Introduction to Metabolic Control Analysis

Herbert M. Sauro
University of Washington
Seattle, WA

Ambrosius Publishing

Copyright ©2013-2015 Herbert M. Sauro. All rights reserved.
First Edition, version 0.22
Published by Ambrosius Publishing and Future Skill Software
www.analogmachine.org

Typeset using LaTeX 2_ε, TikZ, PGFPlots, WinEdt
and 11pt Math Time Professional 2 Fonts

Limit of Liability/Disclaimer of Warranty: While the author has used his best efforts in preparing this book, he makes no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. The advice and strategies contained herein may not be suitable for your situation. Neither the author nor publisher shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. No part of this book may be reproduced by any means without written permission of the author.

ISBN 10: x-xxxx-xxxx-x (ebook)
ISBN 13: xxx-x-xxxxxxx-x-x (ebook)
ISBN 10: x-xxxxxxx-x-x (paperback)
ISBN 13: xxx-x-xxxxxxx-x-x (paperback)

Printed in the United States of America.

Mosaic image modified from Daniel Steger's Tikz image (http://www.texample.net/tikz/examples/mosaic-from-pompeii/

Front-Cover: Metabolic pathway image from JWS online (Jacky Snoep) with permission. The pathway depicts the glycolytic pathway from Lactococcus lactis using the Systems Biology Graphical Notation (SBGN). Ref: Hoefnagel, Hugenholtz and Snoep, 2002, Time dependent responses of glycolytic intermediates in a detailed glycolytic model of Lactococcus lactis during glucose run-out experiments. Mol. Biol. Reports 29, 157-161

Contents

Preface ix

I Preliminary Topics **1**

1 Cellular Networks **5**

1.1 Overall Organization 6
1.2 Network Representation 7
1.3 Metabolic Networks 8
1.4 Protein Networks 12
1.5 Gene Regulatory Networks 20
1.6 Genome Sizes 23
1.7 *E. coli* 25
Further Reading 31
Exercises 33

2 Stoichiometric Networks **35**

2.1 Mass-Balance Equations 35
2.2 Stoichiometry Matrix 41
2.3 Reversiblity 43
2.4 Network Types 45
2.5 The System Equation 49

2.6 Jarnac . . . 51
Further Reading . . . 51
Exercises . . . 52

3 How Systems Behave **57**
3.1 System Behavior . . . 57
3.2 Equilibrium . . . 57
3.3 Steady State . . . 60
3.4 Transients . . . 64
3.5 Setting up a Model in Software . . . 65
3.6 Effect of Different Kinds of Perturbations . . . 65
3.7 Sensitivity Analysis . . . 70
3.8 Robustness and Homeostasis . . . 71
3.9 Stability . . . 72
Further Reading . . . 77
Exercises . . . 77
Jarnac Scripts . . . 78

4 Flux Constraints **83**
4.1 Flux Constraints . . . 83
4.2 Flux Balance Laws . . . 84
4.3 Determined Systems . . . 87
4.4 Flux Balance Analysis . . . 100
4.5 Isotopic Flux Measurements . . . 109
Further Reading . . . 117

II Metabolic Control Analysis **123**

5 Elasticities **127**

5.1 Introduction . . . 127
5.2 Elasticity Coefficients . . . 131
5.3 Mass-action Kinetics . . . 141
5.4 Local Equations . . . 145
5.5 General Elasticity Rules . . . 147
5.6 Summary . . . 151
Further Reading . . . 153
Exercises . . . 153

6 Introduction to Biochemical Control **157**
6.1 Control, Responses and Regulation . . . 157
6.2 Control Coefficients . . . 158
6.3 Distribution of Control . . . 164
6.4 Connectivity Theorems . . . 179
6.5 Response Coefficients . . . 186
6.6 Canonical Control Coefficients . . . 188
6.7 Computing Control Equations . . . 190
6.8 Regulatory Coefficients . . . 192
Further Reading . . . 196
Exercises . . . 196
Appendix 6.A Jarnac Scripts . . . 199

7 Understanding Metabolism **203**
7.1 Introduction . . . 203
7.2 Early Quantitative Efforts . . . 204
7.3 Prevailing Ideas . . . 205
7.4 A Modern Understanding of Metabolism . . . 207
7.5 Operating Principles . . . 207
7.6 Characteristics of a bootleneck? . . . 210

7.7 'Excess' Enzymes . . . 210
7.8 Why are regulated enzymes regulated? . . . 210
Further Reading . . . 210

8 Measuring **213**
8.1 Introduction . . . 213
8.2 Using Classical Genetics . . . 214
8.3 Genetic Engineering . . . 215
8.4 Titration by Inhibitors . . . 216
8.5 Double Modulation Technique . . . 219
8.6 Reconstitution Methods . . . 222
Further Reading . . . 222
Exercises . . . 222

9 Linear Pathways **223**
9.1 Basic Properties . . . 223
9.2 Product Insensitive Steps and Fast Reactions . . . 226
9.3 NonLinear Kinetics . . . 231
9.4 Front Loading . . . 238
9.5 Optimal Allocation of Protein . . . 240
Further Reading . . . 242

10 Negative Feedback **245**
10.1 Historical Background . . . 245
10.2 Simple Quantitative Analysis . . . 250
10.3 Negative Feedback in Biochemical Systems . . . 254
10.4 Robustness and Supply/Demand . . . 261
10.5 Instability . . . 262
Further Reading . . . 262

11 Branched and Cyclic Systems **265**
11.1 Branched Pathways . . . 265
11.2 Implicit Differentiation . . . 273
11.3 Futile or Substrate Cycles . . . 274
Jarnac Scripts . . . 279

12 Moiety Conservation Laws **281**
12.1 Moiety Constraints . . . 281
12.2 Moiety Conserved Cycles . . . 285
12.3 Basic Theory . . . 287
12.4 Computational Approaches . . . 292
12.5 Summary . . . 305
Further Reading . . . 305

