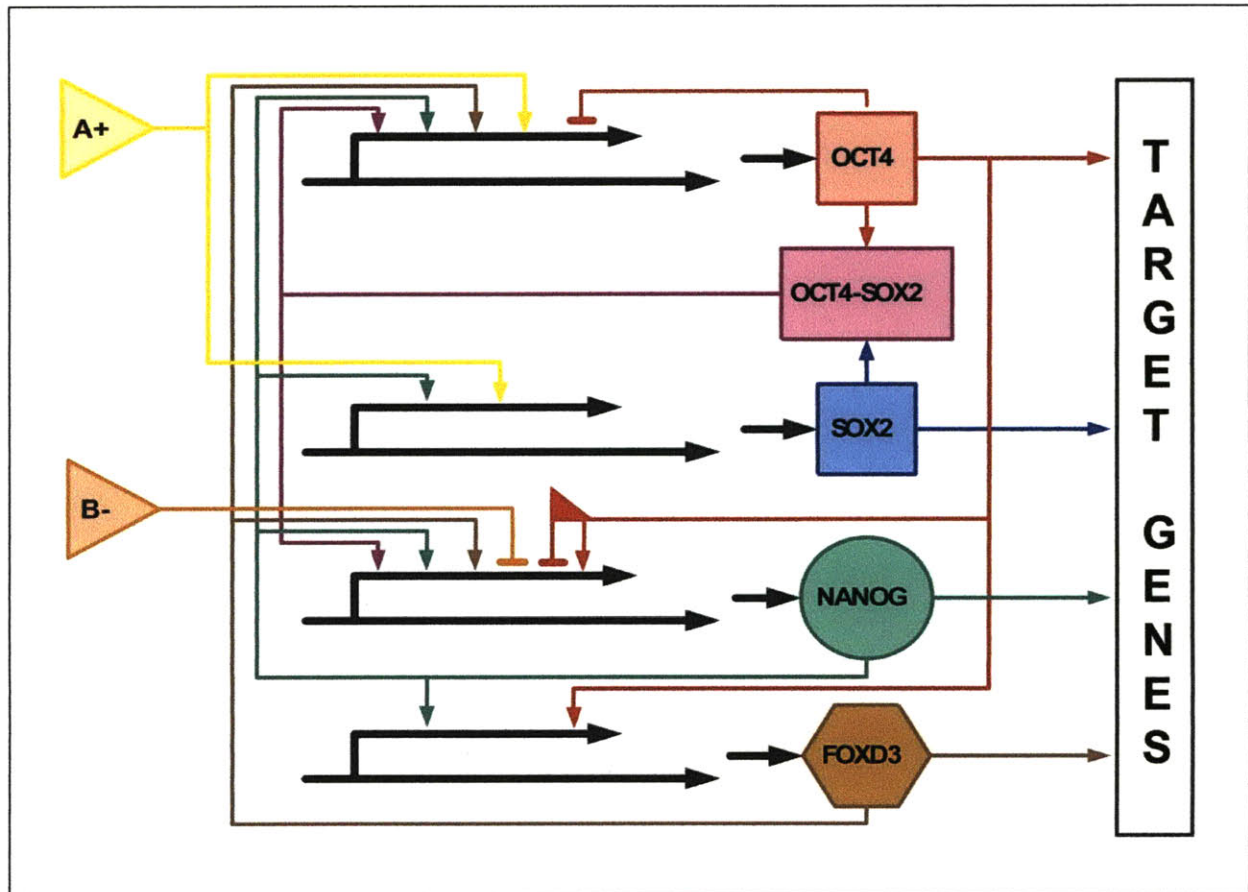
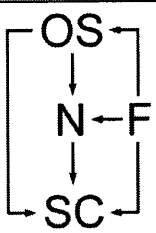
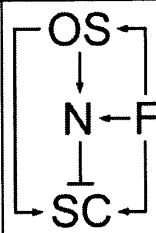
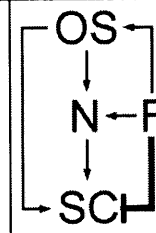
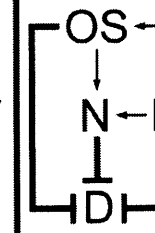
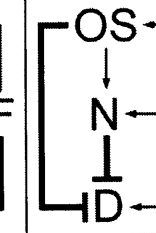
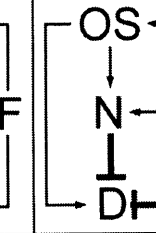


Figure 4.6: The core transcriptional network of ES, including FoxD3 interactions.

Below the steady state concentration, Oct4 activates Nanog and opposite is seen when the Oct4 concentration is lower than this threshold. On the other hand, FoxD3 activates Nanog to reduce the effect of repression due to excess Oct4. Then FoxD3 and Nanog enhance Oct4 expression which in turn represses itself and Nanog, leading to a negative feedback regulation loop to control its own concentration, keeping the Oct4 values at steady state and maintaining ES cell pluripotent state. The presence of FoxD3 changes the proposed FFL-like architecture for the regulation of target genes box, leading to a four-node network between Oct4-Sox2, Nanog, FoxD3 and the target genes for pluripotency and differentiation. Three different combinations

between activation and repression are proposed in this work for stem-cellness and specialization, as shown in Table 4.1.

Table 4.1: Proposed regulation of target genes box with FoxD3. Three different activation (A) – repression (R) logics from 1 to 3 are presented for both self-renewal (SC) and differentiation (D) genes. OS: Oct4-Sox2, N: Nanog, F: FoxD3. Notice that in SC2 network, Nanog is proposed as a weak repressor (WR).

	SC1	SC2	SC3	D1	D2	D3
OS	A	A	A	R	R	A
N	A	WR	A	R	R	R
F	A	A	R	R	A	R
Network						
parameter = 0	-----	m_2	m_3	o_1, o_3	o_1	o_3

In the first logic for stem cell target genes (SC1), all the three transcription factors are activating the genes required for pluripotency, in contrast to the first logic for the differentiation genes (D1), where all of them are repressing these genes. SC2 differs from SC1 because of Nanog being a weak repressor of the final genes, and SC3 has FoxD3 a strong repressor. D2 differs from D1 because of FoxD3 which acts as an activator of the target differentiation genes, and D3 considers Oct4-Sox2 as an activator. Whether the mechanism responsible for pluripotency or differentiation follows any of the three proposed logics or combinations thereof for each of them is not known. Once again, we propose using a Pareto-optimal transcriptional regulatory network framework to elucidate and predict the most likely interaction between transcription factor and target genes.

4.3.1 Mass Balance Equations

The mass balance equations for Oct4 and Nanog now include the interaction of FoxD3, F :

$$\frac{d[O]}{dt} = \frac{\eta_1 + a_1[A_+] + a_2[OS] + a_3[OS][N] + a_4[F]}{1 + \gamma_1 + b_1[A_+] + b_2[OS] + b_3[OS][N] + b_4[F] + b_5[O]} - \alpha_1[O] - k_{1c}[O][S] + k_{2c}[OS] \quad (4.14)$$

$$\frac{d[N]}{dt} = \frac{\eta_3 + e_1[OS] + e_2[OS][N] + e_3[F] + e_4[O]}{1 + \gamma_3 + f_1[OS] + f_2[OS][N] + f_3[B_-] + f_4[O] + f_5[F]} - \alpha_3[N] \quad (4.15)$$

As the transcription behavior of Nanog by Oct4 is concentration dependent, here we propose a Gaussian functionality (Figure 4.7) for the free energy of binding of Oct4 in the promoter region of Nanog, e_4 . At $[O]$ required for keeping the system in the self-renewal state, e_4 is at its maximum value and at this $[O]$ concentration Oct4 behaves as an activator of Nanog. For higher or lower concentrations, this parameter approaches to zero, and thus Oct4 represses the transcription of Nanog.

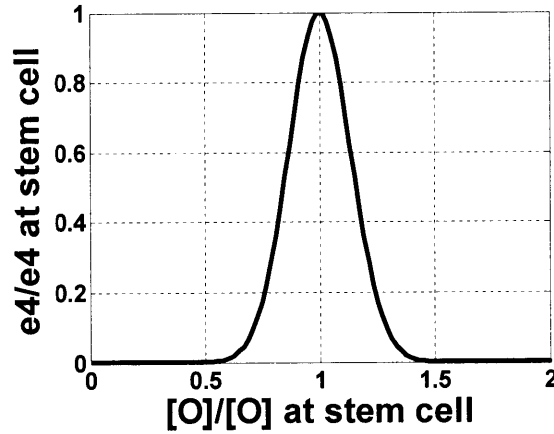


Figure 4.7: Concentration dependence on Oct4 of the free binding energy of this transcription factor on the promoter region of Nanog.

The mass balance equation for FoxD3, F can be written as:

$$\frac{d[F]}{dt} = \frac{\eta_7 + g_1[O] + g_2[OS][N]}{1 + \gamma_7 + h_1[OS] + h_2[OS][N]} - \alpha_7[F] \quad (4.16)$$

and the general equations for the self-renewal and differentiation target genes are given by:

$$\frac{d[SC]}{dt} = \frac{\eta_5 + m_1[OS] + m_2[OS][N] + m_3[F]}{1 + \eta_5 + n_1[OS] + n_2[OS][N] + n_3[F]} - \alpha_5[SC] \quad (4.17)$$

$$\frac{d[D]}{dt} = \frac{\eta_6 + o_1[OS] + o_3[F]}{1 + \gamma_6 + p_1[OS] + p_2[OS][N] + p_3[F]} - \alpha_6[D] \quad (4.18)$$

Notice that depending on the proposed activation-repression logic, parameters m_i and o_i are zero in Equations (4.17) and (4.18) (Table 4.1).

4.3.2 Steady State Responses

Numerical solutions of the steady state mass balance are shown in Figures 4.8 and 4.9. Notice that the addition of FoxD3 in the transcription network does not effect the switch-like behavior and the observed hysteresis presented by (Chickarmane et al., 2006).

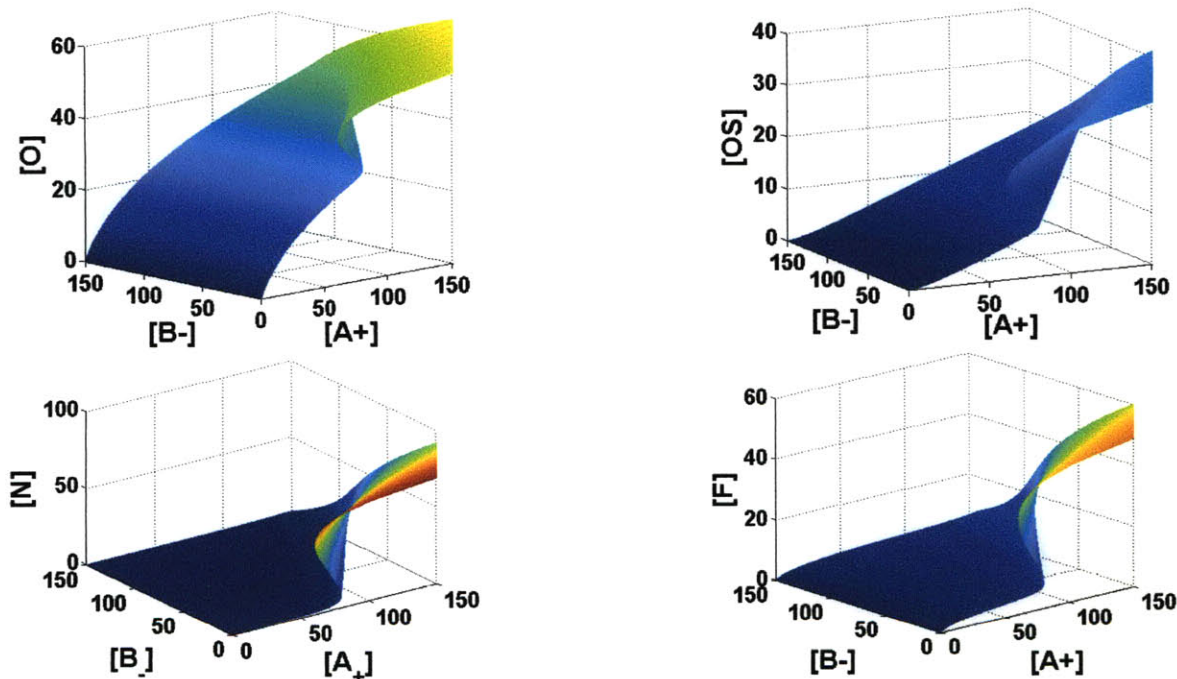


Figure 4.8: Steady state solutions for Oct4 (O), Oct4-Sox2 (OS), Nanog (N) and FoxD3 (F) in the Stem Cell Box.

With the increase in concentration of A_+ , there is a rapid increase in the concentration of Oct4 because of the direct effect of A_+ in the transcription of these factors. As both levels of Oct4 and Sox2 (not shown in the figure, but surface similar to Oct4) increase, Oct4-Sox2 complex formation starts initially at a low rate. The increase in the concentration of O-S complex further allows the recruitment of Nanog in the promoter regions of Oct4 and Sox2, as well as its

own transcription, accelerating the positive feedback mechanism. As both Oct4 and Nanog are produced, they further promote the transcription of FoxD3 which in turn increases the transcription of Oct4, Sox2 and Nanog because of the feedback. At low $[B_-]$, the activation of the positive feedback loops suddenly moves the system to another steady state at a threshold concentration of A_+ , turning ON the switch for the self-renewal target genes and turning OFF the switch for the differentiation genes, in all the proposed logics, shown in Figure 4.9. When the system is at the pluripotent state, decreasing $[A_+]$ turns OFF all the transcription factors and the self-renewal target genes, at the same time the differentiation genes are turned ON. However, the threshold concentration of A_+ is lower than the one required moving the system in the opposite direction, showing the existence of hysteresis. As the concentration of B_- increases, it gets harder to turn ON-OFF the system. Further, as the transcription of Nanog reduces, the effect of the positive feedback loops is diminished. By reducing the concentration of Nanog, the levels of Oct4, Sox2, Oct4-Sox2 and FoxD3 decreases, and thus the system moves towards the differentiated state.

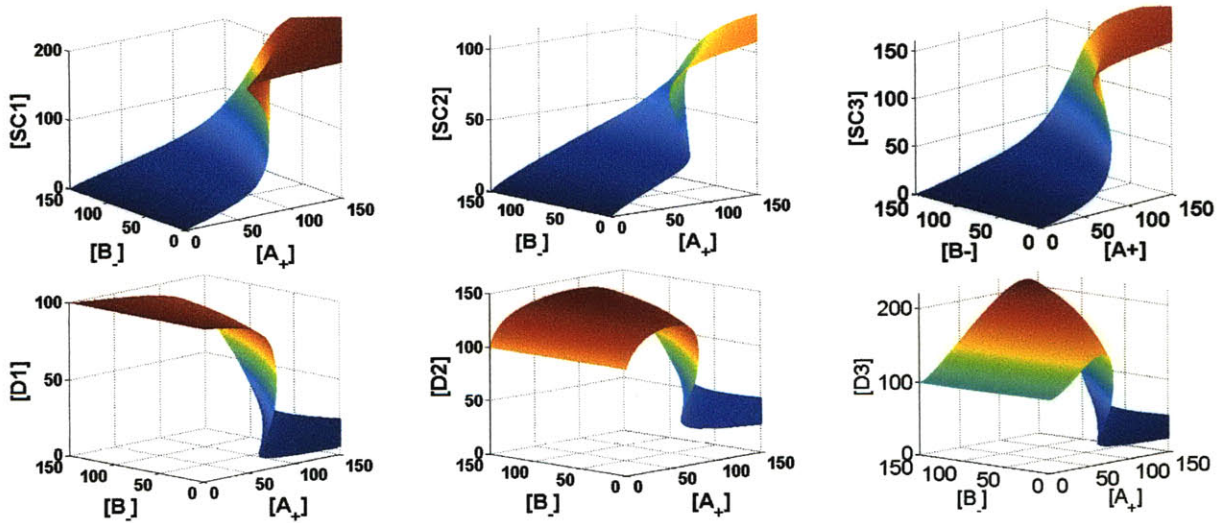


Figure 4.9: Steady state solutions for the proposed activation-repression logics in the Regulation of Target Genes Box.

4.3.3 Pareto Optimality for Prediction of Network Architecture

As explained before, the external inputs A_+ and B_- regulate the transcription in the stem cell box which in turn modifies the expression of the target genes in the regulation of target

genes box. The positive feedback allows the interaction between the regulation of transcription genes box and the stem cell box, as described by the non-linear algebraic equations of the system. As explained earlier, we again hypothesize existence of a Pareto frontier between self-renewal and differentiation, independent of the nature of interaction existing between Oct4-Sox2, Nanog, FoxD3 and the target genes. We again postulate that the cellular systems work in an optimal manner that must be thermodynamically favorable for the entire system. Clearly, the most favorable network should present the maximal Pareto surface, which also represents the favorable energetic state for the cell. Thus, there is a tradeoff between the external inputs that allow the cell to make the decision to go from a pluripotent state to a fully matured or differentiated state of the cell. In addition, as expected from the switch-like behavior, if the cell is in the stem cell mode, the target genes for self renewal must be completely ON and those for differentiation must be fully OFF. The opposite is true for the differentiated cell. Then, by maximizing both target genes we postulate that valuable information can be obtained from the Pareto frontiers.

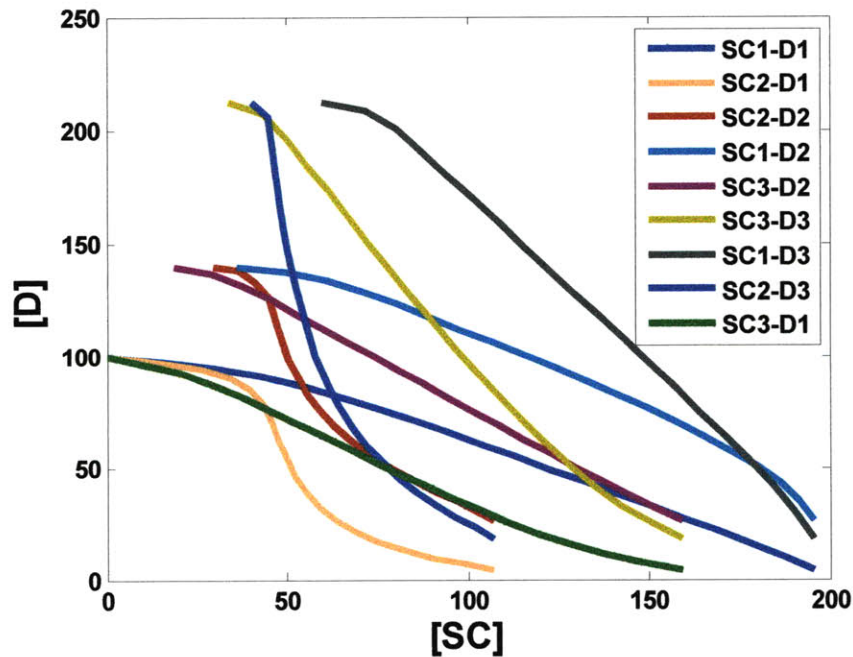


Figure 4.10: Pareto frontiers between self-renewal and differentiation target genes for the proposed activation-repression logic between Oct4-Sox2, Nanog, FoxD3, and the target genes in the regulation

From the proposed activation-repression logics, a total of nine combinations between target genes for self-renewal and differentiation are possible and they are shown in Figure 4.10. Among these, only three of them follow the switch-like behavior: SC1-D1, SC2-D1 and SC3-D1. Using the above mentioned Pareto optimality paradigm, reveals that D1 is the most favorable logic for the differentiation target genes. In D1, Oct4-Sox2, Nanog and FoxD3 are repressors for the differentiation target genes. The combination of D1 with any of the proposed logics for the stem cell state completely turns the genes ON and OFF at the corresponding anchor points. However, if we assume that the maximal Pareto frontier represents a better energetic state for the cell, the maximal Pareto frontier (or optimal energetic state) is given by SC1 (which is the case where Oct4-Sox2, Nanog and FoxD3 are activators for the differentiation target genes). Then, by applying Pareto optimality in the stem cell system, we can observe the switch-like behavior in addition to the maximal Pareto frontier for the combination given by SC1 and D1. For this reason, the most likely architecture of the Regulation of Target Genes Box will be SC1-D1 combination. If the switch-like behavior is not taken into account and only the maximal Pareto frontier is used as criteria, then the two highest tradeoff regions are given by SC1-D2 and SC1-D3. Again, SC1 seems to be the optimal network logic for these transcription factors. Since the Pareto frontiers in SC1-D2 and SC1-D3 overlap, hence, the network architecture logic may utilize in this case a combination of D2 and D3.

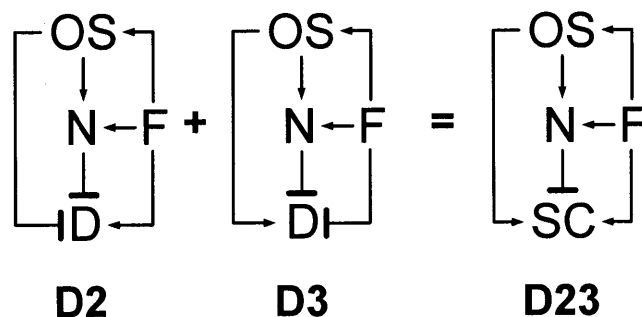


Figure 4.11: New activation-repression logic for the differentiation target genes

Figure 4.11 shows this new proposed logic which we will define as D23. Since, D2 and D3 both have Nanog as a repressor, hence, we propose D23 as the one in which Oct4-Sox2 and FoxD3 will act as activators and Nanog as the repressor of the differentiation target genes. Figure 4.12 shows the Pareto frontier between SC1 and D23. This combination indeed results in the

maximal Pareto frontier which partially turns ON-OFF the system. In this section we showed that the Pareto optimality concept as a strategy can predict the most likely activation-repression logic for a given transcription network. The proposed paradigm is based on the hypothesis that the cellular systems try to maximize expression of target genes with an underlying thermodynamic basis (for example, entropy or total work). Although this concept is intuitive and based on physical principles, it requires validation for variety of systems. Thus, a non-equilibrium transcription network model was further developed .

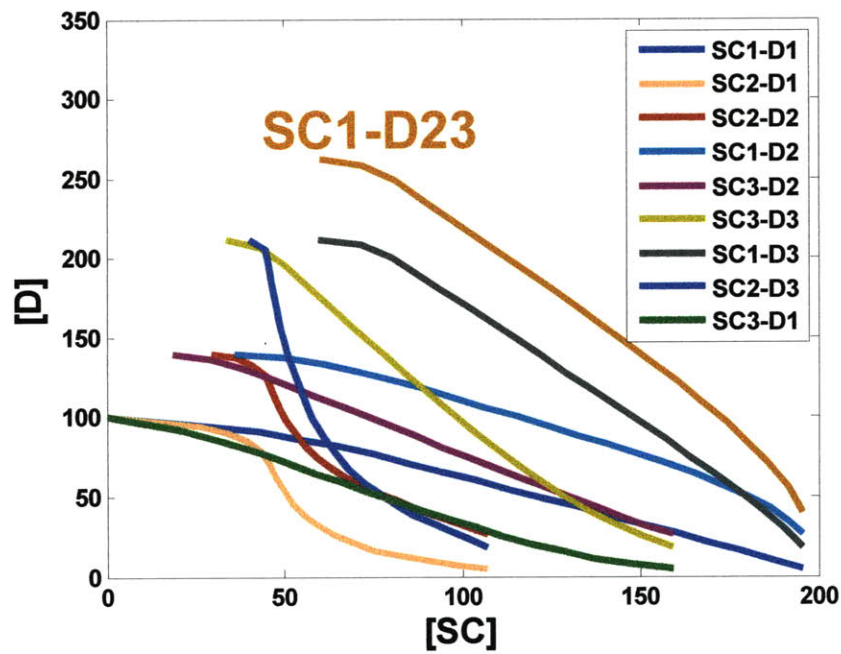


Figure 4.12: Pareto frontiers between self-renewal and differentiation target genes for the proposed activation-repression logic between Oct4-Sox2, Nanog, FoxD3, and the target genes in the regulation of target genes box. The new proposed logic D23 is shown.

5 KINETIC MODELING OF TRANSCRIPTIONAL REGULATORY NETWORKS USING CYCLE CONCEPT

5.1 OVERVIEW

This novel strategy to model the translation and transcription process of any protein Z in a given network not only allows the calculation of the transcription rate but also the estimation of the energy associated with it, in particular, the computation of the heat dissipation rate (HDR). In this approach, each individual transcription factor or the combined set of transcription factors (in case that cooperativity exists) has a cycle associated to it. On the other hand, each cycle is composed of three steps modeled as chemical reactions at nonequilibrium steady state (NESS) which contrasts the basic thermodynamic equilibrium hypothesis in the existing transcriptional regulatory network (TRN) models. At thermodynamic equilibrium steady state, the net flow (J) and chemical potential ($\Delta\mu$) associated to each chemical reaction are null. Since the heat dissipated by each reaction is given by $-J \cdot \Delta\mu$, then at thermodynamic equilibrium the translation and transcription of protein Z is energy free. However, at NESS the net reaction flow is different than zero and its direction is given by the sign of the corresponding chemical potential. If $\Delta\mu < 0$, then $J > 0$, and vice versa. Hence, the heat dissipated by each chemical reaction is always positive, $-J \cdot \Delta\mu > 0$. Notice that $-J \cdot \Delta\mu = 0$ if and only if $\Delta\mu = 0$ and $J = 0$.

5.2 NONEQUILIBRIUM THERMODYNAMICS

Nonequilibrium processes require an external signal flux or driving force to maintain the system far away from equilibrium. As seen in Figure 5.1A, for the mass transfer process, an oxygen concentration gradient between air and water is maintained by wind; in the case of potential energy (Figure 5.1B), external weights can maintain differential nonequilibrium position; in the heat transfer (Figure 5.1C) processes, an external heat source can disturb the equilibrium temperature between the surrounding atmosphere and water; in a chemical reaction (Figure 5.1D), the external reactant flux can drive the reaction system away from equilibrium; and finally, in an enzymatic cyclic reaction (Figure 5.1E), an external substrate flux can drive the cycle in a particular direction. These nonequilibrium processes have their steady states

maintained away from equilibrium because of an external flux and thus, they reach equilibrium without this driving force.

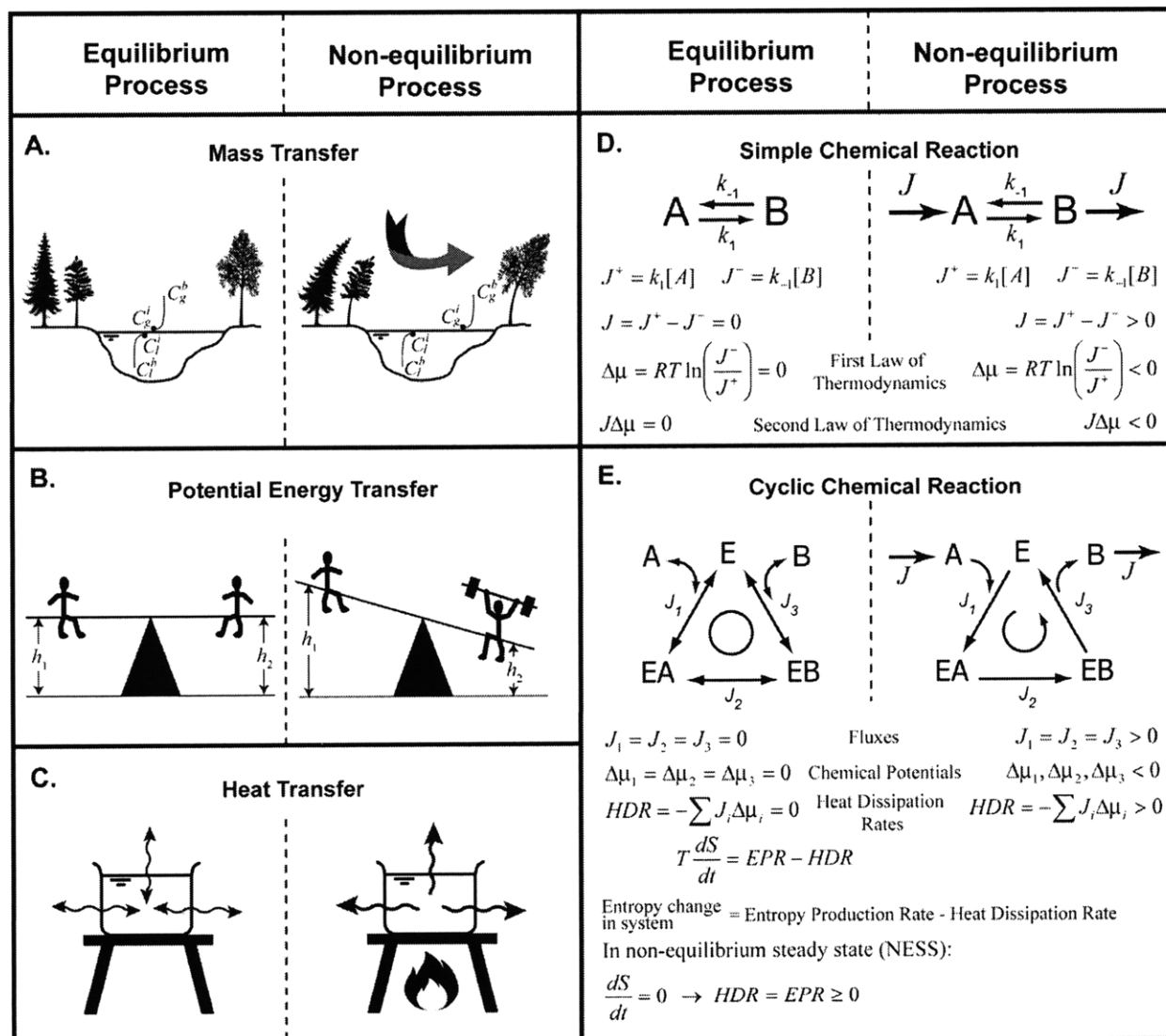


Figure 5.1: Concept of nonequilibrium explained using various physicochemical processes: A) Mass transfer. B) Potential energy transfer. C) Heat transfer. D) Simple chemical reaction. E) Cyclic chemical reaction. Under equilibrium conditions in all processes there is no transport flux and hence work done on the system and the entropy increase of the system is zero. In contrast, in nonequilibrium processes, there is a driving force which maintains the corresponding nonequilibrium steady states and dictates how far from equilibrium the inlet and outlet states can be maintained.

The recent surge of interest in nonequilibrium thermodynamics has strikingly and clearly exposed its role in small-scale systems. Equilibrium systems are generally governed by classical thermodynamics and have reversible work equal to the Gibbs free energy change at constant

temperature and pressure. Nonequilibrium steady-state (NESS) analysis generally is used for irreversible systems and has Gibbs free energy unequal to the work done on system because of heat/work dissipation and hysteresis in these processes. Recently, rather than classical thermodynamics, nonequilibrium thermodynamics has been shown to correlate well in describing behavior of small-scale systems such as biological molecular machines (2-100 nm) (Qian, 2004) which convert chemical energy into work and RNA folding/unfolding involved in biological cells by helicases or ribosomes (Liphardt et al., 2002).

In his seminal work, Jarzynski presented a relation currently defined as Jarzynski's equality (JE):

$$\exp\left(\frac{-\Delta G}{k_B T}\right) = \left\langle \exp\left(-\frac{W}{k_B T}\right) \right\rangle \quad (5.1)$$

where ΔG is the free energy difference on going from a state A to state B, and the right hand side term is the exponential average of infinitesimal nonequilibrium processes. JE allows an estimation of free energy differences under nonequilibrium conditions. The above relation can also be rewritten as:

$$1 = \left\langle \exp\left(-\frac{W - \Delta G}{k_B T}\right) \right\rangle \quad (5.2)$$

where $W - \Delta G = W_{dis}$ is the dissipated work and provides an estimate of work dissipated for a nonequilibrium process.

This brings an extension of the second law of thermodynamics and essentially states that $\langle W_{dis} \rangle \geq 0$ for a nonequilibrium process and thus, presents the estimation of the dissipated work for nonequilibrium states or trajectories moved away from equilibrium in terms of free energy and work. However, most importantly, JE states that the average dissipated work along any trajectory between states A and B is always greater than zero i.e. $\langle W_{dis} \rangle \geq 0$. Based on JE, we hypothesized in current work that because of (a) maintenance requirement, (b) nutrient limitations, and (c) efficient nutrient utilization, biological systems in nature may tend to minimize this dissipated work. However, it is to be noted that this dissipated work will be directly proportional to the mass or raw materials utilized in moving from state A to state B. Hence, we hypothesized that specific dissipation energy which we define in this current work as

the dissipation normalized to its input mass may serve as an indicator for choosing a trajectory to move from state A to state B. This conclusion is similar to the one made in ecosystems development where, among several possible indicators, specific dissipation energy was believed to be the primary criterion for ecosystem maturity (Odum and Pinkerton, 1955). Taking a step forward, we further hypothesized that since structural design or pathways in cellular networks are energy intensive processes, thus, during evolution these designs were converged based on optimal energetic dissipation criterion. Specifically, a structural design will be chosen or selected which has minimal specific dissipation energy. To investigate whether energetic cost criterion is the primary basis for selection of transcriptional motifs we developed a kinetic cyclic transcriptional regulatory network modeling framework. This allows estimation of heat dissipation or energetic cost involved in the transcription-translation process. Similar to the nonequilibrium steady state (NESS) analysis for biochemical reactions (Qian and Beard, 2005), we developed a novel nonequilibrium thermodynamic kinetic formulation for gene transcription. The developed framework was used for estimation of heat dissipation rates for both steady state and dynamic analysis. NESS analysis in biochemical reactions has revealed that chemical potential of nonequilibrium processes is equivalent to heat dissipated by the systems or work done on the system which is equivalent to entropy generated by the system under NESS condition. To address the question of energetic cost involved during transcription we utilized the second law of thermodynamics which states that for any reaction:

$$-J \cdot \Delta\mu \geq 0 \tag{5.3}$$

where J and $\Delta\mu$ are the net flux and chemical potential energy of any reaction, respectively. The first law of thermodynamics establishes that the sum of reaction potentials around any cycle of reactions equals zero. In terms of forward (J_+) and backward (J_-) reaction fluxes, the chemical potential is written as $\Delta\mu = RT \ln(J_- / J_+)$, and thus the second law of thermodynamics can be expressed as

$$-J \cdot \Delta\mu = -RT(J_+ - J_-) \ln\left(\frac{J_-}{J_+}\right) \geq 0 \tag{5.4}$$

which states that the system must dissipate heat and entropy must increase as a result of the work being done on the system through the external fluxes. For equilibrium systems, this is an equality since for these systems $J = \Delta\mu = 0$.

The total heat dissipation rate (*HDR*) of a multiple reaction system is always positive as indicated by

$$HDR = -\sum J_i \cdot \Delta\mu_i \geq 0 \quad (5.5)$$

Observe that $HDR \rightarrow 0$ in the limit as $J_- / J_+ \rightarrow 1$ component-wise while maintaining $J = J_+ - J_-$.

(Qian and Beard, 2005) showed that the time change of the total entropy of the system equals the difference between the entropy production rate (*EPR*) and the heat dissipation rate (*HDR*):

$$T \frac{dS}{dt} = EPR - HDR \quad (5.6)$$

This equation states that, under isothermal conditions, the change of total entropy is either due to entropy created in the system (source term) or heat leaving the system (sink term) (Qian and Beard, 2005). In a nonequilibrium steady state (NESS) $dS/dt = 0$ and thus

$$HDR = EPR \quad (5.7)$$

Equation (5.7) is known as the isothermal Clausius equality. Because $HDR \geq 0$ (Second Law of Thermodynamics), $EPR \geq 0$.

5.3 ACTIVATION OF Z BY X

5.3.1 Mechanism

The mechanism by which the protein *Z* is transcribed by the transcription factor *X* can be divided in a three step process as is shown in Figure 5.2.

Step 1: The transcription factor *X* binds a free DNA site of the promoter region of *Z* (D_Z) to form an occupied DNA site (D_{ZX}):



If k_1 [$s^{-1} \cdot nM^{-1}$] and k_{-1} [s^{-1}] are the forward and backward kinetic constants, respectively, then the net flux for this reaction is given by

$$J_1^{ZX} = k_1^{ZX} [D_Z][X] - k_{-1}^{ZX} [D_{ZX}] \quad (5.8)$$

and the reaction chemical potential is

$$\Delta\mu_1^{ZX} = \mathfrak{R}T \ln \left(\frac{k_{-1}^{ZX} [D_{ZX}]}{k_1^{ZX} [D_Z][X]} \right) = \mathfrak{R}T \ln \left(\frac{[D_{ZX}]}{K_1^{ZX} [D_Z][X]} \right) \quad (5.9)$$

where $\mathfrak{R} [\text{J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}]$ is the gas constant, $T [\text{K}]$ is the absolute temperature and $K_1^{ZX} = k_{-1}^{ZX} / k_1^{ZX}$ $[\text{nM}^{-1}]$ is the association equilibrium constant.

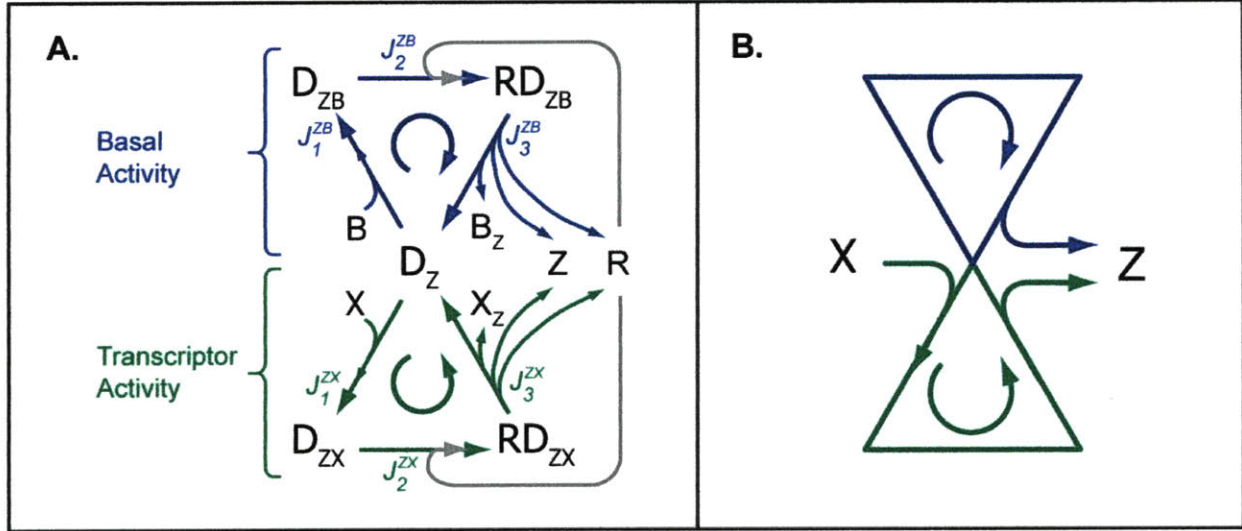


Figure 5.2: Representation of the activation pathway of protein Z by transcription factor X . **A)** Detailed mechanism showing both basal and transcription activity. **B)** Simplified representation of the cyclic concept.

Step 2: RNA polymerase (R) binds to the occupied DNA site (D_{ZX}) and forms a complex with this site (RD_{ZX}):



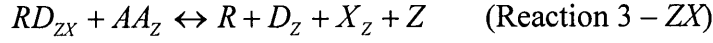
Let $k_2^{0,ZX} [\text{s}^{-1}\cdot\text{nM}^{-1}]$ and $k_{-2}^{ZX} [\text{s}^{-1}]$ the forward and backward kinetic constants for this step, respectively. The net flux for this reaction can be written as follows:

$$J_2^{ZX} = k_2^{0,ZX} [R][D_{ZX}] - k_{-2}^{ZX} [RD_{ZX}] = k_2^{ZX} [D_{ZX}] - k_{-2}^{ZX} [RD_{ZX}] \quad (5.10)$$

where $k_2^{ZX} = k_2^{0,ZX} [R] [\text{s}^{-1}]$ is a pseudo-first order kinetic constant. If $K_2^{ZX} = k_2^{ZX} / k_{-2}^{ZX}$ is the association equilibrium constant for this reaction, then the reaction chemical potential is given by

$$\Delta\mu_2^{ZX} = \mathfrak{R}T \ln \left(\frac{k_{-2}^{ZX} [RD_{ZX}]}{k_2^{ZX} [D_{ZX}]} \right) = \mathfrak{R}T \ln \left(\frac{[RD_{ZX}]}{K_2^{ZX} [D_{ZX}]} \right) \quad (5.11)$$

Step 3: Once RNA polymerase is bound, there is a recruitment of the amino acids (AA_z) required to form the protein Z . As a result of this step, transcription of Z occurs, releasing the RNA polymerase molecule, the free DNA site and the co-activator necessary for this process (X_z):



If $k_3^{ZX} = k_3^{0,ZX} [AA_z] [s^{-1}]$ and $k_{-3}^{ZX} = k_{-3}^{0,ZX} [R] [s^{-1} \cdot nM^{-2}]$ are pseudo-first order kinetic constants, and $K_3^{ZX} = k_3^{ZX} / k_{-3}^{ZX} [nM^2]$ is the association equilibrium constant for this reaction, then the reaction flux and reaction chemical potential are given by

$$J_3^{ZX} = k_3^{ZX} [RD_{ZX}] - k_{-3}^{ZX} [D_z][X_z][Z] \quad (5.12)$$

$$\Delta\mu_3^{ZX} = \Re T \ln \left(\frac{[D_z][X_z][Z]}{K_3^{ZX} [RD_{ZX}]} \right) \quad (5.13)$$

The described steps 1, 2 and 3 are represented as a green triangular cycle in Figure 5.2. An analogous mechanism describes the basal activity (blue cycle in Figure 5.2), in which a basal transcription factor B also binds the free promoter region of Z , initiating the cycle and further transcription of Z . Reactions 1 to 3 and equations (5.8) to (5.13) are identical except for the fact that X is replaced by B in all of them.

5.3.2 Mass Balance Equations

Let J_i and α_i the external intake flux (defined positive is the flux enters the system and negative if exits) in $[nM \cdot s^{-1}]$ and the degradation rate in $[s^{-1}]$ of the species i , respectively. Then, the mass balance is described by the following differential equations:

$$\frac{d[X]}{dt} = J_X - J_1^{ZX} - \alpha_X [X] \quad (5.14)$$

$$\frac{d[B]}{dt} = J_B - J_1^{ZB} - \alpha_B [B] \quad (5.15)$$

$$\frac{d[Z]}{dt} = J_3^{ZX} + J_3^{ZB} + J_Z - \alpha_Z [Z] \quad (5.16)$$

$$\frac{d[X_z]}{dt} = J_3^{ZX} + J_{X_z} - \alpha_{X_z} [X_z] \quad (5.17)$$

$$\frac{d[B_Z]}{dt} = J_3^{ZB} + J_{B_Z} - \alpha_{B_Z}[B_Z] \quad (5.18)$$

$$\frac{d[D_Z]}{dt} = J_3^{ZX} + J_3^{ZB} - J_1^{ZX} - J_1^{ZB} \quad (5.19)$$

$$\frac{d[D_{ZX}]}{dt} = J_1^{ZX} - J_2^{ZX} \quad (5.20)$$

$$\frac{d[D_{ZB}]}{dt} = J_1^{ZB} - J_2^{ZB} \quad (5.21)$$

$$\frac{d[RD_{ZX}]}{dt} = J_2^{ZX} - J_3^{ZX} \quad (5.22)$$

$$\frac{d[RD_{ZB}]}{dt} = J_2^{ZB} - J_3^{ZB} \quad (5.23)$$

In addition to Equations (5.14) to (5.23), the total concentration of DNA sites must be constant at any time:

$$[D_Z]_{tot} = [D_Z] + [D_{ZX}] + [D_{ZB}] + [RD_{ZX}] + [RD_{ZB}] \quad (5.24)$$

5.3.3 General Steady State Solution

By solving Equations (5.19) to (5.23) at steady state, we have

$$J_1^{ZX} = J_2^{ZX} = J_3^{ZX} = J_C^{ZX} \quad (5.25)$$

$$J_1^{ZB} = J_2^{ZB} = J_3^{ZB} = J_C^{ZB} \quad (5.26)$$

being J_C^{ZX} and J_C^{ZB} [nM·s⁻¹] the cyclic fluxes for cycle ZX (shown in green in Figure 5.2) and for cycle ZB (shown in blue in Figure 5.2), respectively. By solving Equations (5.8), (5.10), (5.11) (and their analogous ones for the basal activity), (5.24) to (5.26), a solution for the mass balance of the DNA sites can be found. For $I = X, B$, and number of transcription factors $N_{TF} = 2$ (basal and activator), the following auxiliary variables can be defined:

$$a^{ZI} = k_{-1}^{ZI} k_{-2}^{ZI} + k_{-1}^{ZI} k_3^{ZI} + k_2^{ZI} k_3^{ZI} \quad (5.27)$$

$$b^{ZI} = k_1^{ZI} (k_{-2}^{ZI} + k_3^{ZI}) [I] + k_{-2}^{ZI} k_{-3}^{ZI} [Z] [I_Z] \quad (5.28)$$

$$c^{ZI} = k_1^{ZI} k_2^{ZI} [I] + k_{-3}^{ZI} (k_{-1}^{ZI} + k_2^{ZI}) [Z] [I_Z] \quad (5.29)$$

$$d^{ZI} = a^{ZI} + b^{ZI} + c^{ZI} \quad (5.30)$$

$$d_z = \left(\sum_{I=X,B} \frac{d^{ZI} \prod_{I=X,B} a^{ZI}}{a^{ZI}} \right) - (N_{TF} - 1) \prod_{I=X,B} a^{ZI} \quad (5.31)$$

Then, the fractional site concentrations and the cyclic flux are given by:

$$\frac{[D_Z]}{[D_Z]_{tot}} = \frac{a^{ZX} a^{ZB}}{d_z} = \frac{\prod_{I=X,B} a^{ZI}}{d_z} \quad (5.32)$$

$$\frac{[D_{ZI}]}{[D_Z]_{tot}} = \frac{b^{ZI} \prod_{J=X,B} a^{ZJ}}{a^{ZI} d_z} \quad (5.33)$$

$$\frac{[RD_{ZI}]}{[D_Z]_{tot}} = \frac{c^{ZI} \prod_{J=X,B} a^{ZJ}}{a^{ZI} d_z} \quad (5.34)$$

$$J_C^{ZI} = \frac{\prod_{J=X,B} a^{ZJ}}{a^{ZI} d_z} [D_Z]_{tot} (k_2^{ZI} b^{ZI} - k_{-2}^{ZI} c^{ZI}) \quad (5.35)$$

Notice that when only one cycle is present (for instance, X transcribes in absence of basal transcription rate), $I = X$ and $N_{TF} = 1$, then:

$$d_z = \left(\sum_{I=X,B} \frac{d^{ZI} \prod_{I=X} a^{ZI}}{a^{ZI}} \right) = d^{ZX} \quad (5.36)$$

$$\begin{aligned} [D_Z] &= \frac{a^{ZX}}{d^{ZX}} [D_Z]_{tot} \\ &= \frac{[D_Z]_{tot} (k_{-1}^{ZX} k_{-2}^{ZX} + k_{-1}^{ZX} k_3^{ZX} + k_2^{ZX} k_3^{ZX})}{k_{-1}^{ZX} k_{-2}^{ZX} + k_{-1}^{ZX} k_3^{ZX} + k_2^{ZX} k_3^{ZX} + (k_2^{ZX} + k_{-2}^{ZX} + k_3^{ZX}) k_1^{ZX} [X] + (k_{-1}^{ZX} + k_2^{ZX} + k_{-2}^{ZX}) k_{-3}^{ZX} [X_Z][Z]} \end{aligned} \quad (5.37)$$

$$\begin{aligned} [D_{ZX}] &= \frac{b^{ZX}}{d^{ZX}} [D_Z]_{tot} \\ &= \frac{[D_Z]_{tot} \{k_1^{ZX} (k_{-2}^{ZX} + k_3^{ZX}) [X] + k_{-2}^{ZX} k_{-3}^{ZX} [X_Z][Z]\}}{k_{-1}^{ZX} k_{-2}^{ZX} + k_{-1}^{ZX} k_3^{ZX} + k_2^{ZX} k_3^{ZX} + (k_2^{ZX} + k_{-2}^{ZX} + k_3^{ZX}) k_1^{ZX} [X] + (k_{-1}^{ZX} + k_2^{ZX} + k_{-2}^{ZX}) k_{-3}^{ZX} [X_Z][Z]} \end{aligned} \quad (5.38)$$

$$\begin{aligned}
[RD_{ZX}] &= \frac{c^{ZX}}{d^{ZX}} [D_Z]_{tot} \\
&= \frac{[D_Z]_{tot} \{k_2^{ZX} k_1^{ZX} [X] + k_{-3}^{ZX} (k_{-1}^{ZX} + k_2^{ZX}) [X_Z] [Z]\}}{k_{-1}^{ZX} k_{-2}^{ZX} + k_{-1}^{ZX} k_3^{ZX} + k_2^{ZX} k_3^{ZX} + (k_2^{ZX} + k_{-2}^{ZX} + k_3^{ZX}) k_1^{ZX} [X] + (k_{-1}^{ZX} + k_2^{ZX} + k_{-2}^{ZX}) k_{-3}^{ZX} [X_Z] [Z]}
\end{aligned} \tag{5.39}$$

$$\begin{aligned}
J_C^{ZX} &= \frac{[D_Z]_{tot} (k_2^{ZX} b^{ZX} - k_{-2}^{ZX} c^{ZX})}{d^{ZX}} \\
&= \frac{[D_Z]_{tot} (k_1^{ZX} k_2^{ZX} k_3^{ZX} [X] + k_{-1}^{ZX} k_{-2}^{ZX} k_{-3}^{ZX} [X_Z] [Z])}{k_{-1}^{ZX} k_{-2}^{ZX} + k_{-1}^{ZX} k_3^{ZX} + k_2^{ZX} k_3^{ZX} + (k_2^{ZX} + k_{-2}^{ZX} + k_3^{ZX}) k_1^{ZX} [X] + (k_{-1}^{ZX} + k_2^{ZX} + k_{-2}^{ZX}) k_{-3}^{ZX} [X_Z] [Z]}
\end{aligned} \tag{5.40}$$

For a given concentration of X and B , the steady state concentrations of Z , X_Z and B_Z can now be found by solving Equations (5.16) to (5.18), but analytical expressions lack of simplicity. In order to avoid this problem, an approximate solution is presented in the following section.

5.3.4 Approximate Steady State Solution

Because protein transcription is a highly irreversible process, one can expect that the rate of the forward reaction in each of the three steps has to be much higher than the rate of the corresponding backward reaction. This assumption moves the entire thermodynamic analysis very far from equilibrium and contrasts many other postulated models which equal these rates leading to thermodynamic equilibrium and null net fluxes. If we assume that the order of magnitude of the species concentrations is the same, then by taking $k_i^{ZX} \gg k_{-i}^{ZX}$ for $i = 1, 2, 3$, the net fluxes given by Equations (5.8), (5.10) and (5.13) can be written as

$$J_1^{ZX} \approx k_1^{ZX} [D_Z] [X] \tag{5.41}$$

$$J_2^{ZX} \approx k_2^{ZX} [D_{ZX}] \tag{5.42}$$

$$J_3^{ZX} \approx k_3 [RD_{ZX}] \tag{5.43}$$

Analogous equations can be written for the basal activity, by replacing X by B in the above equations. Recall that even though these reactions are considered highly irreversible, there is still a finite chemical potential associated to each one and given by Equations (5.9), (5.11) and

(5.13). As it is already known, the reaction fluxes are equal at steady state. For $I = X, B$, the auxiliary variables given by Equations (5.27) to (5.29) can be redefined as follows:

$$a^{ZI} \approx k_2^{ZI} k_3^{ZI} \quad (5.44)$$

$$b^{ZI} \approx k_1^{ZI} k_3^{ZI} [I] \quad (5.45)$$

$$c^{ZI} \approx k_1^{ZI} k_2^{ZI} [I] \quad (5.46)$$

Equations (5.30) to (5.35) can be used to find the DNA sites and cyclic fluxes:

$$d_Z \approx k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B] \quad (5.47)$$

$$[D_Z] \approx \frac{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.48)$$

$$[D_{ZX}] \approx \frac{k_1^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} [X] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.49)$$

$$[D_{ZB}] \approx \frac{k_2^{ZX} k_3^{ZX} k_1^{ZB} k_3^{ZB} [B] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.50)$$

$$[RD_{ZX}] \approx \frac{k_1^{ZX} k_2^{ZX} k_2^{ZB} k_3^{ZB} [X] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.51)$$

$$[RD_{ZB}] \approx \frac{k_2^{ZX} k_3^{ZX} k_1^{ZB} k_2^{ZB} [B] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.52)$$

$$J_C^{ZX} \approx \frac{k_1^{ZX} k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} [X] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.53)$$

$$J_C^{ZB} \approx \frac{k_2^{ZX} k_3^{ZX} k_1^{ZB} k_2^{ZB} k_3^{ZB} [B] [D_Z]_{tot}}{k_2^{ZX} k_3^{ZX} k_2^{ZB} k_3^{ZB} + k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX}) [X] + k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB}) [B]} \quad (5.54)$$

It is important to observe that Equations (5.44) to (5.54) are only dependent on the concentrations of X and B . Then, for a given $[X]$ and $[B]$ the partial distribution of the DNA sites and cyclic fluxes are fully determined.

Let us define J_T^Z as the total transcription rate of protein Z in $[\text{nM}\cdot\text{s}^{-1}]$, which is the sum of the transcription rate of Z due to the activity of X , J_T^{ZX} , and the basal activity of B , J_T^{ZB} . As protein Z is being transcribed in the third reaction of the proposed mechanism, then $J_T^{ZX} = J_3^{ZX}$ and $J_T^{ZB} = J_3^{ZB}$. However, it has been shown that at steady state $J_C^{ZX} = J_3^{ZX}$ and $J_C^{ZB} = J_3^{ZB}$. Thus:

$$J_T^Z = J_T^{ZX} + J_T^{ZB} = J_3^{ZX} + J_3^{ZB} = J_C^{ZX} + J_C^{ZB} \quad (5.55)$$

For $I = X, B$, we define :

$$\beta^{ZI} = \frac{k_2^{ZI} k_3^{ZI} [D_Z]_{tot}}{(k_2^{ZI} + k_3^{ZI})}, [\text{nM} \cdot \text{s}^{-1}] \quad (5.56)$$

and

$$\kappa^{ZI} = \frac{k_2^{ZI} k_3^{ZI}}{k_1^{ZI} (k_2^{ZI} + k_3^{ZI})}, [\text{nM}] \quad (5.57)$$

Then, the cyclic fluxes of ZX and ZB given by Equations (5.53) and (5.54) can be written as follows:

$$J_C^{ZX} \approx \frac{\frac{k_2^{ZX} k_3^{ZX} [D_Z]_{tot} [X]}{(k_2^{ZX} + k_3^{ZX})}}{\frac{k_2^{ZX} k_3^{ZX}}{k_1^{ZX} (k_2^{ZX} + k_3^{ZX})} + [X] + \frac{k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB})}{k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX})} [B]} = \frac{\beta^{ZX} [X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}} [B]} \quad (5.58)$$

$$J_C^{ZB} \approx \frac{\frac{k_2^{ZB} k_3^{ZB} [D_Z]_{tot} [B]}{(k_2^{ZB} + k_3^{ZB})}}{\frac{k_2^{ZB} k_3^{ZB}}{k_1^{ZB} (k_2^{ZB} + k_3^{ZB})} + [B] + \frac{k_1^{ZX} k_2^{ZB} k_3^{ZB} (k_2^{ZX} + k_3^{ZX})}{k_1^{ZB} k_2^{ZX} k_3^{ZX} (k_2^{ZB} + k_3^{ZB})} [X]} = \frac{\beta^{ZB} [B]}{\kappa^{ZB} + [B] + \frac{\kappa^{ZB}}{\kappa^{ZX}} [X]} \quad (5.59)$$

In principle, J_T^Z is a function of both $[X]$ and $[B]$. However, if the concentration of the basal activator does not change in the system, then the transcription rate of Z is only a function of its main activator X. The basal and maximum transcription rates, $J_{T,Basal}^Z$ and $J_{T,Max}^Z$, are defined as the transcription rates when $[X] \rightarrow 0$ and $[X] \rightarrow \infty$, respectively, and their values are given by Equations (5.60) and (5.61).

$$J_{T,Basal}^Z = \frac{\beta^{ZB} [B]}{\kappa^{ZB} + [B]} \quad (5.60)$$

$$J_{T,Max}^Z = \beta^{ZX} \quad (5.61)$$

Notice that $J_{T,Basal}^Z$ is constant for a specified $[B]$. In addition, from Equation (5.56), β^{ZX} can be conceptually defined as the *maximal transcription rate* $J_{T,Max}^Z$ of protein Z due to the activity of its transcription factor X. This parameter includes a cooperativity term, ϖ^{ZX} which deals with various input logics. In principle, $\beta^{ZX} = \varpi^{ZX} \beta^Z$ where β^Z is a common maximal transcription

its transcription factor X . This parameter includes a cooperativity term, ϖ^{ZX} which deals with various input logics. In principle, $\beta^{ZX} = \varpi^{ZX} \beta^Z$ where β^Z is a common maximal transcription rate for all the cycles involved in the transcription of Z . Essentially, different β^{ZI} are obtained by incorporation of different cooperativity terms in the cycles, represented by ϖ^{ZI} . This effect was previously described by (Buchler et al., 2003) and in general, ϖ^{ZI} is low for *OR* input logic and high for *AND* input logics. Moreover, the concentration of X required to reach half of $(J_{T,Max}^Z - J_{T,Basal}^Z)$ known as the *activation coefficient* K^X is determined by the following relationship between κ^{ZX} and κ^{ZB} :

$$K^X = \kappa^{ZX} \left(\frac{[B] + \kappa^{ZB}}{\kappa^{ZB}} \right) \quad (5.62)$$

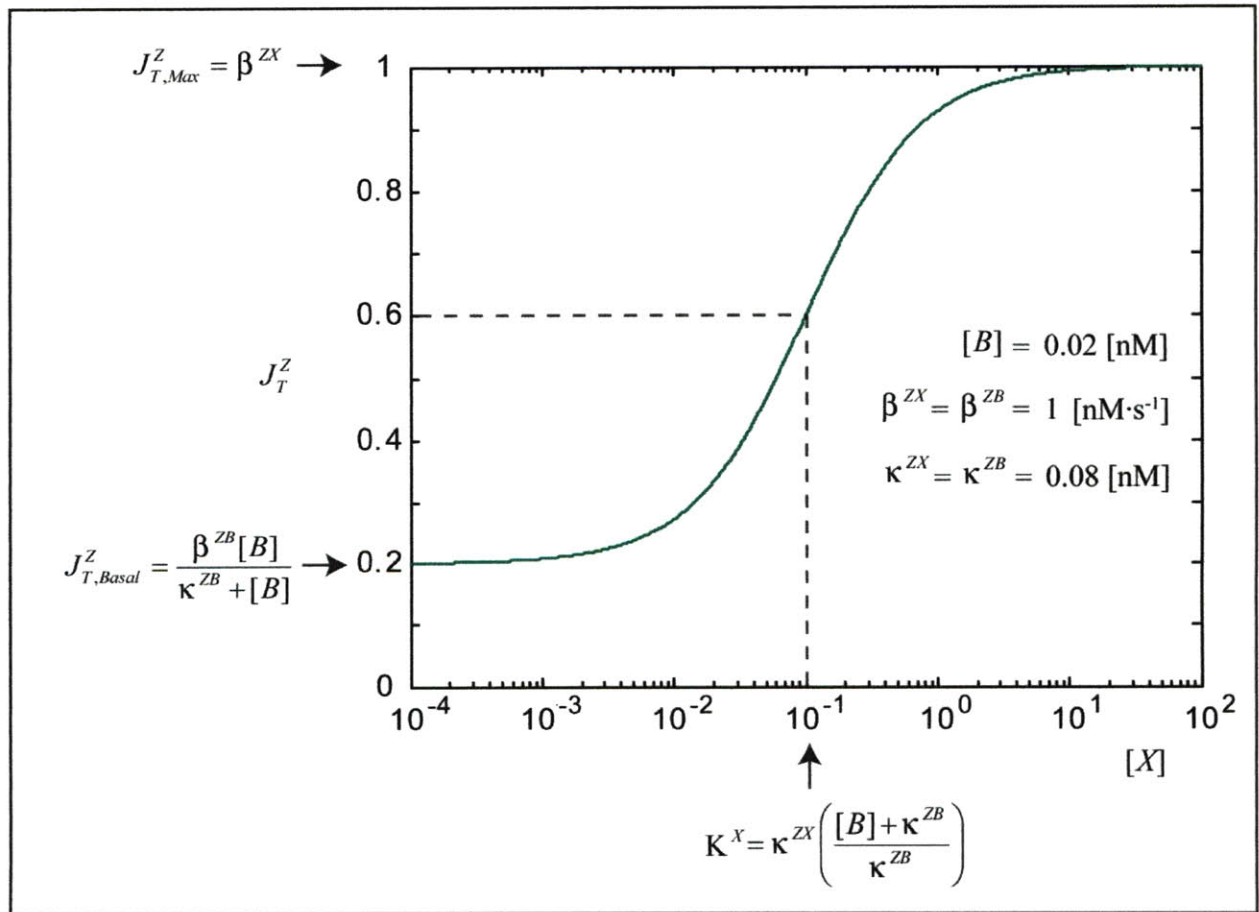


Figure 5.3: Transcription rate of Z as a function of its activator X . Curve follows a first order Hill's function, with maximum transcription rate of 1 [nM·s⁻¹] and half concentration of 0.1 [nM].

Figure 5.3 shows the approximate solution for very irreversible reactions far away from thermodynamic equilibrium. Notice that the transcription rate of Z follows a first order Hill's function with basal transcription activity:

$$J_T^Z = J_{T,Basal}^Z + \frac{(J_{T,Max}^Z - J_{T,Basal}^Z)[X]}{K^X + [X]} \quad (5.63)$$

which is same as the first order Hill's equation.

Since k_3^{ZI} contains the concentration of amino acids required to form the protein Z , the following assumption can be made: $k_3^{ZI} \gg k_2^{ZI}$. Then, for $I = Z, B$:

$$\beta^{ZI} = \frac{k_2^{ZI} k_3^{ZI}}{k_2^{ZI} + k_3^{ZI}} [D_Z]_{tot} \approx k_2^{ZI} [D_Z]_{tot} \quad (5.64)$$

$$\kappa^{ZI} = \frac{k_2^{ZI} k_3^{ZI}}{k_1^{ZI} (k_2^{ZI} + k_3^{ZI})} \approx \frac{k_2^{ZI}}{k_1^{ZI}} \quad (5.65)$$

Because the cyclic fluxes are only a function of X , the close expression for $[Z]$, $[X_Z]$ and $[B_Z]$ can be obtained by solving Equations (5.16) to (5.18):

$$[Z] = \frac{1}{\alpha_Z} \left[(J_{T,Basal}^Z + J_Z) + \frac{(J_{T,Max}^Z - J_{T,Basal}^Z)[X]}{K^X + [X]} \right] \quad (5.66)$$

$$[X_Z] = \frac{1}{\alpha_{X_Z}} \left[J_{X_Z} + \frac{\beta^{ZX}[X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}}[B]} \right] \quad (5.67)$$

$$[B_Z] = \frac{1}{\alpha_{B_Z}} \left[J_{B_Z} + \frac{\beta^{ZB}[B]}{\kappa^{ZB} + [B] + \frac{\kappa^{ZB}}{\kappa^{ZX}}[X]} \right] \quad (5.68)$$

Finally, for a given $[X]$, the rate of consumption of X (J_X) is obtained by solving Equation (5.14):

$$J_X = \frac{\beta^{ZX}[X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}}[B]} + \alpha_X [X] \quad (5.69)$$

When X transcribes Z in absence of the basal activity ($[B] = 0$, $J_{T,Basal}^Z = 0$), then

$$K^X = \kappa^{ZX} \text{ and}$$

$$J_T^Z = \frac{\beta^{ZX} [X]}{\kappa^{ZX} + [X]} \quad (5.70)$$

5.3.5 Energy of Activation

Because these reactions are very irreversible and thus very far apart from chemical equilibrium, the energy dissipated is high. As the reaction progresses, the rate at which energy is being produced is expressed by the product between the flux and chemical potential of such reaction. As no work is being performed by the system, this energy has to equal the heat dissipated by each reaction in order to satisfy the First Law of the Thermodynamics. Then, for this system, the total heat dissipation rate HDR^Z [$J \cdot s^{-1}$] is given by the addition of the heat dissipated by each cycle:

$$HDR^Z = HDR^{ZX} + HDR^{ZB} = - \left(\sum_{i=1}^3 J_i^{ZX} \Delta\mu_i^{ZX} + \sum_{i=1}^3 J_i^{ZB} \Delta\mu_i^{ZB} \right) = - \sum_{I=X,B} \sum_{i=1}^3 J_i^{ZI} \Delta\mu_i^{ZI} \quad (5.71)$$

However, as expressed by Equations (5.25) and (5.26), at steady state:

$$HDR^{ZI} = -(J_1^{ZI} \Delta\mu_1^{ZI} + J_2^{ZI} \Delta\mu_2^{ZI} + J_3^{ZI} \Delta\mu_3^{ZI}) = -J_C^{ZI} (\Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI}) = -J_C^{ZI} \Delta\mu_C^{ZI} \quad (5.72)$$

where

$$\Delta\mu_C^{ZI} = \Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI} = \mathfrak{R}T \ln \left(\frac{[I_Z][Z]}{K_1^{ZI} K_2^{ZI} K_3^{ZI} [I]} \right) \quad (5.73)$$

is defined as the chemical potential of the cycle for $I = X, B$.

Equations (5.71) to (5.73) are general expressions which can be simplified by using the approximations presented in Section 5.3.4. When the analytical solutions for very irreversible reactions are used and there is no external uptake or intake of Z , X_Z and B_Z , then the following equations are valid:

$$\Delta\mu_C^{ZI} = \mathfrak{R}T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} (\beta^{ZX} \kappa^{ZB} [X] + \beta^{ZB} \kappa^{ZX} [B])}{\alpha_Z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} [D_Z]_{tot}^2 \cdot \gamma^{ZI} \right] \quad (5.74)$$

$$\text{where } \gamma^{ZI} = \frac{k_{-1}^{ZI} k_{-2}^{ZI}}{\beta^{ZI} K_3^{ZI} \alpha_{I_Z}}, [\text{nM}^{-3}] \quad (5.75)$$

For $I, J = X, B, I \neq J$:

$$HDR^{ZI} = -\frac{\kappa^{ZJ} \beta^{ZI} [I]}{\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B]} \Re T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} (\beta^{ZX} \kappa^{ZB} [X] + \beta^{ZB} \kappa^{ZX} [B])}{\alpha_z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} [D_z]_{tot}^2 \cdot \gamma^{ZI} \right] \quad (5.76)$$

If $\gamma^{ZX} = \gamma^{ZB} = \gamma^Z$, then the total heat dissipation rate of the transcription process can be simplified as follows:

$$HDR^Z = -\frac{\kappa^{ZB} \beta^{ZX} [X] + \kappa^{ZX} \beta^{ZB} [B]}{\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B]} \Re T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} (\beta^{ZX} \kappa^{ZB} [X] + \beta^{ZB} \kappa^{ZX} [B])}{\alpha_z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} [D_z]_{tot}^2 \cdot \gamma^Z \right] \quad (5.77)$$

Here, we introduce the concept of specific dissipation energy (*SDE*), defined by us as the ratio between the heat dissipation rate to the input mass flux required to keep the system under NESS conditions:

$$SDE^Z = \frac{HDR^Z}{J_X} \quad (5.78)$$

Far from equilibrium, this magnitude can be approximated as:

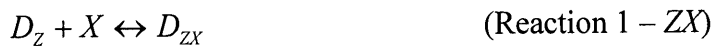
$$SDE^Z = -\frac{\kappa^{ZB} \beta^{ZX} [X] + \kappa^{ZX} \beta^{ZB} [B]}{[X](\kappa^{ZB} \beta^{ZX} + \alpha_x \kappa^{ZX} \kappa^{ZB} + \alpha_x \kappa^{ZB} [X] + \alpha_x \kappa^{ZX} [B])} \times \Re T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} (\beta^{ZX} \kappa^{ZB} [X] + \beta^{ZB} \kappa^{ZX} [B])}{\alpha_z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} [D_z]_{tot}^2 \cdot \gamma^Z \right] \quad (5.79)$$

5.4 REPRESSION OF Z BY X

5.4.1 Mechanism

The mechanism by which the protein Z is repressed by the transcription factor X can be divided in a three step process, similar to the activation case, as shown in Figure 5.4

Step 1: As in the activation case, the transcription factor X binds a free DNA site of the promoter region of protein Z (D_z) to form an occupied DNA site (D_{ZX}):



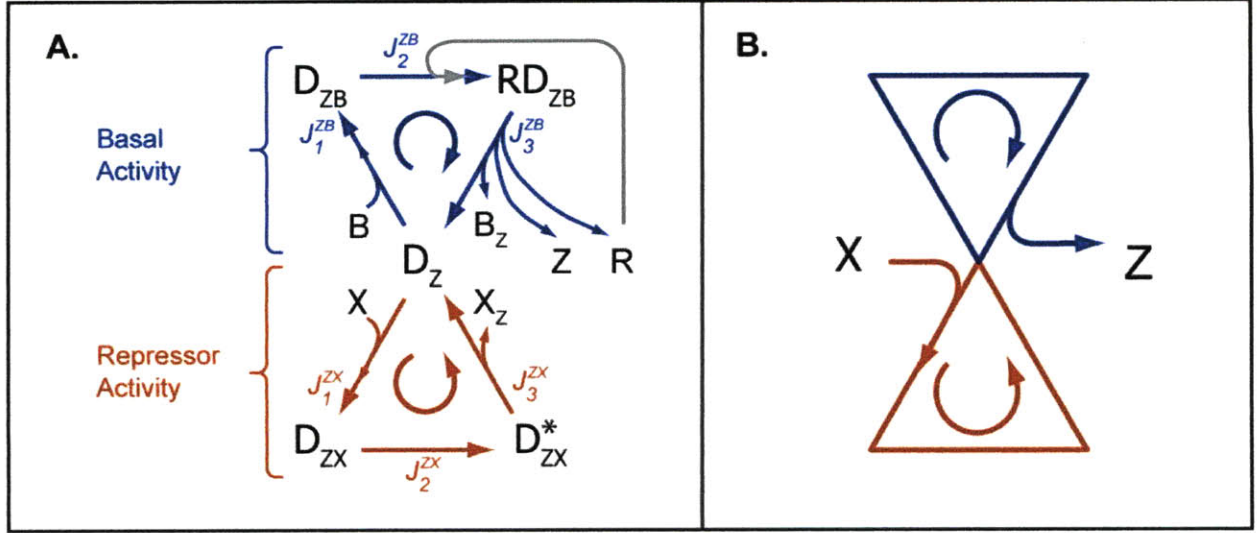


Figure 5.4: Representation of the repression pathway of protein Z by transcription factor X . **A)** Detailed mechanism showing both basal and repressor activity. **B)** Simplified representation of the cyclic concept.

If k_1 [$s^{-1} \cdot nM^{-1}$] and k_{-1} [s^{-1}] are the forward and backward kinetic constants, respectively, then the net flux for this reaction is given by

$$J_1^{ZX} = k_1^{ZX} [D_Z][X] - k_{-1}^{ZX} [D_{ZX}] \quad (5.80)$$

and the reaction chemical potential is

$$\Delta\mu_1^{ZX} = \Re T \ln \left(\frac{k_{-1}^{ZX} [D_{ZX}]}{k_1^{ZX} [D_Z][X]} \right) = \Re T \ln \left(\frac{[D_{ZX}]}{K_1^{ZX} [D_Z][X]} \right) \quad (5.81)$$

where \Re [$J \cdot K^{-1} \cdot mol^{-1}$] is the gas constant, T [K] is the absolute temperature and $K_1^{ZX} = k_{-1}^{ZX} / k_1^{ZX}$ [nM^{-1}] is the association equilibrium constant.

Step 2: Since X is a repressor, RNA polymerase (R) can not bind to the occupied DNA site (D_{ZX}). In turn, the X -bound site changes its configuration into another energetic state (D_{ZX}^*)



Let k_2^{ZX} [s^{-1}] and k_{-2}^{ZX} [s^{-1}] the forward and backward kinetic constants for this step, respectively. The net flux for this reaction can be written as follows:

$$J_2^{ZX} = k_2^{ZX} [D_{ZX}] - k_{-2}^{ZX} [D_{ZX}^*] \quad (5.82)$$

If $K_2^{ZX} = k_2^{ZX} / k_{-2}^{ZX}$ is the association equilibrium constant for this reaction, then the reaction chemical potential is given by

$$\Delta\mu_2^{ZX} = \Re T \ln \left(\frac{k_{-2}^{ZX} [D_{ZX}^*]}{k_2^{ZX} [D_{ZX}]} \right) = \Re T \ln \left(\frac{[D_{ZX}^*]}{K_2^{ZX} [D_{ZX}]} \right) \quad (5.83)$$

Step 3: In this last step, the activated X -bound site releases the free DNA site and the co-repressor necessary for this process (X_Z). Because RNA polymerase is not bound, there is no recruitment of amino acids (AA_Z) and transcription of Z does not proceed.



If k_3^{ZX} [s^{-1}] and k_{-3}^{ZX} [$s^{-1} \cdot nM^{-1}$] are the forward and backward kinetic constants, respectively, and $K_3^{ZX} = k_3^{ZX} / k_{-3}^{ZX}$ [nM^{-1}] is the association equilibrium constant for this reaction, then the reaction flux and reaction chemical potential are given by

$$J_3^{ZX} = k_3^{ZX} [D_{ZX}^*] - k_{-3}^{ZX} [D_Z][X_Z] \quad (5.84)$$

$$\Delta\mu_3^{ZX} = \Re T \ln \left(\frac{[D_Z][X_Z]}{K_3^{ZX} [D_{ZX}^*]} \right) \quad (5.85)$$

The described steps 1, 2 and 3 are represented as a red triangular cycle in Figure 5.4. The basal activity (blue cycle in Figure 5.4), in which a basal transcription factor B binds the free promoter region of Z , initiating the cycle and further transcription of Z , is described in Section 5.3.

5.4.2 Mass Balance Equations

Let J_i and α_i the external intake flux (defined positive is the flux enters the system and negative if exits) in [$nM \cdot s^{-1}$] and the degradation rate in [s^{-1}] of the species i , respectively. Then, the mass balance is described by the following differential equations:

$$\frac{d[X]}{dt} = J_X - J_1^{ZX} - \alpha_X[X] \quad (5.86)$$

$$\frac{d[B]}{dt} = J_B - J_1^{ZB} - \alpha_B[B] \quad (5.87)$$

$$\frac{d[Z]}{dt} = J_3^{ZB} + J_Z - \alpha_Z[Z] \quad (5.88)$$

$$\frac{d[X_Z]}{dt} = J_3^{ZX} + J_{X_Z} - \alpha_{X_Z}[X_Z] \quad (5.89)$$

$$\frac{d[B_Z]}{dt} = J_3^{ZB} + J_{B_Z} - \alpha_{B_Z}[B_Z] \quad (5.90)$$

$$\frac{d[D_Z]}{dt} = J_3^{ZX} + J_3^{ZB} - J_1^{ZX} - J_1^{ZB} \quad (5.91)$$

$$\frac{d[D_{ZX}]}{dt} = J_1^{ZX} - J_2^{ZX} \quad (5.92)$$

$$\frac{d[D_{ZB}]}{dt} = J_1^{ZB} - J_2^{ZB} \quad (5.93)$$

$$\frac{d[D_{ZX}^*]}{dt} = J_2^{ZX} - J_3^{ZX} \quad (5.94)$$

$$\frac{d[RD_{ZB}]}{dt} = J_2^{ZB} - J_3^{ZB} \quad (5.95)$$

Notice that Equation (5.88) differs from Equation (5.16) by eliminating the J_3^{ZX} term, since the transcription of Z is only due to the basal activity. Also, Equation (5.94) is analogous to Equation (5.23) but conceptually D_{ZX}^* is not the same as RD_{ZX} since RNA polymerase is not longer recruited by the transcription factor X . All the other mass balance equations are identical than in the activation case.

In addition to Equations (5.86) to (5.95), the total concentration of DNA sites must be constant at any time:

$$[D_Z]_{tot} = [D_Z] + [D_{ZX}] + [D_{ZB}] + [D_{ZX}^*] + [RD_{ZB}] \quad (5.96)$$

5.4.3 General Steady State Solution

By solving Equations (5.91) to (5.95) at steady state, we have that

$$J_1^{ZX} = J_2^{ZX} = J_3^{ZX} = J_C^{ZX} \quad (5.97)$$

$$J_1^{ZB} = J_2^{ZB} = J_3^{ZB} = J_C^{ZB} \quad (5.98)$$

being J_C^{ZX} and J_C^{ZB} [$\text{nM}\cdot\text{s}^{-1}$] the cyclic fluxes for cycle ZX (shown in red in Figure 5.4) and for cycle ZB (shown in blue in Figure 5.4), respectively. By solving Equations (5.80), (5.81), (5.83) (and their analogous basal activity equations explained in Section 5.3), (5.96) to (5.98), a

solution for the mass balance of the DNA sites can be found. For $I = X, B$, and number of transcription factors $N_{TF} = 2$ (basal and activator), the following auxiliary variables can be defined:

$$a^{ZI} = k_{-1}^{ZI} k_{-2}^{ZI} + k_{-1}^{ZI} k_3^{ZI} + k_2^{ZI} k_3^{ZI} \quad (5.99)$$

$$b^{ZI} = k_1^{ZI} (k_{-2}^{ZI} + k_3^{ZI}) [I] + k_{-2}^{ZI} k_{-3}^{ZI} [P_Z] [I_Z] \quad (5.100)$$

$$c^{ZI} = k_1^{ZI} k_2^{ZI} [I] + k_{-3}^{ZI} (k_{-1}^{ZI} + k_2^{ZI}) [P_Z] [I_Z] \quad (5.101)$$

$$d^{ZI} = a^{ZI} + b^{ZI} + c^{ZI} \quad (5.102)$$

$$d_Z = \left(\sum_{I=X,B} \frac{d^{ZI} \prod_{I=X,B} a^{ZI}}{a^{ZI}} \right) - (N_{TF} - 1) \prod_{I=X,B} a^{ZI} \quad (5.103)$$

where $[P_Z] = \begin{cases} [Z], & \text{if } I = B \text{ (or } I \text{ is an activator)} \\ 1, & \text{if } I = X \text{ (or } I \text{ is a repressor)} \end{cases}$

Then, the fractional site concentrations and the cyclic flux are given by:

$$\frac{[D_Z]}{[D_Z]_{tot}} = \frac{a^{ZX} a^{ZB}}{d_Z} = \frac{\prod_{I=X,B} a^{ZI}}{d_Z} \quad (5.104)$$

$$\frac{[D_{ZI}]}{[D_Z]_{tot}} = \frac{b^{ZI} \prod_{J=X,B} a^{ZJ}}{a^{ZI} d_Z} \quad (5.105)$$

$$\frac{[D_{ZX}^*]}{[D_Z]_{tot}} = \frac{c^{ZX} a^{ZB}}{d_Z} \quad (5.106)$$

$$\frac{[RD_{ZB}]}{[D_Z]_{tot}} = \frac{c^{ZB} a^{ZX}}{d_Z} \quad (5.107)$$

$$J_C^{ZI} = \frac{\prod_{J=X,B} a^{ZJ}}{a^{ZI} d_Z} [D_Z]_{tot} (k_2^{ZI} b^{ZI} - k_{-2}^{ZI} c^{ZI}) \quad (5.108)$$

For a given concentration of X and B , the steady state concentrations of Z , X_Z and B_Z can now be found by solving Equations (5.88) to (5.90), but analytical expressions lack of simplicity. In order to avoid this problem, an approximate solution is presented in the following section.

5.4.4 Approximate Steady State Solution

The analysis presented in section can be used to find the approximated solutions. In fact, Equations (5.41) to (5.44) are the same with the exception of the term RD_{ZX} which has to be replaced by D_{ZX}^* in the repression scenario.

When X is a repressor, the total transcription rate of Z, J_T^Z [nM·s⁻¹], equals the basal cyclic flux at steady state:

$$J_T^Z = J_T^{ZB} = J_3^{ZB} = J_C^{ZB} \quad (5.109)$$

By defining β^{ZI} and κ^{ZI} (for $I = X, B$) as in Equations (5.56) and (5.57), Equations (5.58) and (5.59) are still valid:

$$J_C^{ZX} \approx \frac{\beta^{ZX}[X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}}[B]} \quad (5.110)$$

$$J_C^{ZB} \approx \frac{\beta^{ZB}[B]}{\kappa^{ZB} + [B] + \frac{\kappa^{ZB}}{\kappa^{ZX}}[X]} \quad (5.111)$$

As $[X] \rightarrow 0$, the basal transcription rate $J_{T,Basal}^Z$ follows the same relationship as in the activation case (the basal transcription rate is independent of the activator or repressor nature of X):

$$J_{T,Basal}^Z = \frac{\beta^{ZB}[B]}{\kappa^{ZB} + [B]} \quad (5.112)$$

However, as expected, the basal transcription rate corresponds to the maximum transcription rate. As $[X] \rightarrow \infty$, the transcription rate of Z decreases and tends to zero:

$$J_{T,Min}^Z = 0 \quad (5.113)$$

Moreover, the concentration of X required to reach half of $(J_{T,Basal}^Z - J_{T,Min}^Z) = J_{T,Basal}^Z$, K^X , is determined by the same relationship between κ^{ZX} and κ^{ZB} shown in Equation (5.62):

$$K^X = \kappa^{ZX} \left(\frac{[B] + \kappa^{ZB}}{\kappa^{ZB}} \right) \quad (5.114)$$

Figure 5.5 shows the approximate solution for very irreversible reactions far away from thermodynamic equilibrium. Notice that the transcription rate of Z follows a first order Hill's function with basal transcription activity:

$$J_T^Z = \frac{J_{T,Basal}^Z}{1 + \frac{[X]}{K^X}} \quad (5.115)$$

Equations (5.64) to (5.69) are valid mass balance solutions but (5.66) must be changed as a response to the new mass balance of Z:

$$[Z] = \frac{1}{\alpha_Z} \left[J_Z + \frac{K^X J_{T,Basal}^Z}{K^X + [X]} \right] \quad (5.116)$$

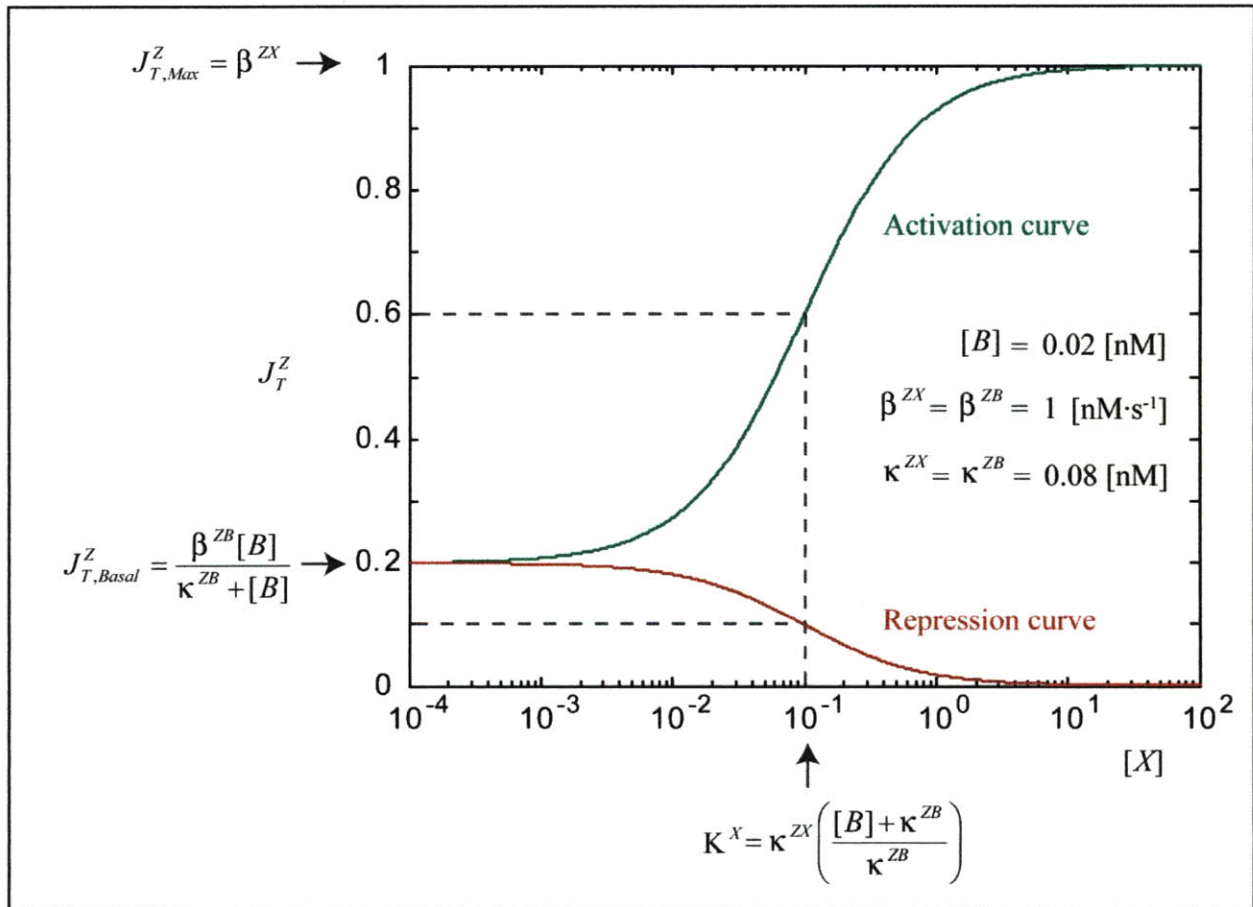


Figure 5.5: Transcription rate of Z as a function of its repression X. Curve follows a first order Hill's function, with maximum transcription rate given by the basal activity of 0.2 $[nM \cdot s^{-1}]$ and half concentration of 0.1 $[nM]$.

5.4.5 Energy of Repression

Equations (5.71) and (5.72) express the general concept of the heat dissipation rate hdr^Z [$J \cdot s^{-1}$] associated with any transcription process and thus they are valid for the repression scenario. However, the cyclic chemical potential equation given by (5.73) must be changed since the overall reactions are different:

$$\Delta\mu_C^{ZI} = \Delta\mu_1^{ZI} + \Delta\mu_2^{ZI} + \Delta\mu_3^{ZI} = \mathfrak{R}T \ln \left(\frac{[I_Z][P_Z]}{K_1^{ZI} K_2^{ZI} K_3^{ZI} [I]} \right) \quad (5.117)$$

$$\text{where } [P_Z] = \begin{cases} [Z], & \text{if } I = B \text{ (or } I \text{ is an activator)} \\ 1, & \text{if } I = X \text{ (or } I \text{ is a repressor)} \end{cases}$$

When the analytical solutions for very irreversible reactions are used and there is no external uptake or intake of Z , X_Z and B_Z , then the following equations are valid:

$$\Delta\mu_C^{ZX} = \mathfrak{R}T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} [D_Z]_{tot}^2}{(\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])} \cdot \gamma^{ZX} \right] \quad (5.118)$$

$$\Delta\mu_C^{ZB} = \mathfrak{R}T \ln \left[\frac{(\kappa^{ZX})^2 \kappa^{ZB} \beta^{ZB} [B] [D_Z]_{tot}^2}{\alpha_Z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} \cdot \gamma^{ZB} \right] \quad (5.119)$$

where γ^{ZI} is defined by Equation (5.75).

The heat dissipation rate of the cycles can be written as follows:

$$HDR^{ZX} = - \frac{\kappa^{ZB} \beta^{ZX} [X]}{\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B]} \mathfrak{R}T \ln \left[\frac{\kappa^{ZX} \kappa^{ZB} [D_Z]_{tot}^2}{(\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])} \cdot \gamma^{ZX} \right] \quad (5.120)$$

$$HDR^{ZB} = - \frac{\kappa^{ZX} \beta^{ZB} [B]}{\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B]} \mathfrak{R}T \ln \left[\frac{(\kappa^{ZX})^2 \kappa^{ZB} \beta^{ZB} [B] [D_Z]_{tot}^2}{\alpha_Z (\kappa^{ZX} \kappa^{ZB} + \kappa^{ZB} [X] + \kappa^{ZX} [B])^2} \cdot \gamma^{ZB} \right] \quad (5.121)$$

Based on Equations (5.120) and (5.121), an analytical expression for the specific dissipation energy defined by Equation (5.78) can be obtained.

5.5 TRANSCRIPTIONAL REGULATION OF Z BY TWO TRANSCRIPTION FACTORS X AND Y

In this section, we will analyze the transcription (activation and repression) of the protein Z by two transcription factors X and Y .

5.5.1 Mechanism

The basic mechanism by which two transcription factors activate or repress certain protein transcription is the same three step process previously described for activation (Section 5.3) and for repression (Section 5.4). Both X and Y (and basal transcription factor B) compete for a free DNA site in the promoter region of Z no matter whether X (or Y) is an activator or a repressor. However, in addition to the individual binding to D_Z , there is also a joint interaction between X and Y , so in principle the complex XY may be considered as an extra transcription factor whose concentration is given by the product between the concentration of X and Y . In other words, $[XY] = [X][Y]$. The activator or repressor nature of this complex depends on two factors:

1. The independent activation – repression nature of X and Y .
2. The logic used to transcribe Z which can be AND or OR.

In order to explain this last concept, let us assume that both X and Y are activators of Z . In principle, protein Z should be transcribed at least from the individual binding of X and Y . However, this is not true. If both X and Y are required to transcribe Z (AND logic), then when X (or Y) bind by its own, transcription cannot proceed. Then, even though X (or Y) is an activator, it will bind to the site but without releasing Z , acting as a repressor. Therefore, transcription of Z can only proceed from the XY cycle (and from the basal cycle ZB which by definition is always transcribing Z), where both are present. This XY complex will bind, recruit RNA polymerase and finally transcribe (and translate) Z . On the other hand, if only one of the two transcription factors is required (OR logic) for further formation of Z , then the joint interaction is not longer necessary and the individual binding as well as the complex XY binding, will lead to formation of Z , as shown in Figure 5.6, where a simple representation of the four cycles is presented in a general way. Notice that the cycles and entering and leaving fluxes for ZX , ZY and ZXY are black. The activator and repressor behaviors are presented as green and red arrows, respectively, according to the logic and independent nature of X and Y as shown in the right-side tables. When the cycle ZI behaves as a repressor (for $I = X, Y, XY$), the transcription flux J_3^{ZI} is not present (recall Figure 5.4), and thus the corresponding arrow is not shown in Figure 5.6.

5.5.2 Mass Balance Equations

The mass balance equations can be easily derived from Sections 5.3 and 5.4, and these will depend on the transcription logic showed in Figure 5.6 and on the individual behavior of X and Y . However, no matter the species mass balance, the total DNA sites must be constant at steady state. Based on the equations presented in Sections 5.3 and 5.4, the general solution for the fractional site occupation and cyclic sites is given below for $I = B, X, Y$ and XY , and $N_{TF} = 4$:

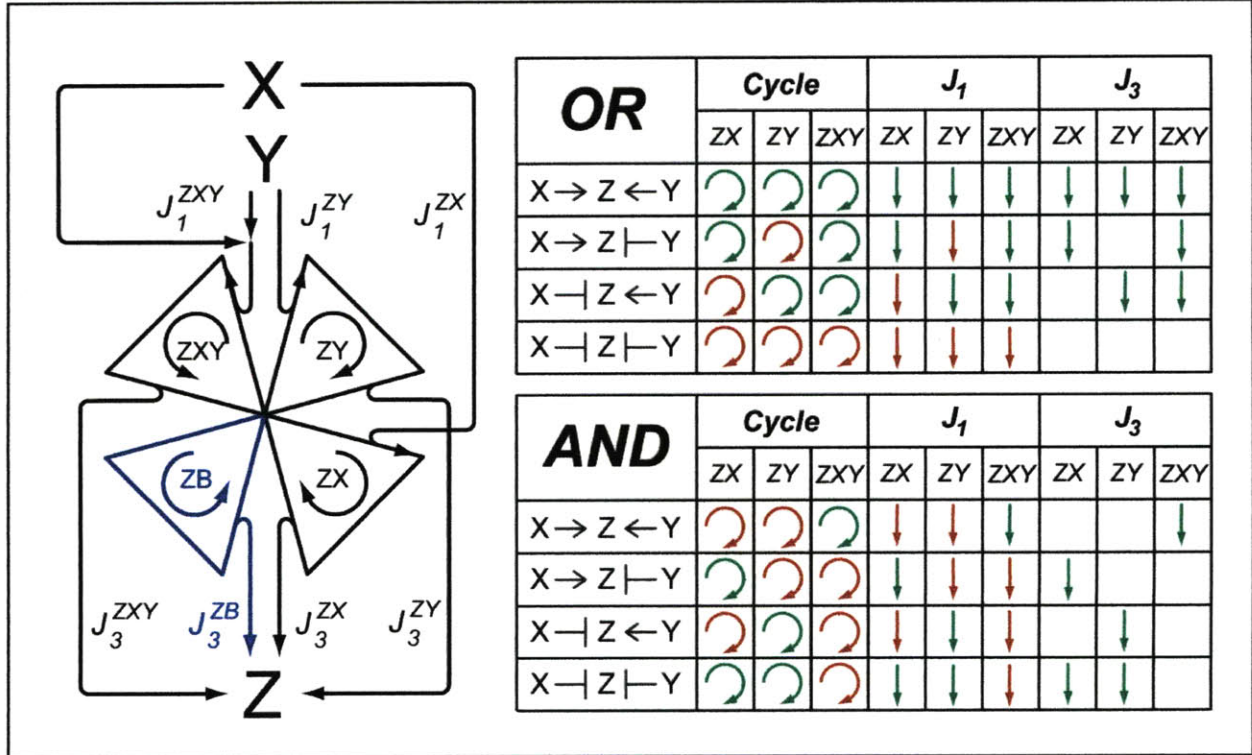


Figure 5.6: Schematic representation of the three-step transcription process of Z by X and Y . Black arrows become green (activator behavior) or red (repressor behavior) based on the individual activation or repressor nature of X and Y , and the transcription logic AND or OR. When a transcription factor acts as a repressor, the corresponding output flux of Z , J_3 , is no present and not shown in the right-side tables.

$$a^{ZI} = k_{-1}^{ZI} k_{-2}^{ZI} + k_{-1}^{ZI} k_3^{ZI} + k_2^{ZI} k_3^{ZI} \quad (5.122)$$

$$b^{ZI} = k_1^{ZI} (k_{-2}^{ZI} + k_3^{ZI}) [I] + k_{-2}^{ZI} k_{-3}^{ZI} [P_Z] [I_Z] \quad (5.123)$$

$$c^{ZI} = k_1^{ZI} k_2^{ZI} [I] + k_{-3}^{ZI} (k_{-1}^{ZI} + k_2^{ZI}) [P_Z] [I_Z] \quad (5.124)$$

$$d^{ZI} = a^{ZI} + b^{ZI} + c^{ZI} \quad (5.125)$$

where $[P_Z] = \begin{cases} [Z], & \text{if } I \text{ is an activator} \\ 1, & \text{if } I \text{ is a repressor} \end{cases}$

$$d_z = \left(\sum_{I=X,B} \frac{d^{ZI} \prod_{I=X,B} a^{ZI}}{a^{ZI}} \right) - (N_{TF} - 1) \prod_{I=X,B} a^{ZI} \quad (5.126)$$

$$\frac{[D_Z]}{[D_Z]_{tot}} = \frac{\prod_{I=B,X,Y,XY} a^{ZI}}{d_z} \quad (5.127)$$

$$\frac{[D_{ZI}]}{[D_Z]_{tot}} = \frac{b^{ZI} \prod_{J=B,X,Y,XY} a^{ZJ}}{a^{ZI} d_z} \quad (5.128)$$

$$\frac{[D_{ZI}^{Node 3}]}{[D_Z]_{tot}} = \frac{c^{ZI} \prod_{J=B,X,Y,XY} a^{ZJ}}{a^{ZI} d_z} \quad (5.129)$$

where $D_{ZI}^{Node 3} = \begin{cases} RD_{ZI}, & \text{if } I \text{ is an activator} \\ D_{ZI}^*, & \text{if } I \text{ is a repressor} \end{cases}$

$$J_C^{ZI} = \frac{\prod_{J=B,X,Y,XY} a^{ZJ}}{a^{ZI} d_z} [D_Z]_{tot} (k_2^{ZI} b^{ZI} - k_{-2}^{ZI} c^{ZI}) \quad (5.130)$$

If the irreversibility condition is satisfied, and the reaction fluxes and chemical potentials are very far apart from the thermodynamic equilibrium, the cyclic fluxes can be approximated as follows:

$$J_C^{ZI} = \frac{\beta^{ZI} [I]}{\kappa^{ZI} + [I] + \sum_{J=B,X,Y,XY} \frac{\kappa^{ZI}}{\kappa^{ZJ}} [J]}, \quad I \neq J \quad (5.131)$$

where β^{ZI} and κ^{ZI} are defined by Equations (5.56) and (5.57). If $k_3^{ZI} \gg k_2^{ZI}$, then Equations (5.64) and (5.65) may be used to define these parameters. Figure 5.7 presents the transcription rate surface for both logics OR and AND, using the approximated solution.

For Boolean input logic *OR*, $J_T^Z = J_C^{ZB} + J_C^{ZX} + J_C^{ZY} + J_C^{ZXY}$. If both basal activity and the transcription rate from the combined activity of *X* and *Y* are zero (i.e., β^{ZXY} is very low and κ^{ZXY} is very high), the expression for the total transcription rate of *Z* given by Equation (5.131) equals the relationship given by (Mangan and Alon, 2003):

$$J_T^Z = \frac{\beta^{ZX} \frac{[X]}{\kappa^{ZX}}}{1 + \frac{[X]}{\kappa^{ZX}} + \frac{[Y]}{\kappa^{ZY}}} + \frac{\beta^{ZY} \frac{[Y]}{\kappa^{ZY}}}{1 + \frac{[Y]}{\kappa^{ZY}} + \frac{[X]}{\kappa^{ZX}}} \quad (5.132)$$

In Boolean input logic *AND*, transcription rate of *Z* is given by the basal activity and the combined interaction of *X* and *Y*: $J_T^Z = J_C^{ZB} + J_C^{ZXY}$. If basal activity is zero and $\kappa^{ZXY} = \kappa^{ZX} \kappa^{ZY}$, then the expression for J_T^Z (Equation 5.131) also equals the relationship given by (Mangan and Alon, 2003):

$$J_T^Z = \frac{\beta^{ZX} \frac{[X]}{\kappa^{ZX}}}{1 + \frac{[X]}{\kappa^{ZX}}} \cdot \frac{\beta^{ZY} \frac{[Y]}{\kappa^{ZY}}}{1 + \frac{[Y]}{\kappa^{ZY}}} \quad (5.133)$$

It is important to mention that, independently of the activation-repression behavior of *X*, its input flux is always given by the same equation:

$$J_X = \frac{\beta^{ZX} [X]}{\kappa^{ZX} + [X] + \frac{\kappa^{ZX}}{\kappa^{ZB}} [B] + \frac{\kappa^{ZX}}{\kappa^{ZY}} [Y] + \frac{\kappa^{ZX}}{\kappa^{ZXY}} [X][Y]} + \frac{\beta^{ZXY} [X][Y]}{\kappa^{ZXY} + [X][Y] + \frac{\kappa^{ZXY}}{\kappa^{ZB}} [B] + \frac{\kappa^{ZXY}}{\kappa^{ZX}} [X] + \frac{\kappa^{ZXY}}{\kappa^{ZY}} [Y]} + \alpha_X [X] \quad (5.134)$$

5.5.3 Energy of Transcription

In general and at steady state, the total heat dissipation rate for this four-cycle system can be written as follows:

$$HDR^Z = - \sum_{I=B,X,Y,XY} J_C^{ZI} \Delta\mu_C^{ZI} \quad (5.135)$$

where

$$\Delta\mu_C^{ZI} = \Re T \ln \left(\frac{[I_Z][P_Z]}{K_1^{ZI} K_2^{ZI} K_3^{ZI} [I]} \right) \quad (5.136)$$

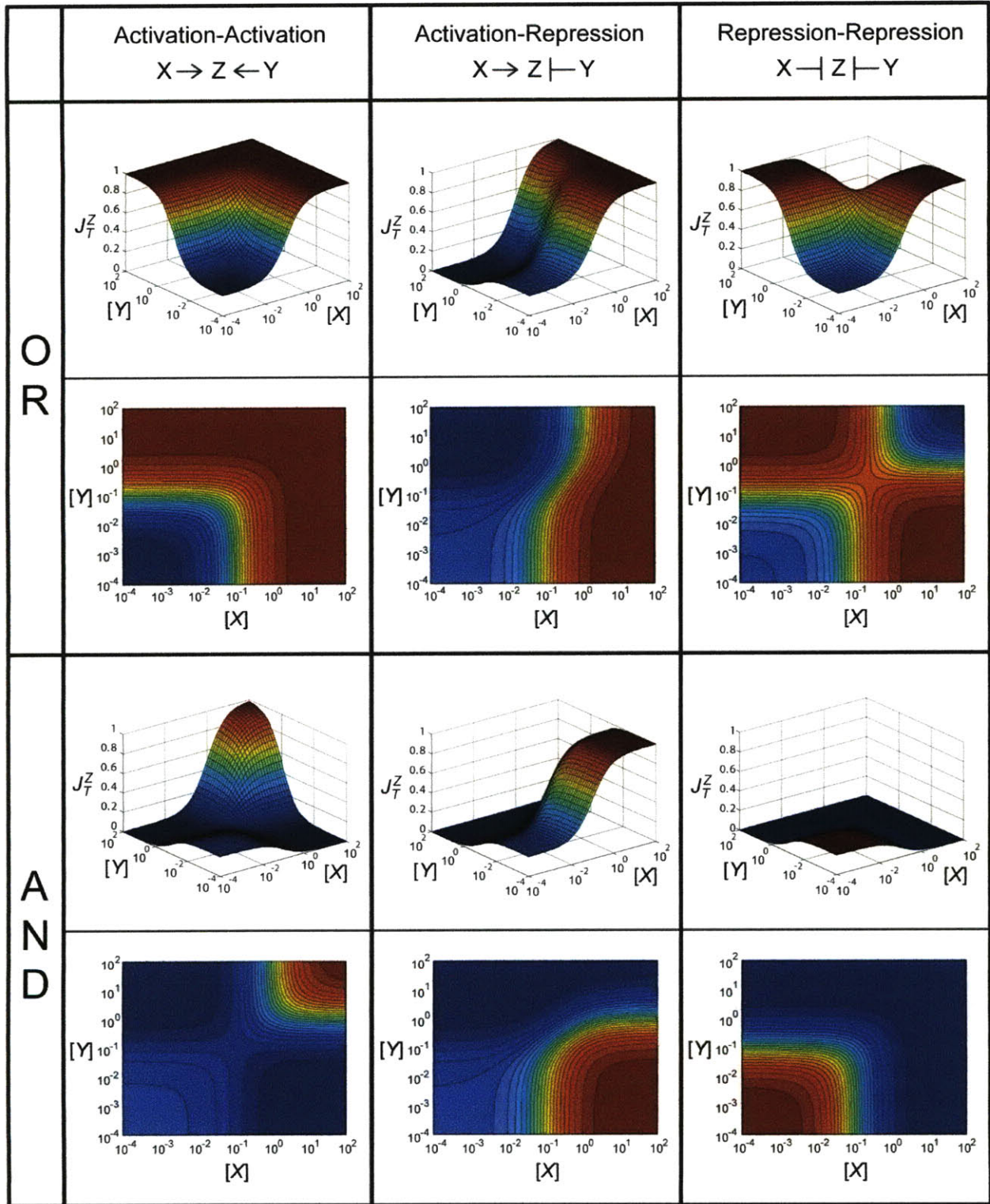


Figure 5.7: Transcription rate of Z when two transcription factors X and Y are present. The transcription surfaces and plane projections are shown for both OR and AND logics in three different cases: activator-activator, activator-repressor and repressor-repressor. $[B] = 0.02$ [nM], $\beta^{ZI} = 1$ [nM·s⁻¹] and $\kappa^{ZI} = 0.08$ [nM], for $I = B, X, Y,$ and XY .

Analytical solutions for HDR^Z get more complicated and logic dependent in this scenario, thus numerical solutions are required to calculate this value and the specific dissipation energy SDE^Z defined by Equation (5.78). For the latest meaning, Equations (5.131) to (5.136) must be used.

6 GLOBAL THERMODYNAMICS CONTROL THE EVOLUTION OF TRANSCRIPTIONAL REGULATORY NETWORKS

6.1 OVERVIEW

Transcriptional motifs are small regulatory interaction patterns that regulate biological functions in highly-interacting cellular networks. Recently, attempts have been made to explain the significance of transcriptional motifs through dynamic function. However, fundamental questions remain unanswered. Why are certain transcriptional motifs with similar dynamic function abundant while others occur rarely? What are the criteria for topological generalization of these motifs into complex networks? Here, we present a novel paradigm that combines nonequilibrium thermodynamics with multiobjective-optimality for network analysis. This yields evidence that dissipative energetics is the underlying criteria used during evolution for motif selection and that biological systems during transcription tend towards evolutionary selection of subgraphs which produces minimum specific heat dissipation, thereby explaining the abundance/rare occurrence of some motifs. Significantly, the proposed energetic hypothesis uncovers a mechanism for environmental selection of motifs, provides explanation for topological generalization of subgraphs into complex networks and enables identification of new functionalities for rarely occurring motifs. The presented insights may establish global thermodynamic analysis as backbone in designing and understanding complex networks systems, such as metabolic and protein interaction networks.

6.2 INTRODUCTION

Network motifs are the basic building blocks of complex networks and are the smallest overrepresented repeated subgraphs occurring commonly in both man-made large-scale networks (such as the world-wide web) and complex natural networks (such as cellular networks) (Alon, 2007; Balazsi et al., 2005; Barabasi, 2005; Milo et al., 2004; Milo et al., 2002; Oikonomou and Cluzel, 2006). Motifs in transcriptional regulatory networks (TRNs) have numerous functions that help maintain phenotypes. The three-node feed-forward loop (FFL) motifs: (1) are among the most abundant and conserved TRNs; (2) are the smallest repeated interacting unit between genes/operon and transcription factors, and maintain gene regulation; and (3) have dynamical

functions such as pulse generation, response delays, and noise filtering (Mangan and Alon, 2003; Mangan et al., 2006). TRN motifs of various types including FFL, BIFAN, and single input module (SIM) (Figure 6.1) have been found to occur in real networks in organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*.

There has been a continued struggle in understanding the common basis of occurrence of these motifs in network biology. Insights about the wiring of these network motifs could explain their evolutionary selection criteria, could uncover the mechanism behind TRNs evolution, and could decipher the basis behind the coordination of regulatory processes. Intense interest in explaining the selection of one motif over another and the frequency of occurrence of these network motifs in various organisms and TRNs has focused attention on the structural and dynamical basis of these motifs (Barabasi, 2005; Mangan and Alon, 2003; Mangan et al., 2006; Vazquez et al., 2004). Despite the advances in the identification of the mechanisms for the natural occurrence of these motifs, structural and dynamical functional bases have failed to provide an understanding of the properties and the density of occurrence of these motifs for various network systems. Thus, there has been the lack of a universal basis which can describe the underlying mechanisms behind the occurrence of these motifs, as well as the way in which these motifs encode functional information and the way that both the dynamical function and the topological generalization may evolve (Vazquez et al., 2004).

Here, we postulate that the abundance of certain motifs in a network can be predicted based on a conceptual framework that integrates nonequilibrium thermodynamics with Pareto-optimality of the biological functions to be carried out by the motif. We present an energetic-cost (defined herein as specific dissipation energy) theory that can explain which network motif has a higher probability of selection under a given environment, as well as the topological generalization of the subgraphs compared to other circuit designs. Through the developed framework, we have tried to answer the questions of why evolution converges to the same network motifs in TRNs and what advantage these selected network motifs offer as compared to other subgraphs for both steady state and dynamic analyses. Our study also demonstrates that network analyses using often-ignored energetics enables identification of new functionalities for rarely occurring motifs.