13 Moiety Conserved Cycles **307**
13.1 Moiety Conserved Cycles . . . 307
13.2 Species Limits . . . 308
13.3 Ultrasensitivity . . . 310
13.4 Saturation . . . 312
13.5 Sequestration . . . 312
13.6 Zero-Order Ultrasensitivity . . . 313

Appendix A Kinetics in a Nutshell **325**

Appendix B Enzyme Kinetics in a Nutshell **331**

Appendix C Math Fundamentals **341**
C.1 Notation . . . 341
C.2 Short Table of Derivatives . . . 342
C.3 Logarithms . . . 343

C.4 Partial Derivatives 343
C.5 Differential Equations 345
C.6 Taylor Series 346
C.7 Total Derivative 348
C.8 Eigenvalues and Eigenvectors 349
Further Reading 351

Appendix D Control Equations 353
D.1 Linear Pathways 353
D.2 Cycles 356
D.3 Branches 356

Appendix E Modeling with Python 359
E.1 Introduction to Python 360
E.2 Describing Reaction Networks using Antimony 366
E.2.1 Initialization of Model Values 369
E.3 Using libRoadRunner in Python 369
E.3.1 Time Course Simulation 371
E.3.2 Plotting Simulation Results 372
E.3.3 Applying Perturbations to a Simulation 373
E.3.4 Steady State and Metabolic Control 374
E.3.5 Other Model Properties of Interest 377
E.4 Generating SBML and Matlab Files 377
E.5 Exercise 378

Appendix F Answers to Questions 381

References 383

History 391

Preface

This book is an introduction to control in biochemical pathways. The book should be suitable for undergraduates in their early (Junior, USA, second year UK) to mid years at college. The book can also serve as a reference guide for researchers and teachers.

The latest edition together with free software and other material can be found at www.analogmachine.org and the research site, www.sys-bio.org.

This is a book I've wanted to write for many years. It covers the basics of metabolic control analysis, that is the framework developed originally by Heinrich, Kacser, Burns and Savageau that helps us think about how biochemical networks operate. There are a number of topics missing from this editions. These include control in complex branched systems, and a thorough review of different flux balance approaches. These topics will be included in subsequent editions. This edition also lack examples taken from the research literature, this again will be remedied in a subsequent edition.

As with my earlier text book on Enzyme Kinetics for Systems Biology I have decided to publish this book myself via a service called Createspace that is part of Amazon. Over the years I've had many offers from publishers to publish text books but have found the contracts they offer to be far too restrictive. Two restrictions in particular stand out, the loss of copyright on the text as well as any figures but more problematic in today's environment, the inability to rapidly update the text when either errors are found or new material needs to be added. With today's print on demand technology there is no reason for these restrictions.

There are many people and organizations who I should thank but foremost must be my infinitely patient wife, Holly, who has put up with the many hours I have spent working alone in our basement or long hours at the department and who contributed significantly to editing this book. I am also most grateful to the National Science Foundation and the National Institutes of Health who paid my summer salary so that I could allocate the time to write, edit and research. I would also like to thank the many undergrad-

uates, graduates and colleagues who have directly or indirectly contributed to this work. In particular I want to thank my two teachers, David Fell and the late Henrik Kacser who I had the privilege to work with as a graduate student and postdoctoral fellow. I had many hours of fruitful and stimulating conversations with Jannie Hofmeyr, Athel Cornish-Bowden, Jim Burns and last but not least Luis Azerenza. In early 2000s I had the good fortune to be introduced to control theory by Brian Ingalls, John Doyle and Mustafa Khammash, all three of which had a significant influence on my understanding of control theory and had that rare knack of explaining complex ideas. More recently I should thank my graduate students, in particular Frank Bergmann (author of SBW) who was (and is) a brilliant programmer, and Deepak Chandran (author of TinkerCell) who developed a very deep understanding of how networks operate. I thank them for they dedication and steadfast enthusiasm while they worked in my lab.

Thanks to the authors of the TeX system, MikTeX (2.9), TikZ (2.1), PGF-Plots (1.5.1) and WinEdt (6.0) for making available such amazing tools to technical authors.

Finally, I should thank Michael Corral (http://www.mecmath.net/) and Mike Hucka (sbml.org) whose LaTeXwork inspired some of the styles I used in the text.

July 2013
Seattle, WA

HERBERT M. SAURO

Other topics to add:

1. How to experimentally determine control coefficients
2. Supply/Demand Theory
3. Metabolic Engineering strategies
4. Operational characteristics

Part I

Preliminary Topics

Preface to Part I

The first few chapters provides a foundation for part II of the book which is where we start to discuss metabolic control analysis. For those already familiar with many of these topics Part I can be omitted. However, it does emphasizes some important concepts which even seasoned practitioners might be unaware of.

1
Cellular Networks

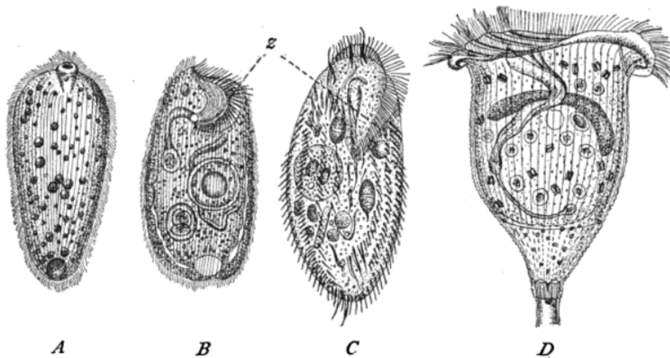


Figure 1.1 Types of Ciliates: From The Protozoa, Gary Nathan Calkins, Macmillan, 1910

The study of cellular networks is one of the defining characteristics of systems and synthetic biology. Such networks involve the coordinated interaction of thousands of molecules that include nucleic acids, proteins, metabolites and other small molecules. Descriptions of these elaborate

networks can be found in text books, on wall charts and more recently in databases such as EcoCyc, RegulonDB, KEGG or STRING – Table 1.1.