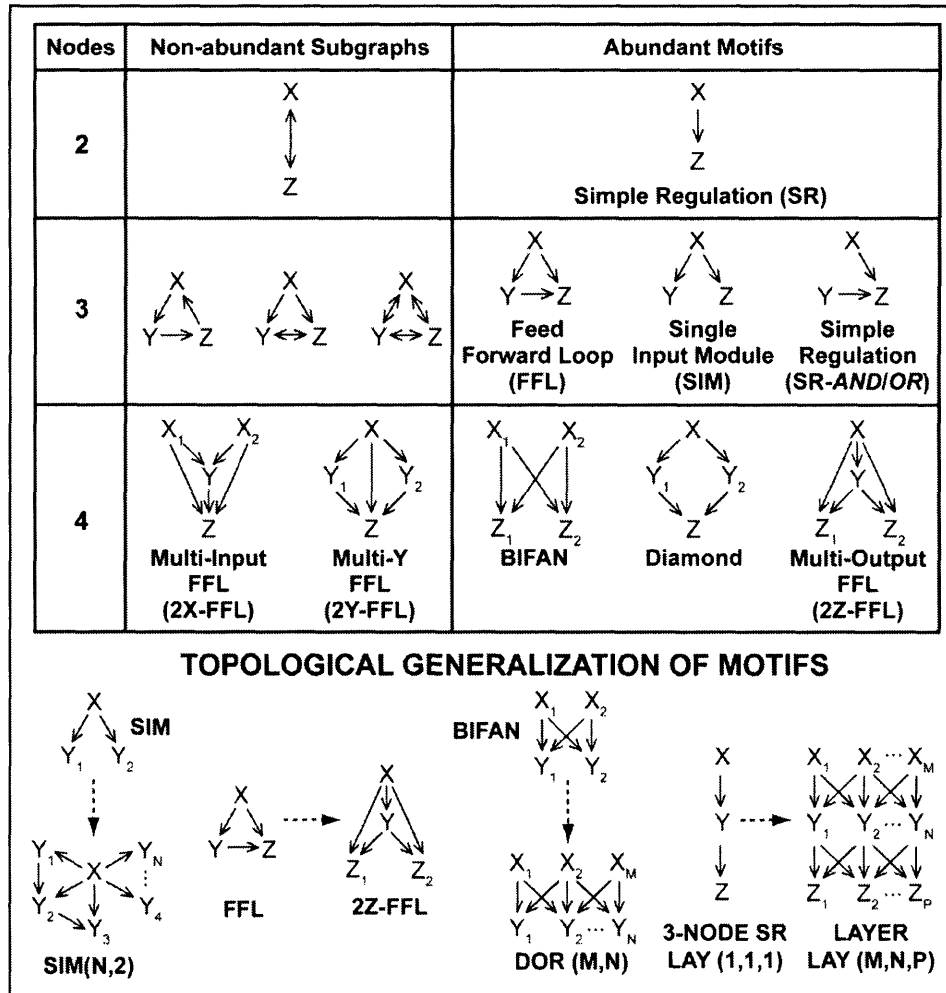


Figure 6.1 Description of Network Motifs and nonequilibrium TRNs. Patterns of subgraphs and network motifs found in *E. coli* and *S. cerevisiae* networks. FFL is the common three node subgraph, BIFAN is the common four node subgraph, SR is the common simple regulation motif. Higher order FFLs (multi-output, 2Z-FFL for four node TRNs) and diamond motif are other common four node subgraphs in these organisms. Also shown are the topologically generalized networks obtained from some of the commonly found relatively simple network motifs

6.3 DESCRIPTION OF NONEQUILIBRIUM TRNs.

Nonequilibrium processes require an external signal, flux or driving force to maintain the system far away from equilibrium (Section 5.2). The recent surge of interest in nonequilibrium thermodynamics has clearly exposed its role in small-scale systems such as biological molecular machines (Qian, 2004) and RNA folding/unfolding (Bustamante et al., 2005; Liphardt et al., 2002; Qian and Beard, 2005; Vilar and Rubi, 2001; Yin et al., 1999). Equilibrium systems are generally governed by classical thermodynamics and have reversible work equal to the Gibbs free energy change at constant temperature and pressure. Nonequilibrium steady-state (NESS)

analysis generally is used for irreversible systems and has Gibbs free energy unequal to the work done on the system because of heat/work dissipation and hysteresis (Rothschild et al., 1980). It has been shown earlier that the average work dissipated along any trajectory between different nonequilibrium states is always positive (Chernyak et al., 2005). This result led to the hypothesis in the current work that heat dissipation normalized to the input mass flux, defined by us as specific dissipation energy (*SDE*), should be minimal for biological systems and systems with lowest *SDE* should win out during evolutionary selection. Specifically, the dissipated work will be directly proportional to the input mass flux or raw materials utilized in moving from state A to state B, hence, normalized heat dissipation may serve as a criterion for choosing a trajectory to move from state A to state B. There is a striking parallel between the proposed energetic hypothesis in our work and the one proposed for ecosystem development where they report that minimization of specific dissipation is a primary criterion for evolution of ecological systems (Ludovisi et al., 2005). Recent NESS analysis in biochemical reactions (Qian, 2004; Qian and Beard, 2005) has revealed that the chemical potential of nonequilibrium processes is equivalent to heat dissipated by the system or work done on the system which in turn is equivalent to entropy generated by the system under NESS conditions. To evaluate the energetic cost involved during transcription, we utilized previous NESS analysis and developed a novel nonequilibrium thermodynamic kinetic formulation for gene transcription (Chapter 4, Figure 6.2). The developed nonequilibrium cyclic TRN model (a) explicitly deals with as many interactions as required with no limit on interactions (activation and repression) (Sections 5.3 and 5.4), (b) uses a competitive binding scheme for Boolean input logics (Section 5.5), (c) can be easily generalized to complex networks, and (d) provides estimation of the *SDE* involved for a gene-transcription factor (TF) combination during a transcription-translation process. Various mechanisms involved during TF-based protein synthesis (TF-DNA binding energetic (Darling et al., 2000; Jana et al., 2000; Scarpulla, 2002; Seredick and Spiegelman, 2004; Shea and Ackers, 1985; Whitson et al., 1986), mRNA binding effecting translation (Liphardt et al., 2002; Walton et al., 2002), structural changes (Xia et al., 2003), chromatin conformations (Russo et al., 1995), and post-translational changes (Dzeja and Terzic, 2003)) may contribute towards energetics of cellular regulation. However, for the sake of simplicity, heat dissipation obtained from the cyclic TRN model developed here, lumps these steps into reactions having pseudo rate constants which integrate these steps by factoring appropriate concentrations.

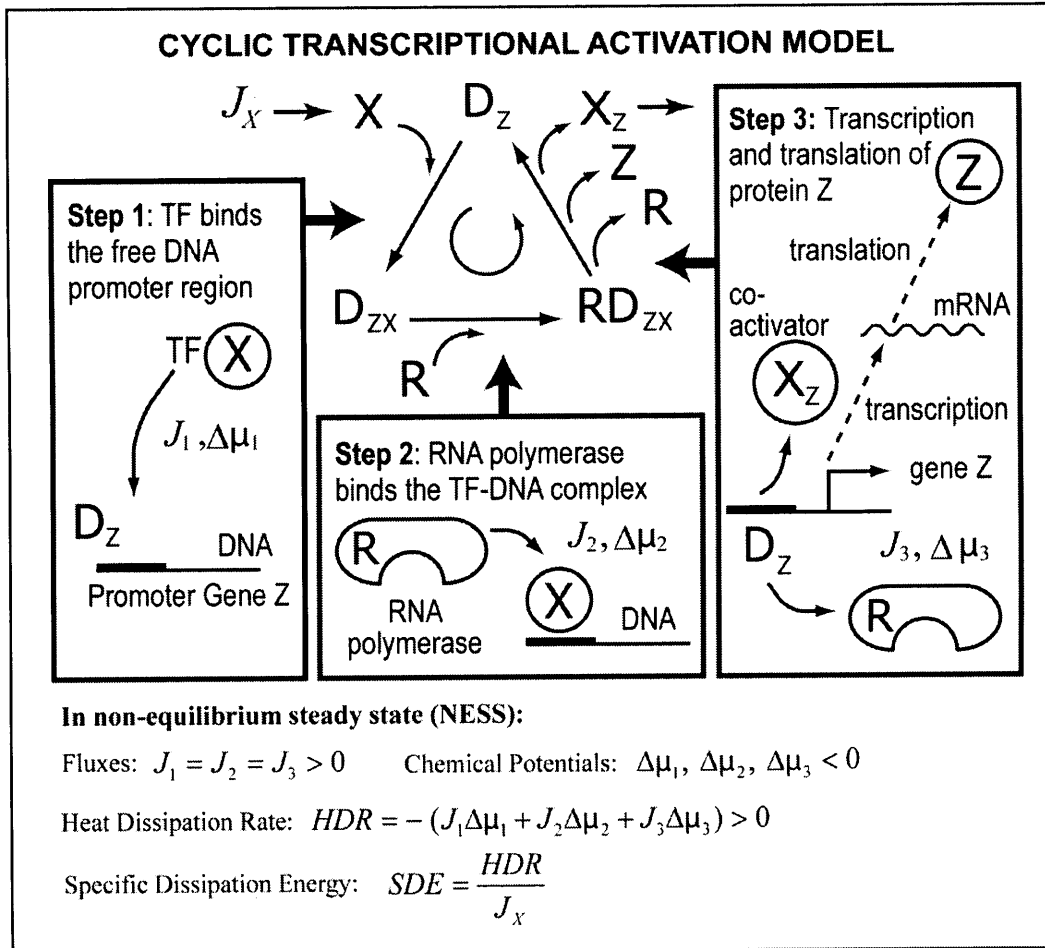


Figure 6.2: Nonequilibrium representation of a cyclic transcriptional activation schematic and estimation of its dissipative energetic (Chapter 4).

6.4 HYPOTHESIS: SPECIFIC DISSIPATION ENERGY SHOULD BE MINIMAL FOR BIOLOGICAL SYSTEMS UNDER OPTIMAL CONDITIONS.

We developed a Pareto thermodynamic criterion that couples the nonequilibrium thermodynamic analysis for TRNs with Pareto-optimal solutions for attaining biological functions of a motif (see Methods, Section 6.10.1). Cellular systems perform an array of regulatory, homeostatic, and phenotypic functions thus: they exhibit tradeoff between proliferation and differentiation (Nagrath et al., 2007), cellular functions and growth (Dekel and Alon, 2005; Kalisky et al., 2007; Zaslaver et al., 2006), and cellular functions and robustness (Savageau and Freter, 1979). Therefore, a multiobjective optimal approach (where tradeoff between several objectives has to be attained simultaneously) is necessary when seeking optimal

functional analysis of these systems (Nagrath et al., 2007). A multiobjective solution is said to be Pareto-optimal if there are no other solutions that can better satisfy all of the objectives simultaneously (Sections 2.3.4 and 3.3.3). Similar to cellular processes, transcription also works using the Pareto principle, i.e. simultaneous attainment of maximal conditions for higher transcription occurs at the cost of robustness or a cellular function (Figure 6.3A).

We first wanted to assess whether there is any correlation between Pareto-optimal transcriptional fluxes and *SDE* defined as the ratio of total heat dissipated by the system to the input mass flux. Existing paradigms (Itzkovitz and Alon, 2007; Itzkovitz et al., 2003) indicate that biological systems operate under optimal environmental conditions with optimal selection and utilization of existing resources. Using the developed energetic hypothesis, we found that at Pareto-optimal environmental surface, *SDE* is always the minimum for both metabolic networks and TRNs (Figure 6.3B, Section 6.10.1). It is to be noted that in both metabolic and transcriptional networks production of a metabolite or protein was used as the network objective. *SDE* steadily decreased when moving throughout the feasible space along the vector from minimum transcriptional fluxes to the Pareto-optimal transcriptional fluxes for the one and two-cycle metabolic network, and the FFL and feedback loop subgraph in TRN systems. From this analysis we can conclude that maximal feasible transcriptional rates are based on the optimal utilization of available energetic resources and that at the maximal Pareto-optimal condition the *SDE* is minimal therefore, natural systems operating under Pareto-optimal conditions will always have lowest specific energetic cost. We tried various architectures of TRN motifs and found this phenomenon to be true in all cases. Although some of the recent studies (Dekel and Alon, 2005; Zaslaver et al., 2004) have demonstrated that biological systems operate under optimal selection and utilization of resources, based on our results, we feel that missing the energetics analysis at the optimal conditions has led to an incomplete understanding of network motif analysis. Importantly, our optimality framework provides a rationale for how cells integrate optimal selection and utilization of resources with energetic cost. Paradoxically, our optimality results elicit that energetic cost is minimal at the maximal/optimal resource utilization conditions although it was not part of cellular objectives being optimized. Thus, rather than operate at optimal conditions, cellular regulatory systems simultaneously use energetic cost minimization as the underlying basis, perhaps to operate under globally optimal conditions.

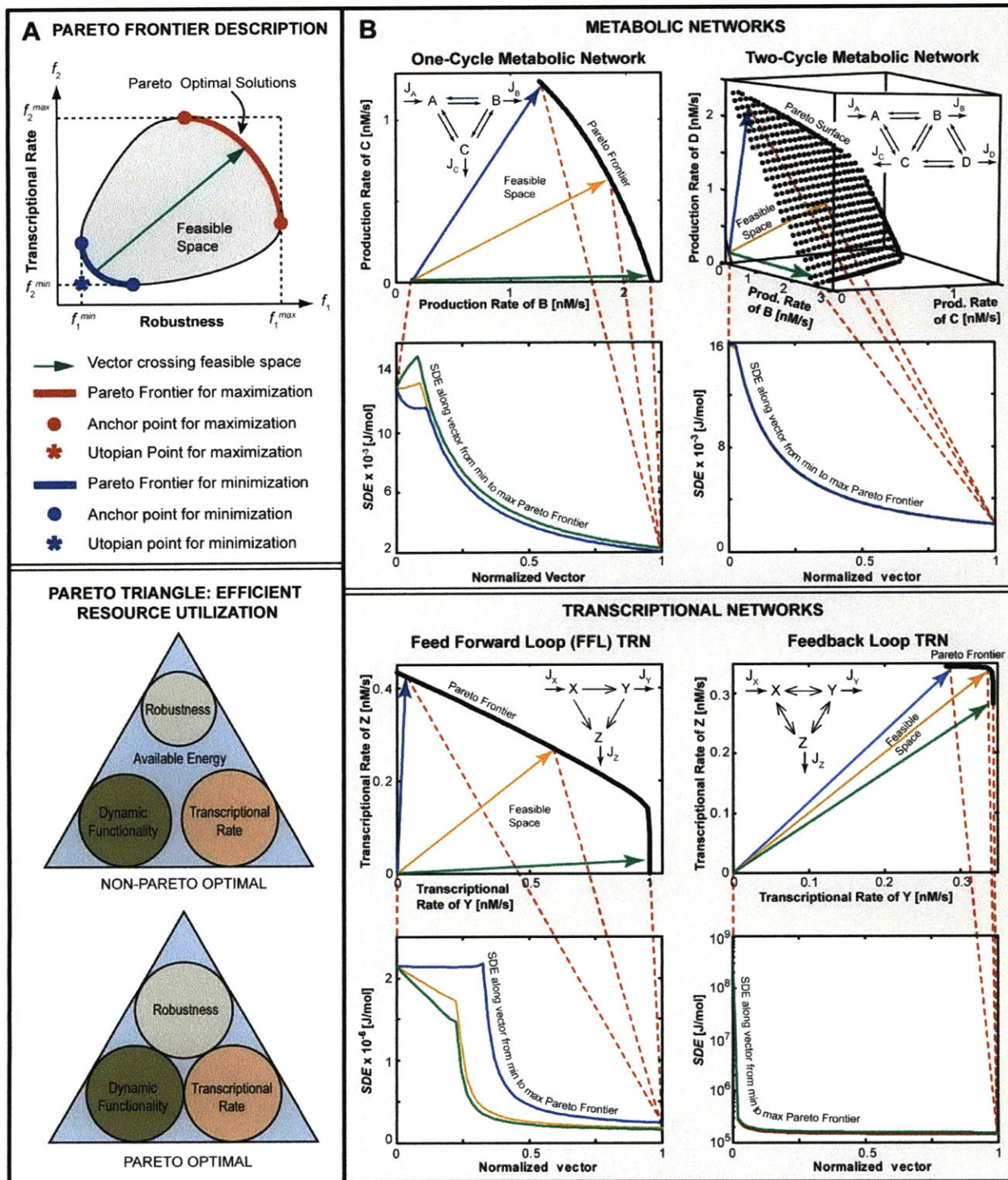


Figure 6.3 Optimal SDE serves as the basis for efficient resource utilization in metabolic and transcriptional networks. **A)** Pareto frontiers for a bi-objective maximization and minimization problem (Nagrath et al., 2007). Pareto solution is one where any improvement in one objective can only take place at the cost of another objective. A Pareto set is a set of Pareto-optimal solutions. If objective functions f_1 (for instance, robustness) and f_2 (for example, transcriptional rate) alone are individually maximized, then the optimal values are f_1^{max} (point P_1) and f_2^{max} (point P_2), respectively. Here g_1^* and g_2^* are the anchor values for design objectives g_1 and g_2 , respectively. The ideal or Utopian solution (g_1^*, g_2^*) obtained by the individual maximization of the objective functions is

generally not a feasible solution of the multiobjective optimization problem. The arc P_1P_2 is defined as the Pareto frontier containing multiobjective-optimal or tradeoff solutions. Pareto-optimal and non-Pareto utilization of available energy and resources are shown as the corresponding circular area in the triangle for three objectives of a cellular TRN. **B)** Pareto frontiers for biochemical networks were obtained by considering the maximization of rate of synthesis of B and C (and D) as the objectives whereas for TRNs the maximization of protein production rates of Y and Z were considered as the appropriate objectives. Both, biochemical networks and TRNs follow the Pareto dominance rule. SDE was found to be minimal at the Pareto frontier compared to the non-Pareto feasible solutions. Following the solutions on the line vector from Pareto frontier of minimization to the Pareto frontier of maximization there is a constant decrease in the SDE and it is minimal at the Pareto frontier optimal transcriptional rates.

We further studied the implications of energetic cost on network structure and motifs. Thus, after establishing SDE as a valid energetic cost concept, we investigated whether natural biological systems utilize this as a rule of selection and thus lead to high abundance (also defined as density or occurrence frequency) of certain subgraphs. FFLs can be classified as coherent (Type-1 to 4) or incoherent (Type-1 to 4). Multiple lines of evidence (Kalir et al., 2005; Kashtan et al., 2004; Ma et al., 2004; Mangan et al., 2006; Mangan et al., 2003) suggest that coherent and incoherent Type-1 FFL motifs are abundant in both prokaryotic *E. coli* and the eukaryotic *S. cerevisiae* TRN systems and other FFL motifs rarely occur in these small organisms irrespective of whether genes or operons were used as nodes. However, no explanation has yet been provided for the rare occurrence of certain FFL types and the abundance of a few motifs (Mangan and Alon, 2003; Mangan et al., 2006; Mangan et al., 2003), and no universal selection parameter has yet been found that can explain functional representation of these motifs with their network structure completely. To answer these questions, we hypothesized that SDE itself is the universal correlate used by biological systems for selection during evolution. To evaluate this, we first obtained the Pareto-averaged SDE (the average SDE over the maximal Pareto frontier between transcription rates of proteins Y and Z) for all eight FFL motifs for both *AND* and *OR* logics (Figure 6.4A). Pareto-averaged SDE is found to be the lowest for highly abundant FFL motifs and highest for rarely occurring network motifs (Section 6.10.2-3). Significantly, we observed that Pareto-averaged SDE for both input logics (*AND* and *OR*) correlates inversely with the abundance of the network motif as experimentally determined (Mangan and Alon, 2003) for microorganisms (*E. coli* and *S. cerevisiae*) as seen in Figure 6.4B (Spearman's correlation coefficient = -0.43 ; $p = 0.007$).

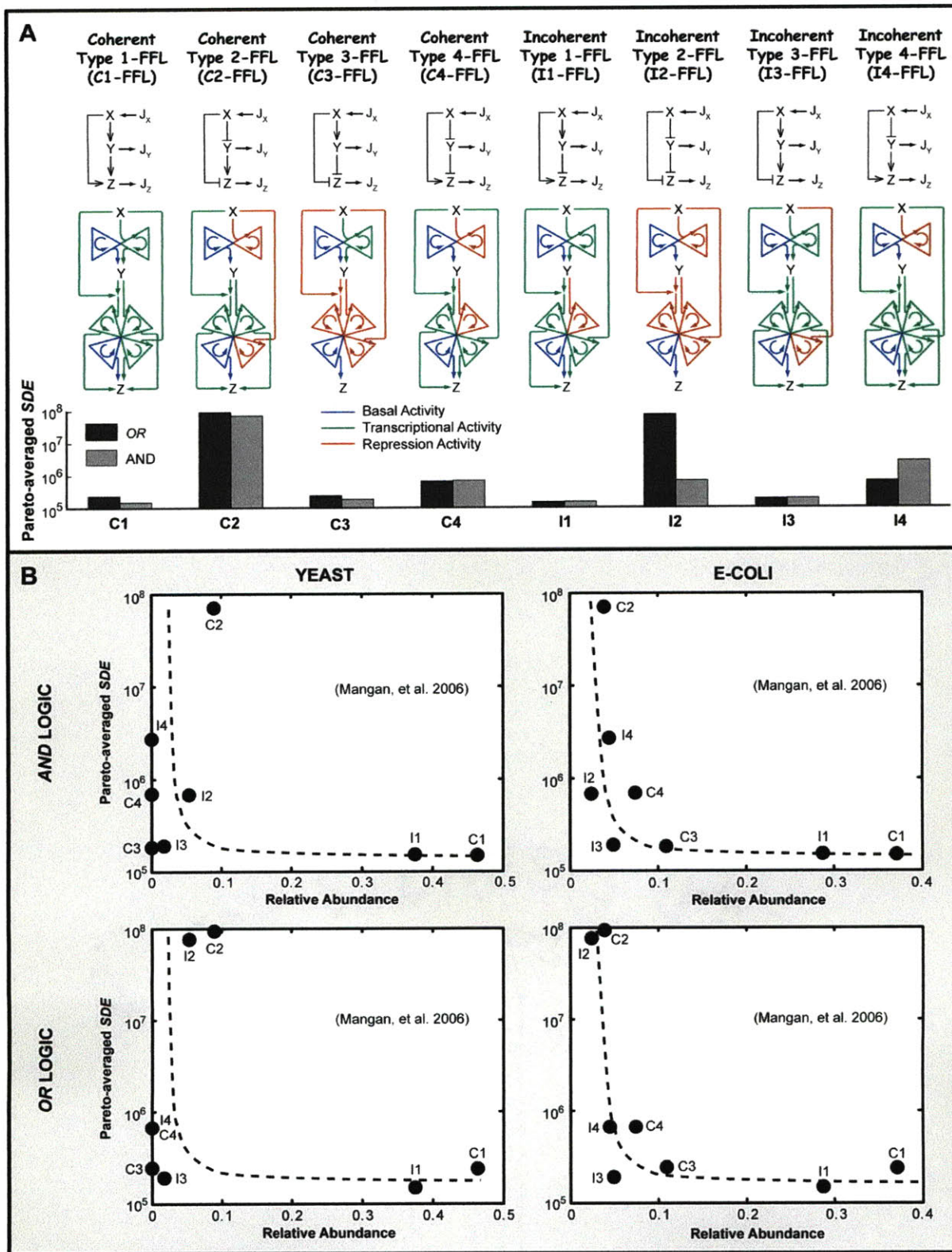


Figure 6.4: Optimal *SDE* predicts the network motif abundance. **A)** Optimal *SDE* of eight coherent and incoherent FFL motifs at the Pareto frontiers with two input functions *OR* and *AND*. The input functions integrate the incoming signals at the promoter of gene *Z*. Here, arrow denotes activation and symbol \perp denotes repression. The bar graph

shows the averaged *SDE* over the Pareto frontier of a FFL motif. The Pareto frontier was obtained between transcription fluxes J_Y and J_Z . The symbolic representation of the corresponding cycles involved is shown below each FFL in the figure and FFL types are marked C and I for coherent and incoherent. **B)** Correlation of averaged Pareto-optimal *SDE* with relative abundance of FFL motifs for *E. coli* and *S. cerevisiae* TRNs. Correlation is done for network motif relative abundance data obtained from experimentally verified *E. coli* (Shen-Orr et al., 2002) and *S. cerevisiae* (Milo et al., 2002) databases by considering gene (Mangan and Alon, 2003) or operon (Ma et al., 2004) (**Supplementary Figs. 10-12**) as nodes. A statistically significant negative correlation (Spearman correlation coefficient = -0.43 and p value = 0.007) between the Pareto-averaged *SDE* and the published data of relative abundance of FFL motifs (Mangan and Alon, 2003) was observed.

Notably, similar *SDE*-network motif abundance correlation is obtained when genes were used as nodes (Ma et al., 2004) (Section 6.10.3) and for virtually all parameter choices satisfying energetic constraints. The preceding results show that during evolution minimal *SDE* may have been an important criteria used for development of highly organized complex TRNs. The success of predicting abundance or the regulatory role of network motifs depends on various factors which may be probably limiting cells in different stages of their life cycle, as well as their evolutionary history. Our results elicit that although energetics necessarily may not be the rate limiting factor, at some period of evolution it might have been the “bottleneck”.

6.5 SDE LINKS NETWORK TOPOLOGY WITH DYNAMIC FUNCTIONALITY AND EVOLUTIONARY ADAPTATIONS.

The dynamical function has been used to explain the occurrence frequency of motifs (Alon, 2007; Mangan and Alon, 2003; Prill et al., 2005); however, it fails to do so for functionally similar motifs. We hypothesized that *SDE* can be used as the mapping function between network topology and dynamic output (Supplementary Text 6). To achieve this goal, we first estimated *SDE* for a dynamic scenario where a phenotype exhibited by an incoherent FFL motif has been shown both theoretically and experimentally (Mangan and Alon, 2003; Mangan et al., 2006; Mangan et al., 2003; Zaslaver et al., 2006; Zaslaver et al., 2004) in the galactose system of *E. coli* to be of pulse generation and response acceleration. Both Type-1 (I1-FFL) and Type-4 (I4-FFL) incoherent FFLs with *AND* input logic were found to have similar dynamic phenotype for an ON-OFF step change in the input signal X (Figure 6.5A, Section 6.10.4). Thus, the current paradigm of associating dynamic functionality with frequency of occurrence of motifs does not explain the variations in abundance density since both FFL types had pulse

generation and response acceleration as function with similar magnitude. However, if energetic cost SDE is compared for this input perturbation scenario, then a difference of at least one order of magnitude between time-averaged SDE ($\langle SDE \rangle$) of I4-FFL and I1-FFL is seen. The energetic hypothesis and SDE seems to perform better than dynamical function in explaining abundance/rare occurrence of incoherent FFLs.

Coherent FFLs have been shown both theoretically (Mangan and Alon, 2003) and experimentally in the arabinose system of *E. coli* (Mangan et al., 2003) to have the sign-sensitive delay as the major dynamic function. Sign-sensitive delay is defined as the response delay when compared to simple regulation (SR) (recall Figure 6.1) and depends on the sign of the input step (ON or OFF). For a coherent FFL-AND this delay is expected to be during an ON step but not during an OFF step where response is similar to SR. Both, coherent Type-1 (C1-FFL) and coherent Type-4 (C4-FFL) FFLs were seen to have similar sign-sensitive delay functionality (Figure 6.5A, Section 6.10.4); hence, it cannot clearly establish the rare occurrence of C4-FFL compared to C1-FFL. For a step input in activator X , as expected, the delay with respect to SR was marginally pronounced in C1-FFL than C4-FFL (Figure 6.5A). Strikingly, we found that $\langle SDE \rangle$ was significantly lower for C1-FFL than C4-FFL with AND logic thus consistent with their abundance.

We further asked whether SDE can correlate dynamic functionality with abundance for higher order four-node FFLs. Previous theoretical and experimental analysis (Kashtan et al., 2004) of the protein synthesis scheme in the flagella system of *E. coli* revealed that the only abundant network motif is the multi-output FFL (2Z-FFL) compared to multi-input (2X-FFL) and multi-Y (2Y-FFL) subgraphs. The preceding results explain the abundance of 2Z-FFL over other FFL generalizations based on the existence of dynamic function of the temporal order First-In-First-Out (FIFO) of protein expression. Our model aptly predicted FIFO temporal order for 2Z-FFL and revealed that SDE for 2Z-FFL was lowest among other possible generalizations, such as 2X-FFL and 2Y-FFL, throughout the time interval of exponential pulse input (Figure 6.5B, Section 6.10.5). This clearly demonstrates SDE as an underlying basis for classifying generalized motifs based on environmental perturbations.

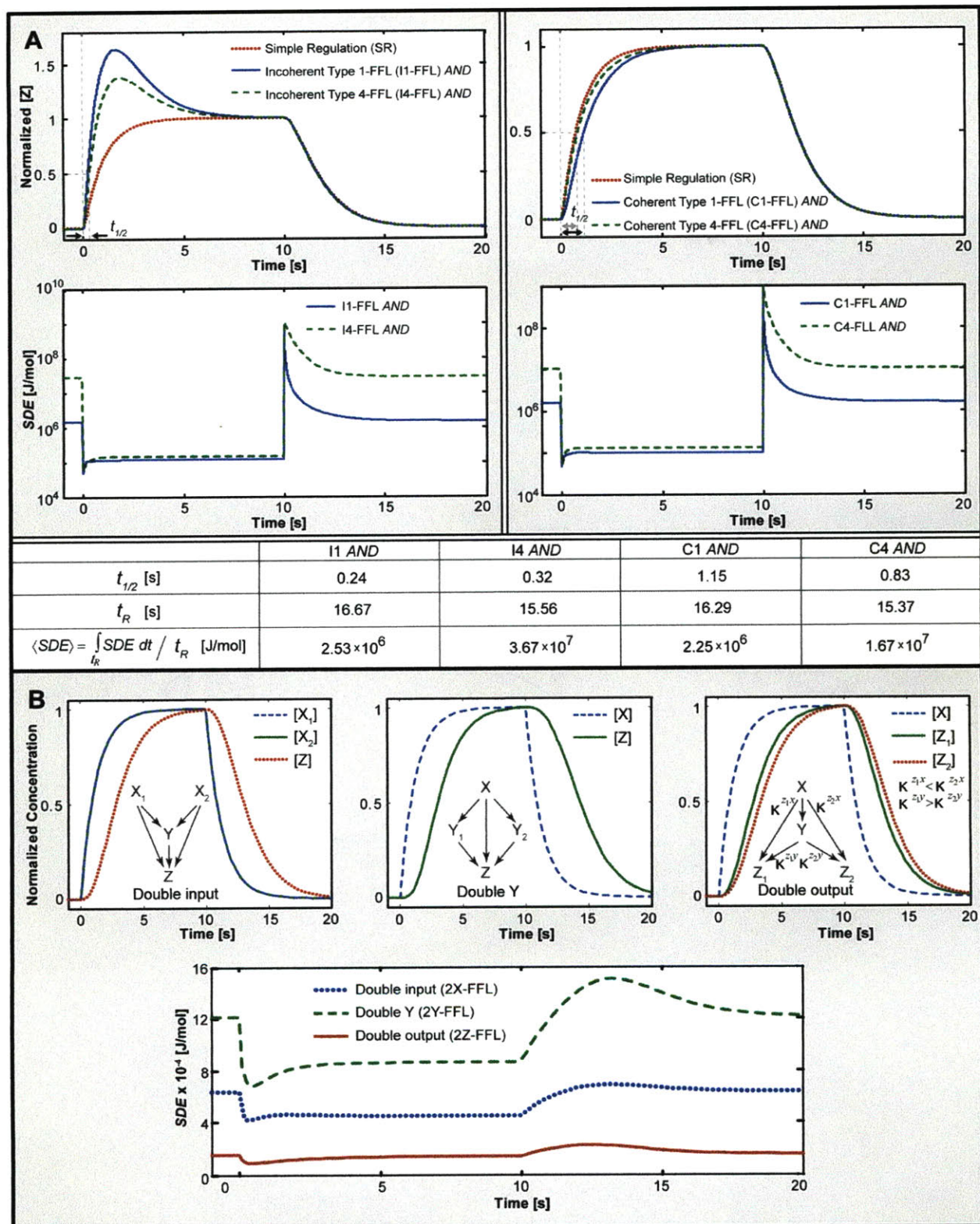


Figure 6.5 Selection of network motifs for dynamic functionality utilizes energetics as the underlying basis. **A)** Comparison between I1-FFL and I4-FFL; and C1-FFL and C4-FFL with AND logic for an input pulse in input X . I1-FFL and I4-FFL with AND logic motifs have pulse generation and response acceleration in ON step as the dynamic function. $t_{1/2}$ is the time to reach 50% of the steady state and response time, t_R is the time to attain the final steady

state. $t_{1/2}$ is slightly lower for I1-FFL than I4-FFL but *SDE* of I1-FFL is significantly lower (10×) than I4-FFL. C1-FFL and C4-FFL with *AND* logic motifs have delay in ON step with marginally higher delay for C1-FFL. *SDE* is significantly lower (10×) for C1-FFL than for C4-FFL. **B**) Comparison of energetics for dynamics in double-input (2X-FFL), double-Y (2Y-FFL), double-output (2Z-FFL) generalizations of FFLs. First-in-First-Out (FIFO) order in a 2Z-FFL (with *OR* logic) is obtained with a specific order of κ ($\kappa^{Z_1X} < \kappa^{Z_2X}$, $\kappa^{Z_1Y} > \kappa^{Z_2Y}$) and for an exponential input pulse. *SDE* for 2Z-FFL is lower than 2X-FFL which in turn is lower than 2Y-FFL. In all (**A** and **B**) cases, similar dynamic and *SDE* responses were obtained when the kinetic parameters were simultaneously increased or decreased by two orders of magnitude.

6.6 SDE PREDICTS THE SELECTION OF MOTIFS UNDER VARYING ENVIRONMENTS

We next analyzed whether *SDE* may predict the environmental selection of motifs and whether cells can use *SDE* as the master sensor for optimal and evolutionary tuning of protein expression. Based on the current understanding in the arabinose sugar catabolism system of *E. coli* for *AND* logic, simple regulation (SR-*AND*) is never selected over C1-FFL for an input pulse environment with a high probability of shorter pulse durations (Dekel and Alon, 2005; Dekel et al., 2005). However, SR-*AND* is favorably selected over C1-FFL for an input with longer pulse durations. To investigate the influence of energetic cost on environment selection, we used similar definitions of objectives and fitness function as identified previously (Dekel and Alon, 2005; Dekel et al., 2005). One of the objectives defined as *benefit* is of relative increase in cellular growth rate because of production of protein *Z* which can be formulated based on Michaelis-Menten kinetics as *cell growth rate* = $\delta[Z]/(L+[Z])$, where δ [nM/s] is the maximal growth rate per *Z*, and *L* is the Michaelis-Menten constant of *Z*. The other objective competing to *benefit* and is the *cost* of production of protein *Z* and is proportional to rate of depletion of resources used to produce protein *Z*, thus shown as: *cell growth reduction rate* = ηJ_T^Z , where η is the reduction in growth rate per molecule of *Z* produced and J_T^Z is the transcription rate of *Z*. Using Pareto-optimal paradigm we obtained the Pareto frontier between *benefit* and inverse of *cost* for both C1-FFL *AND* and SR-*AND* by having the pulse amplitude, pulse duration t_X , and the Michaelis-Menten constant *L* as optimizing variables (Figure. 6.6A). This Pareto frontier as mentioned previously indicates the maximal efficient utilization conditions used in cellular systems for various input pulse durations and fluctuating environments. Remarkably, we found that *SDE* can be a correlate for environmental selection of motifs utilizing the Pareto frontier condition. For shorter pulse durations, *SDE* is higher for SR-*AND* than for C1-FFL, indicating

unfavorable selection of SR-AND because of high energetic cost, whereas the reverse is true at higher pulse durations (Figure 6.6A).

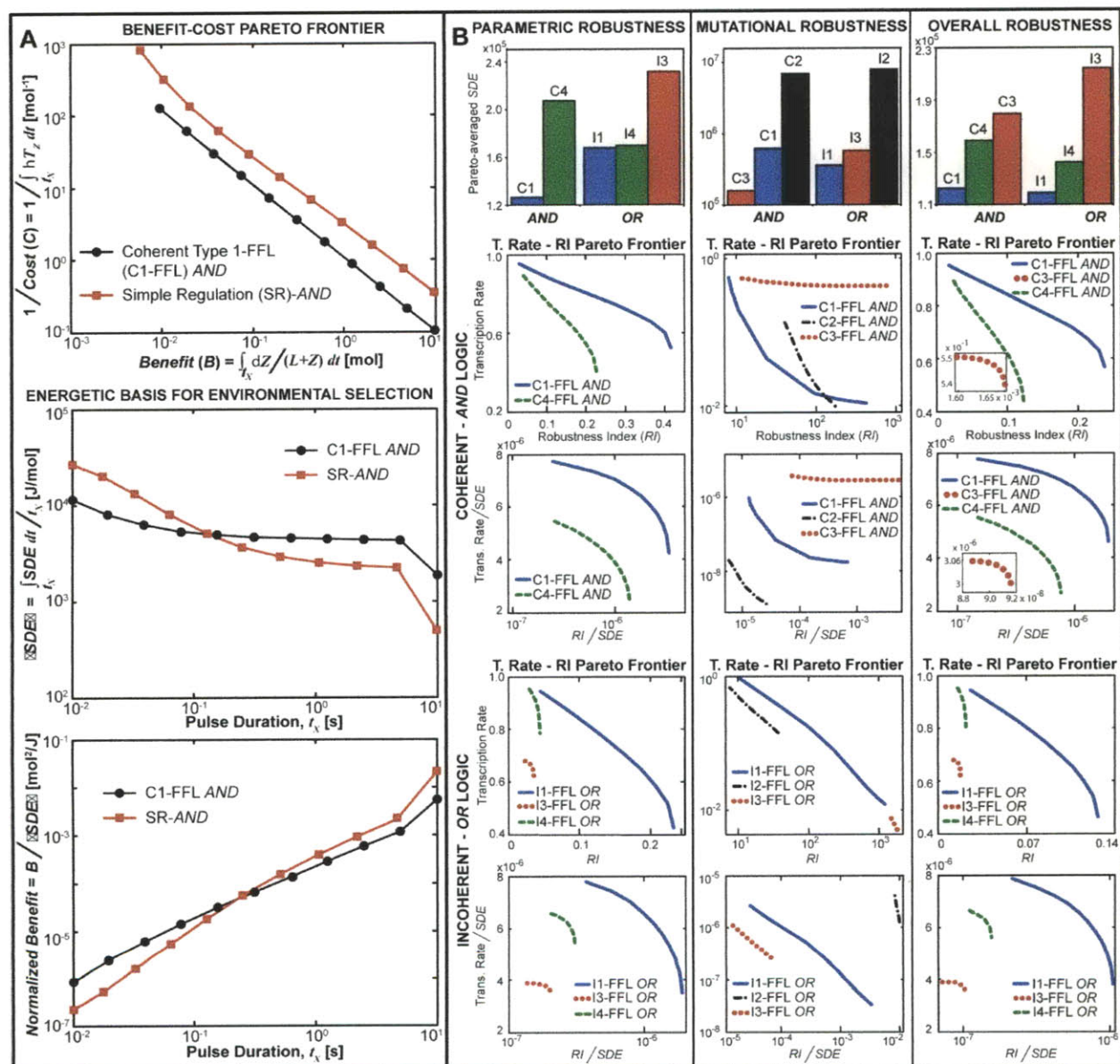


Figure 6.6 Using energetic-cost criterion for environmental selection and predicting motif functionalities. **A)** Environmental selection of motifs. Shown are the Pareto frontiers between *benefit* which is relative increase in cellular growth rate ($\delta[Z]/(L+[Z])$, where δ [nM/s] is the maximal growth rate advantage per *Z* protein, and *L* is the Michaelis-Menten constant), and the inverse of *cost* (which is cell growth reduction rate = η/J_Z^Z , where J_Z^Z is the transcription rate of *Z* and η is the reduction in growth rate per molecule of *Z* transcribed) for C1-FFL AND and SR-AND for various pulse durations. Comparison of C1-FFL AND with SR-AND for various pulse durations along the cost-benefit Pareto frontier shows an energetic advantage (lower $\langle SDE \rangle$) in selecting C1-FFL AND for short pulse durations and vice versa for SR-AND at longer pulse durations. Comparison of C1-FFL AND with SR-AND for various pulse durations along the cost-benefit Pareto frontier shows a *normalized benefit* (ratio of *benefit* over $\langle SDE \rangle$) advantage (higher ratio) in selecting C1-FFL for short pulse durations and vice versa for SR AND at longer pulse

durations. **B)** Prediction of robustness as functionality in rarely occurring motifs. Tradeoffs between cellular TRN objectives of transcription rate and robustness were observed for both coherent and incoherent FFLs by obtaining corresponding Pareto frontiers. Coherent and incoherent FFLs were analyzed for *AND* and *OR* input logic, respectively, for parametric robustness, mutational robustness and combinations of both. For parametric robustness function, Pareto-averaged *SDE* for highly abundant C1-FFL was lower than rarely occurring C4-FFL and similarly *SDE* for I1-FFL was lower than I4-FFL. For mutational robustness function, *SDE* for C3-FFL (another rarely occurring FFL) was lowest and thus is contrary to the abundance-energetic cost paradigm (higher the abundance lower the *SDE*). In overall robustness, both parametric and mutational variations were combined and obtained *SDE* data generally follows abundance-energetic cost paradigm. In general, higher span of Pareto frontiers between cellular TRN objectives of transcription rate and robustness (also defined as *benefit function*) indicate efficient resource utilization and lower *SDE*. *Normalized benefit* (i.e., *benefit* / energetic cost) indicate higher *benefit function* at low energetic cost. For parametric robustness, C1-FFL has highest *normalized benefit function*. Rarely occurring C3-FFL has highest *normalized benefit* of robustness and transcription rate among coherent FFLs. Similarly, I2-FFL was found to have highest normalized robustness. *Normalized benefit functions* for overall robustness environment mostly followed the abundance-energetic cost paradigm.

Additionally, *SDE*'s variation with pulse duration for SR-*AND* was remarkably similar to integrated growth rate variation with pulse duration for SR-*AND* (Dekel and Alon, 2005; Kalisky et al., 2007; Zaslaver et al., 2004). This energetic based selection criteria is not only consistent with recent findings when used as standalone measure but is also holds true when *benefit* is normalized with *cost* to obtain a measure that indicates higher benefit at lower cost. To investigate whether *SDE* was the underlying cost, we evaluated the *normalized benefit* (ratio of *benefit* over time-averaged *SDE*, $\langle SDE \rangle$) for various pulse durations along the Pareto frontier. We found that *benefit normalized* with energetic cost is higher for FFL at shorter pulse durations and higher for SR-*AND* at higher pulse durations thus suggesting that motifs having higher normalized benefit will be favored under particular environment for selection. We also analyzed the environmental selection of C1-FFL over C4-FFL with *AND* logic (Section 6.10.6). Consistent with abundance data we found that C1-FFL is energetically favorable over C4-FFL for all pulse durations. Thus, these results let us conclude that *SDE* can be the underlying basis not only for predicting dynamic phenotype but also for motif selection in varying environments.

6.7 *SDE* PREDICTS NEW FUNCTIONALITIES OF MOTIFS

Having shown that the energetic cost, in combination with optimality theory, can be used to identify the motif selection during varying environments, we moved on to predict new functionalities of motifs using typical regulatory interaction objectives such as robustness and transcription rate, along with the underlying objective of minimal energetic cost. The difference between abundance of various motifs is directly correlated with their functionalities, thus rarely

occurring motifs may owe their existence to unidentified functionalities. Cellular TRN systems often have robustness as their objectives, and tradeoffs between robustness and cellular functions imposed (Barkai and Leibler, 1997; Carlson and Doyle, 2000; Martin and Wagner, 2008; Morohashi et al., 2002; Rao et al., 2004; Rao et al., 2002; Stelling et al., 2004a; Stelling et al., 2004b) on complex genetic circuits may play a dominant role during evolutionary selection of motifs. Robustness index (RI) was defined here as the objective function which cellular decision machinery utilizes for selecting motifs among topologically and architecturally different but equally capable of higher transcription during varying environments of internal and external perturbations, mutations and non-genetic changes. Thus, we hypothesized that motifs which have high RI and transcription rate despite the high fluctuations during varying environments will emerge as preferred during evolutionary selection. Further, the higher the TRN objectives (i.e., robustness and transcription rate) normalized by the underlying energetic cost, the higher the fitness advantage of one FFL motif over another under a fluctuating environment. Typical environment fluctuations that were considered included parametric variations, mutational variations, and a combination of both. We conducted large-scale simulations of varying environments across fluctuations spanning two orders of magnitudes and thus imposed a variety of environments which are generally difficult to study in naturally evolved biological networks. We found that transcription rate and robustness form a Pareto frontier during both parametric fluctuating environment and mutational variations for various motifs (Figure 6.6B). The highly abundant motifs (for coherent FFLs is C1-FFL and for incoherent FFLs is I1-FFL) displayed lower Pareto-averaged SDE and had a higher *benefit (or objective) function* advantage for both, transcription rate and robustness over rarely occurring motifs (for coherent FFLs, C4-FFL; and for incoherent FFLs, I3 and I4-FFLs). As hypothesized, both C1-FFL and I1-FFL had higher fitness advantage, i.e. *normalized benefit function* (ratio of *objective function* over energetic cost) over rarely occurring FFLs throughout the parameter space and Pareto frontier. In a particularly challenging environment of mutational variations, we found that C3-FFL is more favorable over other coherent FFLs and a rarely occurring motif, I2-FFL, has higher robustness *benefit function* over other incoherent FFLs. Our results indicate that C3-FFL has higher normalized robustness and transcription rate throughout the parameter space over other coherent FFLs for mutational variations and likewise I2-FFL has higher normalized robustness advantage over other incoherent FFLs. More precisely, if parametric robustness is the desired functionality then C1-

FFL is favorable over other coherent FFLs, thus explaining their abundance. On the other hand for coherent FFLs, if robustness against mutational changes is desired, then C3-FFL is energetically favorable and has higher *normalized benefit function* for both transcriptional rate and robustness transcriptional objectives. This is significant, because so far no functionality has been found for C3-FFL. Similarly, the results of incoherent FFLs highlight an unidentified functionality for rarely occurring motif, I2-FFL, chiefly the high robustness for mutational changes (I2-FFL has higher normalized robustness although it has lower normalized transcription rate), which was not captured by current TRN models because of ignored energetics. For the third simulation group of overall robustness (both parametric and mutational) (Figure 6.6B), motifs generally followed the abundance principle. In general, our results reveal how functionalities and selection behavior emerges by combining energetics in network motifs. Importantly, current mass-balance based optimality paradigm will not be able to capture these functionalities because essential thermodynamic constraints are ignored.

6.8 *SDE* AS THE BASIS FOR TOPOLOGICAL GENERALIZATION OF MOTIFS

We further hypothesized that evolution converges toward a single basis of *SDE* for topological generalization of subgraphs. Addition of each new topological pattern can occur in a random fashion or can be selected by evolution based on nutrient resources and environment. In the latter case, a new level of biological organization is made having emergent properties governed by the physical laws and evolutionary selection constraints. We first analyzed the generalization of single input module (SIM) patterns (Figure 6.7A, Section 6.10.7). With the addition of nodes there is a steady decrease in *SDE*, which becomes nearly constant around SIM with 13 nodes. Although still unexplained, previous results (Shen-Orr et al., 2002) have revealed that SIM occurs in TRNs only for input nodes greater than 13. We observed that there is no further energetic benefit after increasing the number of nodes beyond 13 which explains the minimum requirement of 13 nodes in SIM.

Next, we analyzed various generalizations of SIM. Earlier studies reported agreement between analytical predictions and the measured subgraph counts of *E. coli* and *S. cerevisiae* TRNs (Vazquez et al., 2004). We hypothesized that *SDE* could explain the reason behind the atypical frequency of occurrence of generalized SIM subgraphs. To elucidate this, a

corresponding energetic phase diagram (Figure 6.7B, Section 6.10.8) using the presented *SDE* analysis was estimated for steady state of subgraphs. Strikingly, we found that the energetic phase diagram was similar to the subgraph phase diagram obtained using abundance data (Vazquez et al., 2004). In both phase diagrams, two types of subgraph patterns emerge: type-I (below the white line), which are highly abundant and have favorable lower *SDE*; and type-II (above the white line), which have higher *SDE* and are rarely occurring in TRNs of both *E. coli* and *S. Cerevisiae* (Vazquez et al., 2004). Remarkably, this shows that energetic demand can be used as a basis for the topological generalization of the simplest unit or motif in a complex network. Coupled with the observation previously made that scale free networks form hubs and that there is a preferential addition of new linkages to the nodes with the higher linkages, the presented energetic phase diagram explains that this phenomenon occurs because of favorable energetics. Throughout the phase diagram (Figure 6.7B), any addition of linkages to the input node decreases the energy compared to the addition of a linkage at other nodes. In summary, our *SDE* data for these subgraphs uncovers a mechanism for formation of these complex networks and is significant because, to date, there has been no functional basis or explanation provided for topological generalization of these subgraphs.

We next developed a new pulsed energetics topological generalization prediction (PETGP) framework which allows predicting generalizations of subgraph patterns without prior knowledge about function and characterization of a subgraph pattern (Section 6.10.9). To validate this for known generalizations, we assumed no knowledge about any relationship between gene-activator linkages or any dynamic behavior; hence, all the linkages had equal transcription parameters. The first step in PETGP analysis requires sending a small pulse input around its κ value (activation coefficient, here defined as the characteristic concentration of the subgraph) whereby energetic cost *SDE* is estimated. PETGP analysis is similar to the residence-time-distribution analysis, in which an input tracer pulse is used to predict the characteristics about chemical reacting systems in a reactor. Similar to Gibbs energy change of any reaction (higher $-\Delta G$ is favorable for a reaction to proceed forward spontaneously) we found that higher $|\Delta\langle SDE \rangle|$ leads to favorable generalizations (Figure 6.7C).

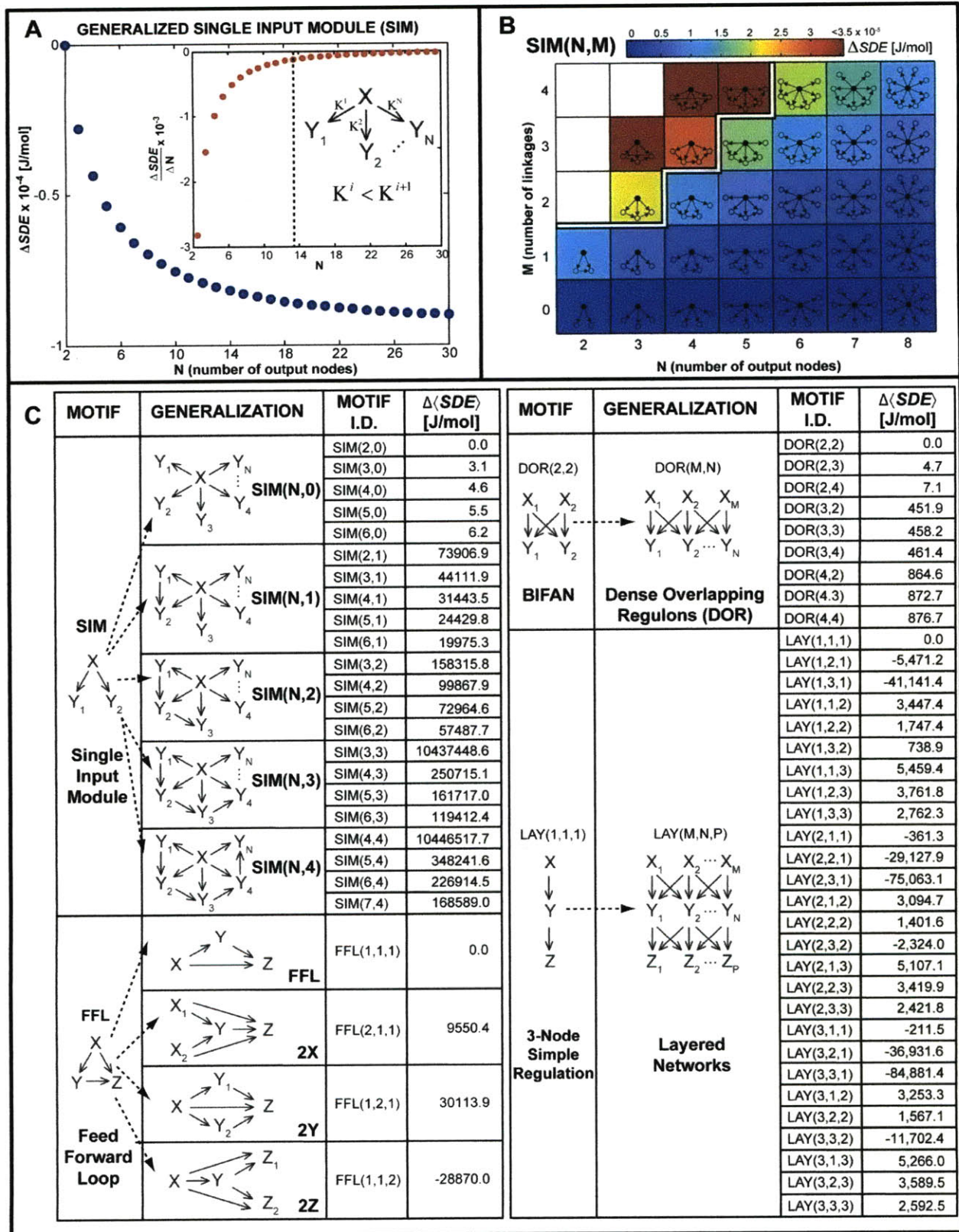


Figure 6.7: Energetics dictates the subgraphs generalization **A)** *SDE* decreases with increasing linkages during generalization of SIM. In the inset in the top panel, the rate of decrease of *SDE* becomes nearly constant around 13 output nodes. **B)** Energetics based organization of directed subgraphs in energetic phase diagram. Generalization

based on the number of nodes (N) and the number of links (M) is similar to the predicted subgraph phase diagram organization based on abundant subgraphs (Vazquez et al., 2004). The stepped white line separates the lower SDE type I subgraphs (below the line) from the higher SDE type II subgraphs. Some of subgraphs for the same number of nodes and linkages can have various topologies based on directionality of linkages and only representative values of SDE s are shown. The background color is proportional to the relative subgraph being SIM with 2 output nodes the base one. For directionality, FFL-OR with a forward basis structure was implemented. The critical energetic difference or ΔSDE was chosen to be $2 \cdot 10^5$ J/mol for this classification. C) PETGP organizes and predicts generalization of a subgraph into large-scale topological assembly of networks (Kashtan et al., 2004). Various input pulses were studied and only representative exponential pulse time-averaged SDE data are shown. The generalization strategy is presented for SIM, FFL, BIFAN or dense overlapping regulons (DOR) and Layered generalization of three-layered SIM with OR input logic. SDE is presented relative to the base case for each generalization with the base case normalized to zero value. Highly favorable subgraphs have high negative $\Delta\langle SDE \rangle$ compared to unfavorable subgraphs with positive $\Delta\langle SDE \rangle$ when compared to the base subgraphs.