1.1 Overall Organization

Biological networks can be organized into three broad groups (Fig. 1.5), gene regulatory, protein and metabolic networks. In the metabolic group, small molecules are chemically transformed by enzymes. These molecules – or metabolites – serve either as energy sources or as building blocks for more complex molecules, particularly polymers such as polysaccharides, nucleic acids and proteins.

Table 1.1 Online *E. coli* resources

Online Resource	URL
EcoCyc	http://ecocyc.org/
RegulonDB	http://regulondb.ccg.unam.mx/
KEGG	http://www.genome.jp/kegg/
STRING	http://string.embl.de/

The protein networks constitute a major part of the decision making and nano-machine apparatus of a cell. We can divide the decision making protein networks into two subgroups. One subgroup involves transcription factor proteins that regulate gene expression, forming what are called the gene regulatory networks (GRNs). The second subgroup, constitutes the signalling pathways that integrate information about the external and internal environment and modulate both the metabolic and gene regulatory networks.

The metabolic, protein and gene regulatory networks each have a characteristic mode of operation and differ by the molecular mechanisms employed and the operating time scales. In general metabolic networks operate on the smallest time scale, followed by protein networks and gene regulatory networks.

This picture is of course a simplified view, for example it omits the extensive RNA network that may be present particularly in eukaryotic cells and protein signalling networks are involved in a variety of other related functions including for example cytoskeleton control and cell cycle regulation. In addition, there is considerable overlap between the different systems with the gene regulatory networks and protein control networks interlinked [5].

1.2 Network Representation

There are different ways to represent cellular networks depending on how the information will be used and what kinds of questions are asked. Traditionally cellular networks have been described using a **stoichiometric** formalism. Such networks are mechanistic in nature, consistent with the laws of mass conservation and will often include kinetic laws describing transformations of species from one form to another through binding/unbinding or molecular reorganization. In recent years an alternative representation, which might be termed **non-stoichiometric**, has gained significant popularity with the advent of high-throughput data collection. Non-stoichiometric networks, of which there are a great variety, include interaction networks which describe the relationship, usually via some physical interaction but sometimes also functional, between molecular species or functional entities such as genes or proteins. Non-stoichiometric networks are by their nature more course grained compared to stoichiometric networks but their study has proved to be very popular due in large part to the availability of vast new data sources. That, coupled with the unprecedented interest in networks in general has made the study of non-stoichiometric networks an intellectually fruitful area of study.

In this book we will be primarily concerned with stoichiometric networks.

1.3 Metabolic Networks

The first cellular networks to be discovered were the metabolic pathways such as Glycolysis in the 1930s and the Calvin cycle in the 1940s. The first metabolic pathways were elucidated by a combination of enzymatic inhibitors and the use of radioisotopes such as Carbon-14. The Calvin cycle for example was discovered by following the fate of carbon when algae were exposed to ^{14}C-labeled CO_2. With the development of microbial genetics significant progress was also made in uncovering other pathways by studying mutants and complementing different mutants of a given pathway to determine the order of steps. The reaction steps in a metabolic pathway are catalysed by enzymes and we now know there are 1000s of enzymes in a given organism catalyzing a great variety of pathways. The collective sum of all reaction pathways in a cell is referred to as metabolism and the small molecules that are interconverted are called the metabolites.

Traditionally, metabolism is classified into two groups, anabolic (synthesis) and catabolic (breakdown) metabolism. Coupling between the two metabolic groups is achieved through cofactors of which a great variety exist although two widely distributed cofactors include the pyridine nucleotides in the form of NAD^+ and $NADP^+$ and the adenine nucleotides in the form of ATP, ADP and AMP. These cofactors couple redox and phosphate respectively by forming reactive intermediates that enables catabolism to drive anabolism. A primary catabolic process is cellular respiration where starting molecules such as glucose are oxidized in a step wise fashion. The energy released is captured in the form of ATP and the oxidized product, water and carbon dioxide are released as waste. The ATP can be used in turn to drive anabolic processes such as amino acid or lipid biosynthesis. In general metabolic pathways tend to be regulated via allosteric regulation. This is where a metabolite can regulate the reaction rate of an enzyme by binding to a site on the enzyme other than the catalytic site. Such interactions form a network of feedback and feedforward regulation. Figure 1.3 shows a metabolic pathway of glycolysis from Lactococcus lactis. On the left, glucose enters the cell which is then converted in a series of reactions to ethanol and a variety of other small molecules.

Metabolic networks are by far the fastest (excluding ion transfer mech-

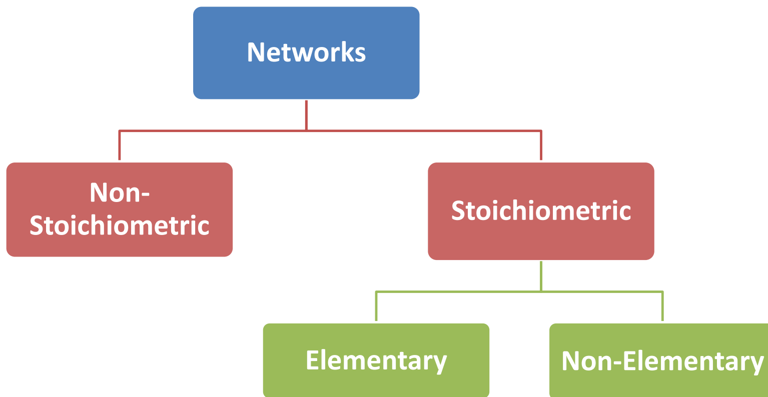


Figure 1.2 Cellular networks are often represented using two common ways, non-stoichiometric and stoichiometric. Non-stoichiometric networks are characterized by a lack of stoichiometric information and mass conservation. Stoichiometric networks are classified according whether they are elementary or not. Elementary networks are those where the reactions cannot be broken into simpler forms. Non-elementary networks may have one or more reaction steps which represent an aggregate of two or more elementary reactions, the aggregation being dependent on some particular assumptions such as quasi-steady state or equilibrium.

anisms) in terms of their response to perturbations and can operate in a time scale from microseconds to seconds. This reflects the need to rapidly adjust the supply of molecular building blocks and energy as supply and demand fluctuate. Physically the rapid response of metabolic networks is achieved by allosteric control where the fast diffusion of small molecules can bind and alter the activity of selected enzymes extremely rapidly.

Figure 1.4 shows a section of the glycolytic pathway which converts glucose to pyruvate with the production of ATP and NADH. The diagram also shows the many negative and positive feedback and feedforward regulatory loops in glycolysis. Not all of these are present in all organisms, however many are. Note the six regulatory signals that converge on 6-Phosphofructose-1-kinase (also known as phosphofructokinase) and Fruc-

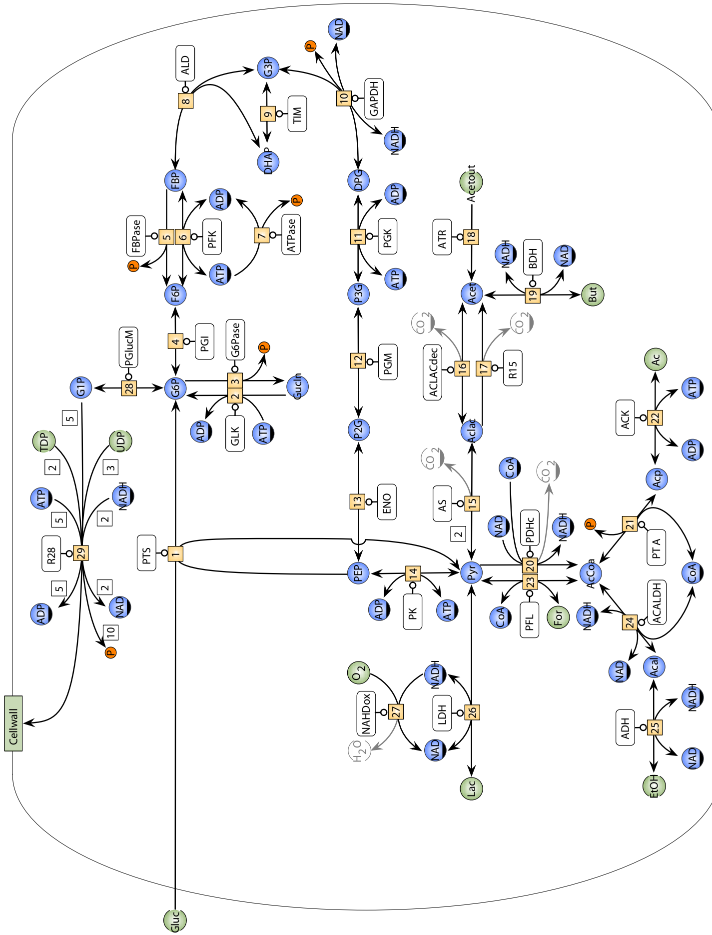


Figure 1.3 Metabolic Pathway: Metabolic pathway image from JWS on-line (Jacky Snoep) with permission. The pathway depicts the glycolytic pathway from Lactococcus lactis using the Systems Biology Graphical Notation (SBGN) [52, 36]

tose Bisphosphatase (Labelled 2 and 3). One of the aims of metabolic control analysis is to quantify regulation and to understand the operational principles of such control.

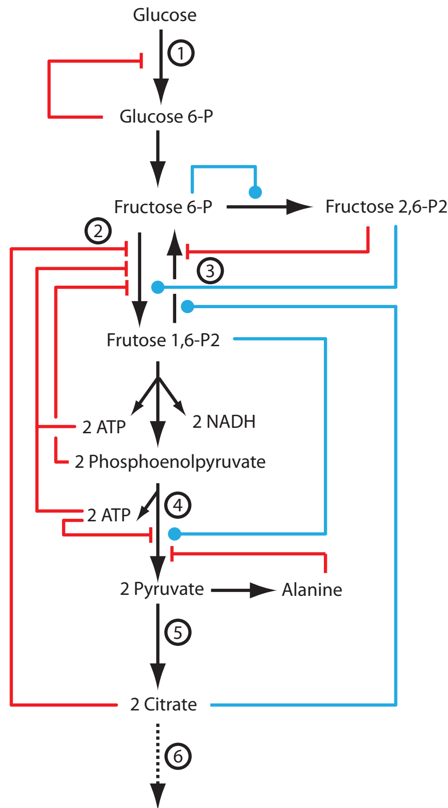


Figure 1.4 A section of glycolysis with negative and positive regulation shown. 1. Hexokinase; 2. 6-Phosphofructose-1-kinase; 3. Fructose Bisphosphatase; 4. Pyruvate Kinase; 5. Entry to Citric Acid Cycle; 6. To oxidative Respiration.

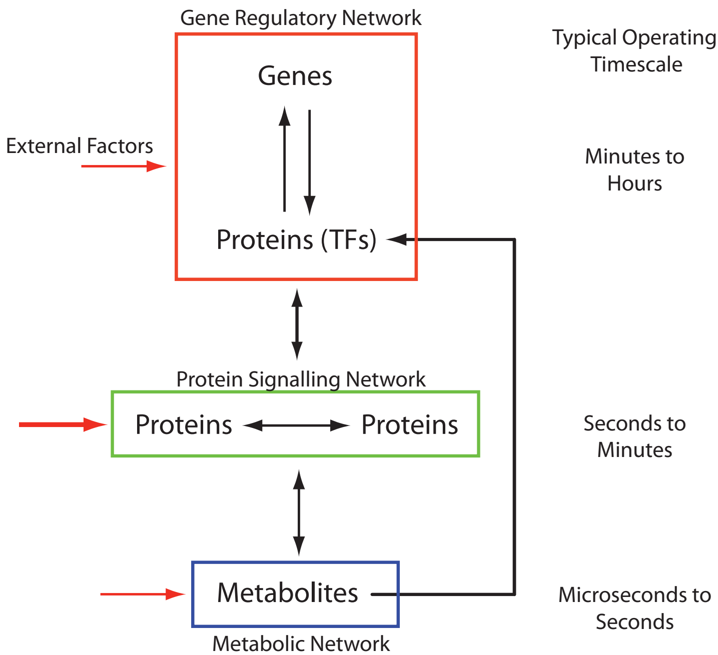


Figure 1.5 Network Overview. The figure illustrates the three main network layers, metabolic, protein and gene. TF – Transcription Factors

1.4 Protein Networks

Protein networks are by far the most varied networks found in biological cells. They range from proteins involved in controlling gene expression, the cell cycle, coordinating and processing signals from the external and internal environments, to highly sophisticated nano-machines such as parts of the ribosome or the bacterial flagella motor.