Using the PETGP framework we correctly predicted possible generalizations of different subgraphs. We observed that 2Z-FFL had higher $|\Delta\langle SDE \rangle|$ than both 2X-FFL and 2Y-FFL thus suggesting that 2Z-FFL is the only motif occurring in TRNs (Figure 6.7C). The other commonly occurring motif (besides FFL) is the so-called BIFAN. Experimentally determined motif abundance data in *E. coli* has revealed that adding more input nodes during BIFAN generalization makes them unfavorable for occurrence (Kashtan et al., 2004). For example, in *E. coli* TRNs (3X-3Y) is less abundant (21 occurrences) than (2X-2Y) which is highly abundant (242 occurrences). Remarkably, we noticed that the addition of a new input node to create a higher input pattern in BIFAN increases $\Delta\langle SDE \rangle$ (Figure 6.7C). Therefore, our model predicts an unfavorable generalization and low frequency of occurrence for input node addition.

Next, we investigated the generalization of a three-node SR in a layered architecture (Figure 6.7C). First, we found that for a three-layered network the most favorable subgraph was the 3X-3Y-1Z configuration. This tapered architecture prevalent in neural networks may be abundant because it is energetically beneficial in TRNs. Second, we found that for one-input node generalizations, LAY(1,3,1) is the most favorable subgraph compared to the other eight possible configurations. This is consistent with earlier findings in TRNs of *E. coli* where it was found that the number of nodes in layer-0, 1, and 2 were 76, 233 and 87, respectively (Kashtan et al., 2004). These results also show that Diamond architecture motifs are energetically favorable since LAY(1,2,1), a Diamond motif, was energetically favorable and is abundant in neural TRNs in *C. elegans* (Milo et al., 2002). We obtained similar observations for different input types (step, exponential, Gaussian and ramp pulse) and for a wide range of pulse durations (Section 6.10.9).

Thus, we demonstrate that the generalization results obtained using PETGP analysis are invariant with respect to pulse durations and pulse types when perturbations are made around the κ of a motif.

6.9 DISCUSSION

The energetic-cost theory presented here clearly indicates that *SDE* may be a suitable basis for evolutionary selection of one motif over another and could provide an explanation for the rare occurrence of various network motifs. Our analysis indicates that the Pareto-optimality principle, when combined with NESS analysis, leads to energetically efficient solutions for transcription. The underlying energetic-cost criterion, *SDE*, for Pareto-optimal conditions is a measure that reflects maximal transcription at the lowest energetic demand. Beyond its application as a functional basis in TRN motifs, the Pareto-optimal *SDE* concept may also lead to an optimal and energetically efficient design of synthetic gene circuits. Further validation of this concept for protein and metabolic networks is required to confirm its generality; however, the corresponding abundance data for these networks is unavailable. The finding that energetic cost may be used as an underlying basis for evolutionary selection of a motif among motifs having similar dynamic functionality is of major significance. The overwhelming diversity of possible dynamical functions with highly-interactive biological networks limits effective learning from experimental data alone. Network analyses using knowledge of the often ignored energetics may greatly reduce the hypothesis space, enabling identification of new functionalities of dynamically perturbed large-scale networks. Further, the developed dynamic PETGP framework may be used not only for analyzing motifs in complex networks but also for designing complex synthetic networks. Appropriate identifications of cellular objectives involved in evolutionary decision making may provide a potentially novel approach to identify optimal environmental conditions and therefore, as a stand-alone strategy, may provide a more efficacious simultaneous prediction and validation strategy for biological networks.

6.10 METHODS

6.10.1 Pareto Dominance Concept Correlates Dissipative Energetics for Metabolic Networks and TRN Motifs

Pareto Dominance-SDE: One-Cycle Metabolic Network

The parameters and constraints used in this metabolic network are shown in detail in Figure 6.8. The steady state mass balances for metabolites A , B and C are given by the following equations:

$$\frac{d[A]}{dt} = J_A - J_1 - J_3 = 0 \quad (6.1)$$

$$\frac{d[B]}{dt} = J_1 + J_2 - J_B = 0 \quad (6.2)$$

$$\frac{d[C]}{dt} = J_3 - J_2 - J_C = 0 \quad (6.3)$$

where the internal reaction fluxes are $J_1 = k_1[A] - k_{-1}[B]$, $J_2 = k_2[C] - k_{-2}[B]$ and $J_3 = k_3[A] - k_{-3}[C]$. Notice that Equations (6.1) to (6.3) can be written in the form:

$$A_{eq}\vec{x} = b_{eq} \quad (6.4)$$

$$\text{where } A_{eq} = \begin{bmatrix} -(k_1 + k_3) & k_{-1} & k_{-3} & 1 & 0 & 0 \\ k_1 & -(k_{-1} + k_{-2}) & k_2 & 0 & -1 & 0 \\ k_3 & k_{-2} & -(k_{-3} + k_2) & 0 & 0 & -1 \end{bmatrix}$$

and $\vec{x} = [[A] \ [B] \ [C] \ J_A \ J_B \ J_C]^T$ and $b_{eq} = [0 \ 0 \ 0]^T$.

Equation $A_{eq}\vec{x} = b_{eq}$ represents the mass balance (or stoichiometric) constraint for this metabolic system. If $K_i = k_i/k_{-i}$, for $i = 1, 2, 3$, then the chemical potential associated to each reaction is:

$$\Delta\mu_1 = \mathfrak{R}T \ln\left(\frac{[B]}{K_1[A]}\right) \quad (6.5)$$

$$\Delta\mu_2 = \mathfrak{R}T \ln\left(\frac{[B]}{K_2[C]}\right) \quad (6.6)$$

$$\Delta\mu_3 = \mathfrak{R}T \ln\left(\frac{[C]}{K_3[A]}\right) \quad (6.7)$$

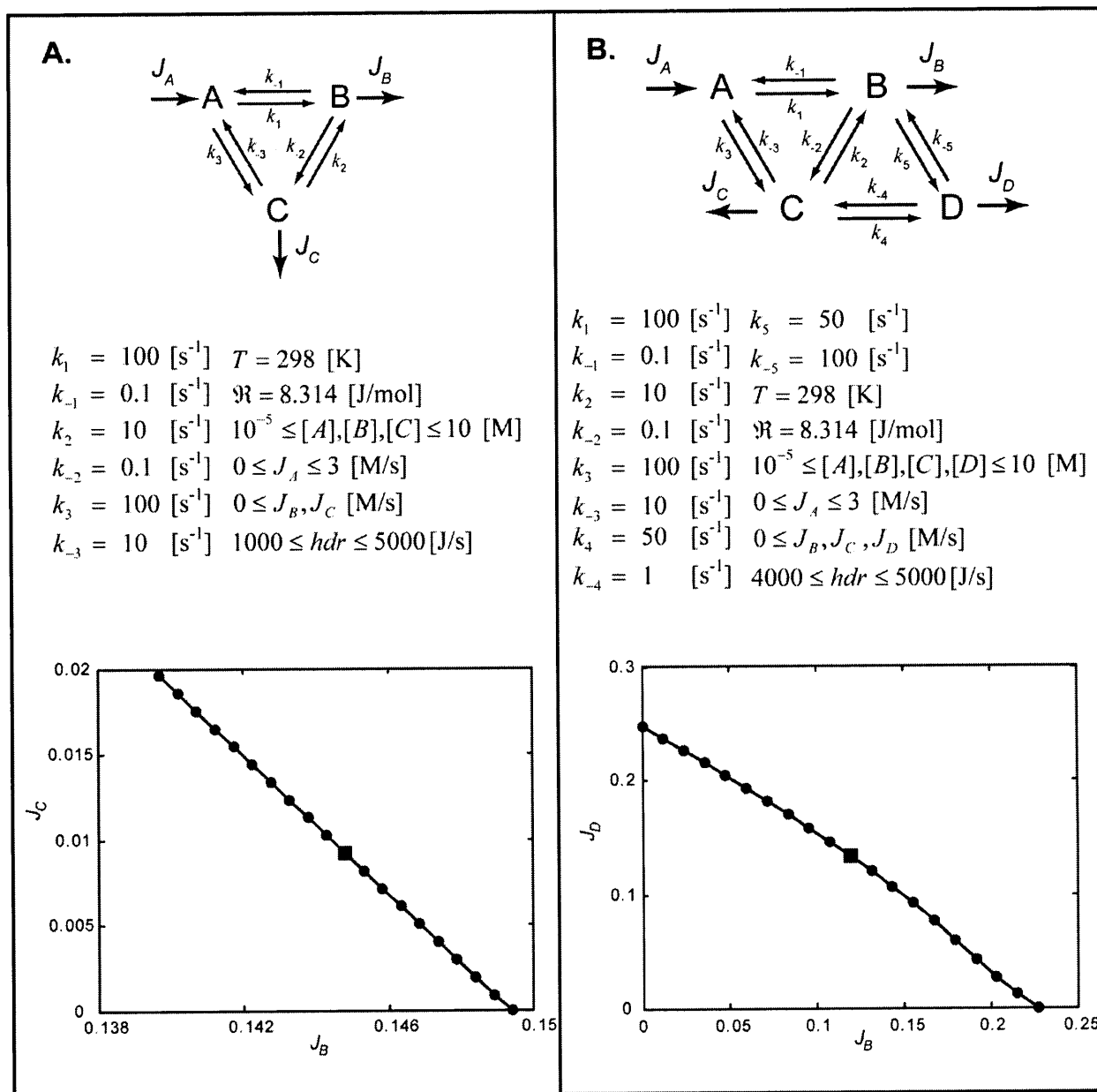


Figure 6.8: Metabolic networks shown in Figure 6.3B. **A)** Three-node metabolic network and Pareto frontier for minimization of J_B and J_C . This curve represents the minimum tradeoff region between these external fluxes and for simplicity it has been reduced into a single point (square). **B)** Four-node metabolic network and Pareto frontier for minimization of J_B , J_C and J_D . This curve represents the minimum tradeoff region between these external fluxes and for simplicity it has been reduced into a single point (square).

For this particular system, in which the net fluxes are assumed to diverge from A , the chemical potentials must satisfy $\Delta\mu_1 - \Delta\mu_2 - \Delta\mu_3 = 0$ which is equivalent to $K_1 = K_2K_3$. From Figure 6.8, $K_1 = 1000$, $K_2 = 100$ and $K_3 = 10$, and then the chemical potential of the cycle is zero. The HDR and the SDE (Section 5.2) for this system can be calculated as follows:

$$HDR = -(J_1 \cdot \Delta\mu_1 + J_2 \cdot \Delta\mu_2 + J_3 \cdot \Delta\mu_3) \quad (6.8)$$

$$SDE = \frac{HDR}{J_A} \quad (6.9)$$

Notice that the term $J_i \cdot \Delta\mu_i$ is always less or equal than zero (Section 5.2). However, because the energy available for the system is finite, the following energetic constraint in the Second Law of the Thermodynamics is imposed:

$$HDR_{lb} \leq HDR \leq HDR_{ub} \quad (6.10)$$

As the production rate of B and C are desired to be maximized, the trade-off between these two fluxes leads to a Pareto frontier. Basically, the optimization problem can be summarized as follows:

$$\text{Maximize}_{\vec{x}} (J_B, J_C)$$

subject to:

$$A_{eq} \vec{x} = \vec{b}_{eq}$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

Once the maximal Pareto frontier is obtained, the minimal Pareto frontier is determined by minimizing both J_B and J_C which also conveys to a Pareto frontier as shown in Figure 6.8A. For simplicity, because this Pareto frontier is very small, it has been represented in Figure 6.3B as a single point, corresponding to the middle point of this optimal curve shown as a square in Figure 6.8A. When the minimal and maximal optimal regions are found, a vector can be traced from the minimal Pareto frontier to any point on the Pareto frontier for maximization (as shown in blue, green and red in Figure 6.3B). As the coordinates of this vector are known in the $J_B J_C$ space, it can be equally divided in n points where point i has the coordinates (J_B^i, J_C^i) . Then, each of these points can be fully characterized by minimizing the input flux J_A in order to maximize the fixed “benefit” (given by the values of J_B^i and J_C^i) at the minimal “cost” (given by J_A). This optimization problem can be described as follows:

for $i = 1 : n$

$$\text{Minimize}_{\vec{x}} (J_A)$$

subject to:

$$A_{eq} \vec{x} = \vec{b}_{eq}$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$J_B = J_B^i, J_C = J_C^i$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

end

As seen in Figure 6.3B, the *SDE* decreases as the vector moves from the minimal to the maximal Pareto frontier demonstrating that the Pareto frontier for maximization is energetically optimal for the system. *It is important to note that at the Pareto frontier (of both one-cycle and two-cycle metabolic networks) there is optimal utilization of nutrient resources for a certain transcriptional rate. SDE being minimal at the optimal production indicates optimal production with efficient energy utilization. This is consistent with recent findings indicating that biological systems operate under optimal environmental conditions with optimal selection and utilization of resources* (Dekel and Alon, 2005) (see Section 6.4).

Pareto Dominance-SDE: Two-Cycle Metabolic Network

Figure 6.8B shows the parameters and constraints used in this metabolic network. The mass balance equation for *A* is given by (6.1). However, for *B*, *C* and *D* are given by the following equations:

$$\frac{d[B]}{dt} = J_1 + J_2 - J_5 - J_B = 0 \quad (6.11)$$

$$\frac{d[C]}{dt} = J_3 - J_2 - J_4 - J_C = 0 \quad (6.12)$$

$$\frac{d[D]}{dt} = J_4 + J_5 - J_D = 0 \quad (6.13)$$

where the internal fluxes for J_1 , J_2 and J_3 are the same as the one cycle network, and $J_4 = k_4[C] - k_{-4}[D]$ and $J_5 = k_5[B] - k_{-5}[D]$. Equations (6.11) to (6.13) can be cast into a matrix form as Equation (6.4) with

$$A_{eq} = \begin{bmatrix} -(k_1 + k_3) & k_{-1} & k_{-3} & 0 & 1 & 0 & 0 & 0 \\ k_1 & -(k_{-1} + k_{-2} + k_5) & k_2 & k_{-5} & 0 & -1 & 0 & 0 \\ k_3 & k_{-2} & -(k_{-3} + k_2 + k_4) & k_{-4} & 0 & 0 & -1 & 0 \\ 0 & k_5 & k_4 & -(k_{-4} + k_{-5}) & 0 & 0 & 0 & -1 \end{bmatrix}$$

and $\vec{x} = [[A] \ [B] \ [C] \ [D] \ J_A \ J_B \ J_C \ J_D]^T$ and $b_{eq} = [0 \ 0 \ 0 \ 0]^T$.

The chemical potential associated to reactions 1, 2 and 3 are given by Equations (6.5) to (6.7). For reactions 4 and 5, it can be expressed as follows:

$$\Delta\mu_4 = \mathfrak{R}T \ln \left(\frac{[D]}{K_4[C]} \right) \quad (6.14)$$

$$\Delta\mu_5 = \mathfrak{R}T \ln \left(\frac{[D]}{K_5[B]} \right) \quad (6.15)$$

In order to have null chemical potential of the cycles, these relationships must satisfy: $\Delta\mu_1 - \Delta\mu_2 - \Delta\mu_3 = 0$ and $\Delta\mu_4 - \Delta\mu_2 - \Delta\mu_5 = 0$. These equations are equivalent to $K_1 = K_2K_3$ and $K_4 = K_2K_5$. From Figure 6.8B, both equalities are verified.

For this metabolic network, the production rate of B , C and D is desired to be maximized and the Pareto surface is obtained. This optimization problem can be summarized as follows:

$$\text{Maximize}_{\vec{x}} (J_B, J_C, J_D)$$

$$\text{subject to: } A_{eq} \vec{x} = \vec{b}_{eq}$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

Once the maximal Pareto frontier is obtained, the minimal Pareto frontier is determined by minimizing J_B , J_C and J_D . For this metabolic system, this region is not a Pareto surface as

expected, but a frontier in the plane $J_B J_D$ for $J_C = 0$ as shown in Figure 6.8B. This Pareto frontier is reduced and it has been represented in Figure 6.3B as a single point, corresponding to the middle point of this optimal curve shown as a square in Figure 6.8B. As in the one-cycle metabolic network, a vector can be traced from the minimal point to any point on the Pareto surface for maximization (as shown in blue, green and red in Figure 6.3B). As the coordinates of this vector are known in the $J_B J_D J_C$ space, it can be equally divided in n points where point i has the coordinates (J_B^i, J_C^i, J_D^i) . Then, each of these points can be fully characterized by minimizing the input flux J_A in order to maximize the fixed “benefit” (given by the values of J_B^i , J_C^i and J_D^i) at the minimal “cost” (given by J_A). This optimization problem can be described as follows:

$$\text{Minimize}_{\vec{x}} (J_A) \text{ (for } i = 1 : n)$$

subject to:

$$A_{eq} \vec{x} = \vec{b}_{eq}$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$J_B = J_B^i, J_C = J_C^i, J_D = J_D^i$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

Pareto Dominance-SDE: FFL TRN Motif

The FFL TRN corresponds to a C1-FFL with an *OR* input logic. Chapter 5 describes the detailed TRN model, mass balance and thermodynamics. Essentially, the FFL network is composed of six cycles as shown in Figure 6.9A: two of them are involved in the transcription of Y by X , and four in the transcription of Z by both X and Y . The first two cycles correspond to cycle YA (transcription of Y from its basal activator A) and cycle YX (transcription of Y from the activator X), and the last four cycles are cycle ZB (transcription of Z from its basal activator B), cycle ZX (transcription of Z by the activator X), cycle ZY (transcription of Z by activator Y) and cycle ZXY (transcription of Z by both X and Y together). The kinetic constants used in Figure 6.3B are shown in Figure 6.9A.

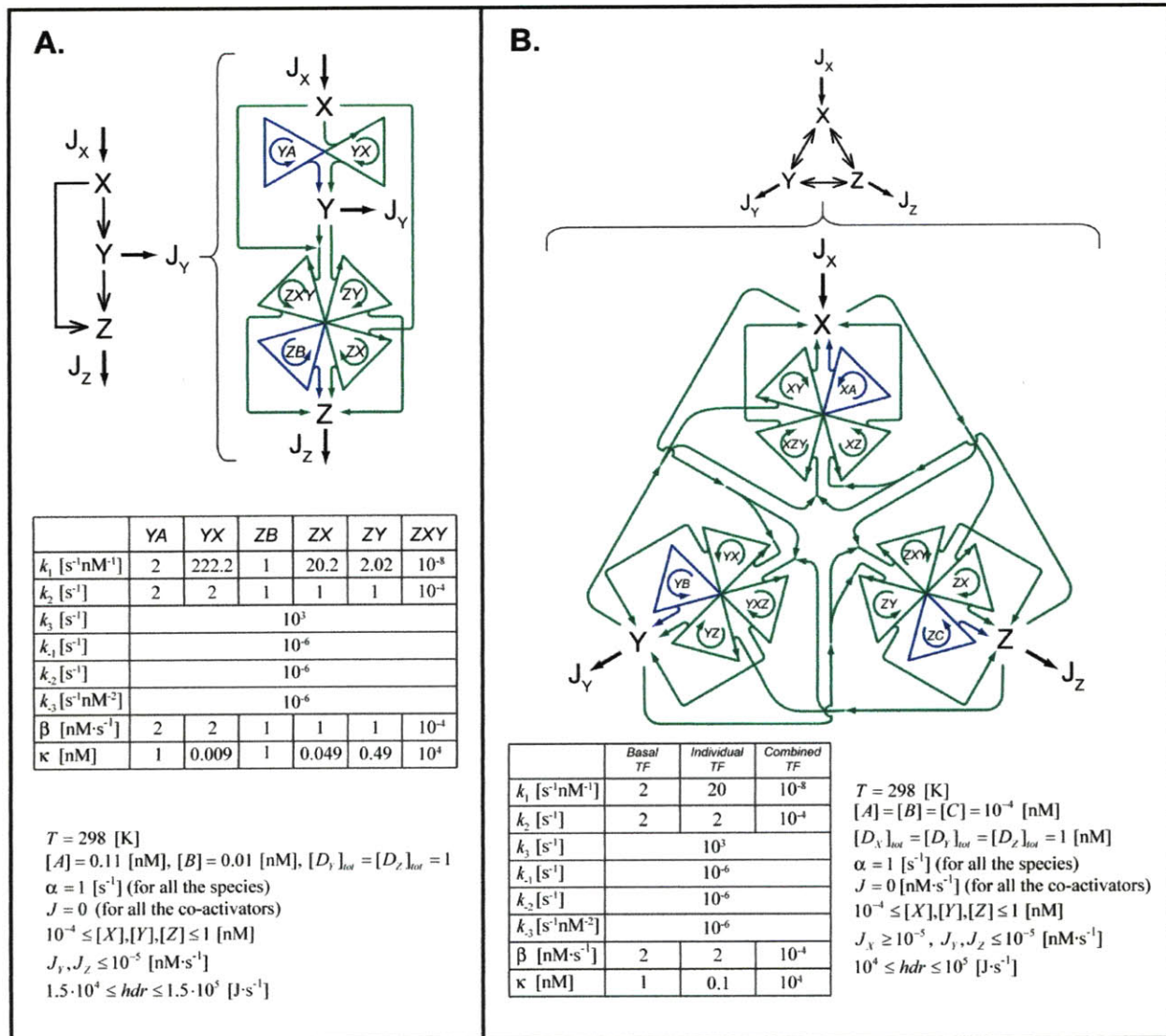


Figure 6.9: Transcriptional Networks shown in Figure 6.3B. **A)** The Coherent Type-1 FFL with OR input logic. Parameters and optimization constraints are shown. **B)** The three-node transcription network with complete feedback and OR input logic. Parameters and optimization constraints are shown. Basal TF refers to the basal activator (blue cycle), individual TF refers to only one transcription factor (for instance, X or Y in the transcription of Z), and combined TF refers to the joint interaction between the individual activators (for instance, XY in the transcription of Z).

The basal cycles are colored blue and the other cycles are green, meaning that there is active transcription from all of them. In fact, since this network obeys an OR input logic for transcription of Z (notice that transcription of Y does not need any input logic, because only one transcription factor is present), transcription of Z can be seen from all of the six cycles. Observe that the basal kinetic parameters and thus the basal maximal transcription rates and activation coefficients were selected in order to have $J_{T,Basal}^Y = 0.1 J_{T,Max}^Y$ and $J_{T,Basal}^Z = 0.001 J_{T,Max}^Z$. In

addition, the kinetic parameters of the other cycles were found in order to get the following activation coefficients: $K^{YX} = 0.01$, $K^{ZX} = 0.05$, $K^{ZY} = 0.5$, $K^{ZXY} = 10^4$ [nM]. The first three values are based on (Mangan and Alon, 2003). In this analysis, the combined contribution of XY is reduced in order to observe the individual effect of X and Y . By making β^{ZXY} low and K^{ZXY} high, the transcription rate from the ZXY cycle is negligible. Notice as well that the maximal transcription rate of Y ($\beta^{YX} = 2$) is twice the maximal transcription rate of Z ($\beta^{ZX} = 1$ and $\beta^{ZY} = 1$). This was done in order to have concentrations of Y of the same order of magnitude of Z since Y is being consumed when transcribes Z . The maximal consumption rate of Y corresponds to the maximal transcription rate of that cycle ($\beta^{ZY} = 1$), and then the net transcription rate of Y when fully transcribes Z is 1 [nM·s⁻¹].

As both external fluxes of Y and Z are desired to be maximized when X is externally provided, there is a tradeoff region between these two objective functions: when more Y is taken out of the system (higher J_Y), the concentration of Y goes down and its contribution in the further transcription of Z decreases, leading to lower J_Z , and vice versa. This Pareto frontier is shown in Figure 6.3B and the optimization problem can be expressed as follows:

$$\text{Maximize}_{\vec{x}} (J_Y, J_Z)$$

subject to:

$$f(\vec{x}) = 0$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

where $f(x)$ represents the nonlinear mass balance equations at steady state and $x = [[X] J_Y J_Z]^T$. Recall that both J_Y and J_Z are negative since these fluxes are leaving the system. *It is important to note that at the Pareto frontier (in both FFL and Feedback TRNs) there is optimal utilization of nutrient resources for a certain metabolic fluxes. SDE being minimal at the optimal production indicates optimal production with efficient energy utilization.* This is in concurrence with the recent laboratory evolution experiments where evolutionary adaptation of *E. coli* towards

optimal metabolic fluxes has been shown and the key role of fitness criteria in evolutionary selection of certain mutations (Ibarra et al., 2002).

The minimal Pareto frontier can be obtained by simultaneously minimizing J_Y and J_Z . For this transcription network, this region is confined to a single point characterized by the vector $x = [10^{-4}, -10^{-5}, -10^{-5}]^T$. A vector in the space $J_Y J_Z$ can be traced from the minimal point to any point on the Pareto frontier for maximization (as shown in blue, green and red in Figure 6.3B). As the coordinates of this vector are known, it can be equally divided in n points where point i has the coordinates (J_Y^i, J_Z^i) . Then, each of these points can be fully determined by minimizing the input flux J_X in order to maximize the fixed “benefit” (given by the values of J_Y^i and J_Z^i) at the minimal “cost” (given by J_X). This optimization problem can be described as follows:

$$\text{Minimize}_{\vec{x}} (J_X) \text{ (for } i = 1 : n)$$

subject to:

$$f(\vec{x}) = 0$$

$$HDR_{lb} \leq HDR \leq HDR_{ub}$$

$$J_Y = J_Y^i, J_Z = J_Z^i$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

Pareto Dominance-SDE: Feedback TRN loop subgraph

The Feedback TRN loop subgraph corresponds to a three-node system with complete feedback and an *OR* input logic as shown in Figure 6.9B. This network is composed of twelve cycles: four of them are involved in the transcription of X by both Y and Z , another four cycles in the transcription of Y by X and Z , and four in the transcription of Z by both X and Y . The first four cycles correspond to cycle XA (transcription of X from its basal activator A), cycle XY (transcription of X from the activator Y), cycle XZ (transcription of X by Z), and cycle XYZ (transcription of X by both Y and Z). The next four cycles involve cycle YB (transcription of Y from its basal activator B), cycle YX (transcription of Y from the activator X), cycle YZ (transcription of Y by Z), and cycle YXZ (transcription of Y by both X and Z). The last four cycles

are cycle ZC (transcription of Z from its basal activator C), cycle ZX (transcription of Z by the activator X), cycle ZY (transcription of Z by activator Y) and cycle ZXY (transcription of Z by both X and Y together). As being depicted throughout this work, the basal cycles are colored blue and the other cycles are green, meaning that there is active transcription from all of them. Transcription of Z occurs from all of the nine cycles because the OR input logic.

Notice that the basal kinetic parameters and thus the basal maximal transcription rates and activation coefficients were selected in order to have $J_{T,Basal}^I = 10^{-4} \cdot J_{T,Max}^I$, for $I = X, Y, Z$. In addition, the kinetic parameters of the other cycles were found in order to get the following activation coefficients: $K^{individual} = 0.1$ and $K^{combined} = 10^4$ [nM]. $K^{individual}$ refers to the individual action of a single transcription factor (cycles XY, XZ, YX, YZ, ZX and ZY) and $K^{combined}$ refers to the combined action of two transcription factors (cycles XYZ, YXZ , and ZXY). The first activation coefficient value is based on (Mangan and Alon, 2003). As in the FFL TRN, $\beta^{combined}$ and $K^{combined}$ are low and high, respectively and $\beta^{individual} = 2$.

For this TRN, the minimal Pareto frontier is confined to a single point characterized by the vector. $x = [1.961 \cdot 10^{-4}, -10^{-5}, -10^{-5}]^T$.

6.10.2 Pareto Optimal SDE for FFL TRN Motifs

Figure 6.4A shows the Pareto averaged SDE for all the FFLs with AND and OR input logics, as depicted in Figure 6.10. FFLs can be divided into Coherent-Type 1 to 4 (C1-FFL to C4-FFL) and Incoherent-Type 1 to 4 (I1-FFL to I4-FFL). Independently of the Boolean input logic or the Coherent or Incoherent structure, these networks are composed of six cycles as shown in Figure 6.10: two of them involved in the transcription of Y by X , and four in the transcription of Z by both X and Y . The first two cycles correspond to cycle YA (transcription of Y from its basal activator A) and cycle YX (transcription of Y from the activator X), and the last four cycles are cycle ZB (transcription of Z from its basal activator B), cycle ZX (transcription of Z by the activator X), cycle ZY (transcription of Z by activator Y) and cycle ZXY (transcription of Z by both X and Y together). In Supplementary Fig. 10, the basal cycles (YA and ZB) are shown in blue, and the other cycles are either green if the transcription factor behaves as an activator, or red if the transcription factor acts as a repressor. In the latter case, no transcription occurs and then there is no arrow leaving the cycle.

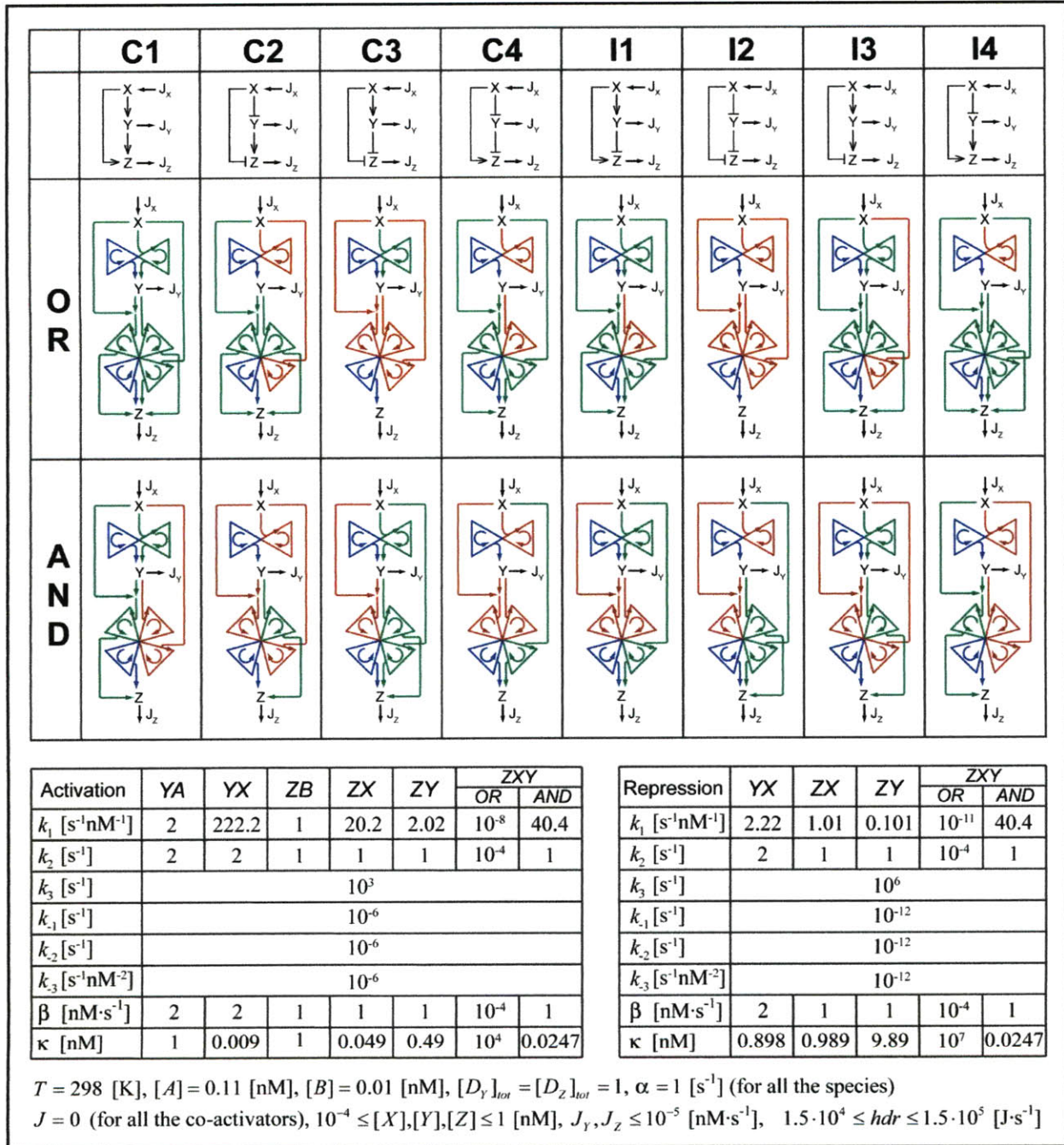


Figure 6.10: Cyclic representation of the Feed Forward Loops (FFL) with input logics *AND* and *OR*. FFL Coherent Type 1 to 4 and Incoherent Type 1 to 4 are shown. Kinetic parameters for activation and repression, as well as the optimization parameters are shown here.

Notice that the basal kinetic parameters and thus the basal maximal transcription rates and activation coefficients were selected in order to have $J_{T,Basal}^Y = 0.1 \cdot J_{T,Max}^Y$,

$J_{T,Basal}^Z = 0.001 \cdot J_{T,Max}^Z$. In addition, the kinetic parameters of the other cycles were found in order to get the following activation coefficients:

For activation:

$K^{YX} = 0.01$, $K^{ZX} = 0.05$, $K^{ZY} = 0.5$, $K^{ZXY} = 10^4$ (*OR* input logic) and $K^{ZXY} = 0.025$ (*AND* input logic) [nM].

For repression:

$K^{YX} = 1$, $K^{ZX} = 1$, $K^{ZY} = 10$, $K^{ZXY} = 10^7$ (*OR* input logic) and $K^{ZXY} = 0.025$ (*AND* input logic) [nM].

For both Boolean input logics, values of K^{YX} , K^{ZX} , and K^{ZY} are based on (Mangan and Alon, 2003). For this analysis, the combined contribution of *XY* for *OR* input logic is desired to be minimal in order to observe the individual effect of *X* and *Y*. By making β^{ZXY} low and K^{ZXY} high, the transcription rate from the *ZXY* cycle is negligible. On the other hand, when the input logic is *AND*, the contribution from the *ZXY* cycle has to be as important as the one from the individual transcription factors. In this case, $\beta^{ZXY} = 1$ is same as $\beta^{ZX} = 1$ and $\beta^{ZY} = 1$ and for the activation case $K^{ZXY} = K^{ZX}K^{ZY}$, as previously suggested (Mangan and Alon, 2003). Observe that the maximal transcription rate of *Y* ($\beta^{YX} = 2$) is twice the maximal transcription rate of *Z*. This was done in order to have concentrations of *Y* of the same order of magnitude of *Z* since *Y* is being consumed when transcribes *Z*. The maximal consumption rate of *Y* corresponds to the maximal transcription rate of that cycle ($\beta^{ZY} = 1$), and then the net transcription rate of *Y* when fully transcribes *Z* is 1 [nM·s⁻¹].

As both external fluxes of *Y* and *Z* are desired to be maximized when *X* is externally provided, there is a tradeoff region between these two objective functions. The first step in generating Figure 6.4A is to get these Pareto frontiers without constraining the total *HDR* (infinite energetic source) In other words, first we need to analyze the feasible space by satisfying the mass balance constraints. The energy-unconstrained Pareto frontiers are shown in Figure 6.11A-B and the optimization problem can be stated as:

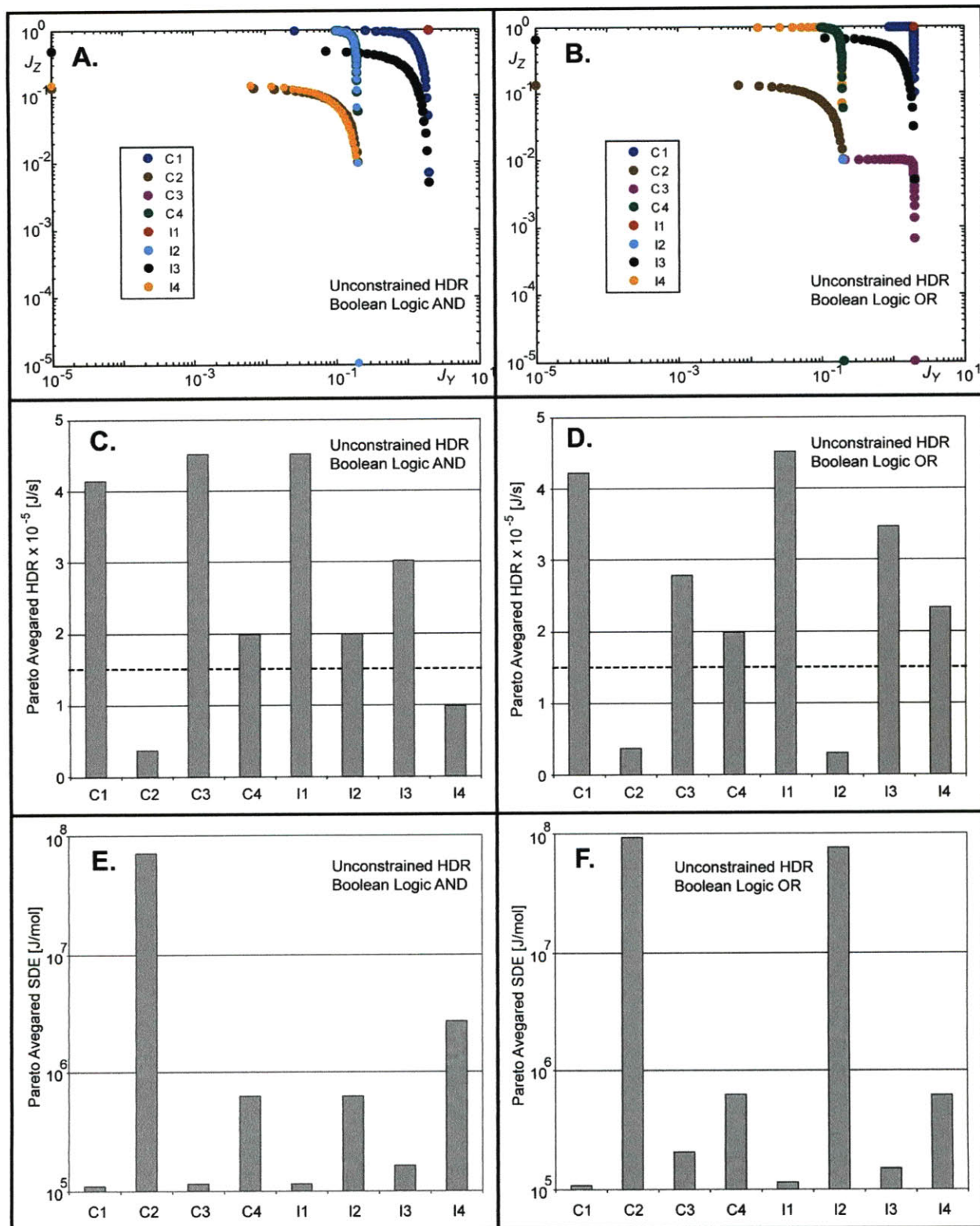


Figure 6.11: A-B) HDR-unconstrained Pareto frontiers for the 8 FFLs with AND and OR input logics. C-D) HDR averaged along the Pareto frontiers for both AND and OR input logics. E-F) SDE averaged along the Pareto frontiers for both AND and OR input logics.

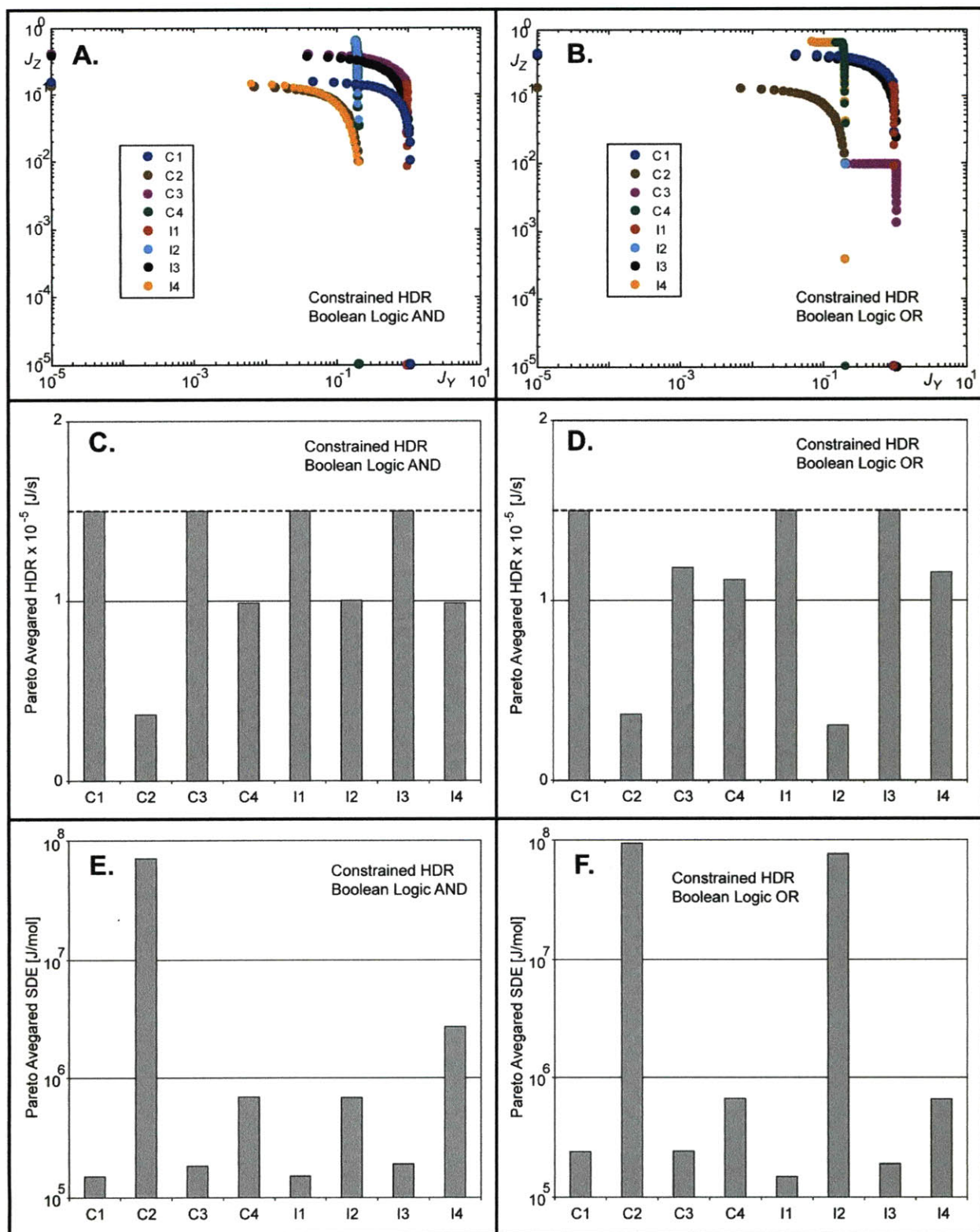


Figure 6.12: A-B) HDR-constrained Pareto frontiers for the 8 FFLs with AND and OR input logics. C-D) HDR averaged along the Pareto frontiers for both AND and OR input logics. E-F) SDE averaged along the Pareto frontiers for both AND and OR input logics.

$$\text{Maximize}_{\vec{x}} (J_Y, J_Z)$$

subject to:

$$f(\vec{x}) = 0$$

$$\vec{x}_{lb} \leq \vec{x} \leq \vec{x}_{ub}$$

where $f(x)$ represents the nonlinear mass balance equations at steady and $x = [[X] J_Y J_Z]^T$. Recall that both J_Y and J_Z are negative since these fluxes are leaving the system. Figure 6.11C-D shows the Pareto averaged *HDR* when the energy of the system is unconstrained and Figure 6.11E-F shows the Pareto averaged *SDE* for both input logics. Notice that even in the energy-unconstrained problem, C1-FFL and I1-FFL present with the lowest values of *SDE* for both input logics. C2-FFL and I2-FFL have the highest *SDE* in the *OR* input logic and C2-FFL and I4-FFL have the highest *SDE* in the *AND* input logic. It is important to note that *similar SDE-abundance correlations were obtained when the kinetics parameters were simultaneously increased or decreased by 3 orders of magnitude.*

The next step is to constrain the energy based on the energy-unconstrained problem. The lower and upper limits for *HDR* are shown in Figure 6.10. These Pareto frontiers are shown in Figure 6.12A-B, and correspond to those reported in Figure 6.4A. Figure 6.12C-D shows the Pareto averaged *HDR* when the energy of the system is now constrained and Figure 6.12E-F shows the Pareto averaged *SDE* for both input logics. Once again, *similar SDE-abundance correlations were obtained for changes in 3 orders of magnitude of the kinetics parameters.*

Finally, all the Pareto frontiers were obtained by the Normalized Constraint Method (Section 3.3.4) (Nagrath et al., 2007).

6.10.3 *SDE* Correlation With Frequency of Occurrence of TRN Motif

Figure 6.4B compares the results obtained in Figure 6.4A with the abundance data for *Escherichia coli* and *Saccharomyces cerevisiae*. From (Ma et al., 2004; Mangan and Alon, 2003). This data is summarized in Table 6.1. A correlation of averaged Pareto-optimal *SDE* with

relative abundance of FFL motifs for *E. coli* and *S. cerevisiae* TRNs from (Ma et al., 2004) is shown in Figure 6.13.

Table 6.1: Data used in Figure 6.4B.

FFL	Relative Abundance <i>S. cerevisiae</i> (Mangan & Alon, 2003)	Relative Abundance <i>E. coli</i> (Mangan & Alon, 2003)	Relative Abundance <i>E. coli</i> (Ma et al, 2004)	Pareto Averaged SDE Boolean logic OR HDR Constrained	Pareto Averaged SDE Boolean logic AND HDR Constrained
C1	0.464	0.371	0.550	2.375E+05	1.496E+05
C2	0.089	0.040	0.058	9.366E+07	7.066E+07
C3	0.000	0.109	0.015	2.389E+05	1.825E+05
C4	0.000	0.074	0.062	6.656E+05	6.922E+05
I1	0.375	0.287	0.247	1.481E+05	1.503E+05
I2	0.054	0.025	0.019	7.608E+07	6.802E+05
I3	0.018	0.050	0.017	1.886E+05	1.894E+05
I4	0.000	0.045	0.033	6.590E+05	2.696E+06

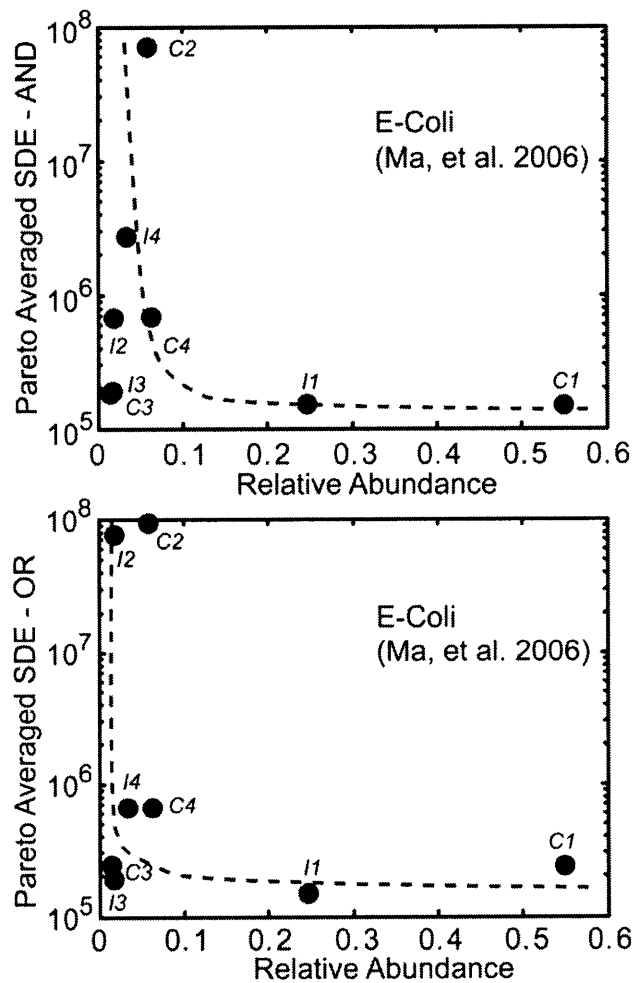


Figure 6.13: Correlation of averaged Pareto-optimal SDE with relative abundance of FFL motifs for *E. coli* and *S. cerevisiae* TRNs (Ma et al., 2004).

6.10.4 Correlating *SDE* Function with Dynamical Functions of Network Motifs

Correlation between dynamical function and energetic cost of FFLs during selection of motifs in evolution.

The dynamic responses of two pairs of FFLs with *AND* input logic (I1-FFL – I4-FFL and C1-FFL– C4-FFL,) are compared in Figure 6.5A for an ON-OFF step change in the concentration of X from 10^{-4} to 1 [nM]. At times $t < 0$, the system is allowed to reach the initial steady state (ss_i) for $[X] = 10^{-4}$ [nM]. At $t = 0$ the ON step change in the concentration of X occurs and is maintained at $[X] = 1$ [nM] for 10 [s]. This time is enough for the system to reach a new steady state (ss_f). At $t \geq 10$ [s], the OFF step change is induced and the concentration of X is again brought to 10^{-4} [nM] and kept at this value for another 10 [s]. After this, the system attains the initial steady state (ss_i). Notice that in these FFLs, the transcription factor X always activates Z in the direct transcriptional pathway. In the indirect pathway (X activating or repressing Z by transcribing the intermediate Y first), different combinations between activation and repression are considered. The dynamic response of FFLs is compared with the dynamic response of simple regulation (SR) $X \rightarrow Z$. The parameters used in Figure 6.5A for the FFLs are the same as used in Figure 6.2A and 6.3B, and are shown in Figure 6.10. The SR case is formed by only two cycles: cycle ZB and ZX . The parameters used for these cycles are the same for their corresponding cycles in the FFLs. In addition, notice that the lower and upper limits in the step change in the concentration of X correspond to the lower and upper constraints in the Pareto frontiers used in Figure 6.4A. In all of the responses, the concentrations were normalized from 0 to 1 as shown in Equation (6.16) where C represents any concentration, and $[C]_{ss_i}$ and $[C]_{ss_f}$ are the concentrations of C at ss_i and ss_f , respectively.

$$[C]_{normalized} = \frac{[C] - [C]_{ss_i}}{[C]_{ss_f} - [C]_{ss_i}} \quad (6.16)$$

Dynamic response of I1-FFL and I4-FFL with *AND* input logic

As theoretically explained (Mangan and Alon, 2003) and experimentally demonstrated in the galactose system of *E.coli* (Mangan and Alon, 2003; Mangan et al., 2006; Mangan et al., 2003; Zaslaver et al., 2006; Zaslaver et al., 2004), I1-FFL with *AND* input logic generates a pulse and speeds up the response time only in the ON phase. I4-FFL with *AND* input logic also

accelerates the response in this sign sensitive manner and generates a pulse response (Mangan and Alon, 2003). Both dynamical behaviors are predicted by the nonequilibrium transcriptional model presented in this work, as shown in Figure 6.14: during the ON response, I1-FFL and I4-FFL speed up the response by generating a pulse that overshoots the final steady state (ss_f) when compared to the SR, but during the OFF response, I1-FFL, I4-FFL, and SR have similar trajectory. Here, we quantify the accelerated response as the time required to reach 50% of the final steady state during the ON response ($t_{1/2}$) as seen in Figure 6.14. For I1-FFL and I4-FFL, these times are $t_{1/2}^{I1-FFL} = 0.24$ and $t_{1/2}^{I4-FFL} = 0.32$ [s], respectively. Observe that the desired dynamic functionality (accelerated ON response compared to SR) is present in both I1-FFL and I4-FFL, but is only little better for I1-FFL. Our kinetic TRN model allowed us to calculate the dissipated energy required to fulfill the dynamic requirement during the entire response time (t_R) or time required for the system to return to the initial steady state within a 1% of error ($t_R^{I1-FFL} = 16.7$ and $t_R^{I4-FFL} = 15.6$ [s]). Figure 6.14 shows the dynamic responses of SDE for I1-FFL and I4-FFL. When SDE is time averaged during the response time, we can find the total averaged specific dissipated energy $\langle SDE \rangle$ for both systems: $\langle SDE \rangle^{I1-FFL} = 2.53 \cdot 10^6$ and $\langle SDE \rangle^{I4-FFL} = 3.67 \cdot 10^7$ [J/mol]. *Importantly, the mapping between both dynamical functionalities is robust to variations in the kinetic parameters. In fact, similar dynamical and SDE responses were obtained when the kinetics parameters were simultaneously increased or decreased by 2 orders of magnitude*

Dynamic response of C1-FFL and C4-FFL with AND input logic

As theoretically explained (Mangan and Alon, 2003) and experimentally demonstrated in the arabinose system of *E.coli* (Mangan et al., 2003), the C1-FFL with AND logic presents a sign sensitive delay for ON steps of the external signal X compared to the SR. Similarly, C4-FFL with AND logic exhibits this delay during the ON response but not during the OFF step change (Mangan and Alon, 2003). Both dynamical behaviors are predicted by our kinetic model of TRNs, as shown in Figure 6.15: during the ON response, C1-FFL and FFL get initially delayed compared to the simple regulation, but during the OFF response, C1-FFL, C4-FFL, and SR follow the same trajectory. For C1-FFL and C4-FFL, the delay times are $t_{1/2}^{C1-FFL} = 1.15$ and $t_{1/2}^{C4-FFL} = 0.83$ [s], respectively.