Protein networks can be studied on different levels, broadly classified as either stoichiometric or non-stoichiometric networks. The non-stoichiometric networks can be as simple as considering the physical associations between different proteins (often through the formation of protein complexes). Such networks, also termed interaction networks, have been elucidated

largely with the help of high-throughput methods. An interaction is formed if two proteins, A and B are known to associate.

Another descriptive level involves functional and stoichiometric networks formed from a consideration of specific stoichiometric binding events, covalent modification (most notably phosphorylation) and degradation. Here two protein A and B might form a complex with a specific stoichiometric relationship and with a specific association constant.

Protein-Protein Networks

Work on uncovering protein networks has be ongoing since the 1950s and considerable detail has accumulated on many different pathways across different organisms. Traditional methods, though laborious [18, 17] have been used extensively to gain detailed knowledge on phosphorylation sites, protein structure, the nature of membrane receptors and the constitution and function of protein complexes. More recently high-throughput methods, though more course grained, have been used to elucidate large swaths of protein-protein interaction networks. For example, in yeast, large scale studies have identified approximately 500 different protein complexes [25, 50] and their relationships to each other.

A popular high-throughput technique that has been used to uncover protein-protein interaction networks is the Yeast two-hybrid method [22, 64] but other methods such as phage display [79, 29] and particularly affinity purification and mass spectrometry have also been employed [25, 50]. The Yeast two-hybrid method (Figure 1.6) is based on the idea that eukaryotic transcriptional activators consist of two domains, a DNA binding domain (DB) and an activation domain (AD). The activation domain is responsible for recruiting the RNA polymerase to begin transcription. What is remarkable is that the two domains do not have to be covalently linked in order to function correctly but merely need to be in close proximity. It is this property that is the basis of the Yeast two-hybrid method.

Let us assume we wish to know whether two proteins, X and Y interact with each other. In the two-hybrid method, protein X is fused with the DB domain (known as the bait protein) and the second protein, Y, is fused with the AD domain (known as the prey protein). These two fused proteins

are now expressed in Yeast and if the two proteins, X and Y, interact in some way they will also bring the DB and AD domains close to each other resulting in an active transcriptional activator. If the gene downstream of the DNA binding sequence is a reporter gene, then the interaction of X and Y can be detected.

A common reporter gene is the lacZ gene which codes for β-galactosidase and which produces a blue coloring in Yeast colonies through the metabolism of exogenously supplied X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside).

There are some caveats with the Yeast two-hybrid method. Although two proteins may be observed to interact, the protein in their natural setting may not be expressed at the same time or may be expressed but in different compartments. In addition using the method to identify interactions between non-yeast proteins may be invalid because of the alien environment in the yeast cell. As with many high-through-put methods caution is advised when interpreting the data.

Using techniques such as Yeast two-hybrid, one of the first interaction graphs to be published was the protein interaction graph of Saccharomyces cerevisiae [83, 43]. Subsequent analysis of this map was conducted by Jeong et al. [44] and included 1870 proteins nodes and 2240 interaction edges. Such graphs give a birds-eye view of protein interactions (Fig. 1.7).

Signalling and Control Networks

Many protein-protein networks operate as signal processing networks and are responsible for sensing external signals such as nutritional (for example by changes in glucose levels) or cell to cell signals such as insulin. Other signalling networks include control networks that are concerned with monitoring and coordinating internal changes, the most well known of these includes the cell cycle control network. Many external signals act by binding to cell-surface receptor proteins such as the large family of receptor tyrosine kinases and G-protein coupled receptors [49]. Once a signal is internalized through the cell-surface receptors, other proteins, including protein kinases and phosphatases continue to process the signal often in coordination with other signaling networks. Eventually the signalling pathway

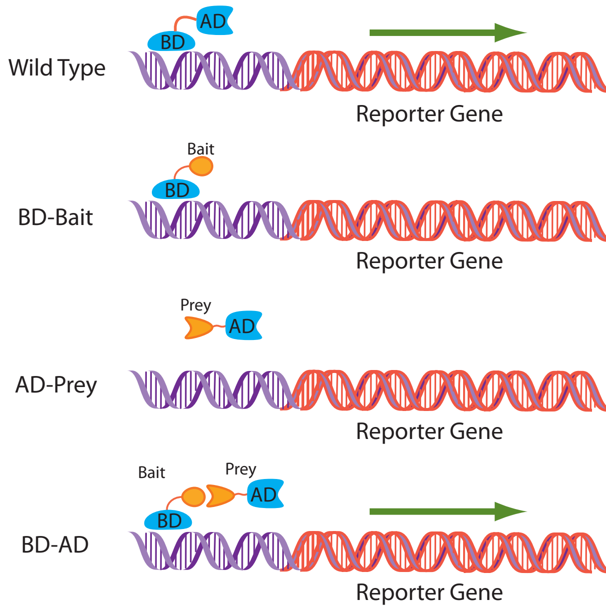


Figure 1.6 Yeast two-hybrid. The wild-type transcription fact is composed of two domains, BD and AD. Both are essential for transcription. Two fusion proteins are made, BD-Bait and AD-Prey. Bait and Prey are two proteins under investigation. If the two protein, Bait and Prey interact bringing BD and AD together resulting in a viable transcription fact that can be used to express a reporter gene.

terminates on target proteins that leads to a change in the cell's behavior. Such targets can include a wide variety of processes such as metabolic pathways, ion channels, cytoskeleton, motor proteins and gene regulatory proteins.

The molecular mechanisms employed by signalling and control pathways include covalent modification, degradation and complex formation. Covalent modification, in particular, is a common mechanism used in signaling networks and includes a variety of different modifications such as phosphorylation, acetylation, methylation, ubiquitylation, and possibly

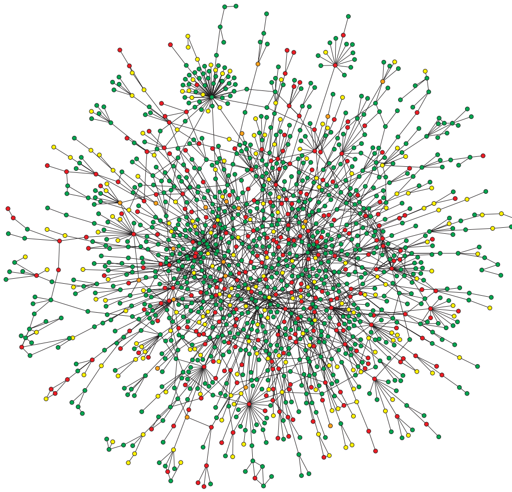


Figure 1.7 The poster child of interaction networks, one of the early Yeast protein interaction networks generated from yeast two-hybrid measurements. Each node represents a protein and each edge an interaction. In addition the graph nodes have been annotated so that red nodes indicate lethal phenotypic effect if removed, green non-lethal, orange slow growth and yellow unknown. Adapted from Barabási and Oltvai [4] but originally published in arXiv and Nature [44]

others [8]. As a result the structure and computational abilities [74] of such networks is likely to be extremely elaborate. It has been estimated from experimental studies that in *E. coli*, 79 proteins can be phosphorylated [54] on serine, threonine and tyrosine side groups whereas in Yeast, 4000 phosphorylation events involving 1,325 different proteins have been recorded [65].

The cell cycle control network is an excellent example of a sophisticated protein control network that coordinates the replication of a biological cell. The cell cycle includes a number of common molecular mechanisms that are found in many other protein networks. These mechanisms can be grouped into three broad types, they include, phosphorylation, degradation and complex formation. Phosphorylation is a common mechanism for

changing the state of a protein and involves phosphorylation on a number of sites on the protein surface including serine/threonine and tyrosine. In prokaryotes, histidine, arginine or lysine can also be phosphorylated. Phosphorylation is mediated by kinases. For example the Human genome may have over 500 kinase encoding genes [55]. The effect of phosphorylation is varied but generally it causes the protein undergoing phosphorylation to change catalytic activity, to change the protein's 'visibility' to other proteins or to mark the protein for degradation. For example, src is a tyrosine kinase protein involved in cell growth. It has two states, active and inactive; when active it has the capacity to phosphorylate other proteins. Deactivation of src is achieved by phosphorylation of a tyrosine group on the C-terminal end of the protein. Dephoshorylation of the tyrosine group by tyrosine phosphatases results in the activation of the protein.

Phosphorylation can also be used to inactivate enzymes such as glycogen synthase by the glycogen synthase kinase 3 protein. In the Yeast cell cycle, the protein Wee1 is phosphorylated and inactivated by the complex Cdc2-Cdc13. Active Wee1 in turn (that is the unphosphorylated form) can inactivate Cdc2-Cdc13 by phosphorylating the Cdc2 subunit.

In addition to changing the activity of proteins, phosphorylation can also be used to mark proteins for degradation. For example, the protein Rum1 that is part of the Yeast cell cycle control network can be phosphorylated by Cdc2-Cdc13. Once phosphorylated, Rum1 is degraded. Degradation itself is an important mechanism used in protein networks and allows proteins to be rapidly removed from a network according to the state of the cell. Degradation is usually mediated by ubiquitylation. For example, Cdc2-Cdc13, via Ste9 and APC is marked for degradation by ubiquitylation (Rum1 is similarly processed once phosphorylated). Once marked this way, such proteins can bind to the proteasome where they are degraded. Finally, binding of one protein to another can change the target protein's activity or visibility. An example of this is the inactivation of Cdc2-Cdc13 by Rum1. When unphosphorylated Rum1 binds to Cdc2-Cdc13, the resulting complex is inactive.

Different combinations of these basic mechanisms are also employed. For example, phosphorylation of complexes can lead to the dissociation of the complex, or the full activity of a protein requires multiple phospho-

rylation events. Although signalling networks can appear highly complex and varied, most of them can be reduced to the three fundamental mechanisms, covalent modification, selective degradation and complex formation (Fig 1.8).

These examples highlight fundamental mechanisms by which protein control networks can be assembled into sophisticated decision making systems.

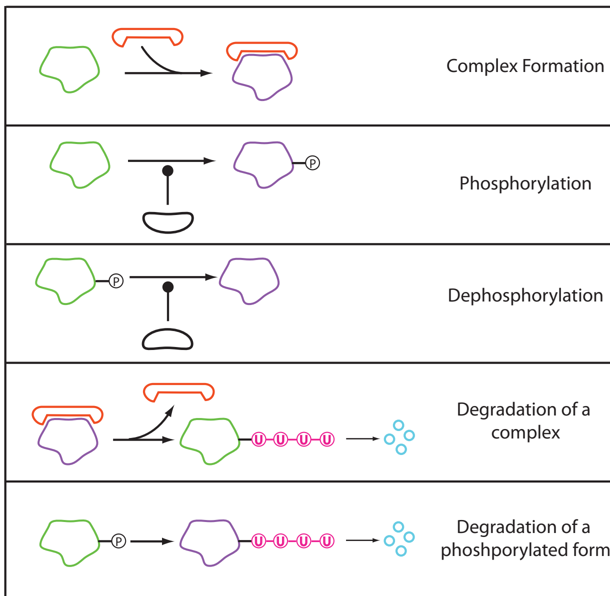


Figure 1.8 Fundamental Protein Mechanisms

In higher eukaryotic cells, particulary human, around 2% of the protein-coding part of the genome is devoted to encoding protein kinases, with perhaps 10% of the coding region dedicated to proteins involved in signalling networks. It has also been suggested that possibly as much as 30% of all cellular proteins in yeast and human can be phosphorylated [14].