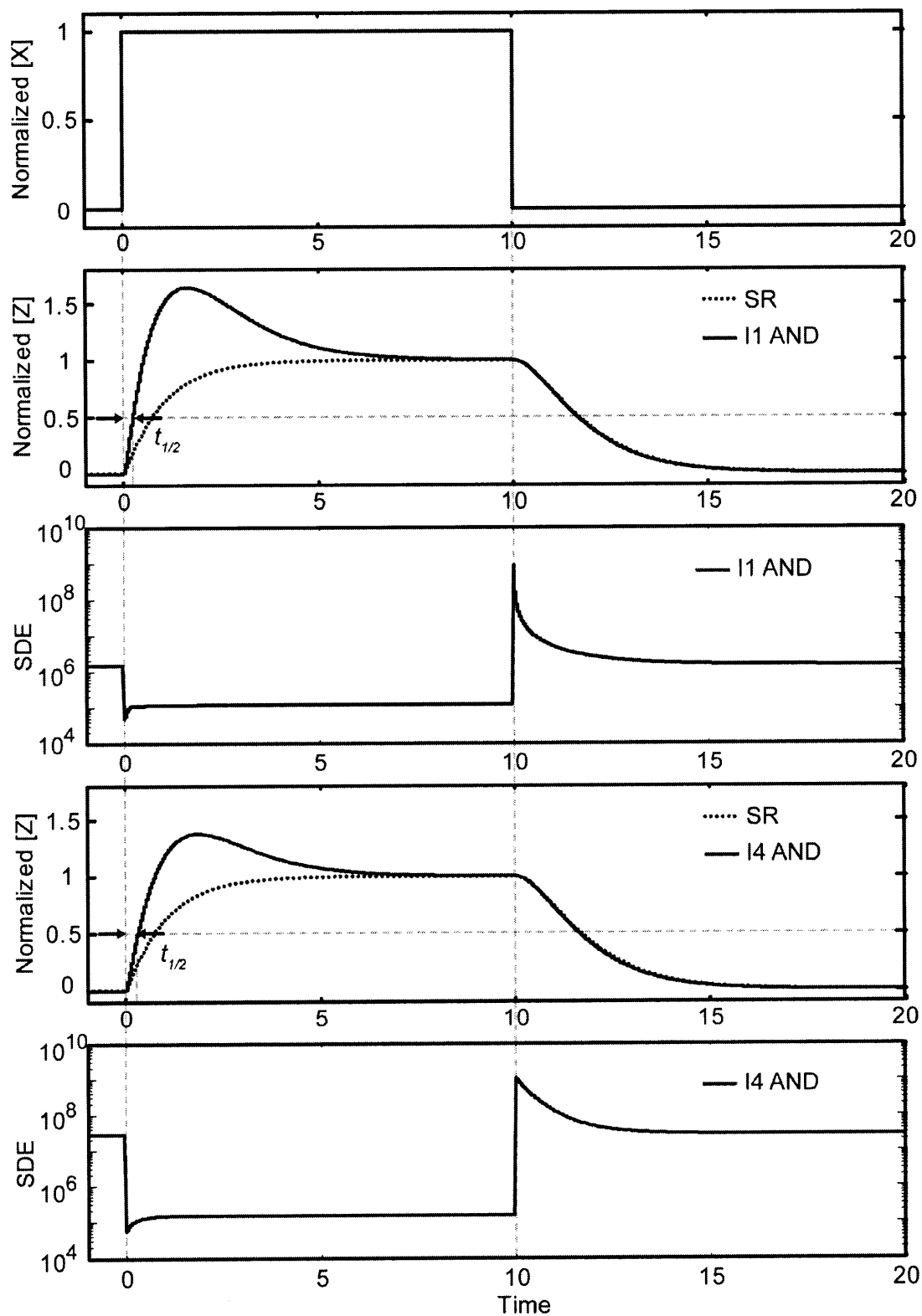


Figure 6.14: Dynamic responses of $[Z]$ and SDE for I1-FFL and I4-FFL with *AND* input logic to an ON-OFF step change of $[X]$. The accelerated response time $t_{1/2}$ is the time to reach half of the ON response in the concentration of Z .

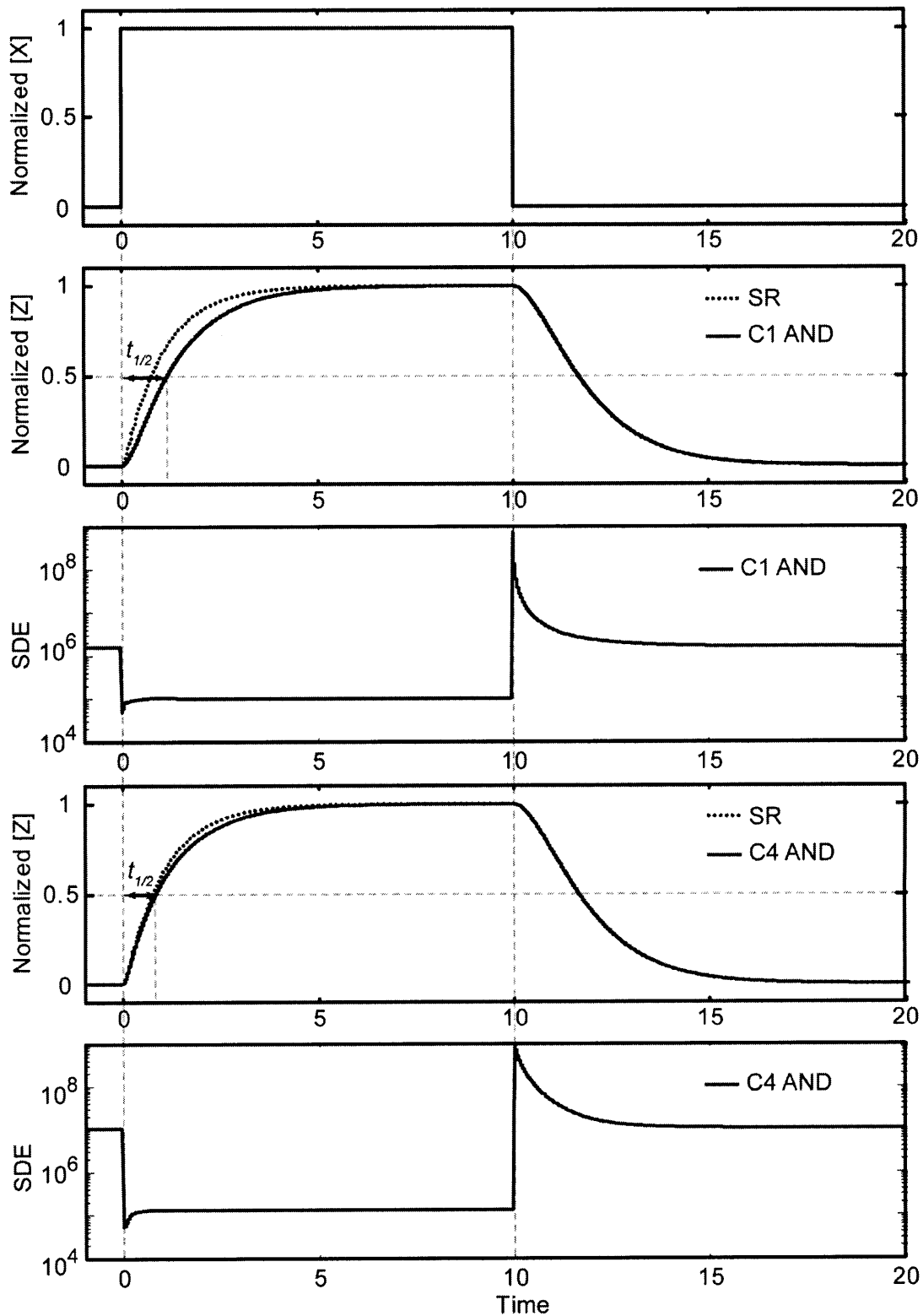


Figure 6.15: Dynamic responses of $[Z]$ and SDE for C1-FFL and C4-FFL with *AND* input logic to an ON-OFF step change of $[X]$. The delayed time $t_{1/2}$ is the time to reach half of the ON response in the concentration of Z .

Observe that the desired “functional dynamic” (delayed ON response compared to SR) is present in both C1 AND and C4 AND, but is better for C1 AND. SDE averaging through the response time ($t_R^{C1-AND} = 16.3$ and $t_{1/2}^{I4-AND} = 15.4$ [s]) shows $\langle SDE \rangle^{C1-FFL} = 2.25 \cdot 10^6$ and $\langle SDE \rangle^{C4-FFL} = 1.67 \cdot 10^7$ [J/mol]. Once again, *the mapping between both dynamical functionalities is robust to variations in the kinetic parameters (when simultaneously increased or decreased by 2 orders of magnitude).*

6.10.5 Dynamical Characterization of Higher Order FFLs Using Energetic Cost

The dynamic responses of three generalizations of the FFLs with OR input logic (double input or 2X-FFL, double Y or 2Y-FFL and double output or 2Z-FFL) are compared in Figure 6.5B for an ON-OFF exponential change in the concentration of X from 10^{-1} to 1 [nM]:

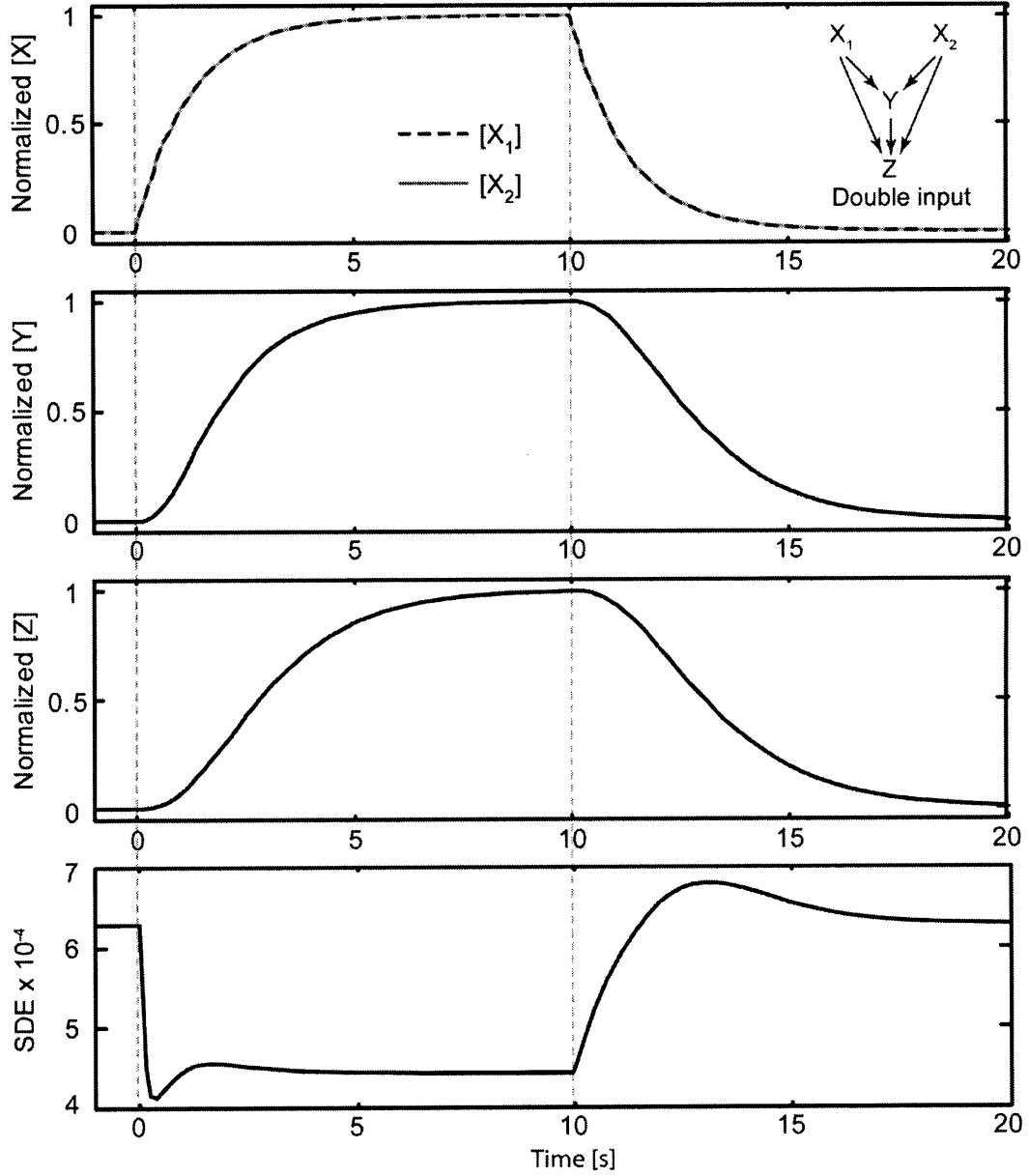
$$[X] = \begin{cases} 0.1 & , t < 0 \\ 0.1 + (1 - 0.1)(1 - e^{-0.8t}) & , 0 \leq t < 10 \\ 0.1 + (1 - 0.1)e^{-0.8(t-10)} & , t \geq 10 \end{cases} \quad (6.17)$$

In all the responses, the concentrations were normalized from 0 to 1 by using Equation (6.16).

Dynamic response of double input (2X-FFL)

Essentially, the 2X-FFL is composed of twelve cycles: four of them are involved in the transcription of Y by X_1 and X_2 , and eight in the transcription of Z by X_1 , X_2 and Y . The first four cycles correspond to cycle YA (transcription of Y from its basal activator A), cycle YX_1 (transcription of Y by X_1), cycle YX_2 (transcription of Y by X_2), and cycle YX_1X_2 (transcription of Y by the joint action of X_1 and X_2). The other eight cycles are cycle ZB (transcription of Z from its basal activator B), cycle ZX_1 (transcription of Z by X_1), cycle ZX_2 (transcription of Z by X_2), cycle ZY (transcription of Z by activator Y), ZX_1X_2 (transcription of Z by both X_1 and X_2), ZX_1Y (transcription of Z by both X_1 and Y), ZX_2Y (transcription of Z by X_2 and Y), and cycle ZX_1X_2Y (transcription of Z by the joint actions of X_1 , X_2 and Y). Notice that the basal kinetic parameters shown in Supplementary Fig. 16 were selected in order to have minimal basal transcription rates:

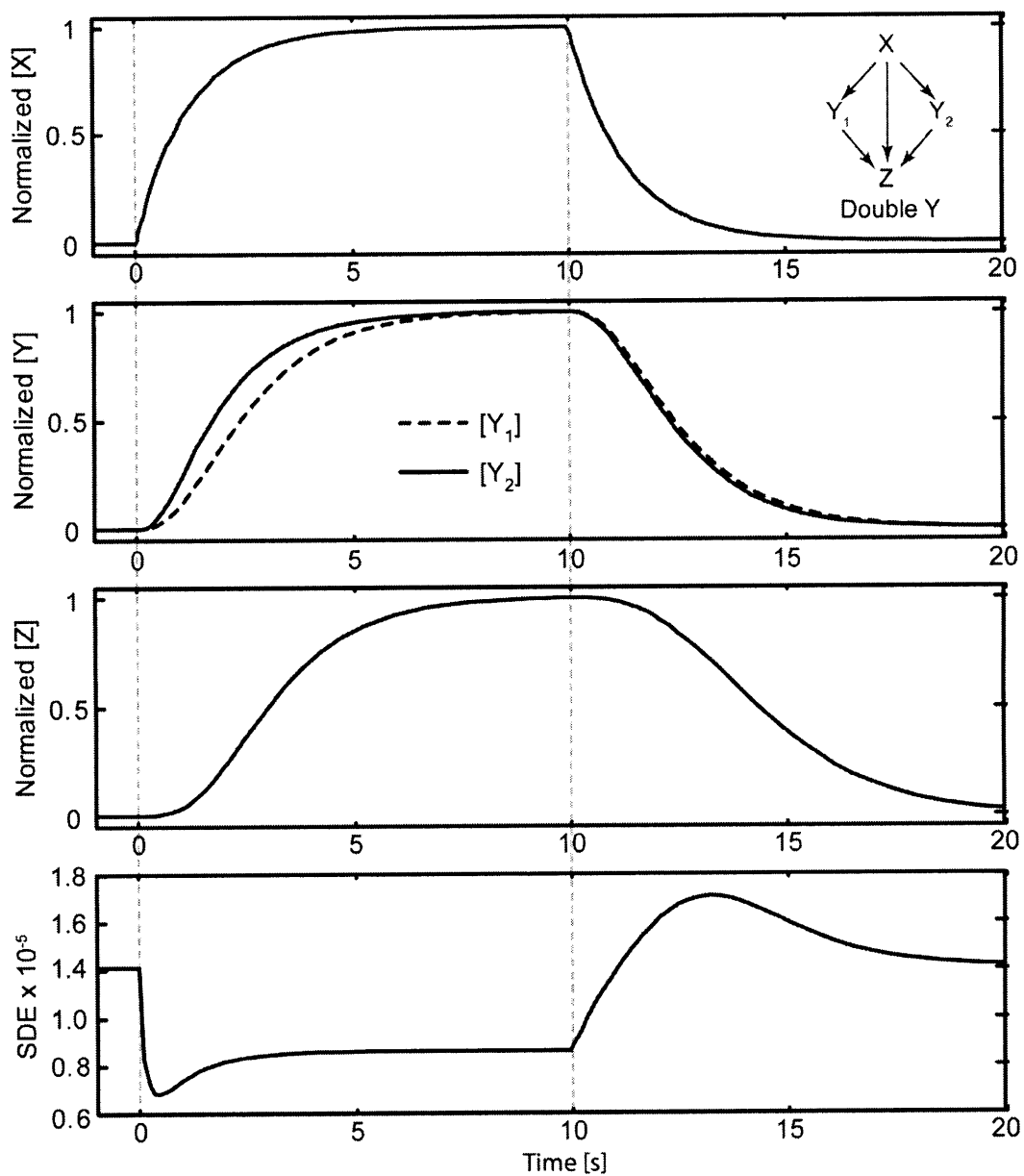
$$J_{T,Basal}^Y = J_{T,Basal}^Z = 10^{-6} \text{ [nM}\cdot\text{s}^{-1}\text{]}.$$



	YA	YX ₁	YX ₂	YX ₁ X ₂	ZB	ZX ₁	ZX ₂	ZY	ZX ₁ X ₂	ZX ₁ Y	ZX ₂ Y	ZX ₁ X ₂ Y
k_1 [s ⁻¹ nM ⁻¹]	1	0.02	2	10 ⁻⁸	1	0.1	0.01	0.01	10 ⁻⁸	10 ⁻⁸	10 ⁻⁸	10 ⁻⁸
k_2 [s ⁻¹]	1	2	2	10 ⁻⁴	1	1	1	1	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
k_3 [s ⁻¹]	10 ³											
k_{-1} [s ⁻¹]	10 ⁻⁶											
k_{-2} [s ⁻¹]	10 ⁻⁶											
k_{-3} [s ⁻¹ nM ⁻²]	10 ⁻⁶											
β [nM·s ⁻¹]	1	2	2	10 ⁻⁴	1	1	1	1	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
κ [nM]	1	100	1	10 ⁴	1	10	100	100	10 ⁴	10 ⁴	10 ⁴	10 ⁴

$T = 298$ [K], $[A] = [B] = 10^{-6}$ [nM], $[D_Y]_{tot} = [D_Z]_{tot} = 1$ [nM], $\alpha = 1$ [s⁻¹], $J = 0$ (for all the species)

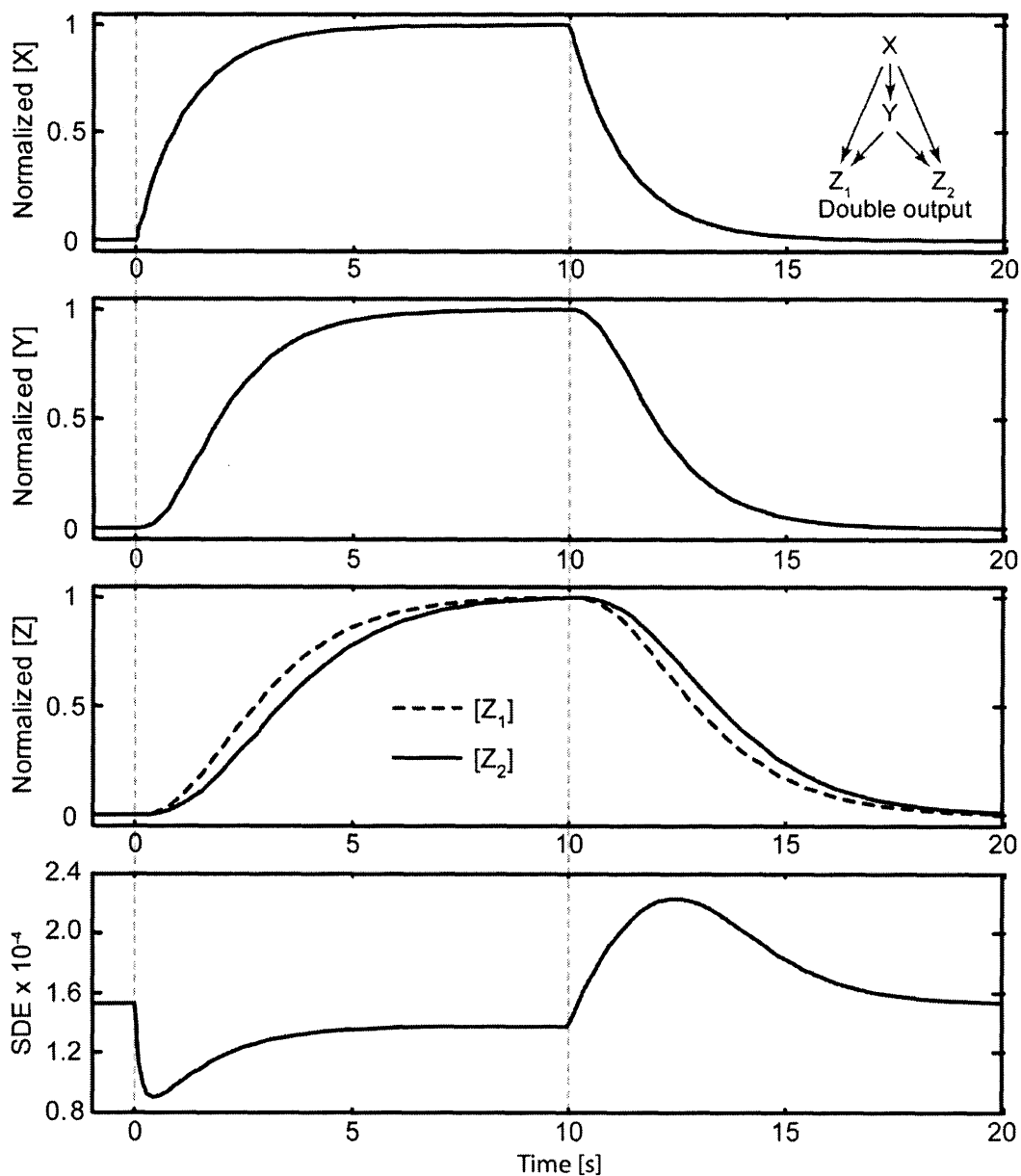
Figure 6.16: Dynamic responses and kinetic parameters of the double input (2X-FFL) generalization of FFL.



	Y_1A_1	Y_1X	Y_2A_2	Y_2X	ZB	ZX	ZY ₁	ZY ₂	ZXY ₁	ZXY ₂	ZY ₁ Y ₂	ZXY ₁ Y ₂
k_1 [s ⁻¹ nM ⁻¹]	1	0.02	1	2	1	0.1	0.01	1	10 ⁻⁸	10 ⁻⁸	10 ⁻⁸	10 ⁻⁸
k_2 [s ⁻¹]	1	2	1	2	1	1	1	1	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
k_3 [s ⁻¹]	10 ³											
k_4 [s ⁻¹]	10 ⁻⁶											
k_5 [s ⁻¹]	10 ⁻⁶											
k_6 [s ⁻¹ nM ⁻²]	10 ⁻⁶											
β [nM·s ⁻¹]	1	2	1	2	1	1	1	1	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
κ [nM]	1	100	1	1	1	10	100	1	10 ⁴	10 ⁴	10 ⁴	10 ⁴

$T = 298$ [K], $[A_1] = [A_2] = [B] = 10^{-6}$ [nM], $[D_{Y_1}]_{tot} = [D_{Y_2}]_{tot} = [D_Z]_{tot} = 1$ [nM], $\alpha = 1$ [s⁻¹], $J = 0$ (for all the species)

Figure 6.17: Dynamic responses and kinetic parameters of the double Y (2Y-FFL) generalization of FFL.



	YA	YX	Z,B ₁	Z,X	Z,Y	Z,XY	Z ₂ B ₂	Z ₂ X	Z ₂ Y	Z ₂ XY
k_1 [s ⁻¹ nM ⁻¹]	1	0.02	1	0.1	0.01	10 ⁻⁸	1	0.01	1	10 ⁻⁸
k_2 [s ⁻¹]	1	2	1	1	1	10 ⁻⁴	1	1	1	10 ⁻⁴
k_3 [s ⁻¹]	10 ³									
k_{-1} [s ⁻¹]	10 ⁻⁶									
k_{-2} [s ⁻¹]	10 ⁻⁶									
k_3 [s ⁻¹ nM ⁻²]	10 ⁻⁶									
β [nM·s ⁻¹]	1	2	1	1	1	10 ⁻⁴	1	1	1	10 ⁻⁴
κ [nM]	1	100	1	10	100	10 ⁴	1	100	1	10 ⁴

$T = 298$ [K], $[A] = [B_1] = [B_2] = 10^{-6}$ [nM], $[D_{i,ov}] = [D_{Z1,ov}] = [D_{Z2,ov}] = 1$ [nM], $\alpha = 1$ [s⁻¹], $J = 0$ (for all the species)

Figure 6.18: Dynamic response and kinetic parameters of the double output (2Z-FFL) generalization of FFL.

As explained in Chapter 5, as the basal activity approaches to zero, $\kappa \rightarrow K$ for any cycle. Then, the κ values shown in Figure 6.16 correspond to the activation coefficients K . Throughout this work, the combined contributions of two or more transcription factors with *OR* input logic are desired to be reduced in order to appreciate the individual effect of the individual activators. This is done by reducing β and increasing K . In order to have concentrations of Y of the same order of magnitude of Z , the maximal transcription rates of Y ($\beta^{YX_1} = \beta^{YX_2} = 2$) are twice the maximal transcription rates of Z ($\beta^{ZX_1} = \beta^{ZX_2} = \beta^{ZY} = 1$). Finally, as seen in Figure 6.16, the exponential input in Equation (6.17) was assigned for both X_1 and X_2 .

Dynamic response of double Y (2Y-FFL)

The 2Y-FFL is composed of twelve cycles: two of them are involved in the transcription of Y_1 by X ; two involved in the transcription of Y_2 by X ; and eight in the transcription of Z by X , Y_1 and Y_2 . The first two cycles correspond to cycle Y_1A_1 (transcription of Y_1 from its basal activator A_1) and cycle Y_1X (transcription of Y_1 by X); the next two cycles are Y_2A_2 (transcription of Y_2 by A_2), and cycle Y_2X (transcription of Y_2 by X). The other eight cycles are cycle ZB (transcription of Z from its basal activator B), cycle ZX (transcription of Z by X), cycle ZY_1 (transcription of Z by Y_1), cycle ZY_2 (transcription of Z by activator Y_2), ZXY_1 (transcription of Z by both X and Y_1), ZXY_2 (transcription of Z by both X and Y_2), ZY_1Y_2 (transcription of Z by Y_1 and Y_2), and cycle ZXY_1Y_2 (transcription of Z by the joint actions of X , Y_1 and Y_2). The kinetic parameters are shown in Figure 6.17 and follow the same *OR* input logic as for the double input network and are based on the previous case. In fact, the parameters used in the FFL composed of X , Y_1 and Z in the 2Y-FFL are same as those in the FFL composed of X_1 , Y and Z in the 2X-FFL. Furthermore, to facilitate the comparison of the network structures, the parameters involved in the transcription of Y_2 by X in the double intermediate network are same as those involved in the transcription of Y by X_2 . In addition, the following rule applies for the three generalizations of the FFLs: the parameters used in the transcription of I by J_k are same as those involved in the transcription of I_k by J , where $I, J = X, Y, Z$ and $k = 1, 2$

Dynamic response of double output (2Z-FFL)

The 2Z-FFL is composed of ten cycles: two of them are involved in the transcription of Y by X ; four involved in the transcription of Z_1 by X and Y ; and four in the transcription of Z_2 by X and Y . The first two cycles correspond to cycle YA (transcription of Y from its basal activator A) and cycle YX (transcription of Y by X); the next four cycles are Z_1B_1 (transcription of Z_1 by its basal activator B_1), cycle Z_1X (transcription of Z_1 by X), cycle Z_1Y (transcription of Z_1 by Y) and Z_1XY (transcription of Z_1 by the joint action of X and Y). The last four cycles are cycle Z_2B_2 (transcription of Z_2 by its basal activator B_2), cycle Z_2X (transcription of Z_2 by X), cycle Z_2Y (transcription of Z_2 by Y) and Z_2XY (transcription of Z_2 by the joint action of X and Y). The kinetic parameters described in Figure 6.18 for *OR* input logic are related to the two other generalizations as explained in the 2Y-FFL case: the parameters in cycle Z_2X and Z_2Y are same than those in cycles ZX_2 (2X-FFL) and ZY_2 (2Y-FFL), respectively. In addition, as presented by (Kashtan et al., 2004), the following relationship is satisfied: $\kappa^{Z_1X} < \kappa^{Z_2X}$ and $\kappa^{Z_1Y} > \kappa^{Z_2Y}$. In concurrence with previous results (Kashtan et al., 2004), there is a temporal order (First-In-First-Out, FIFO) in the expression of Z_1 and Z_2 (Figure 6.18). This sequence in the protein synthesis has been studied in the flagella system of *E. coli* (Kalir and Alon, 2004). It has been postulated that this dynamical property makes the 2Z-FFL a network motif, being the most abundant among the three cases analyzed. In addition, the energetic changes in the dynamic response of the 2Z-FFL motif are much smaller than those in the 2X-FFL (which in turn presents smaller changes than in the 2Y-FFL), being energetically more robust than the two other generalizations of FFLs.

6.10.6 Prediction of Environmental Selection of Motifs Based on Energetic Cost

Dynamic response of C1-FFL and Simple Regulation (SR) with AND input logic

In Figure 6.6A, the dynamic responses of C1-FFL and SR-AND ($X \rightarrow Z \leftarrow Y$) are compared for an ON step change in the concentration of X from an initial concentration X_i to a final concentration X_f . This step change is maintained for a pulse duration t_X that varies from 10^{-1} to 10 [s].

As described by (Dekel and Alon, 2005; Dekel et al., 2005), the production of protein Z has two opposite effects in the overall cell growth. When Z is transcribed, is then used as an

enzyme that potentiates further cell growth. This cell growth rate can be described by a Michaelis-Menten reaction:

$$\text{cell growth rate} = \frac{\delta[Z]}{L+[Z]} \quad (6.18)$$

where δ [nM/s] is the maximal growth rate per Z , and L is the Michaelis-Menten constant of Z . The second effect is a reduction in the growth rate due to utilization of resources required for other protein transcription. This rate is proportional to the transcription rate of Z , J_T^Z :

$$\text{cell growth reduction rate} = \eta J_T^Z \quad (6.19)$$

where η is the reduction in growth rate per molecule of Z produced. When a step pulse is induced in X , the total cell growth or *benefit* and the total cell growth reduction or *cost* can be calculating by integrating the dynamic response during the time at which the pulse is acting:

$$\text{benefit} = \int_0^{t_x} \frac{\delta[Z]}{L+[Z]} dt \quad (6.20)$$

$$\text{cost} = \int_0^{t_x} \eta J_T^Z dt \quad (6.21)$$

A Pareto frontier can be obtained between *benefit* and *cost* for both C1-FFL *AND* and SR-*AND* as shown in Figure 6.6A by having X_i , X_f , t_x and L as decision variables in the ranges $10^{-2} \leq X_i \leq 10^{-1}$ [nM], $10^2 \leq X_f \leq 10^3$ [nM], $10^{-2} \leq t_x \leq 10$ [s] and $10^{-1} \leq L \leq 1$ [nM]. The anchor points represent maximal *benefit* and minimal *cost* (or maximal $1/\text{cost}$).

As an underlying cost, we propose that the time-averaged *SDE* ($\langle SDE \rangle$) may allow us to predict the environmental selection between C1-FFL and SR-*AND*:

$$\langle SDE \rangle = \frac{\int_0^{t_x} SDE(t) dt}{t_x} \quad (6.22)$$

When the *benefit* at the Pareto frontier is normalized by its corresponding $\langle SDE \rangle$ and plotted against the decision variable t_x , we can see that C1-FFL is selected over SR-*AND* for a critical pulse $t_x^C \approx 0.1$ [s]. Observe that t_x^C is in the short pulse duration range ($t_x^C < 1/\alpha$, with $\alpha = 1$ [s⁻¹]) as predicted by (Dekel and Alon, 2005; Dekel et al., 2005).

The parameters used in C1-FFL are the same used in Figure 6.4 and 6.5A and for SR-*AND*, the parameters correspond to those used in the transcription of Z in C1-FFL. For *benefit*,

we used $\delta = 1$ [nM/s] and for cost we used $\eta = 1$. Notice that δ and η do not play any role in the optimization of both *benefit* and *cost*.

Dynamic response of C1-FFL and C4-FFL with AND input logic

In a very similar way, the dynamic response of C1-FFL and C4-FFL with *AND* input logic is compared for an ON-OFF step change in the concentration of *X*. The parameters used here are the same used in Figure 6.4A. Figure 6.19 shows that C1-FFL is energetically favorable over C4-FFL for all pulse durations.

Pareto-optimal robustness analysis

Robustness, as defined here, is with respect to change in the cellular transcription rate. Robustness index (*RI*) was hypothesized here as the measure which cellular decision machinery utilizes for selecting motifs among topologically and architecturally different but equally capable of higher transcription during variable environments having internal and external environmental perturbations, mutations, and non-genetic changes. *RI* used here is similar to the one defined previously (Stelling et al., 2004a).

$$RI = \left(\sum_{k=1}^{n_T} \sum_{i=1}^{p_j} \left[\frac{1}{x_i} \frac{\partial x_i(t_k, t_0)}{\partial p_j} \right]^2 \right)^{1/2} \quad (6.23)$$

where overall sensitivity is determined by integration over discrete time and calculated over vector element for parameters varied p_j . Thus higher the *RI* higher is robustness of cellular system in maintaining a particular function. We adopted systematic approach relying on multi-dimensional sections of parameter space where all possible combinations of appropriate parameters were varied to evaluate particular robustness. For instance, discretizing the parameter space for C1-FFL into 25 dimensions required varying in total 20 reaction parameters (3 rate constants per cycle with 6 cycles in each FFL and two total binding sites). For parametric robustness, we varied the forward kinetic constants of the kinetic TRN model (k_1 , k_2 and k_3 for each cycle), and for mutational robustness we varied the available binding site parameter ($[D]_{tot}$, for all cycles involved during *Y* and *Z* transcription). For overall robustness, all possible combinations of both parametric and mutational parameters were varied to obtain *RI*. Thus, we varied 20 parameters per FFL motif over 25 grid points.

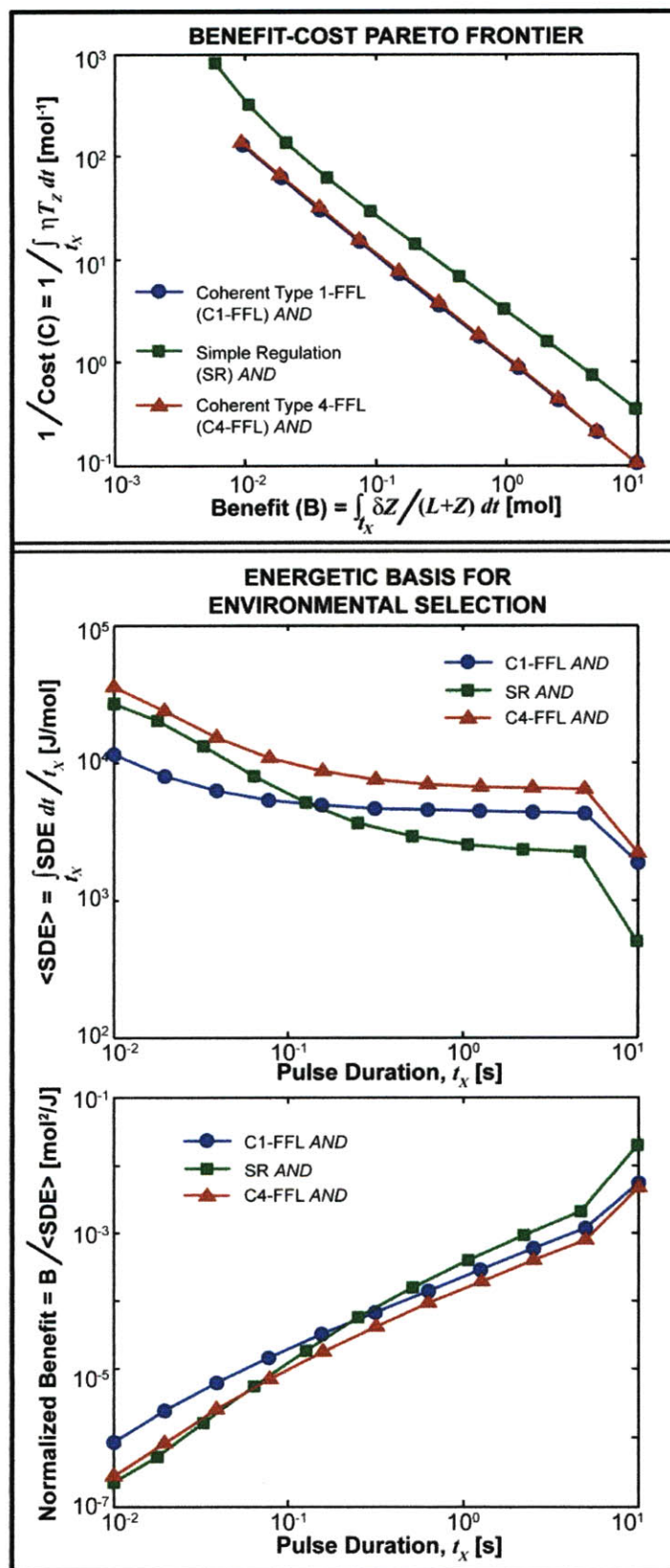


Figure 6.19: Environmental selection of between C1-FFL and C4-FFL with AND input logic.

6.10.7 Energetic Cost as the Underlying Basis for Generalization of Subgraphs to Form Complex Networks

Energetic cost analysis during generalization of single input module (SIM)

Figure 6.7A is constructed based on the Single-Input Module (SIM) seen in Figure 6.20A. In this network motif, the transcription of Y_i is independent of the transcription of Y_j ($i \neq j$) and in fact, the final concentration of Y_i is a function of the concentration of X and the external uptake J_{Y_i} . The transcription of Y_i is composed of two cycles: cycle $Y_i A_i$ (transcription of Y_i by its basal activator A_i) and cycle $Y_i X$ (transcription of Y_i by X). The kinetic parameters used for the basal activity are the same for all proteins Y , and are assigned in order to have a negligible basal transcription rate ($J_{T,Basal}^{Y_i} = 10^{-6}$ [nM·s⁻¹]). This will allow us to observe only the effect of X in the transcription of Y_i . As the basal activity approaches to zero, $\kappa \rightarrow K$, and thus the κ values shown in Figure 6.20A correspond to the activation coefficients K . As experimentally studied in metabolic (Zaslaver et al., 2004) and DNA repair systems (Ronen et al., 2002) of *E. coli*, and theoretically explained by (Shen-Orr et al., 2002), SIMs generate temporal expression programs of Last-In-First-Out (LIFO) order that require $K^i < K^{i+1}$ and ensures an order in the expression of the proteins which guarantees a protein production only when its needed. Notice that this condition is imposed in this work. The values of $k_1^{Y_i X}$ were chosen in order to have $\Delta K = 0.001$ [nM] between the transcription of Y_{i+1} and Y_i . The maximal transcription rates are equal and set as $\beta^i = 1$ [nM·s⁻¹]. As only one transcription factor is present, there is no need of defining a Boolean input logic for the SIM.

The results presented in Figure 6.7A were obtained for $[X] = 0.01$ [nM] and for a constant total external flux $\sum_i J_{Y_i} = 0.01$ [nM·s⁻¹] with $J^{Y_i} = J^{Y_{i+1}}$. For an instance, when only Y_1 and Y_2 were present, $J^{Y_1} = J^{Y_2} = 5 \cdot 10^{-3}$ [nM·s⁻¹]. However, when there are four outputs, $J^{Y_1} = J^{Y_2} = J^{Y_3} = J^{Y_4} = 2.5 \cdot 10^{-3}$ [nM·s⁻¹]. This ensures a constant total external flux which allows us to energetically compare all the topological generalizations of network motifs.

Figure 6.20B shows that there is a constant decrease in the *SDE* with the increase in the number of transcribed proteins which is in agreement with the observations made by (Shen-Orr et al., 2002). Figure 6.20C shows the slope of the curve presented in Figure 6.20B.

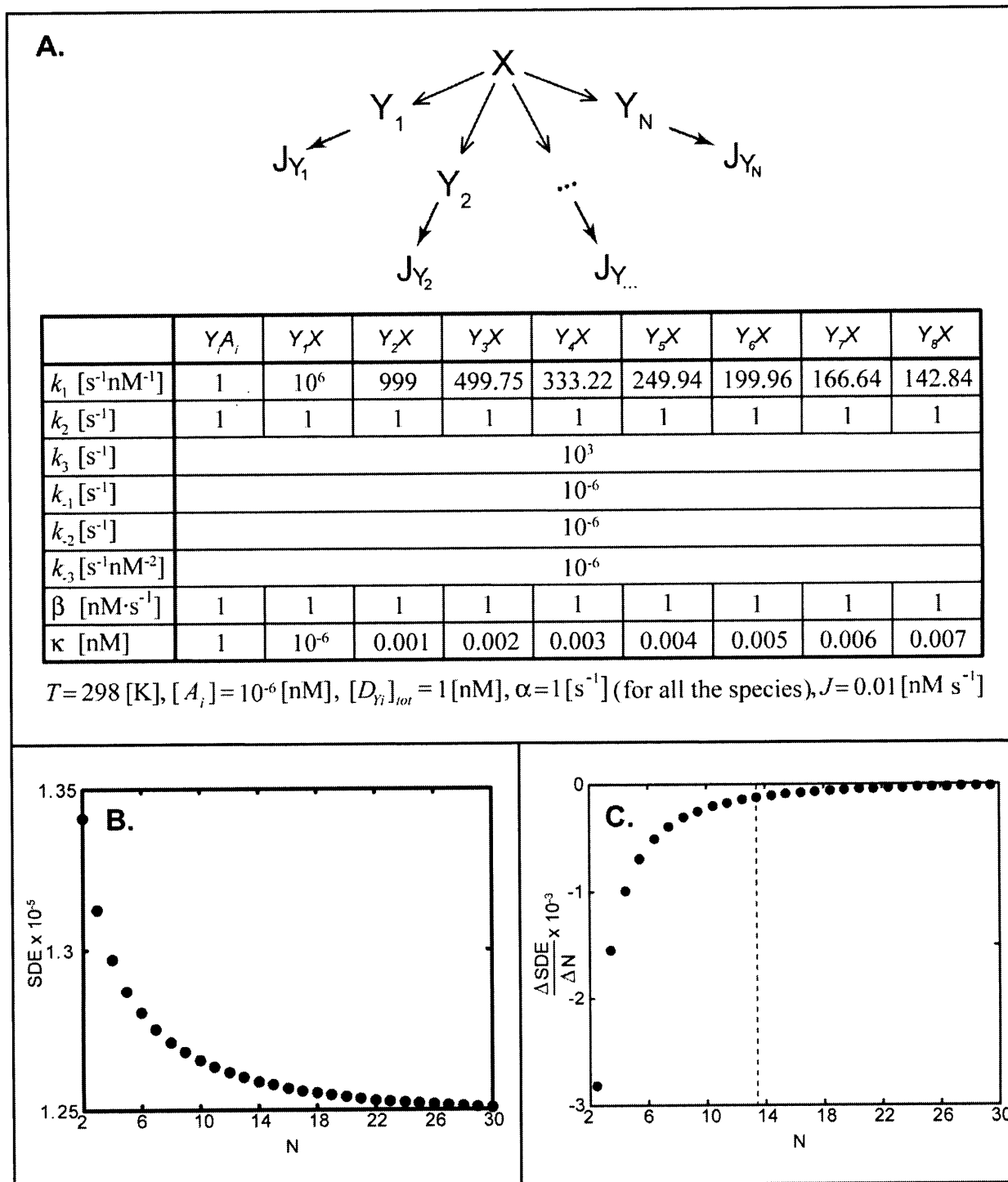


Figure 6.20: The Single-Input Module (SIM). **A)** Parameters used in Figure 6.7A. **B)** SDE as a function of the SIM outputs. **C)** Change of the SDE per change in the number of outputs, as a function of the number of outputs.

As seen in the figure, when SIM has greater than approximately 13 outputs, the slope (indicating the rate of change of SDE with the number of linkages) is almost constant and close

to zero, showing an energetic advantage for $N \geq 13$. This explains that SIM with lower linkages are energetically favorable and also have low frequency of occurrence.

6.10.8 Energetic Phase Diagram Generation for Directional SIM Based Generalizations

First layer: SIM

The *SDE* at steady state is compared for topological generalizations of network motifs. As previously explained, SIM is the simplest topological generalization for $N \geq 2$. The results presented in Figure 6.7A correspond to the bottom row in the matrix shown in Figure 6.7B.

Second Layer: SIM with one linkage

The next layer in Figure 6.7B corresponds to the SIM with one interaction between proteins Y which is always directed from Y_{k-1} to Y_k , Y_k being the protein transcribed by both X and Y_{k-1} . This order is needed to maintain the temporal expression already mentioned in the SIM and, in general, here we assume that Y_{i-1} is always required first than Y_i , and thus Y_{i-1} participates in the transcription of Y_i when these nodes are linked. Based on (Vazquez et al., 2004), the position k of the protein Y is given by:

$$k = \text{round}\left(\frac{N+1}{2}\right), \text{ with } N \geq 2 \quad (6.24)$$

where the function $f = \text{round}(x)$ rounds x towards nearest integer and N is the number of total outputs. For an instance, when $N = 2$, then $k = 2$, and thus Y_2 is transcribed by both X and Y_1 . Notice that this topological generalization becomes a C1-FFL. As another example, if $N = 5$, then $k = 3$ and now Y_3 is transcribed by both X and Y_2 . Because there is only one interaction, Y_1 , Y_2 , Y_4 , and Y_5 are still being transcribed only by X as in the SIM.

Transcription of Y_k is modeled by four cycles: cycle $Y_k A_k$ (transcription of Y_k by its basal activator A_k), cycle $Y_k X$ (transcription of Y_k by X), cycle $Y_k Y_{k-1}$ (transcription of Y_k by Y_{k-1}) and cycle $Y_k X Y_{k-1}$ (transcription of Y_k by the joint interaction of both X and Y_{k-1}). The kinetic parameters used in the cycles $Y_k A_k$ and $Y_k X$ are the same as used in the SIM layer (shown in Figure 6.20). In order to compare the transcriptional effect of X and Y_{k-1} , we have set the parameters of the cycle $Y_k Y_{k-1}$ same as those used in cycle $Y_k X$. As Y_k is transcribed by two transcription factors X and Y_{k-1} , a transcriptional input logic is required. For this work, we have

selected an *OR* input logic. As has been mentioned previously, we make the contribution of this cycle negligible by using low $\beta^{Y_k X Y_{k-1}} = 10^{-4}$ [nM·s⁻¹] and high $K^{Y_k X Y_{k-1}} = 10^4$ [nM]. The results presented in Figure 6.7B were obtained for $[X] = 0.01$ [nM] and for a constant total external flux

$$\sum_i J_{Y_i} = 0.01 \text{ [nM}\cdot\text{s}^{-1}] \text{ with } J^{Y_i} = J^{Y_{i+1}}.$$

Third Layer: SIM with two linkages

We next built more complicated topological network motifs by adding a second interaction to the SIM. As shown in (Vazquez et al., 2004), two proteins Y_k , and Y_m , are now being transcribed by X and Y_{k-1} , and by X and Y_{m-1} , respectively. The positions k and m are given by the following equations:

$$k = \text{round}\left(\frac{N}{2}\right), \text{ for } N \geq 3 \quad (6.25)$$

$$m = \begin{cases} k + 1, & \text{if } (k - N/2) = 0.5 \\ k + 2, & \text{if } (k - N/2) = 0 \end{cases}, \text{ for } N \geq 3 \quad (6.26)$$

For an instance, if $N = 3$, then $k = 2$ and $m = 3$. Therefore Y_2 is transcribed by both X and Y_1 ; Y_3 is transcribed by both X and Y_2 ; and Y_1 is only transcribed by X . Now, if $N = 4$, then $k = 2$ and $m = 4$. Thus Y_2 is transcribed by both X and Y_1 ; Y_4 is transcribed by both X and Y_3 ; and Y_1 and Y_3 are only transcribed by X as in the SIM.

Once again, the transcription of Y_k , and Y_m is presented as a four cycle model. For $i = \{k, m\}$, these cycles are: cycle $Y_i A_i$ (transcription of Y_i by its basal activator A_i), cycle $Y_i X$ (transcription of Y_i by X), cycle $Y_i Y_{i-1}$ (transcription of Y_i by Y_{i-1}) and cycle $Y_i X Y_{i-1}$ (transcription of Y_i by the joint interaction of both X and Y_{i-1}). Parameter assignments and inputs follows the same logic described for the second layer.

Fourth Layer: SIM with three linkages

General case ($N \geq 4$)

When three interactions are added to SIM with $N \geq 4$, proteins Y_k , Y_{k+1} and Y_{k+2} are transcribed by both X and their corresponding predecessor. Position k is assigned by the following logic:

$$k = \text{round}\left(\frac{N-1}{2}\right), \text{ with } N \geq 4 \quad (6.27)$$

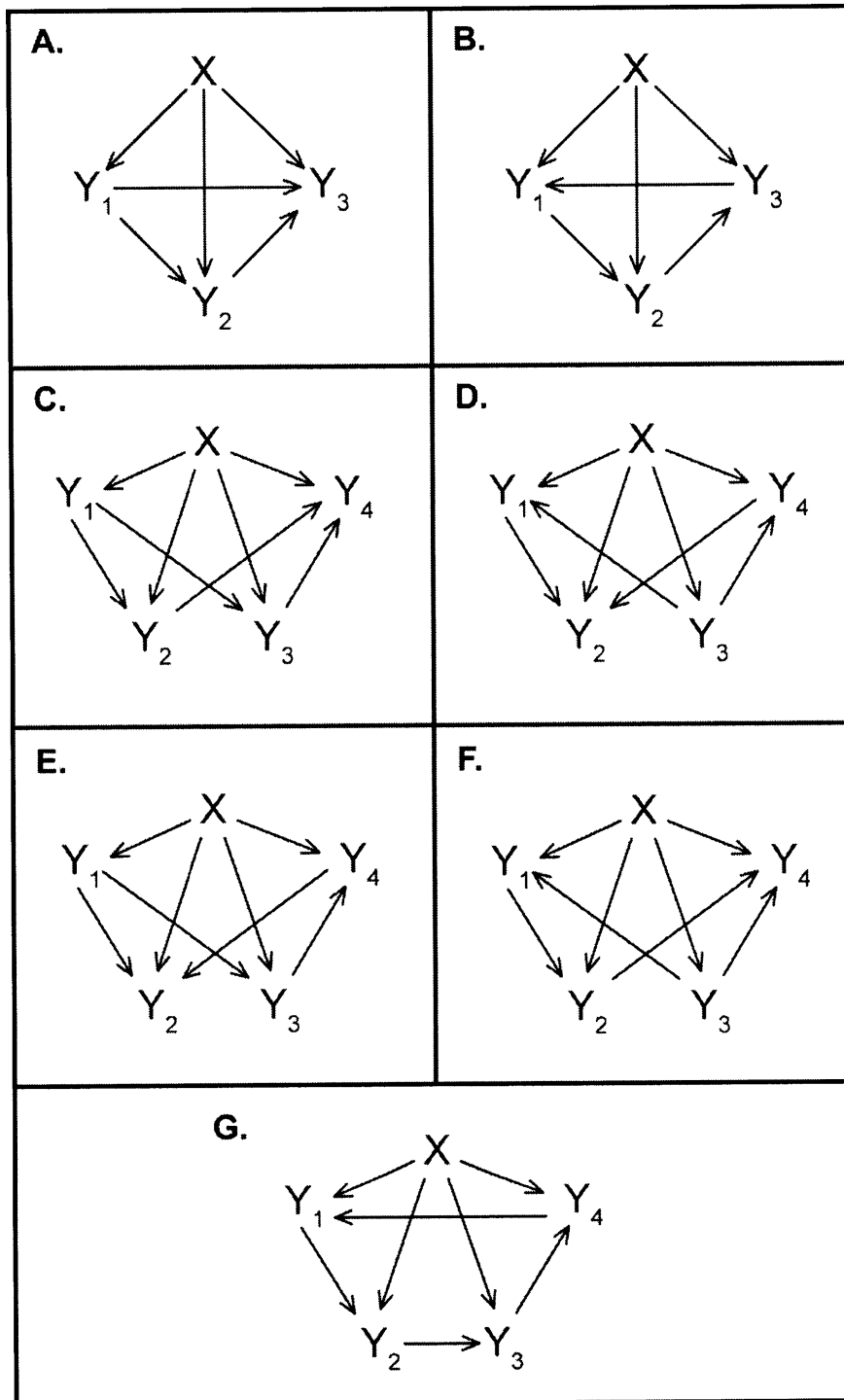


Figure 6.21: A-B) Two possible topological generalization motifs for a 3 output SIM with three interactions: A) Y_1 transcribes Y_3 . B) Y_3 transcribes Y_1 . In both cases, the temporal order is maintained by keeping $Y_1 \rightarrow Y_2$ and $Y_2 \rightarrow Y_3$. C-G) Four possible topological generalization motifs for a 4 output SIM with four interactions. C) Y_1 transcribes Y_3 and Y_2 transcribes Y_4 . D) Y_3 transcribes Y_1 and Y_4 transcribes Y_2 . E) Y_1 transcribes Y_3 and Y_4 transcribes Y_2 . F) Y_3 transcribes Y_1 and Y_2 transcribes Y_4 . In the four cases, the temporal order is maintained by keeping $Y_1 \rightarrow Y_2$ and $Y_3 \rightarrow Y_4$, and no interaction is allowed between Y_2 and Y_3 and between Y_1 and Y_4 , according to (Vazquez et al., 2004). G) Another possible topological generalization motifs for a 4 output SIM with four interactions. The temporal order is maintained by keeping $Y_1 \rightarrow Y_2$, $Y_2 \rightarrow Y_3$, and $Y_3 \rightarrow Y_4$.

Thus, if $N = 4$, then $k = 2$. Therefore Y_2 is transcribed by both X and Y_1 ; Y_3 is transcribed by both X and Y_2 ; Y_4 is transcribed by both X and Y_3 ; and Y_1 is only transcribed by X . Now, if $N = 6$, $k = 3$. Then Y_3 is transcribed by both X and Y_2 ; Y_4 is transcribed by both X and Y_3 ; Y_5 is transcribed by both X and Y_4 ; and Y_1 and Y_6 are only transcribed by X as in the SIM.

As explained before, the *OR* input logic still applies, as well as the four cycle model for Y_k , Y_{k+1} and Y_{k+2} , the parameters assignment, and the inputs.

Special cases ($N = 3$)

When three outputs are present and the following interactions are already fixed by the temporal expression of SIM: $Y_1 \rightarrow Y_2$ and $Y_2 \rightarrow Y_3$, then the only way of adding a third interaction is by adding $Y_1 \rightarrow Y_3$ (Figure 6.21A) or $Y_3 \rightarrow Y_1$ (Figure 6.21B).

In the first case, transcription of Y_1 and Y_2 follow the two and four cycle model already explained. However, transcription of Y_3 involves now eight cycles: cycle Y_3A_3 (transcription of Y_3 by its basal activator A_3), cycle Y_3X (transcription of Y_3 by X), cycle Y_3Y_1 (transcription of Y_3 by Y_1), cycle Y_3Y_2 (transcription of Y_3 by Y_2), Y_3XY_1 cycle (transcription of Y_3 by both X and Y_1), cycle Y_3XY_2 (transcription of Y_3 by both X and Y_2), cycle $Y_3Y_1Y_2$ (transcription of Y_3 by both Y_1 and Y_2), and cycle $Y_3XY_1Y_2$ (transcription of Y_3 by X , Y_1 and Y_2). Because the system follows an *OR* input logic, the kinetic parameters associated to cycles Y_3XY_1 , Y_3XY_2 , $Y_3Y_1Y_2$, and $Y_3XY_1Y_2$ are given in order to have negligible transcription from the joint interaction of two or more transcription factors, and the parameters given to cycles Y_3Y_1 and Y_3Y_2 are same than those assigned to the initial SIM configuration for cycle Y_3X .

In the second case, the transcription of each protein Y is always determined by the *OR* input logic of two transcription factors: X and Y_3 for transcription of Y_1 , X and Y_1 for transcription of Y_2 , and X and Y_2 for transcription of Y_3 . Four cycles are associated to each protein expression and the parameters assignment is same as previously described.

Fourth Layer: SIM with four linkages

General case ($N \geq 5$)

When four interactions are added to SIM with $N \geq 5$, proteins Y_{k-1} , Y_k , Y_m , and Y_{m+1} , are transcribed by both X and their corresponding predecessor. Assignment of positions k and m follow Equations (6.25) and (6.26), respectively, but for $N \geq 5$.

Thus, if $N = 5$, $k = 3$ and $m = 4$. Then Y_2 is transcribed by both X and Y_1 ; Y_3 is transcribed by both X and Y_2 ; Y_4 is transcribed by both X and Y_3 ; Y_5 is transcribed by both X and Y_4 ; and Y_1 is only transcribed by X . Now, if $N = 6$, $k = 3$ but $m = 5$. Then Y_2 is transcribed by both X and Y_1 ; Y_3 is transcribed by both X and Y_2 ; Y_4 is transcribed by both X and Y_3 ; Y_5 is transcribed by both X and Y_4 ; Y_6 is transcribed by both X and Y_5 ; and Y_1 and Y_4 are only transcribed by X as in the SIM. The *OR* input logic still applies, as well as the four cycle model for Y_{k-1} , Y_k , Y_m , and Y_{m+1} , the parameters assignment, and the inputs.

Special cases ($N = 4$)

As presented by (Vazquez et al., 2004) where there is not interaction between Y_2 and Y_3 , and between Y_1 and Y_4 , and by fixing the following interactions according to the temporal expression of SIM: $Y_1 \rightarrow Y_2$ and $Y_3 \rightarrow Y_4$, four possible cases shown in Figure 6.21C-F are possible in order to have a total of four interactions, which kinetic parameters follows the same assignment as already explained for *OR* input logic.

The first case is shown in Figure 6.21C and shows Y_1 being transcribed only by X and thus only 2 cycles describe the model; transcription of Y_2 by X and Y_1 (total of 4 cycles), transcription of Y_3 by X and Y_1 (total of 4 cycles) and transcription of Y_4 by X , Y_2 and Y_3 (total of 8 cycles). The cycles can be easily defined by getting all the possible combinations between the transcription factors.

The second case is shown in Figure 6.21D and shows Y_1 being transcribed by X and Y_3 (4 cycles); transcription of Y_2 by X , Y_1 and Y_4 (total of 8 cycles), transcription of Y_3 only by X (total of 2 cycles) and transcription of Y_4 by X and Y_3 (total of 4 cycles).

The third special case is shown in Figure 6.21E and shows Y_1 being transcribed only by X (2 cycles); transcription of Y_2 by X , Y_1 and Y_4 (total of 8 cycles), transcription of Y_3 by X and Y_1 (total of 4 cycles) and transcription of Y_4 by X and Y_3 (total of 4 cycles).

The fourth special case is shown in Figure 6.21F and shows Y_1 being transcribed by X and Y_3 (4 cycles); transcription of Y_2 by X and Y_1 (total of 4 cycles), transcription of Y_3 only by X (total of 2 cycles) and transcription of Y_4 by X , Y_2 and Y_3 (total of 8 cycles).

In addition to the interactions proposed by (Vazquez et al., 2004), here we proposed a fifth case which is based on Case B from SIM with three linkages as shown in Figure 6.21G. The transcription of each protein Y is always determined by the *OR* input logic of two transcription factors: X and Y_4 for transcription of Y_1 , X and Y_1 for transcription of Y_2 , X and Y_2 for

transcription of Y_3 and X and Y_3 for transcription of Y_4 . Four cycles are associated to each protein expression and the parameters assignment is same as already mentioned.

6.10.9 Pulsed Energetics Topological Generalization Prediction (PETGP) Strategy for Energetics Based Dynamical Analysis for Characterizing Subgraphs

In this section, we propose a novel strategy to predict the abundance of generalized network motifs based on the computation of $\langle SDE \rangle$.

Assumptions

1. The basal transcription activity is neglected. This is achieved by using low concentrations of the basal transcription factor.
2. The generalized network motifs have an *OR* input logic. This implies that the combined interaction of multiple transcription factors is neglected by reducing β and increasing κ values.
3. Only the cycles corresponding to the individual transcription factors are active.
4. No external flux is withdrawn from the network.

The proposed strategy follows three main steps (Figure 6.22):

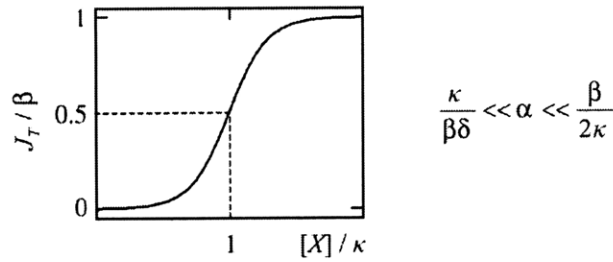
Step 1: Assignment of parameters

For any generalized network motif, *assign equal parameters* for all the existing interactions between the nodes. *It is important to note that the dynamic functionality (and thus the relationship between the network parameters) is not required to be known.* For every linkage, assign any desired κ and β value. For very irreversible reactions (i.e., $k_i \gg k_{-i}$, $i = 1, 2, 3$) and if $k_3 \gg k_2$, the relationship between κ and β and the kinetic parameters is given by:

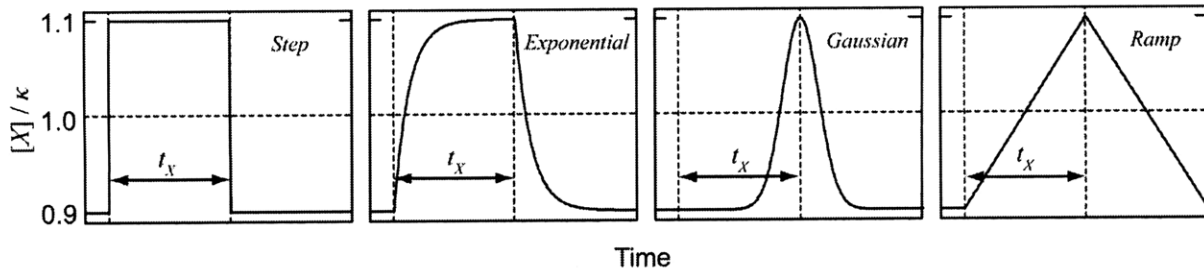
$$\kappa = \frac{k_2}{k_1} \tag{6.28}$$

$$\beta = k_2[D]_{tot} \tag{6.29}$$

STEP 1: For all linkages between nodes, assign equal parameters.



STEP 2: Choose a dynamic input for X around κ with a pulse duration t_X .



STEP 3: Obtain the dynamic energetic response and time average.

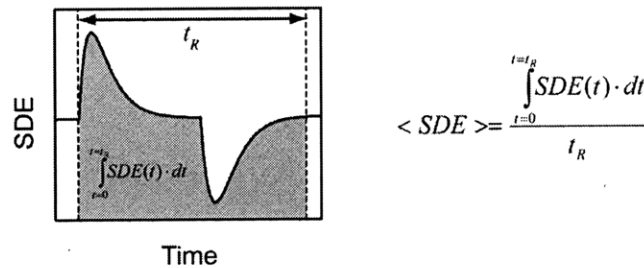


Figure 6.22: Description of the general strategy procedure. The three main steps are: 1) Assignment of the parameters, 2) Selection of the input X , and 3) Calculation of $\langle SDE \rangle$.

Thus, when κ and β are assigned, it allows the determination of the following kinetic parameters: k_1 , k_2 , k_3 and $[D]_{tot}$. In addition, recall that when the basal activity approaches to zero, $\kappa = K$.

Constraints

When all the kinetic parameters are assigned, the network is fully characterized except for the degradation rate α which is the same for all the proteins in the network. However, **this strategy has only one restriction**. It requires the following relationship to be satisfied when this parameter is assigned:

$$\frac{\kappa}{\beta\delta} = \frac{k_{-2}[D]_{tot}}{K_1K_3} \ll \alpha \ll \frac{\beta}{2\kappa} = \frac{k_2^2[D]_{tot}}{2k_1} \quad (6.30)$$

$$\text{where } \delta = \frac{k_3}{k_{-1}k_{-2}k_{-3}[D]_{tot}^2}, [\text{s}^2] \quad (6.31)$$

In Figure 6.7C, the following parameters were used: $k_1 = 10 [\text{s}^{-1}\cdot\text{nM}^{-1}]$, $k_2 = 1 [\text{s}^{-1}]$, $k_3 = 1000 [\text{s}^{-1}]$, $k_{-1} = 10^{-6} [\text{s}^{-1}]$, $k_{-2} = 10^{-6} [\text{s}^{-1}]$, $k_{-3} = 10^{-6} [\text{s}^{-1}\cdot\text{nM}^{-2}]$, $[D]_{tot} = 1 [\text{nM}]$, $\alpha = 10^{-3} [\text{s}^{-1}]$. These values determine $\kappa = 10^{-1} [\text{nM}]$, $\beta = 1 [\text{nM}\cdot\text{s}^{-1}]$, and $\delta = 10^{21} [\text{s}^2]$. Notice that condition (6.30) is satisfied by the current parameters.

Step 2: Selection of the input X

In order to obtain the dynamic response of the *SDE*, a perturbation in the concentration of X must be performed. As seen in Figure 6.22, we propose four different pulses to study the dynamic energetic behavior: step, exponential, Gaussian, and ramp ON-OFF changes. The duration of the ON input is the pulse duration t_X and this time is the same for the OFF response. Because κ is the concentration of X required to reach half of the maximal transcription rate, in this work we define κ as *the characteristic concentration of the network*. We further hypothesize that small perturbations around κ can give us valuable information about the *SDE* of the network. Thus, we select $X_i = 0.9\kappa$ and $X_f = 1.1\kappa$. As seen in Tables 6.2 to 6.5, *$\langle SDE \rangle$ is robust with respect to the shape of the input and the pulse duration*. If the TRN has more than one input X (as in the generalized FFL, DOR and LAY), the same input has to be given to all of them. In Figure 6.7C, the shown results correspond to an ON-OFF step change in the concentration of X with $t_X = 10^4 [\text{s}]$.

Step 3: Calculation of $\Delta\langle SDE \rangle$

Once the input of $[X]$ is given, the dynamic energetic response can be obtained. Here we define the response time t_R , as the time required to the *SDE* to reach 0.1% of the initial state. After calculation of the response time, the integration of *SDE* in time can be performed in order to calculate the $\langle SDE \rangle$:

$$\langle SDE \rangle = \frac{\int_0^{t_R} SDE(t) dt}{t_R} \quad (6.32)$$

The *SDE* of the base subgraph is considered as the reference energetic state for all of its generalized motifs and its $\langle SDE \rangle$ is zero by definition. $\Delta \langle SDE \rangle$ of a generalized network is obtained by subtracting $\langle SDE \rangle$ with a $\langle SDE \rangle$ of the base subgraph. If the generalized network is energetically more favorable than the base subgraph, its $\Delta \langle SDE \rangle$ will be negative. Otherwise, it will take positive values.

Table 6.2: Time averaged *SDE* for the SIM generalization

INPUT	t_x [s]	SIM(2,0)	SIM(3,0)	SIM(4,0)	SIM(5,0)	SIM(6,0)	SIM(2,1)	SIM(3,1)	SIM(4,1)	SIM(5,1)	SIM(6,1)
Step	10000	0.0	3.1	4.6	5.5	6.2	73906.9	44111.9	31443.5	24429.8	19975.3
	1000	0.0	3.0	4.4	5.3	5.9	73225.3	43837.5	31301.2	24344.7	19919.8
	100	0.0	3.1	4.7	5.6	6.3	74052.0	44180.5	31481.5	24453.3	19990.9
	10	0.0	3.1	4.6	5.5	6.2	73437.8	43976.4	31378.3	24388.4	19944.2
	1	0.0	2.9	4.4	5.3	5.8	72315.0	43566.0	31163.0	24253.1	19849.6
Exponential	10000	0.0	3.0	4.5	5.4	6.0	73328.5	44046.5	31342.3	24366.2	19932.1
	1000	0.0	2.9	4.3	5.2	5.8	72329.7	43551.5	31160.0	24259.5	19862.0
	100	0.0	2.8	4.3	5.1	5.7	71947.9	43415.3	31091.8	24218.9	19835.2
	10	0.0	3.0	4.5	5.4	6.0	72955.5	43795.9	31287.9	24336.7	19913.1
	1	0.0	2.9	4.3	5.2	5.8	72126.1	43483.1	31114.3	24219.1	19822.8
Gaussian	10000	0.0	3.0	4.5	5.4	6.0	73205.8	43869.9	31322.4	24357.4	19927.1
	1000	0.0	2.9	4.3	5.2	5.8	72401.5	43577.6	31173.3	24267.5	19867.3
	100	0.0	2.8	4.3	5.1	5.7	71955.1	43419.1	31093.7	24220.1	19836.0
	10	0.0	3.0	4.5	5.4	6.0	72988.0	43793.7	31278.8	24326.2	19902.1
	1	0.0	2.9	4.4	5.2	5.8	72184.4	43507.0	31128.7	24229.6	19831.5
Ramp	10000	0.0	3.0	4.4	5.3	5.9	72899.1	43761.6	31267.8	24324.7	19905.4
	1000	0.0	2.9	4.3	5.2	5.8	72281.1	43536.0	31152.5	24255.2	19859.1
	100	0.0	2.8	4.3	5.1	5.7	71939.7	43413.3	31090.8	24218.4	19834.9
	10	0.0	3.0	4.4	5.3	5.9	72723.8	43700.7	31233.9	24301.0	19886.9
	1	0.0	2.9	4.3	5.2	5.8	72083.1	43469.8	31111.7	24222.0	19829.2

Table 6.2 (Cont): Time averaged *SDE* for the SIM generalization

INPUT	t_x [s]	SIM(3,2)	SIM(4,2)	SIM(5,2)	SIM(6,2)	SIM(3,3)	SIM(4,3)	SIM(5,3)	SIM(6,3)	SIM(4,4)	SIM(5,4)	SIM(6,4)	SIM(7,4)
Step	10000	158315.8	99867.9	72964.6	57487.7	10437448.6	250715.1	161717.0	119412.4	10446517.7	348241.6	226914.5	168589.0
	1000	157664.4	99547.1	72884.3	57519.6	11006229.4	250986.4	161649.3	119539.7	11015380.4	347376.4	227063.9	169181.6
	100	158041.0	99834.4	72952.1	57471.0	11244394.3	248358.9	161272.7	119359.2	11253575.5	341796.9	226047.6	168983.4
	10	155552.0	99153.3	72731.2	57407.6	11265042.2	243993.5	160321.8	119310.5	11274189.2	337872.1	225476.5	169129.5
	1	153606.4	98565.6	72530.0	57356.1	9270150.2	242392.5	159942.4	119292.0	9277658.5	336817.7	225294.5	169215.2
Exponential	10000	156469.5	99553.5	72862.5	57477.9	10668997.6	247387.7	161267.6	119435.6	10677916.7	343563.1	226555.1	168770.4
	1000	154543.1	98837.2	72672.8	57467.7	11080086.5	244662.5	160491.4	119472.0	11089379.8	340348.3	225948.6	169227.3
	100	153724.8	98596.7	72606.7	57467.2	11268325.0	243842.5	160244.0	119499.3	11277417.9	340922.8	225842.7	169436.8
	10	154881.3	98957.0	72684.2	57424.6	11294058.9	243313.9	160189.6	119360.0	11303139.5	337244.8	225454.0	169277.5
	1	153353.3	98487.1	72512.2	57365.2	11296383.4	242132.6	159906.7	119339.9	11305462.0	336577.1	225356.5	169350.3
Gaussian	10000	156534.5	99404.3	72836.1	57478.0	10648680.7	247712.8	161150.6	119444.9	10657823.3	344154.1	226470.7	168806.3
	1000	154706.3	98885.2	72686.9	57468.9	11034513.5	244910.5	160549.2	119471.6	11043566.9	340649.8	225998.5	169196.3
	100	153681.8	98596.2	72606.5	57465.9	11259310.1	243473.6	160221.5	119494.0	11268404.8	339383.7	225773.8	169433.4
	10	155013.7	98967.3	72666.4	57398.7	11292723.7	243523.5	160212.4	119333.5	11301803.6	337383.8	225453.8	169238.8
	1	153440.6	98520.6	72530.7	57378.2	11296749.3	242201.7	159934.1	119355.2	11305829.9	336623.2	225376.3	169362.2
Ramp	10000	155799.7	99206.7	72781.5	57475.8	10785898.6	246532.6	160919.6	119461.9	10794797.7	342615.4	226297.2	168977.4
	1000	154403.7	98806.2	72665.6	57468.0	11085741.7	244404.6	160454.6	119478.8	11094838.5	339971.8	225927.1	169268.1
	100	153677.3	98589.9	72605.1	57466.8	11265822.0	243642.9	160228.6	119498.8	11274912.4	340423.1	225813.4	169442.1
	10	154605.5	98864.1	72649.3	57414.9	11293592.5	243222.9	160171.3	119368.4	11302656.5	337236.0	225469.6	169303.6
	1	153343.5	98500.5	72535.6	57394.1	11296171.0	242157.5	159943.9	119383.2	11305253.2	336618.2	225404.8	169402.0

Table 6.3: Time averaged *SDE* for the FFL generalization

INPUT	t_x [s]	FFL	2X	2Y	2Z
Step	10000	0.0	9550.4	30113.9	-28870.0
	1000	0.0	10295.6	32811.3	-28450.4
	100	0.0	9065.6	31817.5	-29005.2
	10	0.0	8475.3	31775.4	-28511.7
	1	0.0	8565.0	32476.9	-27733.9
Exponential	10000	0.0	9207.9	31542.2	-28423.8
	1000	0.0	9232.9	33386.5	-27925.2
	100	0.0	9273.6	33868.9	-27695.8
	10	0.0	8695.3	32355.3	-28304.0
	1	0.0	8690.8	33085.2	-27676.9
Gaussian	10000	0.0	9332.7	31969.1	-28448.4
	1000	0.0	9241.0	33293.9	-27967.9
	100	0.0	9203.6	33837.7	-27700.3
	10	0.0	8728.6	32346.0	-28306.3
	1	0.0	8683.8	33051.7	-27726.2
Ramp	10000	0.0	9270.7	32796.3	-28263.8
	1000	0.0	9205.9	33533.9	-27895.4
	100	0.0	9240.5	33851.2	-27690.4
	10	0.0	8799.9	32635.1	-28153.5
	1	0.0	8763.6	33198.3	-27692.5

Table 6.4: Time averaged *SDE* for the BIFAN generalization

INPUT	t_x [s]	DOR(2,2)	DOR(2,3)	DOR(2,4)	DOR(3,2)	DOR(3,3)	DOR(3,4)	DOR(4,2)	DOR(4,3)	DOR(4,4)
Step	10000	0.0	4.7	7.1	451.9	458.2	461.4	864.6	872.7	876.7
	1000	0.0	4.5	6.7	451.7	457.8	460.9	887.5	895.3	899.2
	100	0.0	4.8	7.2	376.5	383.0	386.2	740.6	748.8	752.8
	10	0.0	4.7	7.1	375.1	381.5	384.7	738.5	746.5	750.6
	1	0.0	4.4	6.6	385.4	391.4	394.3	754.1	761.6	765.3
Exponential	10000	0.0	4.6	6.8	435.1	441.3	444.4	880.5	888.2	892.1
	1000	0.0	4.3	6.5	397.1	402.8	405.7	776.5	783.8	787.4
	100	0.0	4.2	6.3	385.3	390.9	393.7	760.8	767.8	771.3
	10	0.0	4.5	6.8	375.0	381.1	384.1	739.4	747.0	750.8
	1	0.0	4.3	6.5	379.8	385.6	388.5	745.1	752.4	756.1
Gaussian	10000	0.0	4.5	6.8	425.3	431.4	434.4	820.1	827.7	831.5
	1000	0.0	4.3	6.5	399.4	405.2	408.1	780.2	787.5	791.1
	100	0.0	4.2	6.3	384.4	390.0	392.8	757.7	764.8	768.3
	10	0.0	4.5	6.8	377.8	383.9	386.9	743.1	750.8	754.6
	1	0.0	4.4	6.5	378.2	384.1	387.0	745.9	753.3	757.0
Ramp	10000	0.0	4.5	6.7	414.7	420.7	423.7	803.0	810.5	814.2
	1000	0.0	4.3	6.4	394.9	400.6	403.5	772.6	779.8	783.4
	100	0.0	4.2	6.3	384.5	390.1	392.9	758.8	765.8	769.3
	10	0.0	4.5	6.7	376.4	382.4	385.4	741.0	748.5	752.3
	1	0.0	4.3	6.5	379.1	384.9	387.8	745.2	752.5	756.2

Table 6.5: Time averaged SDE for the Simple Regulation generalization

INPUT	t_x [s]	LAY(1,1,1)	LAY(1,2,1)	LAY(1,3,1)	LAY(1,1,2)	LAY(1,2,2)	LAY(1,3,2)	LAY(1,1,3)	LAY(1,2,3)	LAY(1,3,3)
Step	10000	0.0	-5471.2	-41141.4	3447.4	1747.4	738.9	5459.4	3761.8	2762.3
	1000	0.0	-3030.0	-39671.6	3448.3	1747.1	739.7	5461.2	3762.1	2762.1
	100	0.0	-2308.9	-45380.7	3449.5	1748.8	736.0	5461.9	3764.0	2764.0
	10	0.0	-2411.4	-44303.2	3466.2	1757.5	713.3	5481.3	3779.5	2776.0
	1	0.0	-2776.0	-40958.2	3526.9	1793.9	657.9	5552.7	3842.1	2823.6
Exponential	10000	0.0	-4143.9	-40274.8	3446.8	1746.6	739.1	5458.7	3760.7	2761.1
	1000	0.0	-2353.8	-39148.1	3446.3	1745.5	739.5	5458.2	3759.4	2759.8
	100	0.0	-1867.4	-39036.2	3446.4	1745.2	739.7	5458.5	3759.2	2759.4
	10	0.0	-2125.6	-42874.2	3447.6	1747.0	736.5	5459.9	3761.0	2761.6
	1	0.0	-2199.2	-40391.0	3494.0	1772.9	688.6	5513.6	3806.8	2795.9
Gaussian	10000	0.0	-3849.4	-40625.1	3446.9	1746.6	739.2	5458.8	3760.7	2761.1
	1000	0.0	-2456.5	-39207.4	3446.4	1745.6	739.5	5458.3	3759.5	2759.8
	100	0.0	-1883.3	-39016.0	3446.3	1745.2	739.6	5458.4	3759.1	2759.3
	10	0.0	-2092.4	-42951.0	3453.5	1750.3	731.3	5466.2	3767.2	2766.4
	1	0.0	-2159.2	-40648.8	3482.7	1766.4	696.8	5499.6	3795.2	2788.1
Ramp	10000	0.0	-3100.6	-40273.0	3446.7	1746.2	739.3	5458.7	3760.3	2760.7
	1000	0.0	-2237.8	-39163.6	3446.3	1745.5	739.5	5458.2	3759.4	2759.7
	100	0.0	-1875.4	-39036.5	3446.4	1745.2	739.7	5458.5	3759.2	2759.4
	10	0.0	-2042.8	-42127.3	3451.8	1749.1	733.2	5464.3	3765.1	2764.4
	1	0.0	-2081.5	-40318.6	3474.7	1761.1	708.4	5490.1	3787.0	2781.8

Table 6.5 (Cont): Time averaged SDE for the Simple Regulation generalization

INPUT	t_x [s]	LAY(2,1,1)	LAY(2,2,1)	LAY(2,3,1)	LAY(2,1,2)	LAY(2,2,2)	LAY(2,3,2)	LAY(2,1,3)	LAY(2,2,3)	LAY(2,3,3)
Step	10000	-361.3	-29127.9	-75063.1	3094.7	1401.6	-2324.0	5107.1	3419.9	2421.8
	1000	-416.7	-28802.2	-74192.2	3041.9	1346.6	-784.8	5056.4	3366.5	2367.2
	100	-588.1	-29726.6	-76100.5	2876.5	1180.0	-393.3	4889.4	3201.9	2201.8
	10	-709.3	-29109.9	-75910.6	2810.9	1087.2	-533.1	4828.0	3134.7	2119.4
	1	-778.3	-26446.6	-74823.3	2858.5	1069.7	-1634.9	4888.7	3180.5	2126.6
Exponential	10000	-423.9	-27271.7	-74610.4	3030.9	1337.3	-1406.1	5043.2	3355.0	2356.8
	1000	-518.1	-25869.4	-73963.8	2935.6	1241.0	-258.1	4947.9	3258.2	2259.8
	100	-554.1	-25207.9	-73892.4	2899.6	1204.3	66.6	4912.1	3221.5	2222.9
	10	-626.9	-27812.8	-75247.5	2847.3	1146.3	-219.4	4860.2	3171.1	2169.0
	1	-715.8	-25955.0	-74540.4	2834.7	1088.9	-284.7	4854.5	3154.3	2126.2
Gaussian	10000	-428.8	-27559.8	-74781.0	3025.9	1332.2	-1178.5	5038.2	3349.9	2351.7
	1000	-510.8	-26019.3	-73960.9	2942.9	1248.4	-319.2	4955.2	3265.6	2267.2
	100	-554.6	-25155.4	-73892.8	2898.9	1203.7	55.2	4911.4	3220.9	2222.3
	10	-606.4	-27843.2	-75281.1	2873.8	1168.3	-184.3	4887.7	3197.3	2192.4
	1	-707.5	-26204.3	-74639.3	2819.5	1086.7	-245.4	4838.1	3139.9	2119.3
Ramp	10000	-459.2	-26980.0	-74714.6	2995.2	1301.1	-686.9	5007.4	3318.6	2320.3
	1000	-522.9	-25777.7	-73978.0	2930.7	1236.0	-170.3	4943.0	3253.2	2254.8
	100	-555.3	-25177.8	-73885.3	2898.4	1203.1	62.0	4910.9	3220.4	2221.7
	10	-599.1	-27151.3	-74985.8	2874.5	1171.7	-130.0	4888.0	3197.6	2194.3
	1	-670.6	-25870.7	-74476.5	2840.1	1115.4	-165.7	4857.2	3161.3	2144.8

Table 6.5 (Cont): Time averaged SDE for the Simple Regulation generalization

INPUT	t_x [s]	LAY(3,1,1)	LAY(3,2,1)	LAY(3,3,1)	LAY(3,1,2)	LAY(3,2,2)	LAY(3,3,2)	LAY(3,1,3)	LAY(3,2,3)	LAY(3,3,3)
Step	10000	-211.5	-36931.6	-84881.4	3253.3	1567.1	-11702.4	5266.0	3589.5	2592.5
	1000	-245.6	-37093.5	-84278.1	3224.9	1535.1	-12128.1	5241.0	3561.0	2562.1
	100	-564.2	-37475.8	-85244.7	2918.2	1224.0	-14749.1	4931.2	3254.4	2252.6
	10	-763.4	-37195.9	-85236.5	2813.5	1067.0	-14345.6	4831.5	3146.7	2113.0
	1	-885.1	-35491.0	-84714.2	2822.5	984.2	-12115.6	4856.0	3150.7	2055.4
Exponential	10000	-304.9	-35742.2	-84531.3	3158.2	1471.3	-11027.0	5170.6	3492.9	2495.8
	1000	-446.6	-34874.2	-84110.1	3014.8	1326.2	-10344.9	5027.3	3347.0	2349.6
	100	-493.1	-34592.1	-84069.3	2968.7	1279.0	-10512.3	4981.9	3300.1	2302.3
	10	-630.1	-36215.4	-84806.7	2880.9	1172.2	-13153.8	4894.2	3214.3	2205.6
	1	-762.9	-35117.1	-84507.2	2817.3	1055.5	-11673.6	4838.6	3146.1	2101.1
Gaussian	10000	-316.3	-35909.6	-84667.6	3146.6	1459.3	-11230.3	5159.1	3480.9	2483.7
	1000	-436.0	-34964.6	-84136.4	3025.6	1337.0	-10365.3	5038.0	3357.9	2360.6
	100	-498.1	-34475.6	-84069.1	2963.3	1273.8	-10349.6	4976.0	3294.5	2296.8
	10	-596.9	-36231.3	-84805.8	2915.2	1202.8	-13149.4	4929.7	3248.5	2236.0
	1	-742.9	-35291.4	-84557.7	2810.4	1065.7	-11904.2	4829.0	3140.7	2106.4
Ramp	10000	-361.8	-35536.2	-84509.6	3100.5	1412.7	-11043.6	5112.9	3433.9	2436.7
	1000	-454.5	-34805.7	-84114.6	3006.8	1318.1	-10350.9	5019.2	3338.8	2341.4
	100	-496.3	-34542.1	-84067.1	2965.2	1275.5	-10479.2	4978.3	3296.7	2298.8
	10	-583.2	-35755.8	-84647.6	2916.8	1209.7	-12535.5	4930.7	3249.5	2240.6
	1	-687.7	-35025.8	-84445.1	2845.5	1113.0	-11571.6	4862.4	3175.6	2149.2

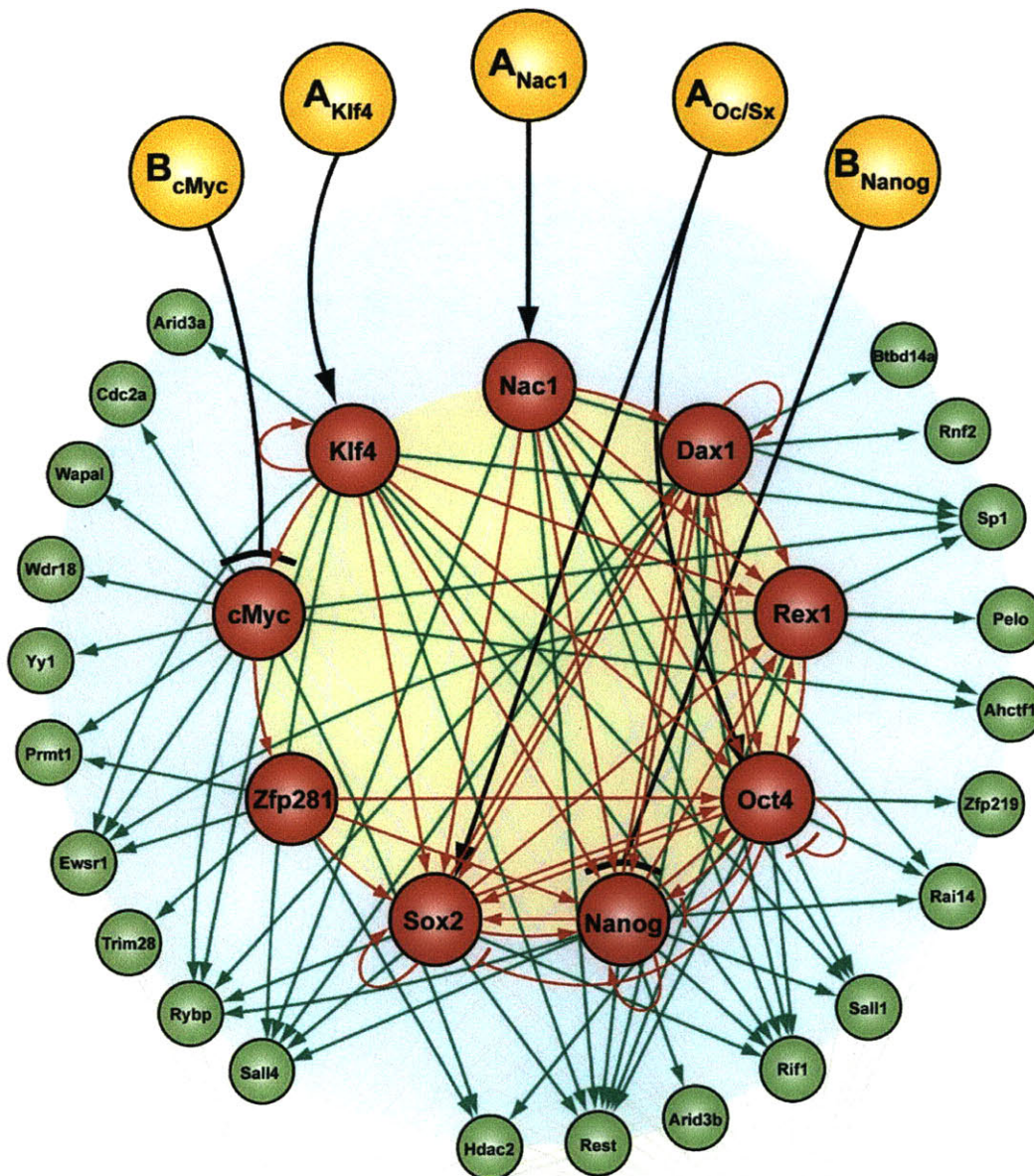
7 ENERGETIC ANALYSIS OF AN EXTENDED PLURIPOTENCY MODEL OF THE TRANSCRIPTIONAL NETWORK OF EMBRYONIC STEM CELLS

7.1 INTRODUCTION

The transcriptional regulatory network model for pluripotency of embryonic stem cells (Figure 7.1) consists of various autoregulatory, feedback and feedforward loops, single input modules and bifan motifs. The ESC regulatory network as seen in figure 7.1 has multilevel network architecture. It can be divided into three levels: (1) Level-0, transcriptional regulatory circuit with three transcription factors Oct4, Sox2 and Nanog; (2) Level-1, inner core of transcriptional circuitry with nine transcription factors Klf4, c-Myc, Zfp281, Sox2, Nanog, Oct4, Rex1, Dax1 and Nac1; and (3) Level-2, combined transcriptional regulatory circuit with target hubs of multiple transcription factors.

Not much has been understood about the occurrence of these motifs and levels in embryonic stem cells network biology. Thus, there has been lack of explanation on how a level architecture in ESC network influences other levels, functional role of various levels, dynamics and functional objectives of ESCs TRN. Moreover, there has been the lack of a regulatory network analysis criterion which can describe the underlying mechanisms behind the occurrence of these layers, motifs, as well as the way in which these motifs play a role in developmental biology and the way that both functional objectives and topological architecture may evolve.

Here, we postulate that the architecture of ESC-TRN can be explained based on a conceptual framework that integrates nonequilibrium thermodynamics with Pareto-optimality of the biological functions to be carried out by the embryonic stem cells. We present an energetic-cost theory that can explain cofactor occupancy as well as the topological arrangement of the ESC network. Through the developed framework, we have tried to answer the questions of why during development certain architecture is favored and what advantage different levels offer for steady state analyses. Our study also demonstrates that ESC network analyses using often-ignored energetics enables identification of new functionalities for various topological arrangements.



**Proliferation
Target Genes**

**Differentiation
Target Genes**

Figure 7.1: Expanded transcriptional regulatory network model for pluripotency of embryonic stem cells. The ESC regulatory network is divided into three levels: (1) Level-0, transcriptional regulatory circuit with three transcription factors Oct4, Sox2 and Nanog; (2) Level-1, inner core of transcriptional circuitry with nine transcription factors Klf4, c-Myc, Zfp281, Sox2, Nanog, Oct4, Rex1, Dax1 and Nac1; and (3) Level-2, combined transcriptional regulatory circuit with target hubs of multiple transcription factors. Arrowhead indicates the direction of transcriptional regulation by transcription factors. Ingoing arrow indicates action of protein as a TF on a gene and outgoing arrow indicates the role of protein as TF. Activators A_{Klf4} , A_{Nac1} and $A_{Oct/Sox}$ moves the ESC cell toward pluripotent state and repressors B_{Nanog} and B_{cMyc} moves the ESC towards differentiated state. Klf4, Sox2, Nanog, and Dax1 TF nodes have positive autoregulation whereas Oct4 TF has concentration dependent negative autoregulation.

7.2 THEORY

7.2.1 Improved Kinetic Transcriptional Regulatory Network (TRN) Model

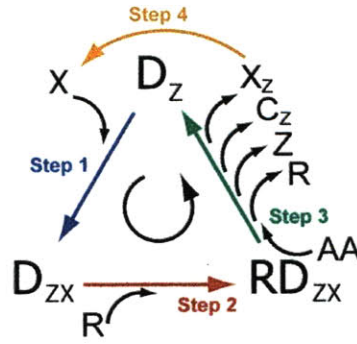
This novel strategy to model the translation and transcription process of any protein Z in a given network not only allows the calculation of the transcription rate but also the estimation of the energy associated with it, in particular, the computation of the heat dissipation rate (HDR). In this approach, each individual transcription factor or the combined set of transcription factors (in case that cooperativity exists) has a cycle associated to it. On the other hand, each cycle is composed of four steps modeled as chemical reactions at nonequilibrium steady state (NESS) which contrasts the basic thermodynamic equilibrium hypothesis in the existing TRN models. At thermodynamic equilibrium steady state, the net flow (J) and chemical potential ($\Delta\mu$) associated to each chemical reaction are null. Since the heat dissipated by each reaction is given by $-J \cdot \Delta\mu$, then at thermodynamic equilibrium the translation and transcription of protein Z is energy free. However, at NESS the net reaction flow is different than zero and its direction is given by the sign of the corresponding chemical potential. If $\Delta\mu < 0$, then $J > 0$, and vice versa. Hence, the heat dissipated by each chemical reaction is always positive, $-J \cdot \Delta\mu > 0$. Notice that $-J \cdot \Delta\mu = 0$ if and only if $\Delta\mu = 0$ and $J = 0$.

Simple activation of protein Z by X : To conceptualize the kinetic TRN model, a simple network consisting of transcription factor (TF) X activating the translation and transcription of protein Z will be used. As there is only one TF, there is only one cycle and thus four steps as explained below (see Figure 7.2).

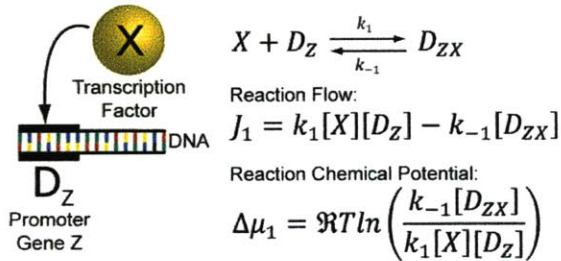
Step 1: TF X binds a free DNA site (D_Z) in the promoter region of gene Z . The free DNA site occupied by X is symbolized by D_{ZX} . If the corresponding chemical reaction $X + D_Z \leftrightarrow D_{ZX}$ has k_1 and k_{-1} as the forward and reverse kinetic constants, then the reaction flow and reaction chemical potential are given by:

$$J_1 = J_1^+ - J_1^- = k_1[X][D_Z] - k_{-1}[D_{ZX}] \quad (7.1)$$

$$\Delta\mu_1 = \mathfrak{R}T \ln \left(\frac{J_1^-}{J_1^+} \right) = \mathfrak{R}T \ln \left(\frac{k_{-1}[D_{ZX}]}{k_1[X][D_Z]} \right) \quad (7.2)$$



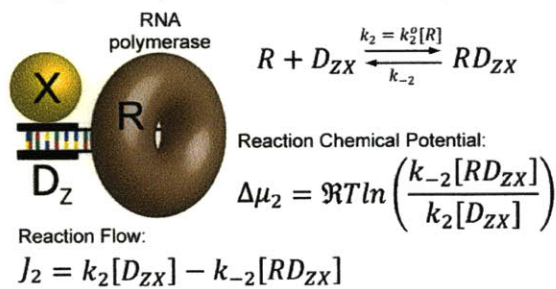
Step 1: TF binds the free DNA promoter region



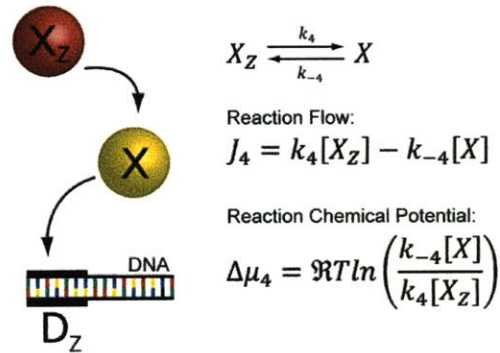
Heat Dissipation Rate (HDR):

$$HDR = -(J_1\Delta\mu_1 + J_2\Delta\mu_2 + J_3\Delta\mu_3 + J_4\Delta\mu_4)$$

Step 2: RNAP binds the TF - DNA complex



Step 4: TF returns to its initial energetic state



Step 3: Transcription and translation of protein Z

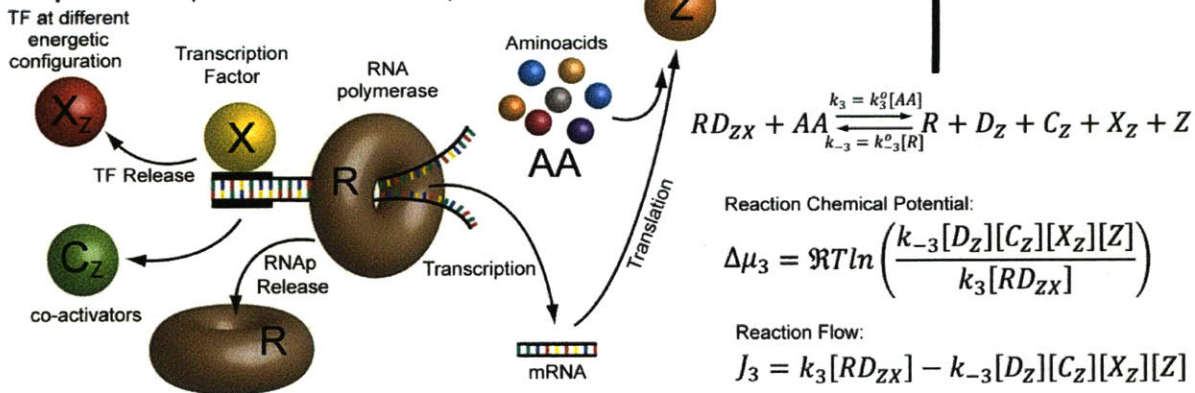


Figure 7.2: Putative kinetic transcriptional regulatory network model and estimation of energetic cost. Here activation of gene Z by transcription factor X is shown.

where \mathfrak{R} is the universal gas constant and T is the temperature. As seen from Equation (7.1), the net reaction flow is the difference between the forward J_1^+ and reverse J_1^- reaction flows.

Step 2: Once the TF X is bound to the promoter region of gene Z , RNA polymerase (R) is recruited to form the complex RD_{ZX} in the following chemical reaction $R + D_{ZX} \leftrightarrow RD_{ZX}$. Under the assumption that $[R] \gg [D_{ZX}]$, the forward kinetic constant can be expressed as $k_2 = k_2^o[R]$, and the net flow and chemical potential for the second step are:

$$J_2 = J_2^+ - J_2^- = k_2[D_{ZX}] - k_{-2}[RD_{ZX}] \quad (7.3)$$

$$\Delta\mu_2 = \mathfrak{R}T \ln \left(\frac{J_2^-}{J_2^+} \right) = \mathfrak{R}T \ln \left(\frac{k_{-2}[RD_{ZX}]}{k_2[D_{ZX}]} \right) \quad (7.4)$$

Step 3: After RNA polymerase is bound, amino acids (AA) are recruited. In a combined translation of mRNA and transcription of protein Z , this step releases the free DNA site, the TF X , co-activators (C_X), and RNA polymerase. However, TF X is released at some different configuration (or energetic state), called X_Z and it is not ready to bound a free DNA site once again. This step can be represented as $RD_{ZX} + AA \leftrightarrow R + D_Z + X_Z + C_Z + Z$. If $[AA] \gg [RD_{ZX}]$, then the forward kinetic constant can be written as $k_3 = k_3^o[AA]$. In addition, at relative high $[R]$, the reverse kinetic constant is given by $k_{-3} = k_{-3}^o[R]$. The net reaction flow and reaction chemical potential are given by:

$$J_3 = J_3^+ - J_3^- = k_3[RD_{ZX}] - k_{-3}[D_Z][C_Z][X_Z][Z] \quad (7.5)$$

$$\Delta\mu_3 = \mathfrak{R}T \ln \left(\frac{J_3^-}{J_3^+} \right) = \mathfrak{R}T \ln \left(\frac{k_{-3}[D_Z][C_Z][X_Z][Z]}{k_3[RD_{ZX}]} \right) \quad (7.6)$$

Step 4: The released TF X_Z returns to its initial energetic state X . Once X is recovered, it can be used again to bind a free DNA site and start a new transcriptional cycle. For this step $X_Z \leftrightarrow X$, the reaction flow and chemical potential are:

$$J_4 = J_4^+ - J_4^- = k_4[X_Z] - k_{-4}[X] \quad (7.7)$$

$$\Delta\mu_4 = \mathfrak{R}T \ln \left(\frac{J_4^-}{J_4^+} \right) = \mathfrak{R}T \ln \left(\frac{k_{-4}[X]}{k_4[X_Z]} \right) \quad (7.8)$$

Because the translation and transcription of Z is an irreversible process, one should expect reactions 1-4 to be highly irreversible ($k_i \gg k_{-i}$). Therefore, the net reaction flows of

steps 1-4 are highly positive (in accordance to NESS) and approximated to the forward reaction flows. The steady state mass balances of the DNA sites (D_Z , D_{ZX} , and RD_{ZX}) reveal that $J_1 = J_2 = J_3 = J_C$, where J_C is the cyclic NESS flow. In addition, the rate at which protein Z is being transcribed corresponds to $J_3 = J_C$. Thus, in this case, J_C is also the transcription rate of Z , TR .

Let us define:

$$\beta = \frac{k_2 k_3 [D_Z]_{tot}}{k_2 + k_3} \quad (7.9)$$

and

$$\kappa = \frac{k_2 k_3}{k_1 (k_2 + k_3)} \quad (7.10)$$

where $[D_Z]_{tot} = [D_Z] + [D_{ZX}] + [RD_{ZX}]$ is the total DNA sites. Then, it can be shown that:

$$J_C = TR = \frac{\beta [X]}{\kappa + [X]} \quad (7.11)$$

Equation (7.11) clearly shows that the transcription rate of protein Z follows a first order Hill's functionality with respect to the concentration of the TF X . This Hill's equation has a maximal transcription rate β ($TR \rightarrow \beta$ when $[X] \rightarrow \infty$) and an activation coefficient κ , defined as the concentration at which half of the maximal transcription rate is obtained, i.e., $TR = \beta/2$ at $[X] = \kappa$.

From the individual mass balances, the concentration of each species can be found:

$$[D_Z] = \frac{1}{k_1} \frac{\beta [X]}{\kappa + [X]} \quad (7.12)$$

$$[D_{ZX}] = \frac{1}{k_2} \frac{\beta [X]}{\kappa + [X]} \quad (7.13)$$

$$[RD_{ZX}] = \frac{1}{k_3} \frac{\beta [X]}{\kappa + [X]} \quad (7.14)$$

$$[Z] = \frac{1}{\alpha_Z} \frac{\beta [X]}{\kappa + [X]} \quad (7.15)$$

$$[C_X] = \frac{1}{\alpha_{C_X}} \frac{\beta [X]}{\kappa + [X]} \quad (7.16)$$

In order to ensure that X is not consumed during transcription but degraded at rate α_X , the constraint $k_4 \gg \alpha_X$ is imposed in the model, leading to $J_4 - J_1 \cong 0$ (J_4 is the rate at which X is being recovered in step 4 after being consumed in step 1 at rate J_1). This constraint gives:

$$[X_Z] = \frac{1}{k_4} \frac{\beta [X]}{\kappa + [X]} \quad (7.17)$$

$$J_4 = J_C = \frac{\beta [X]}{\kappa + [X]} \quad (7.18)$$

Then, the concentrations of each species and the transcription rate of Z can be obtained for a given concentration of the TF X . Computation the reaction chemical potentials given by Equations (7.2), (7.4), (7.6), and (7.8) can be done using Equations (7.12)-(7.17). The heat dissipation rate (HDR) of the i^{th} reaction can then be found as:

$$HDR_i = -J_i \cdot \Delta\mu_i = -J_C \cdot \Delta\mu_i \quad (7.19)$$

And the total heat dissipation rate (HDR) can be obtained as the sum of the individual contribution of the heat dissipated by each reaction:

$$HDR = -J_C \cdot \Delta\mu_C \quad (7.20)$$

where $\Delta\mu_C = \Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 + \Delta\mu_4$ is the cyclic chemical potential. From Equations (7.2), (7.4), (7.6), and (7.8) it can be shown that:

$$\Delta\mu_C = \mathfrak{R}T \ln \left(\frac{k_{-1}k_{-2}k_{-3}k_{-4}[C_Z][Z]}{k_1k_2k_3k_4} \right) \quad (7.21)$$

Notice that if $k_3 \gg k_2$, then from Equations (7.9) and (7.10), $k_2 = \beta/[D_Z]_{tot}$ and $k_1 = \beta/(\kappa[D_Z]_{tot})$. Then:

$$\Delta\mu_C = \mathfrak{R}T \ln \left(\frac{\delta\kappa[C_Z][Z]}{\beta^2} \right) \quad (7.22)$$

with

$$\delta = \frac{k_{-1}k_{-2}k_{-3}k_{-4}[D_Z]_{tot}^2}{k_3k_4} \quad (7.23)$$

Finally, the heat dissipation rate can be expressed as:

$$HDR = \frac{\beta[X]}{\kappa+[X]} \mathfrak{R}T \ln \left(\frac{\beta^2}{\delta\kappa[C_Z][Z]} \right) \quad (7.24)$$

Activation of protein Z by multiple transcription factors: When $N + 1$ transcription factors X_0, X_1, \dots, X_N activate the translation and transcription of Z , there will be N cycles each of those composed of the four steps previously described. If cooperativity exists between two or more TFs, for instance X_1 and X_2 can both individually and jointly activate Z , then a new TF is defined as $[X_1X_2] = [X_1][X_2]$ and an additional cycle is incorporated. As the model assumes competitive binding of the TFs at the available DNA site, $[D_Z]$ is common in the first step of each cycle. Thus, the total DNA sites is given by $[D_Z]_{tot} = [D_Z] + \sum_{j=0}^{N+1} D_{ZX_j} + \sum_{j=0}^{N+1} RD_{ZX_j}$. When the mass balances of the sites are solved at steady state, it can be shown as before that $J_C^{ZX_j} = J_1^{ZX_j} =$

$J_2^{ZX_j} = J_3^{ZX_j}$, where $J_i^{ZX_j}$ is the i^{th} net reaction flow in the corresponding cycle of transcription of Z by X_j . Because the rate at which Z is being transcribed is $\sum_{j=1}^N J_3^{ZX_j}$, then it can be shown that:

$$J_C^{ZX_j} = \frac{\beta^{ZX_j} [X_j]}{\kappa^{ZX_j} \left(1 + \sum_{j=0}^{N+1} \frac{[X_j]}{\kappa^{ZX_j}} \right)} \quad (7.25)$$

$$TR = \sum_{j=1}^N J_C^{ZX_j} = \frac{\sum_{j=0}^{N+1} \frac{\beta^{ZX_j} [X_j]}{\kappa^{ZX_j}}}{1 + \sum_{j=0}^{N+1} \frac{[X_j]}{\kappa^{ZX_j}}} \quad (7.26)$$

where the maximal transcriptional rate of the cycle ZX_j , β^{ZX_j} , and the activation coefficient of the cycle ZX_j , κ^{ZX_j} are defined by Equations (7.9) and (7.10), with $k_i = k_i^{ZX_j}$.

When the following auxiliary variables are defined:

$$a^{ZX_j} = k_2^{ZX_j} k_3^{ZX_j} \quad (7.27)$$

$$b^{ZX_j} = k_1^{ZX_j} k_3^{ZX_j} [X_j] \quad (7.28)$$

$$c^{ZX_j} = k_1^{ZX_j} k_2^{ZX_j} [X_j] \quad (7.29)$$

$$d^{ZX_j} = a^{ZX_j} + b^{ZX_j} + c^{ZX_j} \quad (7.30)$$

$$d_Z = \left(\sum_{j=0}^{N+1} \frac{d^{ZX_j} \prod_{j=0}^{N+1} a^{ZX_j}}{a^{ZX_j}} \right) - N \prod_{j=0}^{N+1} a^{ZX_j} \quad (7.31)$$

then, the concentrations of the individual species can be found by solving the corresponding mass balance equation at steady state:

$$[D_Z] = \frac{\prod_{j=0}^{N+1} a^{ZX_j}}{d_Z} [D_Z]_{tot} \quad (7.32)$$

$$[D_{ZX_j}] = \frac{b^{ZX_j} \prod_{j=0}^{N+1} a^{ZX_j}}{a^{ZX_j} d_Z} [D_Z]_{tot} \quad (7.33)$$

$$[RD_{ZX_j}] = \frac{c^{ZX_j} \prod_{j=0}^{N+1} a^{ZX_j}}{a^{ZX_j} d_Z} [D_Z]_{tot} \quad (7.34)$$

$$[Z] = \frac{TR}{\alpha_Z} \quad (7.35)$$

$$[C_{X_j}] = \frac{J_C^{ZX_j}}{\alpha_{C_{X_j}}} \quad (7.36)$$

Under the assumption that $k_4^{ZX_j} \gg \alpha_{X_{Zj}}$ (to ensure only degradation of X_j):

$$[X_{Zj}] = \frac{J_C^{ZXj}}{k_4^{ZXj}} \quad (7.37)$$

$$J_4^{ZXj} = J_C^{ZXj} \quad (7.38)$$

The chemical potentials of reactions 1-4 for each cycle ZX_j can be calculated by using Equations (7.32)-(7.37) into Equations (7.2), (7.4), (7.6) and (7.8), and the heat dissipation rate of each cycle by using Equation (7.20). Finally, the network HDR corresponds to the sum of the heat dissipated through each cycle.