The actual size of the networks themselves is however even larger that

these numbers suggest because of the significant number of covalent variants and binding permutations. For example, p53, the tumor suppressor protein, has between 17 and 20 phosphorylation sites alone [82]. If every combination were phenotypically significant, though unlikely, that amounts to at least 131,072 different states.

Ptacek and Snyder [66] have published a review on elucidating phosphorylation networks where much more detailed information is given.

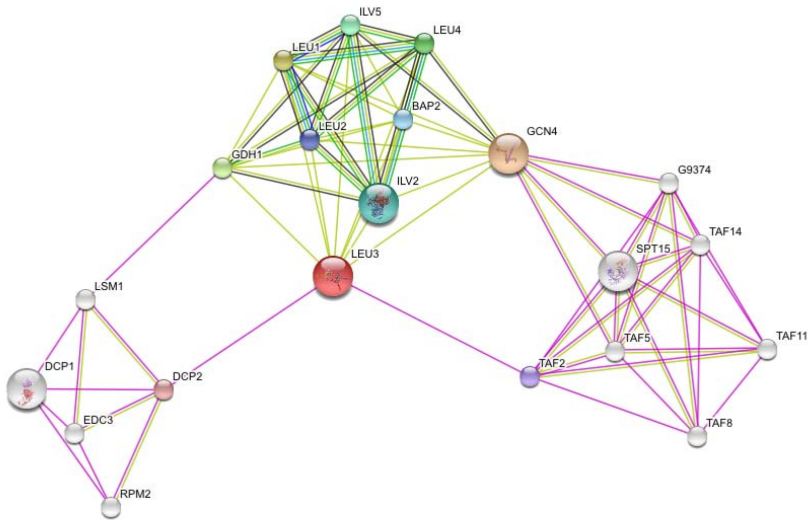


Figure 1.9 A Small Protein-Protein Interaction Map. This image was taken from the STRING web site (Search Tool for the Retrieval of Interacting Genes/Proteins, http://string.embl.de/). The image displays a small segment of the protein interaction map centered around LEU3, the transcription factor that regulates genes involved in leucine and other branched chain amino acid biosynthesis.

1.5 Gene Regulatory Networks

At least in prokaryotes, the control of gene expression is relatively well understood. Transcription factors control gene expression by binding to special upstream DNA sequences called operator sites. Such binding results in the activation or inhibition of gene transcription. Multiple transcription factors can also interact to control the expression of a single gene. Such interactions can emulate simple logical functions (such as AND, OR etc.) or more elaborate computations. Gene regulatory networks can range from a single controlled gene to hundreds of genes interlinked with transcription factors forming a complex decision making circuit. Different classes of transcription factors exist, for example the binding of some transcription factors to operators sites is modulated by small molecules, the classical example being the binding of allolactose (a disaccharide very similar to lactose) to the lac repressor or cAMP to the catabolite activator protein (CAP). Alternatively a transcription factor may be expressed by one gene and either directly modulate a second gene (which could be its own gene) or via other transcription factors integrate multiple signals on to another gene. Additionally, some transcription factors only become active when phosphorylated or unphosphorylated by protein kinases and phosphatases. Like protein signaling and control networks, gene regulatory networks can be elaborate, structurally and computationally.

Significant advances have been made in developing high-throughput methods that can be used to determine protein-gene networks. Of particular interest are ChIP-chip [68, 2] and the more recently developed ChIP-seq [56] screening method – Chromatin immunoprecipitation microarray/Sequencing. ChIP works by treating cells with formaldehyde which crosslinks the DNA to the transcription binding protein if it is bound to the DNA. The cells are then lysed and the DNA fragmented into small 1 kB or less fragments. A specific antibody is now required that will bind to the DNA-binding protein of interest and precipitate the protein and associated DNA fragment. The precipitated DNA pieces are released by reversing the crosslinking. In ChIP-chip, the released DNA pieces are hybridized to a microarray that enables the bound protein to be located on the Genome. A more recent version that is gaining popularity is ChIP-seq. In this pro-

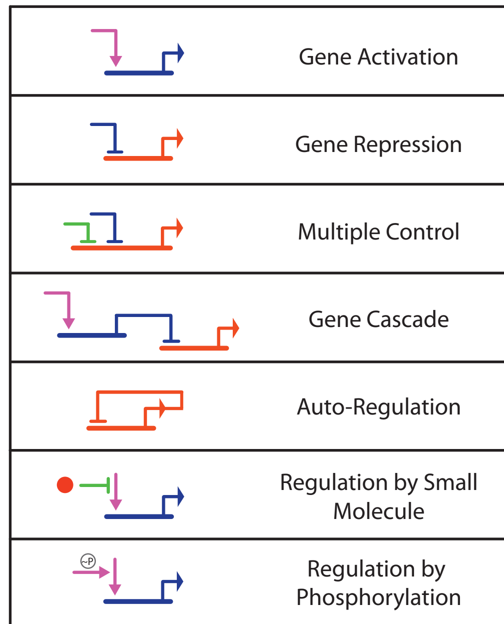


Figure 1.10 Simple gene regulatory patterns

cedure the microarray stage is abandoned and instead the released DNA pieces are sequenced. Once sequenced the location on the Genome can be determined. These methods have been successfully used to determine the gene-protein network of a number of organisms, with Yeast being one of the first [53]. Alternatively other approaches have focused on determining gene-protein networks from literature mining and careful curation or even prediction of putative binding sites.

In general gene regulatory networks are the slowest responding networks in a cell and work from minutes to hours depending on the organism, with bacterial gene regulatory networks tending to operate more rapidly.