Presence of basal activity: When basal activity exists, the model assumes a basal transcription factor B at constant concentration that also competes for the free DNA site D_Z . Basically, this effect is introduced by assigning TF $X_0 = B$ into Equations (7.25)-(7.38).

Simple repression of protein Z by X : The simplest repression model, i.e., TF X represses the translation and transcription activity of protein Z , is composed of two cycles: a basal activity cycle ZB and a repression cycle ZX . Since cycle ZB leads to activation, this cycle is formed by the four steps previously described. On the other hand, the repression cycle has four reactions as well, but with the following modifications due to the repressive nature of X :

Step 1: As in the activation case, X competes with B for D_Z and forms the complex D_{ZX} in the reaction $X + D_Z \leftrightarrow D_{ZX}$. Hence, Equations (7.1) and (7.2) apply, but the forward and reverse kinetic constants k_1 and k_{-1} may have different values.

Step 2: Because X is a repressor, RNA polymerase is not recruited. In turn, the complex D_{ZX} experiences a change into another energetic state and changes its configuration to D_{ZX}^* as $D_{ZX} \leftrightarrow D_{ZX}^*$. Let k_2 and k_{-2} the forward and reverse kinetic constants for this step, respectively, with different values than in the activation case. Then, the net reaction flow and chemical potential for this step are given by:

$$J_2 = J_2^+ - J_2^- = k_2[D_{ZX}] - k_{-2}[D_{ZX}^*] \quad (7.39)$$

$$\Delta\mu_2 = \mathfrak{R}T \ln\left(\frac{J_2^-}{J_2^+}\right) = \mathfrak{R}T \ln\left(\frac{k_{-2}[D_{ZX}^*]}{k_2[D_{ZX}]}\right) \quad (7.40)$$

Step 3: The TF-DNA complex D_{ZX}^* returns to its initial state D_Z by releasing co-repressors (C_X) and the TF X in a different energetic state, X_Z in the corresponding reaction $D_{ZX}^* \leftrightarrow D_Z + C_X + X_Z$. Since RNA polymerase is not bound in the previous step, no amino acids are recruited and protein Z is not formed. However, protein Z is being transcribed from the basal TF B that competes with X for the free D_Z . When no X is present, TF B can bind every available D_Z and transcribe the highest amount of Z . But when the concentration of X increases, TF B sees less D_Z and the availability to drive the cycle ZB and its corresponding cyclic flow decreases. If k_3 and k_{-3} are the forward and reverse kinetic constants for this step, respectively, the repression flow and chemical potential are:

$$J_3 = J_3^+ - J_3^- = k_3[D_{ZX}^*] - k_{-3}[D_Z][C_Z][X_Z] \quad (7.41)$$

$$\Delta\mu_3 = \mathfrak{R}T \ln\left(\frac{J_3^-}{J_3^+}\right) = \mathfrak{R}T \ln\left(\frac{k_{-3}[D_Z][C_Z][X_Z]}{k_3[D_{ZX}^*]}\right) \quad (7.42)$$

The kinetic constants k_3 and k_{-3} are different than in the activation case.

Step 4: Same as in the activation case, the released TF X_Z returns to its initial configuration X and is ready to bind a new free DNA site as $X_Z \leftrightarrow X$. Equations (7.7) and (7.8) are valid but kinetic constants may have different values as in the activation case.

From the individual mass balances of the DNA species and using by $[D_Z]_{tot} = [D_Z] + \sum_{j=0}^1 D_{ZX_j} + \sum_{j=0}^1 RD_{ZX_j}$ with $X_0 = B$ and $X_1 = X$, it can be shown that the equation for the cyclic flow of each cycle given by Equation (7.25) remains the same and $J_C^{ZX_j} = J_1^{ZX_j} = J_2^{ZX_j} = J_3^{ZX_j}$. However, because Z is only transcribed from the third step of the ZB cycle, then:

$$TR = J_C^{ZB} = \frac{\frac{\beta^{ZB}[B]}{\kappa^{ZB}}}{1 + \frac{[B]}{\kappa^{ZB}} + \frac{[X]}{\kappa^{ZX}}} \quad (7.43)$$

When $k_4^{ZX_j} \gg \alpha_{X_{Zj}}$ is assumed for $X_j = B, X$, Equations (7.27)-(7.38) remain unchangeable with exception of Equation (7.34) in which RD_{ZX} becomes D_{ZX}^* in the case of repression.

Generalization of the kinetic TRN model for multiple TFs with activation-repression nature: Let us finally consider the most general case where $N + 1$ transcription factors X_0, X_1, \dots ,

X_N regulate the transcription activity of protein Z through activation and repression. Let us define the following vectors: transcription factor vector $\mathbf{x} = [[X_0] [X_1] \dots [X_N]]^T$; activation/repression logic vector \mathbf{l} whose j^{th} element is 1 if X_j is an activator and 0 if X_j is a repressor (notice that $l_0 = 1$ when basal activity is present, with $X_0 = B$); vector \mathbf{d}_{ZX}^3 whose j^{th} element is $[RD_{ZX}]$ if $l_j = 1$ and $[D_{ZX}^*]$ if $l_j = 0$. Then, the total mass balance for the DNA sites can be written as $[D_Z]_{\text{tot}} = [D_Z] + \sum_{j=0}^{N+1} [D_{ZX}] + \sum_{j=0}^{N+1} d_{ZX,j}^3$ and from the individual mass balance of the DNA sites it can be shown that $J_C^{ZX_j} = J_i^{ZX_j}$ ($i = 1, 2, 3$) with $J_C^{ZX_j}$ given by Equation (7.25). Since transcription of Z occurs only from the activation cycles, then the total transcription rate of Z is given by:

$$TR = \mathbf{j}_C \cdot \mathbf{l} \quad (7.44)$$

where $\mathbf{j}_C = [J_C^{ZX_0} \dots J_C^{ZX_N}]^T$. The individual concentrations of the species can be obtained by using Equations (7.27)-(7.37). However, $[RD_{ZX_j}]$ in Equation (7.34) should be replaced by $d_{ZX,j}^3$. Once the species concentrations are obtained, the chemical potentials of reactions 1 and 4 can be calculated using Equations (7.2) and (7.8), respectively, but the expression of the chemical potentials of reactions 2 and 3 should be reformulated as follows:

$$\Delta\mu_2^{ZX_j} = \mathfrak{R}T \ln \left(\frac{k_{-2}^{ZX_j} d_{ZX,j}^3}{k_2^{ZX_j} [D_{ZX_j}]} \right) \quad (7.45)$$

$$\Delta\mu_3^{ZX_j} = \mathfrak{R}T \ln \left(\frac{k_{-3}^{ZX_j} [D_Z] [C_{ZX_j}] [X_Z] z_j}{k_3^{ZX_j} d_{ZX,j}^3} \right) \quad (7.46)$$

where the j^{th} element of vector \mathbf{z} is 1 if $l_j = 0$ and $[Z]$ if $l_j = 1$.

Based on Equation (7.24), the general expression of the heat dissipation rate is given by:

$$HDR^{ZX_j} = \frac{\beta^{ZX_j} [X_j]}{\kappa^{ZX_j} + [X_j]} \mathfrak{R}T \ln \left(\frac{\beta^{ZX_j} z_j^2}{\delta^{ZX_j} \kappa^{ZX_j} [C_{ZX_j}] z_j} \right) \quad (7.47)$$

7.2.2 General Michaelis-Menten Formulation

In general, the Michaelis-Menten approach assumes binding of the TFs in the promoter region of gene Z at thermodynamic equilibrium and that the total transcription rate is proportional to the concentration of DNA sites bound by RNA polymerase. Let us assume that

the translation and transcription activity of protein Z is regulated by N_A activation TFs (X_j^A), N_R repressive TFs (X_j^R) and RNA polymerase (R), with $N = N_A + N_R$. In a competitive binding X_j^A , X_j^R , or R can bind a free DNA site in the promoter region D_Z for transcription of protein Z , forming a transcription factor-DNA site complex $X_j D_Z$: $X_j + D_Z \leftrightarrow X_j D_Z$, where X_j represents either X_j^A , X_j^R , or R . Let $K_1^{ZX_j}$ be the association equilibrium constant of this step. Then, at thermodynamic equilibrium:

$$K_1^{ZX_j} [X_j] [D_Z] = [X_j D_Z] \quad (7.48)$$

If X_j is an activator ($X_j = X_j^A$), then RNA polymerase (R) is recruited by the activator X_j^A bound to D_Z : $R + X_j^A D_Z \leftrightarrow R X_j^A D_Z$. If $K_2^{ZX_j}$ is the association equilibrium constant of this reaction, then at thermodynamic equilibrium:

$$K_2^{ZX_j} [R] [X_j^A D_Z] = [R X_j^A D_Z] \quad (7.49)$$

If X_j is a repressor for the expression of Z ($X_j = X_j^R$), then RNA polymerase is not bound and translation does not occur. The total number of sites in the promoter region for transcription of Z can be expressed as follows:

$$[D_Z]_{tot} = [D_Z] + [R D_Z] + \sum_{j=1}^{N_A} [X_j D_Z] + \sum_{j=1}^{N_A} [R X_j^A D_Z] = [D_Z] \left(1 + K_1^{ZR} [R] + \sum_{j=1}^{N_A} K_1^{ZX_j} [X_j] + \sum_{j=1}^{N_A} K_2^{ZX_j} [R] [X_j^A] \right) \quad (7.50)$$

As initially mentioned, the transcription rate (TR) of protein Z is proportional to the total number of sites bound by RNA polymerase $TR = k_Z ([X_0 D_Z] + \sum_{j=1}^{N_A} [R X_j^A D_Z])$, with k_Z being the proportionality constant. Using Equations (7.48)-(7.50), it can be shown that:

$$TR = \frac{k_Z [D_Z]_{tot} [R] \left(K_1^{ZR} + \sum_{j=1}^{N_A} K_1^{ZX_j} K_2^{ZX_j} [X_j^A] \right)}{1 + K_1^{ZR} [R] + \sum_{j=1}^{N_A} K_1^{ZX_j} [X_j^A] \left(1 + K_2^{ZX_j} [R] \right) + \sum_{j=1}^{N_R} K_1^{ZX_j} [X_j^R]} \quad (7.51)$$

By defining the following variables, $\eta^{ZR} = k_z[D_Z]_{tot}[R]K_1^{ZR}$, $a^{ZX_j^A} = k_z[D_Z]_{tot}[R]K_1^{ZX_j}K_2^{ZX_j}$, $\gamma^{ZR} = K_1^{ZR}[R]$, $b^{ZX_j^A} = K_1^{ZX_j}(1 + K_2^{ZX_j}[R])$, and $b^{ZX_j^R} = K_1^{ZX_j}$, the transcription rate of protein of Z can be expressed as:

$$TR = \frac{\eta^{ZR} + \sum_{j=1}^{N_A} a^{ZX_j^A} [X_j^A]}{1 + \gamma^{ZR} + \sum_{j=1}^{N_A} b^{ZX_j^A} [X_j^A] + \sum_{j=1}^{N_R} b^{ZX_j^R} [X_j^R]} \quad (7.52)$$

7.2.3 Relationship Between the Kinetic TRN Model and the Michaelis-Menten Formulation

It can be shown that the transcription rate given by the kinetic transcriptional regulatory model [Equation (7.44)] matches the one obtained from thermodynamic equilibrium hypothesis [Equation (7.51)] by having:

$$\alpha^{ZX_j} = \frac{\beta^{ZX_j}}{\kappa^{ZX_j}} = k_1^{ZX_j} [D_Z]_{tot} \quad (7.53)$$

$$b^{ZX_j} = \frac{1}{\kappa^{ZX_j}} = \frac{k_1^{ZX_j}(k_2^{ZX_j} + k_3^{ZX_j})}{k_2^{ZX_j} k_3^{ZX_j}} \quad (7.54)$$

$$\eta^{ZR} = \frac{\beta^{ZB}}{\kappa^{ZB}} = k_1^{ZB} [B][D_Z]_{tot} \quad (7.55)$$

$$\gamma^{ZR} = \frac{1}{\kappa^{ZB}} = \frac{k_1^{ZB}(k_2^{ZB} + k_3^{ZB})[B]}{k_2^{ZB} k_3^{ZB}} \quad (7.56)$$

Based on Equation (7.22), the chemical potential of the cycle ZX_j can be expressed as

$$\Delta\mu_c^{ZX_j} = \mathfrak{R}T \ln \left(\frac{\delta^{ZX_j} b^{ZX_j} [c_{ZX_j}]^{z_j}}{a^{ZX_j}{}^2} \right) \quad (7.57)$$

7.2.4 Total Transcription Rate

The ordinary differential equation describing the mass balance of Z_i , for $i = 1 \dots N$, is given by:

$$f_i = \frac{d[Z_i]}{dt} = TR_i - \alpha_i Z_i \quad (7.58)$$

where TR_i is the transcription rate of Z_i given by either Equation (7.44) (kinetic TRN model) or Equation (7.51) (Michaelis-Menten approach), and α_i is the degradation rate of Z_i . In general, the transcription rate of protein Z_i in a network is a function of the input vector of external activators and repressors (\mathbf{x}) and a function of the vector containing all the parameters of the model (\mathbf{p}). Thus,

$$TR_i = TR_i(\mathbf{x}, \mathbf{p}) \quad (7.59)$$

7.2.5 Robustness Index

In this work, we define the sensitivity of the transcription rate of protein Z_i with respect to the input vector \mathbf{x} (or input sensitivity) as the relative change in the transcription rate of Z_i when the input vector \mathbf{x} is perturbed (or amplified) by a scalar λ_x for a constant set of parameters:

$$S_i^{\mathbf{x}} = \frac{TR_i(\mathbf{x}, \mathbf{p}) - TR_i(\lambda_x \mathbf{x}, \mathbf{p})}{TR_i(\mathbf{x}, \mathbf{p})} \quad (7.60)$$

In an analogous way, we define the parametrical sensitivity of the transcription rate of Z_i as the relative change in TR_i for a perturbation of magnitude λ_p in the parametrical vector \mathbf{p} at constant \mathbf{x} :

$$S_i^{\mathbf{p}} = \frac{TR_i(\mathbf{x}, \mathbf{p}) - TR_i(\mathbf{x}, \lambda_p \mathbf{p})}{TR_i(\mathbf{x}, \mathbf{p})} \quad (7.61)$$

Both input and parametrical sensitivities can be grouped in a sensitivity matrix \mathbf{S} whose columns are the input sensitivity vector and the parametrical sensitivity vector. For a network composed of N transcription factors, the sensitivity matrix is defined as:

$$\mathbf{S} = \begin{bmatrix} S_1^{\mathbf{x}} & S_1^{\mathbf{p}} \\ \vdots & \vdots \\ S_N^{\mathbf{x}} & S_N^{\mathbf{p}} \end{bmatrix} = [\mathbf{S}^{\mathbf{x}} \quad \mathbf{S}^{\mathbf{p}}] \quad (7.62)$$

As a measurement of the network sensitivity, we also define the sensitivity index (SI) of the network as the Frobenius norm (also known as the Euclidian norm) of the sensitivity matrix \mathbf{S} :

$$SI = \|\mathbf{S}\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^2 |S_{ij}|^2} \quad (7.63)$$

Here, we defined the robustness index (RI) as the inverse of the network sensitivity:

$$RI = \frac{1}{SI} \quad (7.64)$$

If the total transcriptional rate of the network barely changes when both perturbations in the inputs and parameters are introduced, then the overall sensitivity is small and the robustness of the network is high (represented by the robustness index). On the other hand, if small changes in the inputs and parameters lead to high relative changes in the total transcription rate, then the network is highly sensitive and thus lacks of robustness.

As the mass balance solutions of the kinetic TRN model reduces to the Michaelis-Menten formalism, and as seen from Equations (7.53) to (7.56), a perturbation of λ_p in \mathbf{p} will induce a perturbation of magnitude λ_p^2 in a^{ZX_j} and no change in b^{ZX_j} . The basal parameters η^{ZR} and γ^{ZR} will get perturbed by λ_p^3 and λ_p , respectively. To avoid this effect in the total transcription rate, only the forward kinetic constant of the first step is affected when \mathbf{p} is amplified by λ_p (i.e., only vector \mathbf{k}_1), all the four backward kinetic constants (vectors \mathbf{k}_{-i}) and the degradation rate vector α .

7.3 RESULTS

7.3.1 Insights From A Basic TRN Model of ESCs or Level-0

This simple model of ESCs (Chickarmane et al., 2006) is composed of three core TFs: Oct4 (O_c), Sox2 (S_x), and Nanog (N_g) (see Figure 7.3A). Both Oct4 and Sox2 are activated by an external TF A_{O_c/S_x} . When A_{O_c/S_x} binds the promoter region of Oct4 (D_{O_c}) and Sox2 (D_{S_x}), translation and transcription of these proteins occur. In a protein-protein interaction, Oct4 and Sox2 bind to form the complex Oct4-Sox2 (O_S) which also acts as a TF of activation of both Oct4 and Sox2, creating a positive feedback loop. This protein complex also binds the free DNA site in the promoter region of the third core TF Nanog (D_{N_g}), leading to transcription of Nanog. Nanog in turn cooperates with the protein complex Oct4-Sox2 to further activate Oct4, Sox2 and itself, forming additional positive feedback loops. Although Nanog can be externally activated, a most efficient activation of the system can be achieved by manipulation of A_{O_c/S_x} . On the other hand, repression or deactivation of the network can be better obtained by externally repressing Nanog through the external repressor B_{N_g} instead of externally repress both Oct4 and Sox4 (Chickarmane et al., 2006).

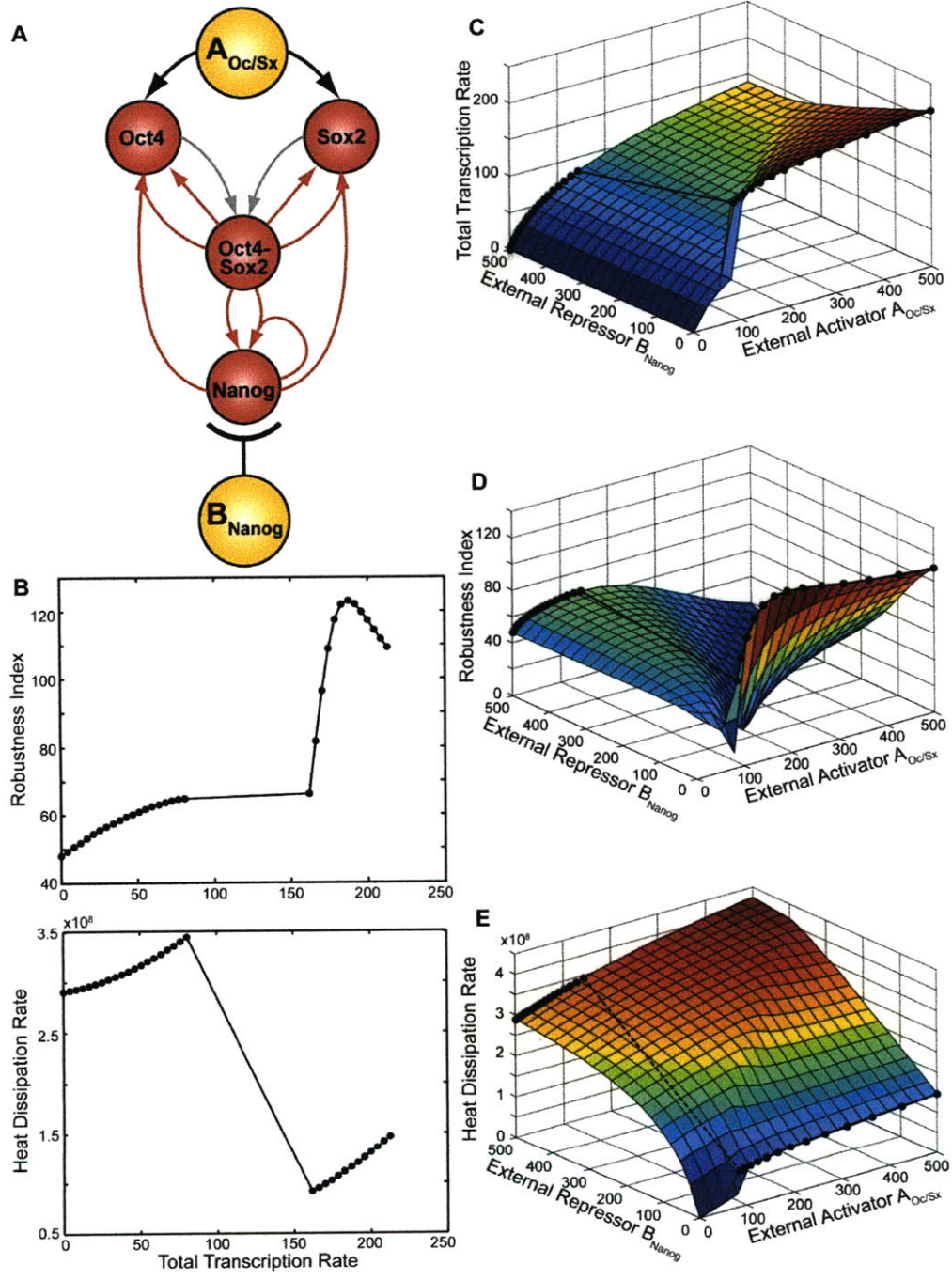


Figure 7.3. Pareto optimality and energetic analysis of Level-0 ESC regulatory circuitry at steady state. **A)** Level-0 circuit consisting of various positive feedback interactions between Oct4, Sox2 and Nanog TFs is shown. **B)** Pareto optimal analysis reveals existence of two separate Pareto frontiers. In pluripotent state, transcriptional rate is maximal and Pareto frontier is between ESC objectives maximal robustness and maximal transcription rate. Second, Pareto frontier exists in differentiated state, i.e. lowest transcription rate region and there the desirability is to minimize transcription and maximize robustness. The corresponding energetic cost (heat dissipation rate) estimation reveals lower cost for max-max Pareto frontier in self-renewal state and higher cost for min-max Pareto frontier in differentiate state. **D)-E)** Corresponding Pareto optimal activator and repressor concentration shows switch in B_{Ng} concentration from low to high when moving from self renewal state to the differentiate state. There is a continual decrease in $A_{Oc/Sx}$ concentration when moving from self-renewal state to differentiate state. Similar behavior was observed for robustness index.

The representation of this network by the kinetic TRN model is based on three sets (one for each core TF) composed of four cycles each one and one biochemical reaction for the protein-protein interaction. Both transcription of Oct4 and Sox2 contain one cycle for basal activator (C_{Oc} and C_{Sx} , respectively), another cycle for the external activator $A_{Oc/Sx}$, a third cycle for activation through OS , and a fourth cycle for the cooperativity interaction $OS \cdot Ng$. Transcription of Nanog also takes into account basal TF activation (C_{Ng}), a cycle for the external repressor B_{Ng} , a cycle for activation by OS , and a fourth cycle for the autoregulatory activation with OS , $OS \cdot Ng$. Formation of the protein complex Oct4-Sox2 is modeled as a single biochemical reaction $Oc + Sx \leftrightarrow OS$ with k_f^{OS} and k_r^{OS} being the forward and reverse kinetic constant for this reaction. Table 7.1 summarizes the parameters of the kinetic TRN model.

Table 7.1: Parameters of the kinetic TRN model for the basic ESCs network. All chemical potentials were obtained at $T = 298.15$ [K]. For the protein-protein interaction $Oc + Sx \leftrightarrow OS$, $k_f^{OS} = 0.05$, $k_r^{OS} = 10^{-3}$, $\alpha_{OS} = 5$. All other degradation rates were set at $\alpha = 1$.

	Transcription of Oct4 (Oc) [D_{Oc}] = 1, [C_{Oc}] = 10^{-4}				Transcription of Sox2 (Sx) [D_{Sx}] = 1, [C_{Sx}] = 10^{-4}				Transcription of Nanog (Ng) [D_{Ng}] = 1, [C_{Ng}] = 10^{-4}			
	C_{Oc}	$A_{Oc/Sx}$	OS	$OSNg$	C_{Sx}	$A_{Oc/Sx}$	OS	$OSNg$	C_{Ng}	B_{Ng}	OS	$OSNg$
k_1	1	1	0.01	0.2	1	1	0.01	0.2	1	10	$5 \cdot 10^{-3}$	0.1
k_2	10^3	910	10	286	10^3	910	10	286	10^3	10^3	5	100
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^9	10^6	10^6
k_4	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-6}	10^{-6}
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-6}	10^{-6}
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-6}	10^{-6}
k_{-4}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}
l	1	1	1	1	1	1	1	1	1	0	1	1

For the basic TRN of ESCs, the input and parameter vectors are given by

$$\mathbf{x} = [A_{Oc/Sx} \ B_{Ng}]^T \quad (7.65)$$

$$\mathbf{p} = [k_1^T \ k_2^T \ k_3^T \ k_4^T \ k_{-1}^T \ k_{-2}^T \ k_{-3}^T \ k_{-4}^T \ k_{OS}^T \ D^T \ C^T \ \alpha^T]^T \quad (7.66)$$

where vectors k_i and k_{-i} ($i = 1 \dots 4$) contain the forward and reverse kinetic parameters of step i for the twelve cycles of the basic TRN of ESC model. Vector k_{OS} contain the forward and reverse kinetic constants for the protein-protein interaction between Oct4 and Sox2. Vectors D , C , and α contain the concentration of total available DNA sites, basal transcription factors and degradation rates, respectively (all of these parameters are specified in Table 7.1). Once vectors

\mathbf{x} and \mathbf{p} are specified, the transcriptional network is fully described. At steady state, $f_i = 0$ and both Z_i and TR_i can be calculated. Then, the total transcription rate for this Oct4/Sox2/Nanog network is calculated as the sum of the individual transcription rates:

$$TR_{tot} = TR_{Oc} + TR_{Sx} + TR_{Ng} \quad (7.67)$$

To calculate the robustness index, perturbations of $\lambda_x = 1.01$ and $\lambda_p = 1.01$ were used. The sensitivity matrix [Equation (7.62)] for this network is:

$$\mathbf{S} = \begin{bmatrix} S_{Oc}^x & S_{Oc}^p \\ S_{Sx}^x & S_{Sx}^p \\ S_{Ng}^x & S_{Ng}^p \end{bmatrix} \quad (7.68)$$

As seen in Figure 7.3B, Pareto optimality analysis of Level-0 ESC network revealed existence of two separate Pareto frontiers. In pluripotent state, transcriptional rate is highest and Pareto frontier is between ESC objectives of maximal robustness and maximal transcription rate. This is not surprising because in the proliferation stage ESCs objective is to have maximal expression of genes related to proliferation and also maintain this high proliferation gene expression under external influences. Second Pareto frontier exists in differentiated state, i.e. lowest transcription rate region and here the desirability is to minimize transcription and maximize robustness. This can be understood from the fact that in differentiated state in ESCs it will be desirable to have both minimum expression of proliferation genes and high robustness. As seen in Figure 7.3 C, the corresponding energetic cost (heat dissipation rate) estimation reveals lower cost for max-max Pareto frontier in self-renewal state and higher cost for min-max Pareto frontier in differentiate state. Interestingly, we observed bistability (Figure 7.3 D) for total transcription rate as a function of external inputs B_{Ng} and $A_{oc/sx}$. Moreover, the corresponding Pareto optimal activator and repressor concentration also shows switch in B_{Ng} concentration from low to high when moving from self renewal state to the differentiate state. Furthermore, there is a continual decrease in $A_{oc/sx}$ concentration when moving from self-renewal state to differentiate state. Similar behavior was observed for robustness index.

7.3.2 The Extended Pluripotent TRN of ESCs. Level-1

This network is composed of two interaction levels: a first level containing nine core TFs (including Oct4, Sox2 and Nanog) with positive feedback loops; and a second level of 22 target genes (TGs) for the nine core TFs (see Figure 7.1).

Level -1 contains two master regulators which are externally activated: Klf4 (Kf) and Nac1 (Nc) activated by A_{Kf} and A_{Nc} , respectively. Klf4 binds the free DNA site in the promoter region of cMyc (Mc), Sox2 (Sx), Nanog (Ng), Oct4 (Oc), Rex1 (Rx), as well as its own promoter region for positive autoregulation (Jiang et al., 2008; Kim et al., 2008). On the other hand, Nac1 activates the transcription of Sox2, Nanog, Oct4, Rex1, and Dax1 (Dx), but does not have positive autocontrol (Kim et al., 2008). TF cMyc plays the role of activating the transcription of a downstream TF, Zfp281 (Zp) which in turn binds the promoter region of Sox2, Nanog, and Oct4 (Kim et al., 2008). This positive effect in the activation of Sox2, Nanog, and Oct4 by cMyc is externally controlled by an external repressor of cMyc, B_{Mc} . Dax1 also contributes to the activation of Oct4, Sox2, and Nanog, in addition to the activation of Rex1 and itself (Kim et al., 2008).

The interactions between Oct4, Sox2, and Nanog are basically the same as in the basic TRN model of ESCs but with two major differences: the complex Oct4-Sox2 (OS) is not explicitly formed through a protein-protein interaction and negative regulations of Oct4 on Sox2, Nanog and itself are added. Because many authors do not agree with the existence of the complex Oct4-Sox2, we have decided not to include this explicit reaction (Chickarmane and Peterson, 2008; Kim et al., 2008). Instead, a cooperativity interaction between Oct4 and Sox2 replaces the explicit transcriptional activity of Oct4-Sox2. Similarly, the positive feedback loop due to the joint interaction of Oct4-Sox2 and Nanog is reformulated as a cooperative binding between Oct4, Sox2 and Nanog. The second modification to the basic TRN model of ESCs takes into account the repressive nature of Oct4 on Sox2, Nanog and itself at low concentrations (Chickarmane and Peterson, 2008; Pan et al., 2006; Pan and Thomson, 2007).

Table 7.2: Transcriptional logic for the Level 1 of the extended pluripotent TRN model. Entry (i, j) of the table represents TF j binding the promoter region of TF i . Element $(i, j) = 0$ if no interaction exists, 1 if TF j is an activator of i , and -1 if TF j is a repressor of TF i .

TF	$A_{Oc/Sx}$	A_{Kf}	A_{Mc}	B_{Ng}	B_{Mc}	Oc	Sx	Ng	Rx	Dx	Mc	Zp	Kf	Nc
Oct4	1	0	0	0	0	± 1	1	1	1	1	0	1	1	1
Sox2	1	0	0	0	0	± 1	1	1	0	1	0	1	1	1
Nanog	0	0	0	-1	0	± 1	1	1	0	1	0	1	1	1
Rex1	0	0	0	0	0	1	1	1	0	1	0	0	1	1
Dax1	0	0	0	0	0	1	1	1	0	1	0	0	0	1
cMyc	0	0	0	0	-1	0	0	0	0	0	0	0	1	0
Zfp281	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Klf4	0	1	0	0	0	0	0	0	0	0	0	0	1	0
Nac1	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 7.3: Parameters for Level 1 of the kinetic TRN model for the extended pluripotent ESCs network. All chemical potentials were obtained at $T = 298.15$ [K] and all degradation rates were set at $\alpha = 10$.

Transcription of Oct4: $[D_{Oc}] = 1, [C_{Oc}] = 10^{-4}$											
	C_{Oc}	$A_{Oc/Sx}$	Kf	Nc	$OcSx$	$OcSxNg$	Rx	Dx	Zp	Oc	
k_1	1	1	0.5	0.5	0.01	0.2	0.5	0.5	0.5	10	
k_2	10^3	10^4	500	500	10	286	500	500	500	10^3	
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^9	
k_4	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-4}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	
l	1	1	1	1	1	1	1	1	1	0	
Transcription of Sox2: $[D_{Sx}] = 1, [C_{Sx}] = 10^{-4}$											
	C_{Sx}	$A_{Oc/Sx}$	Kf	Nc	$OcSx$	$OcSxNg$	Dx	Zp	Oc		
k_1	1	1	0.5	0.5	0.01	0.2	0.5	0.5	10		
k_2	10^3	10^4	500	500	10	286	500	500	10^3		
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^9		
k_4	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6		
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	
k_{-4}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	
l	1	1	1	1	1	1	1	1	1	0	
Transcription of Nanog: $[D_{Ng}] = 1, [C_{Ng}] = 10^{-4}$											
	C_{Ng}	Kf	Nc	$OcSx$	$OcSxNg$	Dx	Zp	Oc	B_{Ng}		
k_1	1	0.25	0.25	0.005	0.1	0.25	0.25	10	10		
k_2	10^3	250	250	5	143	250	250	10^3	$2 \cdot 10^3$		
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^9	10^9		
k_4	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6		
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-12}	
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-12}	
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-12}	10^{-12}	
k_{-4}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	
l	1	1	1	1	1	1	1	1	0	0	
Transcription of Rex1: $[D_{Rx}] = 1, [C_{Rx}] = 10^{-4}$						Transcription of Dax1: $[D_{Dx}] = 1, [C_{Dx}] = 10^{-4}$					
	C_{Rx}	Kf	Nc	Oc	$SxNg$	Dx	C_{Dx}	Nc	Oc	$SxNg$	Dx
k_1	1	0.5	0.5	0.01	0.2	0.5	1	0.5	0.01	0.2	0.5
k_2	10^3	500	500	10	286	500	10^3	500	10	286	500
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6
k_4	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-4}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}
l	1	1	1	1	1	1	1	1	1	1	1
Transcription of cMyc: $[D_{Mc}] = 1, [C_{Mc}] = 10^{-4}$			Transcription of Zfp281: $[D_{Zp}] = 1, [C_{Zp}] = 10^{-4}$			Transcription of Klf4: $[D_{Kf}] = 1, [C_{Kf}] = 10^{-4}$			Transcription of Nac1: $[D_{Nc}] = 1, [C_{Nc}] = 10^{-4}$		
	C_{Mc}	Kf	B_{Mc}	C_{Zp}	Mc	C_{Kf}	A_{Kf}	Kf	C_{Nc}	A_{Nc}	
k_1	1	5	10	1	10	1	1	10	1	10	
k_2	10^3	$5 \cdot 10^3$	10^3	10^3	10^4	10^3	10^4	10^4	10^3	10^4	
k_3	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	
k_4	10^6	10^6	10^9	10^6	10^6	10^6	10^6	10^6	10^6	10^6	
k_{-1}	10^6	10^6	10^6	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	
k_{-2}	10^{-6}	10^{-6}	10^{-12}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	
k_{-3}	10^{-6}	10^{-6}	10^{-12}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	
k_{-4}	10^{-6}	10^{-6}	10^{-12}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	10^{-9}	
l	1	1	0	1	1	1	1	1	1	1	

In addition to the activation of Rex1 by Klf4, Nac1, and Dax1, Rex1 is also activated by Oct4 (Ben-Shushan et al., 1998; Kim et al., 2008) and Sox2 and Nanog (Kim et al., 2008; Shi et al., 2006). Although all of them can bind the promoter region of Rex1, Sox2 cooperates with Nanog to upregulate the transcription of Rex1 (Shi et al., 2006). Thus, the individual interactions of both Sox2 and Nanog on the regulation of Rex1 have been neglected and only the combined transcription has been considered. The key role of Rex1 in the core level is the activation of Oct4 (Kim et al., 2008).

Dax1 is activated through the master regulator Nac1 and by positive autoregulation (Kim et al., 2008) and it is also activated by Oct4 (Sun et al., 2008), Sox2 and Nanog (Kim et al., 2008). As not much is known about Dax1, we have assumed a cooperativity interaction between Sox2 and Nanog as in the case of activation of Rex1. Table 7.2 presents the interaction logic of Level 1:

The transcription of Oct4 is represented by the kinetic TRN model as ten cycles: activation of Oct4 by basal TF C_{Oc} , external activation by $A_{Oc/Sx}$, two cycles for activation through the master regulators Klf4 and Nac1, two cycle for positive autoregulation by Oct4 and Sox2, and by Oct4, Sox2 and Nanog, three cycles for activation of Oct4 by Rex1, Dax1 and Zfp281, and one cycle for autorepression of Oct4.

Both transcription of Sox2 and Nanog include nine cycles with the difference that one of the cycles of Sox2 is activation of Sox2 by $A_{Oc/Sx}$ and one cycle for Nanog is repression through the external repressor B_{Ng} . The other eight remaining cycles are: activation by basal TFs (C_{Sx} and C_{Ng} for Sox2 and Nanog, respectively), two cycles for activation by master regulators, two cycles for positive feedback between Oct4, Sox2 and Nanog, two cycles for activation by Dax1 and Zfp281, and one cycle for repression of Oct4.

The transcription rate of Rex1 is modeled by six cycles, being one of them the activation of Rex1 by Klf4. The remaining five cycles are similar for the transcription of Dax1 which include: activation by basal TFs (C_{Rx} and C_{Dx} for Rex1 and Dax1, respectively), one cycle for activation by Oct4, another cycle for the joint activation of Sox2 and Nanog, and another cycle for activation by Dax1.

Transcription of cMyc occurs from two cycles: basal activity (due to C_{Mc}) and activation by Klf4. A third repressive cycle is included for the external repressor B_{Mc} . Finally, Zfp281 is

transcribed from two cycles, basal TF C_{Zp} and activation from cMyc. The kinetic parameters involved in the Level 1 of the extended pluripotent TRN model of ESCs are summarized in Table 7.3.

In the extended pluripotent TRN model of ESCs, the input and parameter vectors are defined as follows:

$$\mathbf{x} = [A_{Oc/Sx} \ A_{Kf} \ A_{Nc} \ B_{Ng} \ B_{Mc}]^T \quad (7.69)$$

$$\mathbf{p} = [k_1^T \ k_2^T \ k_3^T \ k_4^T \ k_{-1}^T \ k_{-2}^T \ k_{-3}^T \ k_{-4}^T \ D^T \ C^T \ \alpha^T]^T \quad (7.70)$$

Elements in vector \mathbf{p} in Equation (7.70) are the same than in (7.66), but defined for the 49 cycles of this model. Values of \mathbf{p} are given in Table 7.3. The total transcription rate for Level 1 is

$$TR_{tot}^{L1} = TR_{Oc} + TR_{Sx} + TR_{Ng} + TR_{Rx} + TR_{Dx} + TR_{Mc} + TR_{Zp} + TR_{Kf} + TR_{Nc} \quad (7.71)$$

Figure 7.4 shows the steady state responses in concentration of the Level-1 transcription factors for both individual and total transcription rate of transcription factors. As seen in the figure, bistability was observed in concentration of all level 1 transcription factors as a function of various external activators. Interestingly, the expression of proliferation and differentiation target genes also exhibited bistable behavior although in reverse direction. This shows that various external activators/repressors can “switch on” or “switch off” the ESC-TRN system. The ability of various external activators to completely switch off and switch on the developmental cascade is not surprising because each external activator is coupled in turn with various developmental pathways and thus they can independently control the differentiation/proliferation of ESCs.

Input and parametric vectors were perturbed by $\lambda_x = 1.01$ and $\lambda_p = 1.01$, respectively. The sensitivity matrix for the individual transcription factors gets reduced to the vector:

$$\mathbf{S}_i = [S_i^x \ S_i^p], \text{ for } i = Oc, Sx, Ng, Rx, Dx, Mc, Zp, Kf, Nc \quad (7.72)$$

The overall sensitivity matrix for Level 1 is defined as

$$\mathbf{S}_O^{L1} = \begin{bmatrix} S_{Oc}^x & S_{Sx}^x & S_{Ng}^x & S_{Rx}^x & S_{Dx}^x & S_{Mc}^x & S_{Zp}^x & S_{Kf}^x & S_{Nc}^x \\ S_{Oc}^p & S_{Sx}^p & S_{Ng}^p & S_{Rx}^p & S_{Dx}^p & S_{Mc}^p & S_{Zp}^p & S_{Kf}^p & S_{Nc}^p \end{bmatrix}^T \quad (7.73)$$

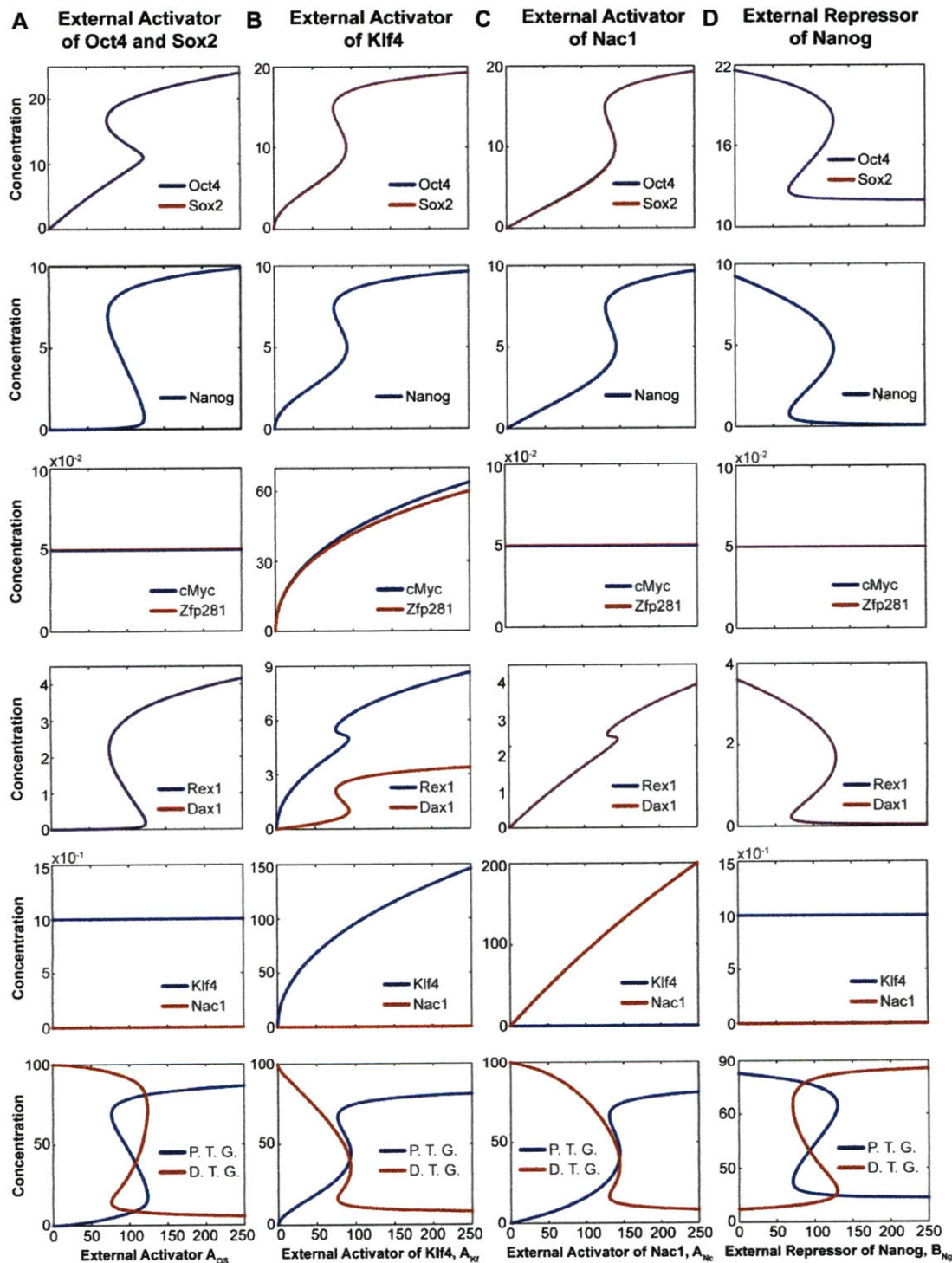


Figure 7.4: Steady state responses of the individual and total transcription rate of transcription factors of Level 1. **A)** Steady state responses as a function of the external activator $A_{Oc/Sx}$. All the other inputs were kept constant at a value of 10^{-3} . **B)** Steady state transcription rates as a function of the external activator A_{Kf} . All the other inputs were kept constant at a value of 10^{-3} . **C)** Steady state response as a function of the external activator A_{Nc} . All the other inputs were kept constant at a value of 10^{-3} . **D)** Steady state transcription rates as a function of the external repressor B_{Kf} . External activator $A_{Oc/Sx}$ was set at 150 and all the other activators were kept constant at a value of 10^{-3} .

Figure 7.5 shows the steady state response of the robustness index. As seen previously for transcription rates, the steady state robustness of the expression of various transcription factors as a function of external activators of Oct4-Sox2, Klf4, Nac1 and external repressors of Nanog exhibits bistable switch with respect to both external activator and repressor concentrations. The switch behavior existence in individual TFs influenced by external signals and also in overall Level-1 system robustness which includes the entire nine inner core TFs.

The total heat dissipation rate is calculated as the sum of the contribution from each of the 49 cycles of level one:

$$HDR = \sum_{i=1}^{49} HDR_i \quad (7.74)$$

where HDR_i is calculated from Eq. (7.20). Figure 7.6 shows the steady state response of HDR. Similar to transcription rate and robustness, energetic cost (Figure 7.6) exhibits switch like behavior as a function of external activator. Energetic cost (heat dissipation rate) increases with increase in both activator and repressor concentrations.

The Pareto frontiers between maximal total transcription rate [given by Equation (7.71)] and the overall Level-1 robustness index [with sensitivity matrix given by (7.73)], is shown in Figure 7.7.

We next analyzed Level-1 architecture of the extended TRN model of ESCs using Pareto optimality based energetic analysis. Similar to Level-0 we observed two different two separate Pareto frontiers. Since, in pluripotent state, transcriptional rate is highest Pareto frontier is obtained between ESC objectives of maximal robustness and maximal transcription rate. Again, this is because in the proliferation stage ESCs objective is to have maximal expression of genes related to proliferation and also maintain this high proliferation gene expression under external influences. Second Pareto frontier exists in differentiated state, i.e. lowest transcription rate region and here the desirability is to minimize transcription and maximize robustness. As seen in Figure 7.7 C, the corresponding energetic cost (heat dissipation rate) estimation reveals lower cost for max-max Pareto frontier in self-renewal state and higher cost for min-max Pareto frontier in differentiate state. Figure 7.D-E shows the variation in external activator and repressor to obtain the corresponding Pareto frontiers.

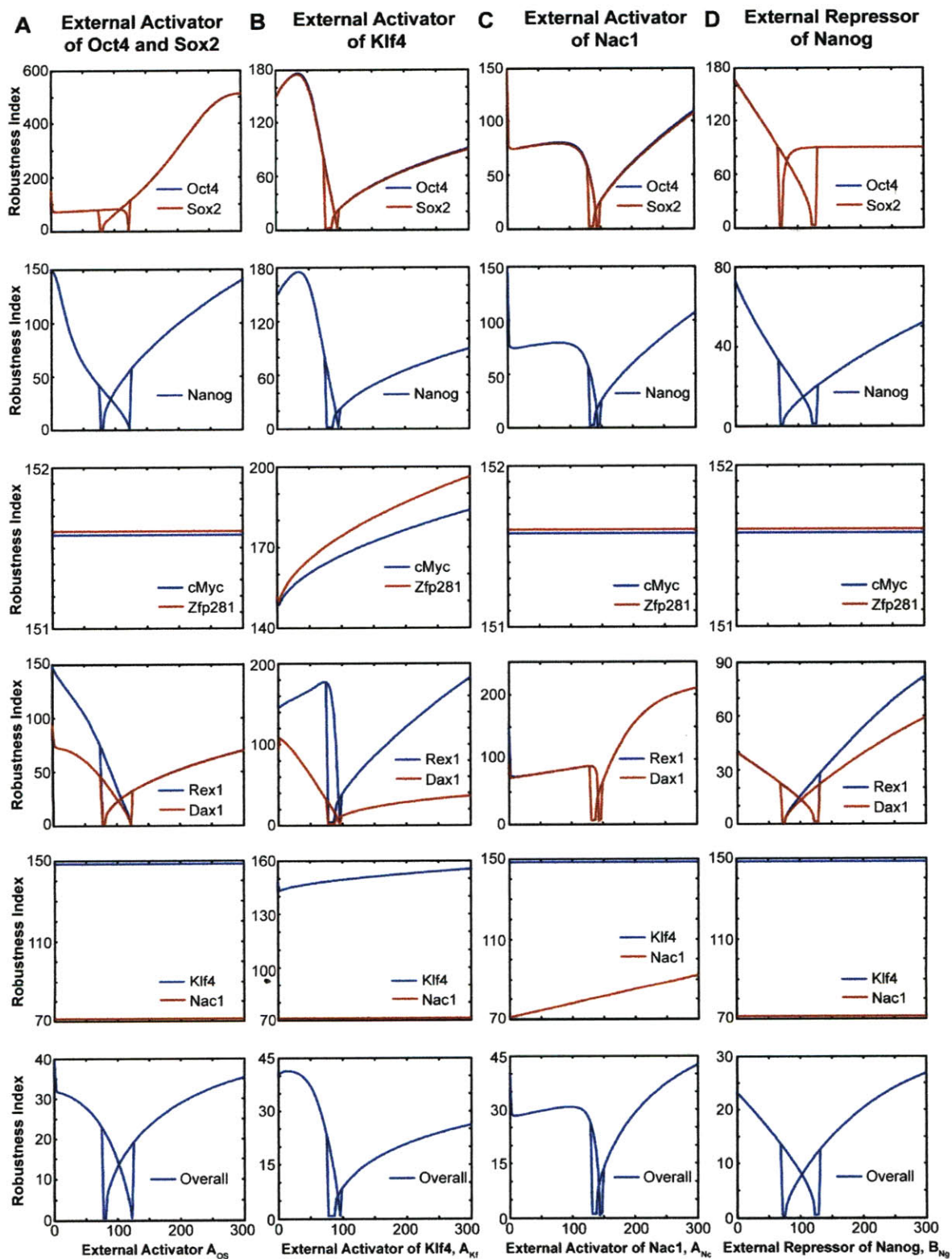


Figure 7.5: Steady State Characteristic of robustness as a function of external activators of Oct4-Sox2, Klf4, Nac1 and external repressors of Nanog. Robustness exhibits bistable switch with respect to both external activator and repressor concentrations. The switch behavior exists in TFs influenced by external signals and also in overall Level-1 system robustness which includes the entire nine inner core TFs. Conditions are same as in Figure 7.4.

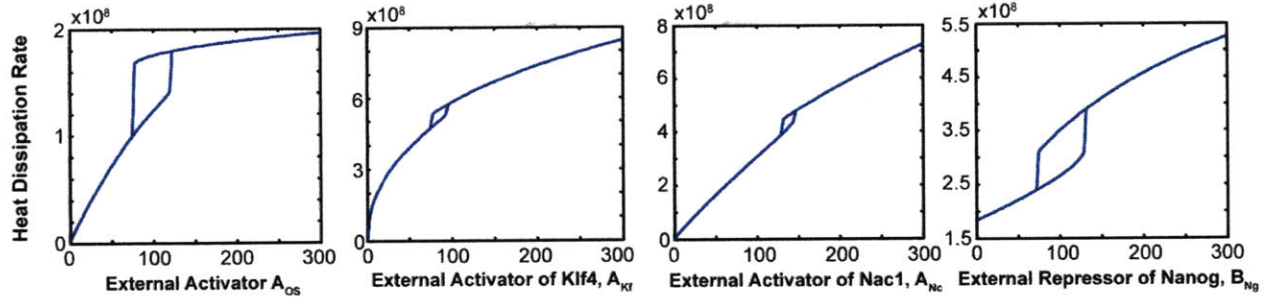


Figure 7.6: Energetic cost exhibits switch like behavior as a function of external activator. Energetic cost (heat dissipation rate) increases with increase in both activator and repressor concentrations. Conditions are same as in Figure 7.4.

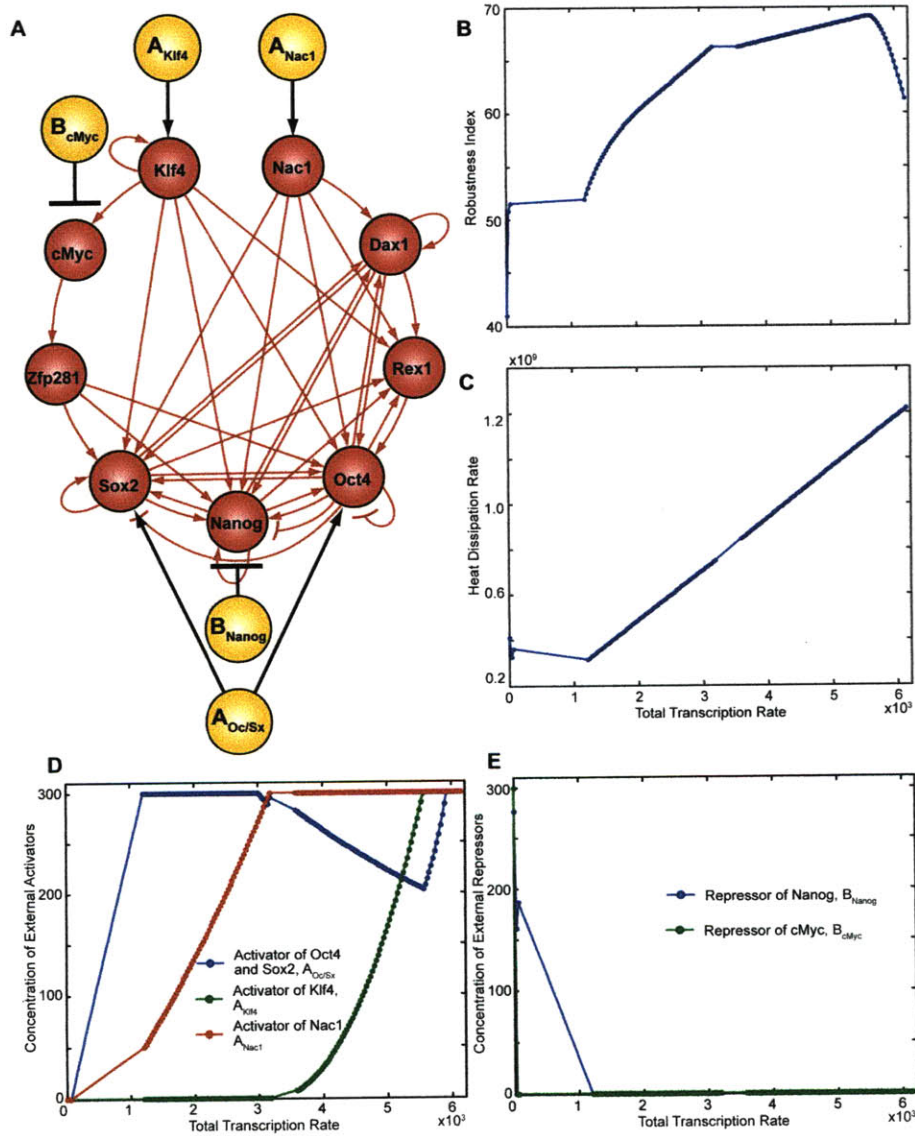


Figure 7.7: Pareto optimality analysis of Level 1 of the extended TRN model of ESCs. **A)** Schematic representation of Level-1. **B)** Pareto frontier between maximal total transcription rate and the maximal overall robustness index of Level 1. The Pareto frontier between minimal total transcription rate and maximal overall robustness index is also

shown. C) Pareto-underlying heat dissipation rate. D)-E) Decision activators (D) and decision repressors (E) of the Pareto solutions shown in B). Conditions are same as in Figures 7.5 and 7.6.

7.3.3 Energetic Cost Explains Regulation of Target Gene Expression by Transcription Factor Occupancy in Level-2

Level-2 This is the level of target genes (TGs) for the nine TFs in the core level or Level 1. Thousands of TGs have been identified for Oct4, Sox2, Nanog, Rex1, Dax1, cMyc, Zfp281, Klf4, and Nac1 but combination of transcriptional network data and protein-protein interaction data identified 22 important TGs (Kim et al., 2008). Table 7.4 shows the TGs in Level 2 as well as the TFs from Level 1 that bind the promoter region of each of them.

Table 7.4: Transcriptional logic for the Level-2 of the extended pluripotent TRN model. Entry (i, j) of the table represents TF j binding the promoter region of TF i . Element $(i, j) = 0$ if no interaction exists, 1 if TF j is an activator of i , and -1 if TF j is a repressor of TF i .

TG	<i>Oc</i>	<i>Sx</i>	<i>Ng</i>	<i>Rx</i>	<i>Dx</i>	<i>Mc</i>	<i>Zp</i>	<i>Kf</i>	<i>Nc</i>	Total TFs
Ahctf1 (<i>Ah</i>)	0	0	0	± 1	0	± 1	0	0	0	2
Arid3a (<i>Aa</i>)	0	0	0	0	0	0	0	± 1	0	1
Arid3b (<i>Ab</i>)	0	0	± 1	0	0	0	0	0	0	1
Btbd14a (<i>Bt</i>)	0	0	0	0	± 1	0	0	0	0	1
Cdc2a (<i>Cd</i>)	0	0	0	0	0	± 1	0	0	0	1
Ewsr1 (<i>Ew</i>)	0	0	0	1	0	1	1	1	0	4
Hdac2 (<i>Hd</i>)	± 1	0	0	0	0	± 1	± 1	0	0	3
Pelo (<i>Pe</i>)	0	0	0	± 1	0	0	0	0	0	1
Prmt1 (<i>Pr</i>)	0	0	0	0	0	± 1	-1	0	0	2
Rai14 (<i>Ra</i>)	± 1	0	± 1	0	0	0	0	0	± 1	3
Rest (<i>Re</i>)	1	1	1	1	1	0	0	1	1	7
Rif1 (<i>Ri</i>)	1	1	1	0	1	0	0	1	1	6
Rnf2 (<i>Rn</i>)	0	0	0	0	± 1	0	0	0	0	1
Rybp (<i>Ry</i>)	0	1	1	0	1	1	0	1	0	5
Sall1 (<i>S1</i>)	1	0	1	0	0	0	0	1	1	4
Sall4 (<i>S4</i>)	0	1	1	0	1	0	0	1	1	5
Sp1 (<i>Sp</i>)	0	0	0	1	0	1	0	1	1	4
Trim28 (<i>Tr</i>)	0	0	0	0	0	0	± 1	0	0	1
Wapal (<i>Wa</i>)	0	0	0	0	0	± 1	0	0	0	1
Wdr18 (<i>Wd</i>)	0	0	0	0	0	± 1	0	0	0	1
Yy1 (<i>Yy</i>)	0	0	0	0	0	± 1	0	0	0	1
Zfp219 (<i>Zf</i>)	± 1	0	0	0	0	0	0	0	0	1

As seen from Table 7.4, TGs in Level 2 are not externally activated nor repressed. Kim et al. (Kim et al., 2008) experimentally observed that those TGs whose promoter regions are bound by few TFs are inactive and those TGs with more than four TFs are highly active during proliferation. We have modeled these observations by explicit repression of those TGs with total

TFs less than four and activation for those TGs with more than four TFs when we analyzed the cofactor occupancy of TFs from Level 1 on Level 2. In particular, these TGs are Ahctf1, Arid3a, Arid3b, Cdc2a, Hdac2, Pelo, Prmt1, Rai14, Rnf2, Trim28, Wapal, Wdr18, Yy1, and Zfp219. For the analysis of Level 1 structure on Level 2 and on the overall extended network, these interactions were assumed as activation linkages with both OR and AND Boolean logics. Thus, the sign \pm was used in Table 7.4. Although not stated in Table 7.4, all the 22 TGs have basal activity.

When we studied cofactor occupancy, we assumed cooperative transcription in Level 2 by all the nine TFs from Level 1. Based on (Kim et al., 2008), this assumption seems to apply when a large number of TFs regulate the translation/transcription of the TGs in Level 2. We have extended this assumption to those TGs transcribed by less than four TFs. Under these considerations, the kinetic TRN model represents the transcription of each of the 22 TGs in Level 2 by only two cycles: basal activity due to basal TF C_i , and either an activation or a repressive cycle for TFs from Level 1. Table 7.5 summarizes the kinetic parameters used in Level 2.

Table 7.5: Parameters for Level 2 of the kinetic TRN model for the extended pluripotent ESCs network. All chemical potentials were obtained at $T = 298.15$ [K] and all degradation rates were set at $\alpha = 10$. Here, TG refers to any of the 22 target genes in Level 2 and k is the number of transcription factors (TFs) from Level 1 bound to the promoter region of TG.

	Transcription of TG for $k \geq 4$ $[D_{TG}] = 1, [C_{TG}] = 10^{-4}$		Transcription of TG for $k < 4$ $[D_{TG}] = 1, [C_{TG}] = 10^{-4}$	
	C_{TG}	$\prod_l^k TF_l$	C_{TG}	$\prod_l^k TF_l$
k_1	1	1	1	1
k_2	10^3	10^4	10^3	10^4
k_3	10^6	10^6	10^6	10^6
k_4	10^6	10^6	10^6	10^6
k_{-1}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-2}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
k_{-4}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
l	1	1	1	0

We compared the Pareto-underlying energetic cost when k transcription factors from Level 1 ($k = 1, \dots, 9$) activate or repress a single target gene in Level 2. In this figure, Boolean logic AND was used for both activation and repression, and the parameters for the basal cycle

and the cycle of the transcription factors from Level 1 are shown in Table 7.5. As seen in Table 7.5, same parameters were used for activation (when four or more transcription factors are bound, $k \geq 4$) and repression ($k < 4$). The input vector of this network is the same as for Level 1 [Equation (7.69)]. The parametric vector is defined by Equation (7.70) but extended to the parameters of the two cycles of Level 2. Perturbations of $\lambda_x = 1.3$ and $\lambda_p = 1$ were used. The objective function total transcription rate was obtained for the entire network as:

$$TR_{tot} = TR_{tot}^{L1} + TR_{TG} \quad (7.75)$$

Where TR_{tot}^{L1} is given by Equation (7.71).

The sensitivity matrix for the calculation of the overall robustness index was defined as:

$$\mathbf{S}_O = \begin{bmatrix} (\mathbf{S}_O^{L1})^T & S_{TG}^x \\ & S_{TG}^p \end{bmatrix}^T \quad (7.76)$$

where \mathbf{S}_O^{L1} is defined by Eq. (7.73).

We studied the effect of having all possible combinations of k transcription factors acting as activators or repressors, resulting in $2 \times C_k^9$ Pareto frontiers between total transcription rate and robustness index, with

$$C_k^9 = \frac{9!}{k!(9-k)!} \quad (7.77)$$

Thus, when one TF bounds the promoter region of the target gene ($k = 1$), there are 9 possible combinations (each of the 9 TF from Level 1) of activation and repression, resulting in 18 Pareto frontiers composed of 5 points. As another example, when 4 TFs from Level 1 regulate the transcription rate of the target gene in Level 2, there are 126 possible combinations. If all the TFs activate or repress the TG, then there are 252 Pareto frontiers. Therefore, we obtained 1022 Pareto frontiers between total transcription rate and overall robustness index.

After the numerical computation of the C_k^9 Pareto frontiers of activation and the C_k^9 Pareto frontiers of repression for each k , the Pareto-underlying benefit over cost was calculated for each Pareto frontier, as the Pareto average of the ratio between the Pareto robustness index and the underlying heat dissipation rate:

$$\left(\frac{B}{C}\right)_i = \frac{1}{N_P} \sum_{j=1}^{N_P} \left(\frac{RI}{HDR}\right)_j \quad (7.78)$$

where N_p is the number of points in the i Pareto frontier (we used $N_p = 5$). Then, the fitness function defined by Eq. (7.78) was compared between activation and the corresponding case of repression. Figure 7.8 shows these results.

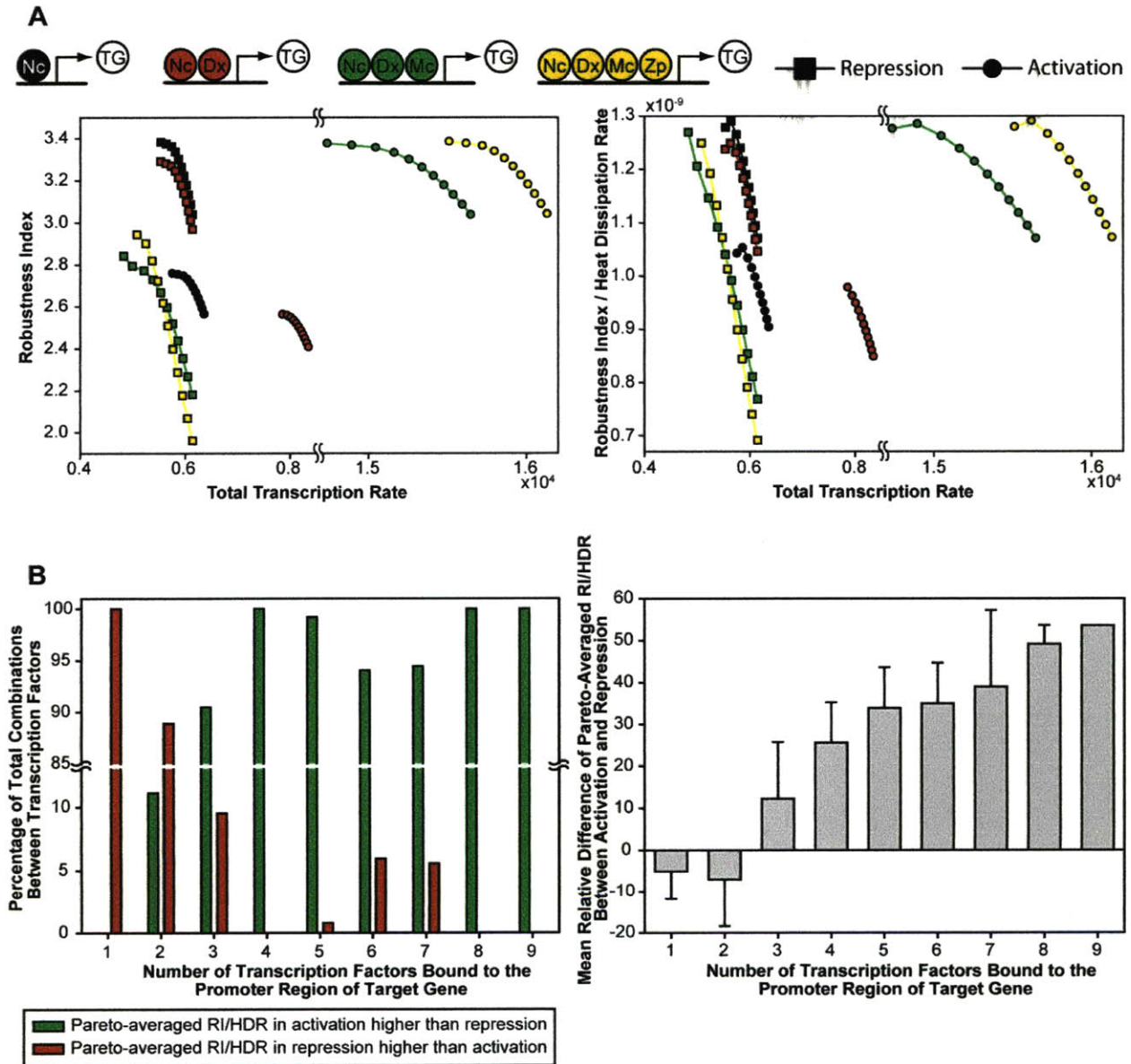


Figure 7.8: Energetic cost explains regulation of target gene expression by transcription factor occupancy. **A)** Pareto frontiers between robustness and transcription rate under self-renewal state was obtained for genes whose promoters are targets occupied by one TF (Nac1), two TF (Nac, Dax), three TFs (Nac, Dax, Myc) and four TFs (Nac, Dax, c-Myc, and ZFp281) for two scenarios: (1) all TFs are activators; (2) all TFs are repressors. Typical case is shown in **A**. **B)** Benefit/cost i.e. Robustness Index/heat dissipation rate was compared for the eight cases. Higher benefit/cost was observed for the cases where all TFs are repressors and TFs are less than 3. For higher than 3 TFs benefit/cost was higher for cases where all TFs are activators. Exhaustive search was done where all possible nCr combinations were considered when all TFs are either activators or repressors. Percentage of cases where repressors or activators are better for normalized benefit is shown.

Figure 7.8 indicates the usage of energetic cost for explaining regulation of target gene expression by cofactor occupancy. First, Pareto frontiers between robustness and transcription rate under self-renewal state was obtained for genes whose promoters are targets occupied by one TF (Nac1), two TF (Nac, Dax), three TFs(Nac, Dax, Myc) and four TFs (Nac, Dax, c-Myc, and ZFp281) for two scenarios: (1) all TFs are activators; (2) all TFs are repressors. We only show the typical case is shown in 7.8 A-B. Next, Benefit/cost i.e. Robustness Index/heat dissipation rate was compared for all of the possible eight cases. We observed higher benefit/cost for the cases when all TFs are repressors and less than 3 TFs are coregulating. For the case where more than 3 TFs coregulate a gene, benefit/cost was higher for cases when all TFs were activators. Exhaustive search was done to obtain all possible nCr combinations of TFs as activators or repressors. Percentage of cases where repressors or activators are better for normalized benefit is shown. Our results indicate higher benefit/cost for i.e. higher functional objective at lower energetic cost for the cases when all TFs are repressors for less than 3 TFs coregulation and higher benefit/cost for the cases when all TFs are activators for more than 4 TFs coregulation.

7.3.4 Analysis of ESC TRN Architecture: Role of Autoregulation, Feedback Loops, and Other Motifs.

To investigate further the role of complex architecture of various motifs and TRN loops in ESCs-TRN we utilized energetic cost and benefit/cost analysis to decipher the utility of various motifs.

Effect of Level-1 on Level-1: Conditions and definitions of total transcription rate and robustness index are same as for Figures 7.5 and 7.6. When an interaction is removed, the parameters from Table 7.3 associated to that linkage become zero, but the other parameters are kept at their values set in Table 7.3.

The normalized Pareto frontiers and normalized Pareto benefit over cost between the total transcription rate of Level 1 (objective function 1 or g_1) and the overall robustness index of the network (objective function 2 or g_2) was obtained by following the next sequence:

1. Find the overall Utopian point **UP**, and the overall Nadir point **NP**, defined as:

$$UP = [\max(g_{1,j}^i) \quad \max(g_{2,j}^i)] = [UP_1 \quad UP_2] \quad (7.79)$$

$$NP = [\min(g_{1,j}^i) \quad \min(g_{2,j}^i)] = [NP_1 \quad NP_2] \quad (7.80)$$

Where $g_{k,j}^i$ is the j Pareto point of objective function k in the i Pareto frontier.

2. Normalize the Pareto solutions with respect to the Utopian and Nadir points:

$$\bar{g}_{1,j}^i = \frac{g_{1,j}^i - NP_1}{UP_1 - NP_1} \quad (7.81)$$

$$\bar{g}_{2,j}^i = \frac{g_{2,j}^i - NP_2}{UP_2 - NP_2} \quad (7.82)$$

3. For each normalized Pareto point, calculate the benefit over cost as

$$\left(\frac{B}{C}\right)_{i,j} = \frac{\bar{g}_{1,j}^i + \bar{g}_{2,j}^i}{HDR_j^i} \quad (7.83)$$

4. Calculate the Pareto benefit over cost as:

$$\left(\frac{B}{C}\right)_i = \sqrt{\sum_{j=1}^{N_p} \left(\frac{B}{C}\right)_{i,j}^2} \quad (7.84)$$

Where N_p is the number of Pareto points in Pareto frontier i .

5. Among the Pareto frontiers, find the maximal and minimal benefit over cost, and normalize their values as:

$$\overline{\left(\frac{B}{C}\right)}_i = \frac{\left(\frac{B}{C}\right)_i - \left(\frac{B}{C}\right)_{min}}{\left(\frac{B}{C}\right)_{max} - \left(\frac{B}{C}\right)_{min}} \quad (7.85)$$

In the analysis of the effect of Level 1 architecture on Level 2 and on the overall network, only activation was considered. When Boolean logic AND was used, each TG is transcribed from two cycles (as explained before). In the case of OR logic, each TG is transcribed from a variable number of cycles which equals the total number of TFs (see Table S4) plus basal activation. For instance, when Rest is activated, the kinetic TRN model represents this transcription as eight cycles. Parameters for each cycles are the same no matter what Boolean logic was used. For the basal cycle, parameters are same as in Table 7.5. For the activation cycles from TFs (either AND or OR), parameters are the same as in transcription of TG for $k \geq 4$ (activation case of Table S5) but with $k_2 = 10^3$.

We next investigated effect of Level 1 architecture on Level 1 using energetic cost approach presented previously. Figure 7.9A shows different network architectures of Level 1. Here, N0 is the basic case that contains all the possible interactions of the network. In N1 architecture, all the feedback and mutual interactions are removed and the nodes are activated (or repressed) only through the master regulators Klf4 and Nac1. This allows us to understand the role of these feedback loops and mutual interactions. In N2 architecture, basic Oct4, Sox2 and Nanog interactions were removed to understand their significance. In N3 architecture, all the positive interactions (including the basic Oct4, Sox2 and Nanog linkages) were removed. In N4, in addition to the positive autoregulatory interactions, the repressive effect to Oct4 was also eliminated. In N5, N6, and N7, the individual effects to Zfp281, Dax1 and Rex1 were removed, respectively.

For the above mentioned various cases ranging from N0-N7 we obtained normalized Pareto frontiers between maximal transcription rate and maximal robustness index (Figure 7.9B). The corresponding normalized Pareto benefit-cost is shown in Figure 7.9C. As seen in the figure, the complete architecture has highest benefit/cost compared to all other architectural designs, thus, indicating the importance of various interactions. Interestingly, our analysis reveals that N3 and N4 i.e. positive interactions play a strong role in maintaining functionality of ESC network during development.

Effect of Level 1 on Level 2: For this analysis we assumed all interactions in Level 2 as activation with OR logic. The parameters of Level 1 are given in Table 7.3 and the parameters of Level 2 are given in the left portion of Table 7.5 ($k \geq 4$) but with $k_2 = 10^3$. As in Figure 7.9, perturbations of $\lambda_x = 1.01$ and $\lambda_p = 1.01$ were used. Since we want to study the effect of Level 1 on Level 2, the total transcription rate given by Equation (7.86) was used.

$$TR_{tot}^{L2} = TR_{Ah} + TR_{Aa} + TR_{Ab} + TR_{Bt} + TR_{Cd} + TR_{Ed} + TR_{Hd} + TR_{Pe} + TR_{Pr} + TR_{Ra} + TR_{Re} + TR_{Ri} + TR_{Rn} + TR_{Ry} + TR_{S1} + TR_{S4} + TR_{Sp} + TR_{Tr} + TR_{Wa} + TR_{Wd} + TR_{Yy} + TR_{Zf} \quad (7.86)$$

The sensitivity matrix was built only with the individual relative change in the transcription rate of the 22 TGs in Level 2 (the 9 TFs from Level 1 were not included in neither of both objectives functions). The normalized Pareto frontiers and normalized benefit over cost were obtained through Equations (7.79) to (7.85). Figure 7.10 shows these results.

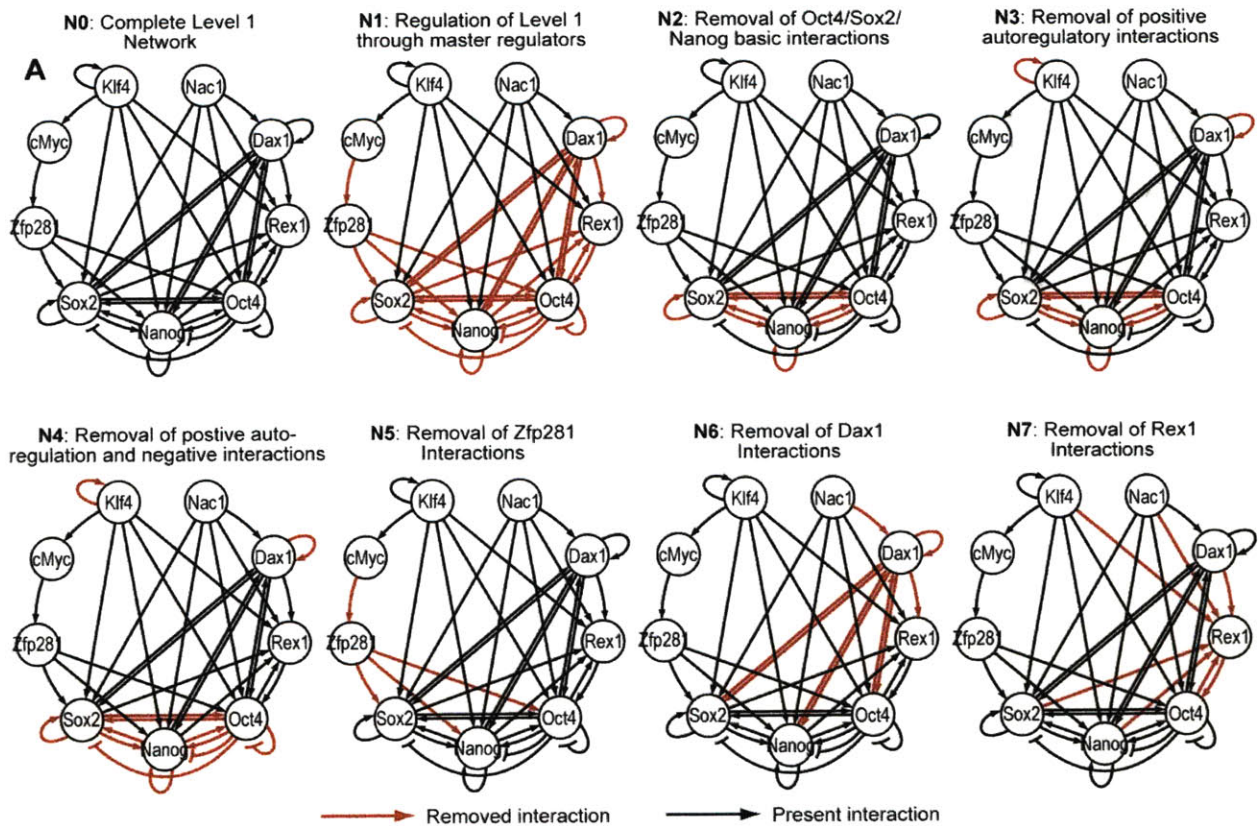
We next analyzed, effect of Level 1 architecture interactions on Level 2 using Pareto-optimal energetic cost approach. Figure 7.10A is similar to Figure 7.9 A but has an additional linkage with the gene of protein networks of Level 2. Figure 7.10 B shows the normalized Pareto Frontiers between maximal transcription rate and maximal robustness index of Level 1 for the different network architectures defined in Figure 7.10 A. The normalized Pareto-benefit over cost for the Level 1 network architectures is shown in Figure 7.10 C. As seen in the figures, again positive autoregulatory interactions and positive feedback loops play an important role for maintain higher functionality in Level 2. This is surprising because it suggests that N3 and N4 interactions are not only important for maintaining higher functionality at lower cost in its individual layer but also for overall network and downstream network cascade.

Effect of Level-1 on the Overall ESC TRN: Same parameters and conditions as in Figure 7.10, but total transcription rate is defined as:

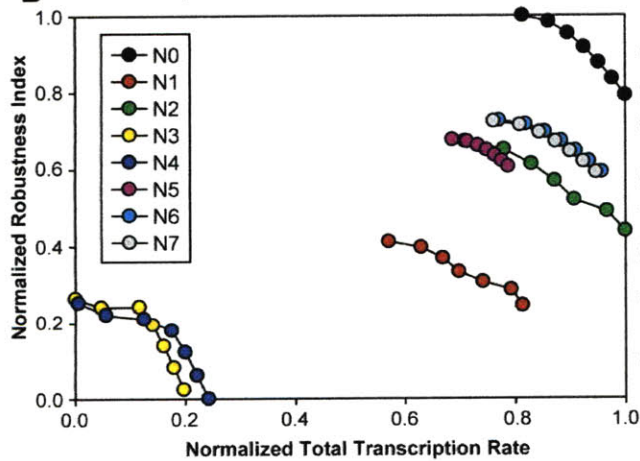
$$TR_{tot}^O = TR_{tot}^{L1} + TR_{tot}^{L2} \quad (7.87)$$

and the sensitivity matrix is contains the individual sensitivities of the 22 TGs in the Level 2 in addition to the 9 TFs in Level 1.

Next, we compared the Effect of Level 1 architecture on the overall network using energetic cost. Again, various network architectures are similar to that in Figure 7.9 but the connectivity with Level 2 TGs can be seen. Here, we see significant role of not only N3 and N4 but also of feedback and oct4-sox2-nanog complex interactions



B Transcription-Robustness Pareto Frontiers



C Pareto Benefit-Cost Analysis

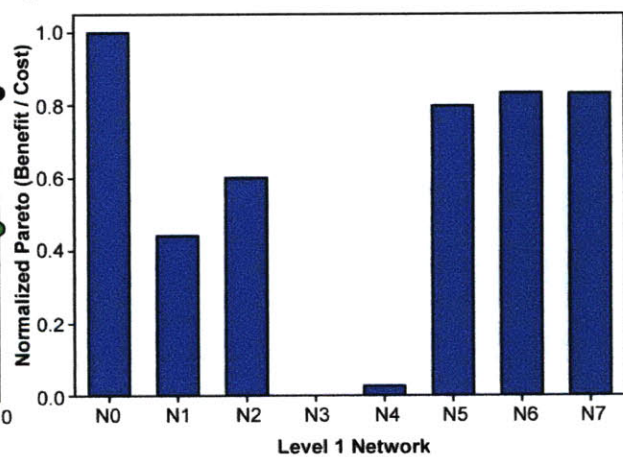


Figure 7.9: Effect of Level 1 architecture on Level 1 energetic cost. **A)** Different network architectures of Level 1. N0 is the basic case that contains all the possible interactions of the network. In N1, all the feedbacks and mutual interactions are removed and the nodes are activated (or repressed) only through the master regulators Klf4 and Nac1. N2 architecture lacks of the basic Oct4, Sox2 and Nanog interactions. In network N3, all the positive interactions (including the basic Oct4, Sox2 and Nanog linkages) were removed. In N4, in addition to the positive autoregulatory interactions, the repressive effect to Oct4 was also eliminated. In N5, N6, and N7, the individual effects to Zfp281, Dax1 and Rex1 were removed, respectively. **B)** Normalized Pareto Frontiers between maximal transcription rate and maximal robustness index of Level 1 for the different network architectures defined in A). **C)** Normalized Pareto-benefit over cost for the Level 1 network architectures.

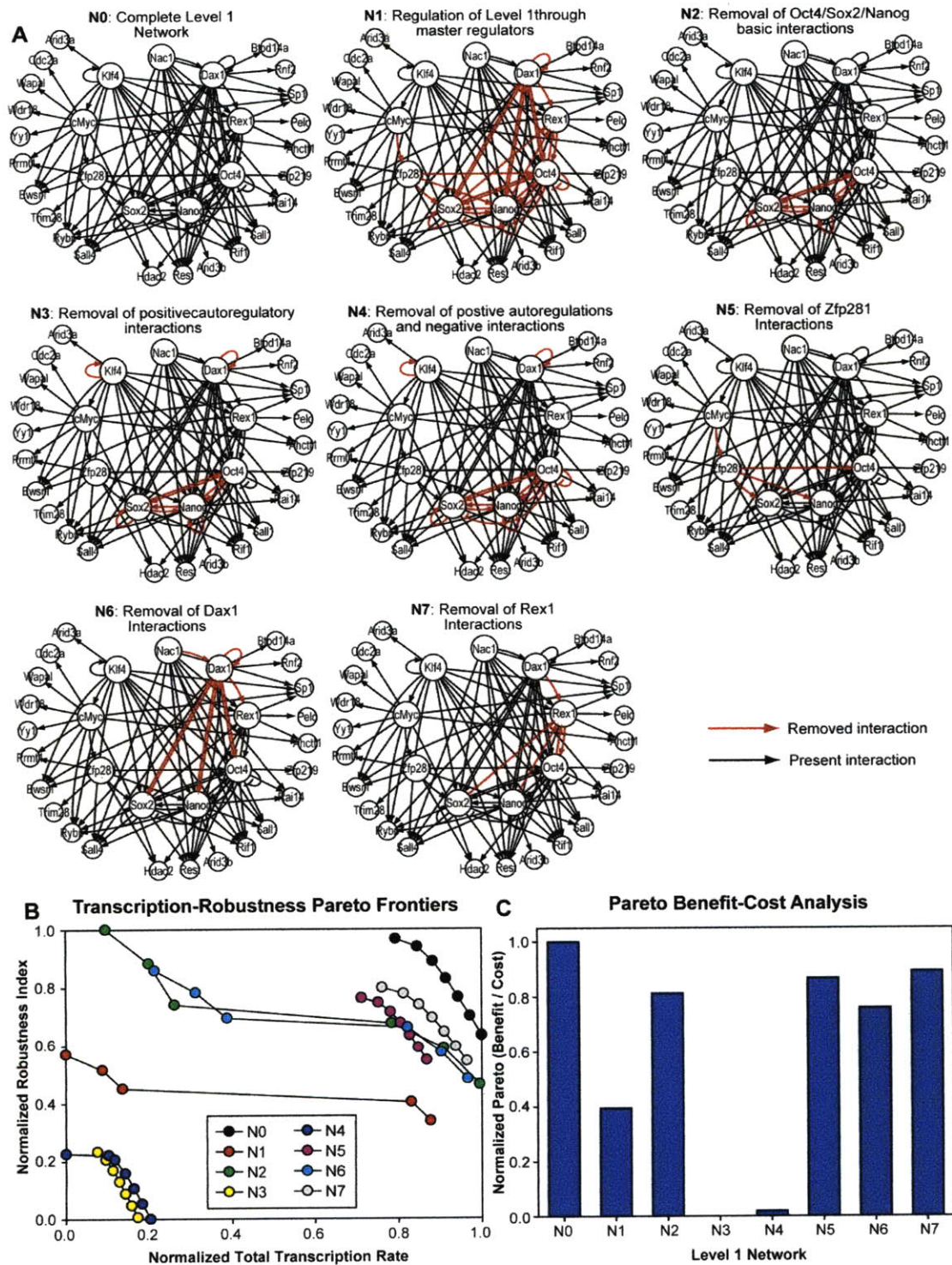


Figure 7.10: Effect of Level 1 architecture on Level 2 energetic cost. **A)** Same as in Figure 7.9 but the connectivity with Level 2 TGs is seen. **B)** Normalized Pareto Frontiers between maximal transcription rate and maximal robustness index of Level1 for the different network architectures defined in A). **C)** Normalized Pareto-benefit over cost for the Level 1 network architectures.

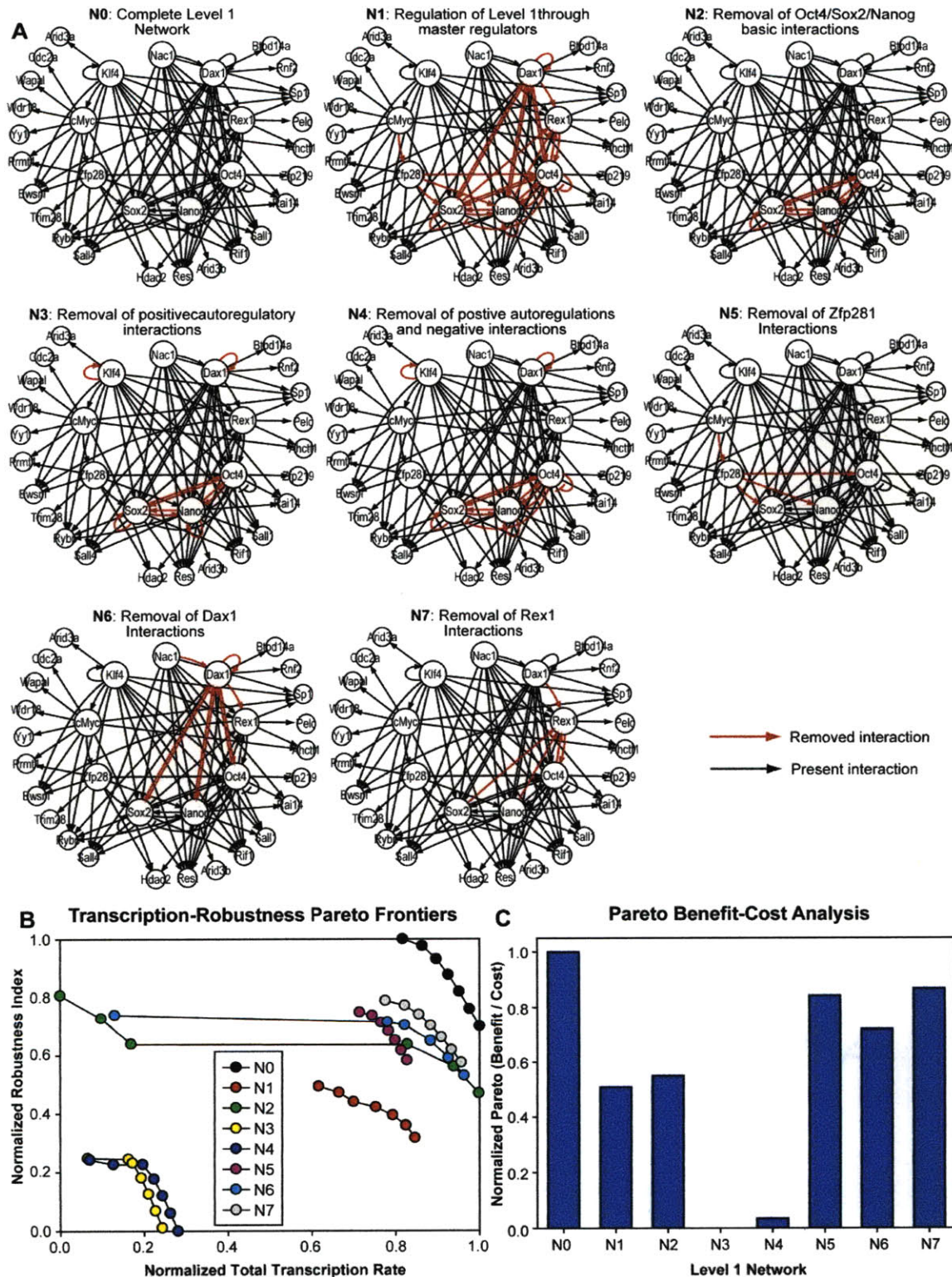


Figure 7.11: Effect of Level 1 architecture on the overall network energetic cost. **A)** Same as in Figure 7.9 but the connectivity with Level 2 TGs is seen. **B)** Normalized Pareto Frontiers between maximal transcription rate and maximal robustness index of Level1 for the different network architectures defined in A). **C)** Normalized Pareto-benefit over cost for the Level 1 network architectures.

7.4 CONCLUSIONS

Here we applied previously developed energetic cost theory to study the architecture of embryonic stem cells. We studied the layered architecture of ESC-TRNs using energetic cost. Our results reveal that feedback loops, positive interactions play an important role in ESCs for maintaining functionality and robustness. Moreover, our energetics based framework could explain both why cofactor occupancy is favored towards repressors for less than 4 TFs coregulation and also towards activators for greater than 4 TFs. In summary, our study was able to predict the functionalities, objectives and experimental observations in the expanded ESC transcriptional regulatory network model.

8 SUMMARY AND FUTURE DIRECTIONS

Mammalian systems perform several different functions in nature and hence, optimization of such systems may involve more than one objective as the goal. For example, hepatocytes perform several different functions as the key component of BAL systems, and these functional objectives are potentially conflicting. As seen in previous chapters, higher albumin synthesis changes the uptake of various metabolites in such a manner that necessarily decreases urea secretion. In order to investigate the trade-offs between these conflicting objectives and to explore available design options, one need to formulate the optimization problem with multiple objectives (vector optimization). Vector optimization obtains a Pareto optimal solution that satisfies the strict constraints imposed by multiple objectives. We developed a constrained multiobjective formulation, LPPFBA approach which overcomes several disadvantages possessed by current algorithms, such as: requiring a priori selection of weights or targets for each of the objective functions which are inadequate in capturing desired preferences; providing a single Pareto solution; inability to generate proper Pareto points for non-convex problems (e.g., the weights method); inability to generate sensitivity information for trade-off and decision making; and no inherent capabilities for design exploration. The LPPFBA approach provides a new effective tool to obtain Pareto optimal solutions. The incorporation of LPP into the standard Metabolic Flux Analysis method enables an unambiguous formulation of an aggregate objective function that facilitates effective multiobjective flux balance analysis for large-scale problems. The presented LPPFBA approach initiates a meaningful step towards infusion of genomic and proteomics data into metabolite perturbations. Importantly, the presented methodology could be employed in various metabolic networks that invariably have multiple objectives (ranging from physiological to design objectives) to be optimized. The combined quantitative and visualization framework presented in this work sets the stage for the development of true optimal solutions for large scale genomics based metabolic network systems.

To incorporate energetic constraints, we developed a NCEFBA platform which is a useful tool for optimality analysis of large scale metabolic networks that are bound to possess multi-objective Pareto optimal solutions. This technique enables the systematic identification of tradeoff situations between various metabolic objectives that characterize a particular cellular phenotype. The addition of FBA to EBA constraints ensures that thermodynamically feasible solutions are obtained. Furthermore, experimental flux data can be easily incorporated into the

analysis, which further reduces the feasible space of fluxes. Although the NCEFBA approach described here was applied to the specific case of hepatocellular metabolism, it can be readily used on any large-scale metabolic network. This study highlights how Pareto optimal solutions may contribute to operating BAL devices, alter the metabolic states of hepatocytes, achieve the desired range of objectives and has relevance for understanding the impact of environmental stress, inducers, hormones and supplements on cellular metabolism.

The energetic-cost theory was further developed in the thesis and which clearly indicates that *energetic cost* may be a suitable basis for evolutionary selection of one motif over another and could provide an explanation for the rare occurrence of various network motifs. Our analysis indicates that the Pareto-optimality principle, when combined with NESS analysis, leads to energetically efficient solutions for transcription. The underlying energetic-cost criterion, *SDE*, for Pareto-optimal conditions is a measure that reflects maximal transcription at the lowest energetic demand. Beyond its application as a functional basis in TRN motifs, the Pareto-optimal *SDE* concept may also lead to an optimal and energetically efficient design of synthetic gene circuits. Further validation of this concept for protein and metabolic networks is required to confirm its generality; however, the corresponding abundance data for these networks is unavailable. The finding that energetic cost may be used as an underlying basis for evolutionary selection of a motif among motifs having similar dynamic functionality is of major significance. The overwhelming diversity of possible dynamical functions with highly-interactive biological networks limits effective learning from experimental data alone. Network analyses using knowledge of the often ignored energetics may greatly reduce the hypothesis space, enabling identification of new functionalities of dynamically perturbed large-scale networks. Further, the developed dynamic PETGP framework may be used not only for analyzing motifs in complex networks but also for designing complex synthetic networks. Appropriate identifications of cellular objectives involved in evolutionary decision making may provide a potentially novel approach to identify optimal environmental conditions and therefore, as a stand-alone strategy, may provide a more efficacious simultaneous prediction and validation strategy for biological networks.

We also developed energetic cost theory to study the architecture of embryonic stem cells and studied the layered architecture of ESC-TRNs using energetic cost. Our results reveal that feedback loops, positive interactions play an important role in ESCs for maintaining

functionality and robustness. Moreover, our energetics based framework could explain both why cofactor occupancy is favored towards repressors for less than 4 TFs coregulation and also towards activators for greater than 4 TFs. In summary, our study was able to predict the functionalities, objectives and experimental observations in the expanded ESC transcriptional regulatory network model.

In future, our theory could be extended for coupled protein and transcriptional regulatory network. For metabolic systems, network architecture theory could be applied to analyze futile cycles and role of various competing metabolic cycles.

9 REFERENCES

- Alon, U., 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet.* 8, 450-61.
- Antoniewicz, M. R., et al., 2007a. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng.* 9, 68-86.
- Antoniewicz, M. R., et al., 2007b. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab Eng.* 9, 277-92.
- Balazsi, G., et al., 2005. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 102, 7841-6.
- Banta, S., et al., 2004. Quantitative effects of thermal injury and insulin on the metabolism of the skeletal muscle using the perfused rat hindquarter preparation. *Biotechnol Bioeng.* 88, 613-29.
- Barabasi, A. L., 2005. Sociology. Network theory--the emergence of the creative enterprise. *Science.* 308, 639-41.
- Barkai, N., Leibler, S., 1997. Robustness in simple biochemical networks. *Nature.* 387, 913-7.
- Beard, D. A., et al., 2004. Thermodynamic constraints for biochemical networks. *J Theor Biol.* 228, 327-33.
- Beard, D. A., et al., 2002. Energy balance for analysis of complex metabolic networks. *Biophys J.* 83, 79-86.
- Beard, D. A., Qian, H., 2005. Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am J Physiol Endocrinol Metab.* 288, E633-44.
- Ben-Shushan, E., et al., 1998. Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Mol Cell Biol.* 18, 1866-78.
- Berthiaume, F., et al., 1996. Effect of extracellular matrix topology on cell structure, function, and physiological responsiveness: hepatocytes cultured in a sandwich configuration. *Faseb J.* 10, 1471-84.
- Bolouri, H., Davidson, E. H., 2002. Modeling transcriptional regulatory networks. *Bioessays.* 24, 1118-29.
- Buchler, N. E., et al., 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A.* 100, 5136-41.
- Burgard, A. P., Maranas, C. D., 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng.* 82, 670-7.
- Bustamante, C., et al., 2005. The nonequilibrium thermodynamics of small systems. *Physics Today.* 58, 43-48.
- Carlson, J. M., Doyle, J., 2000. Highly optimized tolerance: robustness and design in complex systems. *Phys Rev Lett.* 84, 2529-32.
- Chan, C., et al., 2003a. Metabolic flux analysis of cultured hepatocytes exposed to plasma. *Biotechnol Bioeng.* 81, 33-49.
- Chan, C., et al., 2003b. Metabolic flux analysis of hepatocyte function in hormone- and amino acid-supplemented plasma. *Metab Eng.* 5, 1-15.

- Chan, C., et al., 2002. Metabolic pre-conditioning of cultured cells in physiological levels of insulin: generating resistance to the lipid-accumulating effects of plasma in hepatocytes. *Biotechnol Bioeng.* 78, 753-60.
- Chan, C., et al., 2003c. Application of multivariate analysis to optimize function of cultured hepatocytes. *Biotechnol Prog.* 19, 580-98.
- Chernyak, V., et al., 2005. Dynamical generalization of nonequilibrium work relation. *Phys Rev E Stat Nonlin Soft Matter Phys.* 71, 025102.
- Chickarmane, V., Peterson, C., 2008. A computational model for understanding stem cell, trophoderm and endoderm lineage determination. *PLoS ONE.* 3, e3478.
- Chickarmane, V., et al., 2006. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol.* 2, e123.
- Darling, P. J., et al., 2000. Coupled energetics of lambda cro repressor self-assembly and site-specific DNA operator binding I: analysis of cro dimerization from nanomolar to micromolar concentrations. *Biochemistry.* 39, 11500-7.
- Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature.* 436, 588-92.
- Dekel, E., et al., 2005. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys Biol.* 2, 81-8.
- Dunn, J. C., et al., 1991. Long-term in vitro function of adult hepatocytes in a collagen sandwich configuration. *Biotechnol Prog.* 7, 237-45.
- Dzeja, P. P., Terzic, A., 2003. Phosphotransfer networks and cellular energetics. *J Exp Biol.* 206, 2039-47.
- Edwards, J. S., Palsson, B. O., 2000a. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics.* 1, 1.
- Edwards, J. S., Palsson, B. O., 2000b. Robustness analysis of the Escherichia coli metabolic network. *Biotechnol Prog.* 16, 927-39.
- Ibarra, R. U., et al., 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature.* 420, 186-9.
- Itzkovitz, S., Alon, U., 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res.* 17, 405-12.
- Itzkovitz, S., et al., 2003. Subgraphs in random networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 68, 026127.
- Jana, N. K., et al., 2000. A study of energetics of cooperative interaction using a mutant lambda-repressor. *Protein Eng.* 13, 629-33.
- Jiang, J., et al., 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol.* 10, 353-60.
- Kalir, S., Alon, U., 2004. Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell.* 117, 713-20.
- Kalir, S., et al., 2005. A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. *Mol Syst Biol.* 1, 2005 0006.
- Kalisky, T., et al., 2007. Cost-benefit theory and optimal design of gene regulation functions. *Phys Biol.* 4, 229-45.
- Kashtan, N., et al., 2004. Topological generalizations of network motifs. *Phys Rev E Stat Nonlin Soft Matter Phys.* 70, 031909.
- Kim, J., et al., 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell.* 132, 1049-61.

- Knorr, A. L., et al., 2007. Bayesian-based selection of metabolic objective functions. *Bioinformatics*. 23, 351-7.
- Lee, K., et al., 2000. Metabolic flux analysis of postburn hepatic hypermetabolism. *Metab Eng*. 2, 312-27.
- Lee, K., et al., 2003a. Induction of a hypermetabolic state in cultured hepatocytes by glucagon and H₂O₂. *Metab Eng*. 5, 221-9.
- Lee, K., et al., 2003b. Profiling of dynamic changes in hypermetabolic livers. *Biotechnol Bioeng*. 83, 400-15.
- Lee, K., et al., 2004. Identification of optimal classification functions for biological sample and state discrimination from metabolic profiling data. *Bioinformatics*. 20, 959-69.
- Liphardt, J., et al., 2002. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*. 296, 1832-5.
- Ludovisi, A., et al., 2005. The strategy of ecosystem development: specific dissipation as an indicator of ecosystem maturity. *J Theor Biol*. 235, 33-43.
- Ma, H. W., et al., 2004. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res*. 32, 6643-9.
- Mangan, S., Alon, U., 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*. 100, 11980-5.
- Mangan, S., et al., 2006. The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J Mol Biol*. 356, 1073-81.
- Mangan, S., et al., 2003. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol*. 334, 197-204.
- Maria, A., et al., 2003. Linear physical programming for production planning optimization. *Engineering Optimization*. 35, 19-37.
- Martin, O. C., Wagner, A., 2008. Multifunctionality and robustness trade-offs in model genetic circuits. *Biophys J*. 94, 2927-37.
- Messac, A., 2000. From dubious construction of objective functions to the application of physical programming. *Aiaa Journal*. 38, 155-163.
- Messac, A., et al., 2003. The Normalized Normal Constraint Method For Generating The Pareto Frontier. *Structural and Multidisciplinary Optimization*. 25, 86-98.
- Milo, R., et al., 2004. Superfamilies of evolved and designed networks. *Science*. 303, 1538-42.
- Milo, R., et al., 2002. Network motifs: simple building blocks of complex networks. *Science*. 298, 824-7.
- Morohashi, M., et al., 2002. Robustness as a measure of plausibility in models of biochemical networks. *J Theor Biol*. 216, 19-30.
- Nagrath, D., et al., 2007. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng*. 35, 863-85.
- Nagrath, D., et al., 2005. Multiobjective optimization strategies for linear gradient chromatography. *Aiche Journal*. 51, 511-525.
- Nagrath, D., Messac, A., Bequette, B.W., Cramer, S.M., 2005. Multiobjective optimization strategies for linear gradient chromatography *AICHE Journal*. 51, 511-525.
- Odum, H. T., Pinkerton, R. C., 1955. TIMES SPEED REGULATOR - THE OPTIMUM EFFICIENCY FOR MAXIMUM POWER OUTPUT IN PHYSICAL AND BIOLOGICAL SYSTEMS. *American Scientist*. 43, 331-343.
- Oikonomou, P., Cluzel, P., 2006. Effects of topology on network evolution. *Nature Physics*. 2, 532-536.

- Pan, G., et al., 2006. A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *Faseb J.* 20, 1730-2.
- Pan, G., Thomson, J. A., 2007. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.* 17, 42-9.
- Papin, J. A., et al., 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J Theor Biol.* 215, 67-82.
- Prill, R. J., et al., 2005. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.* 3, e343.
- Qian, H., 2004. Motor protein with nonequilibrium potential: Its thermodynamics and efficiency. *Phys Rev E Stat Nonlin Soft Matter Phys.* 69, 012901.
- Qian, H., Beard, D. A., 2005. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys Chem.* 114, 213-20.
- Rao, C. V., et al., 2004. Design and diversity in bacterial chemotaxis: a comparative study in *Escherichia coli* and *Bacillus subtilis*. *PLoS Biol.* 2, E49.
- Rao, C. V., et al., 2002. Control, exploitation and tolerance of intracellular noise. *Nature.* 420, 231-7.
- Ronen, M., et al., 2002. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci U S A.* 99, 10555-60.
- Rothschild, K. J., et al., 1980. Nonequilibrium linear behavior of biological systems. Existence of enzyme-mediated multidimensional inflection points. *Biophys J.* 30, 209-30.
- Russo, I., et al., 1995. Role of H1 in chromatin folding. A thermodynamic study of chromatin reconstitution by differential scanning calorimetry. *Biochemistry.* 34, 301-11.
- Savageau, M. A., Freter, R. R., 1979. Energy cost of proofreading to increase fidelity of transfer ribonucleic acid aminoacylation. *Biochemistry.* 18, 3486-93.
- Savinell, J. M., Palsson, B. O., 1992a. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J Theor Biol.* 154, 421-54.
- Savinell, J. M., Palsson, B. O., 1992b. Network analysis of intermediary metabolism using linear optimization. II. Interpretation of hybridoma cell metabolism. *J Theor Biol.* 154, 455-73.
- Scarpulla, R. C., 2002. Nuclear activators and coactivators in mammalian mitochondrial biogenesis. *Biochim Biophys Acta.* 1576, 1-14.
- Schilling, C. H., et al., 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol.* 184, 4582-93.
- Segre, D., et al., 2005. Modular epistasis in yeast metabolism. *Nat Genet.* 37, 77-83.
- Segre, D., et al., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A.* 99, 15112-7.
- Seredick, S. D., Spiegelman, G. B., 2004. The *Bacillus subtilis* response regulator Spo0A stimulates sigmaA-dependent transcription prior to the major energetic barrier. *J Biol Chem.* 279, 17397-403.
- Shea, M. A., Ackers, G. K., 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J Mol Biol.* 181, 211-30.
- Shen-Orr, S. S., et al., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet.* 31, 64-8.
- Shi, W., et al., 2006. Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *J Biol Chem.* 281, 23319-25.
- Stafford, D. E., et al., 2002. Optimizing bioconversion pathways through systems analysis and metabolic engineering. *Proc Natl Acad Sci U S A.* 99, 1801-6.

- Stelling, J., et al., 2004a. Robustness properties of circadian clock architectures. *Proc Natl Acad Sci U S A.* 101, 13210-5.
- Stelling, J., et al., 2004b. Robustness of cellular functions. *Cell.* 118, 675-85.
- Sun, C., et al., 2008. Stem cell-specific expression of Dax1 is conferred by STAT3 and Oct3/4 in embryonic stem cells. *Biochem Biophys Res Commun.* 372, 91-6.
- Vazquez, A., et al., 2004. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc Natl Acad Sci U S A.* 101, 17940-5.
- Vilar, J. M., Rubi, J. M., 2001. Thermodynamics "beyond" local equilibrium. *Proc Natl Acad Sci U S A.* 98, 11081-4.
- Vo, T. D., et al., 2004. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem.* 279, 39532-40.
- Walton, S. P., et al., 2002. Thermodynamic and kinetic characterization of antisense oligodeoxynucleotide binding to a structured mRNA. *Biophys J.* 82, 366-77.
- Whitson, P. A., et al., 1986. Thermodynamic analysis of the lactose repressor-operator DNA interaction. *Biochemistry.* 25, 3852-8.
- Wong, M. S., et al., 2004. Metabolic and transcriptional patterns accompanying glutamine depletion and repletion in mouse hepatoma cells: a model for physiological regulatory networks. *Physiol Genomics.* 16, 247-55.
- Xia, T., et al., 2003. Context and conformation dictate function of a transcription antitermination switch. *Nat Struct Biol.* 10, 812-9.
- Yang, F., et al., 2005. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metab Eng.* 7, 251-9.
- Yarmush, M. L., Banta, S., 2003. Metabolic engineering: advances in modeling and intervention in health and disease. *Annu Rev Biomed Eng.* 5, 349-81.
- Yin, H., et al., 1999. Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc Natl Acad Sci U S A.* 96, 13124-9.
- Yokoyama, T., et al., 2005. Evolution of intrahepatic carbon, nitrogen, and energy metabolism in a D-galactosamine-induced rat liver failure model. *Metab Eng.* 7, 88-103.
- Young, J. D., et al., 2008. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol Bioeng.* 99, 686-99.
- Zaslaver, A., et al., 2006. Optimal gene partition into operons correlates with gene functional order. *Phys Biol.* 3, 183-9.
- Zaslaver, A., et al., 2004. Just-in-time transcription program in metabolic pathways. *Nat Genet.* 36, 486-91.