The most extensive database on a gene regulatory network is Regulon-DB [42, 24] which is a database on the gene regulatory network of *E. coli*. Reviews that cover in more detail the structure of regulatory networks can

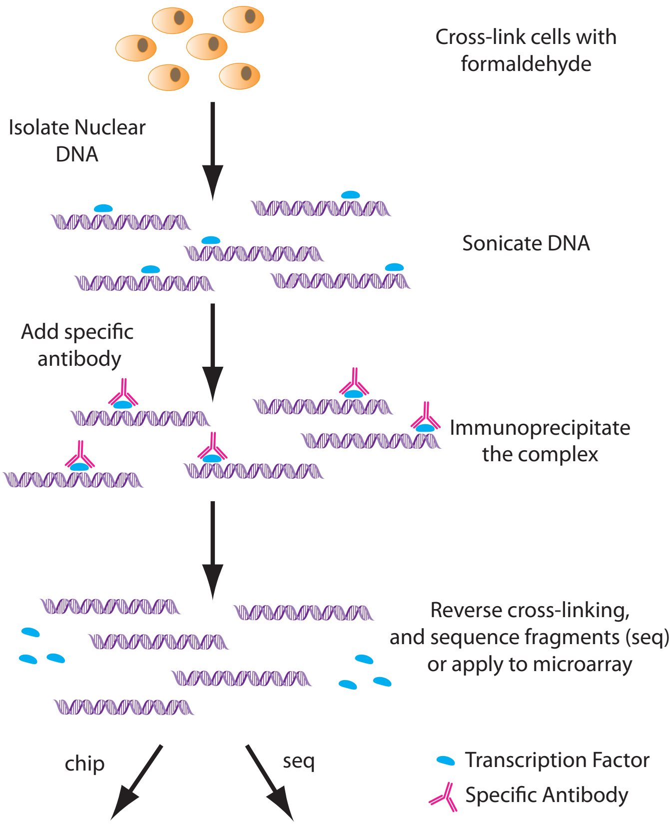


Figure 1.11 ChIP-chip and ChIP-seq methods for determining transcriptional binding sites. Adapted from [56].

be found in the works of Alon [77] and Seshasayee [76].

Although the description of the three main network types may give the impression that they act independently of each other this is most definitely not the case. In general, the different networks will often act together. For example, Figure 1.12 shows a small example taken from Caulobacter [10] showing a mixed gene regulatory and protein network.

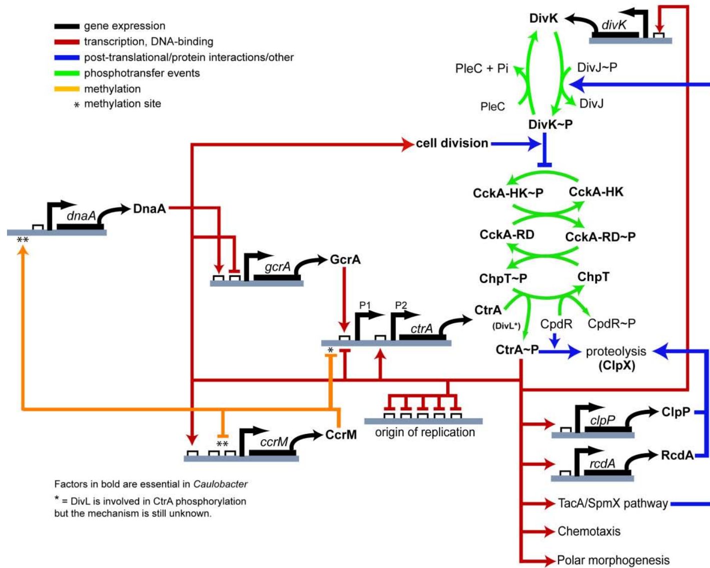


Figure 1.12 Example of a mixed network involving gene regulatory and protein phosphorylation networks in Caulobacter. Blunt ends to regulatory arcs indicate inhibition while arrow ends indicate activation. Image from BioMed Central [10].

1.6 Genome Sizes

How big are cellular networks? To answer that we can look at whole genomes. The sizes of genomes vary considerable from the minuscule 159,662 bases of the symbiotic bacterium called Carsonella ruddii, which

lives off sap-feeding insects, to the Whisk fern with 2.5×10^{11} bases. Some of this size difference is related to the complexity of the organism, simpler organisms requiring fewer genes. However the correlation, although positive is not entirely linear. For example *E. Coli* has roughly 4,300 genes on a genome of size 4.6 Mb, while humans have roughly 25,000 genes on a genome of about 3000 Mb. The *E. Coli* genome is quite dense with roughly 88% of the genome coding for proteins [81] with the remainder being made up of RNA coding, promoter sequences and so on. The human genome on the other hand is very sparse with only about 2% of the genome actually coding for protein. There is ongoing speculation as to why the human genome is so sparse and what the role of the other 98% might be. Some evidence suggests an extensive RNA based regulatory network [58] that is coded in at least some of the non-coding sequences (the so-called junk DNA).

Figure 1.13 shows an example of a small genome from Mycoplasma genitalium. This organism is a small parasitic bacteria that lives in primate genital and respiratory tracts and is the smallest known free-living bacteria. The genome of this organism has 521 genes in total, 482 or these code for protein with the remaining 39 reading frames coding for tRNA and rRNA.

The 482 genes that encode proteins in Mycoplasma genitalium include a wide variety of functions (Figure 1.14) that cover areas such as energy metabolism, replication and the cell envelope. Even for such a small organism there are still eight genes of unknown function.

Eukaryotic genes, especially Human, are also fragmented into segments called exons (coding) and introns (non-coding). This segmentation allows different forms of protein to be derived from the same gene by splicing together different exons. Although the apparent number of genes is of the order of 25,000, alternative splicing probably increases this number significantly [9, 86]. Finally, many proteins, particularly those involved in signalling pathways also have alternative forms due to covalent modification via for example phosphorylation or methylation. This again increases the actual number of states. In others words the number of genes in a genome gives a lower limit to the size of a cellular network, particularly in eukaryotic organisms. The size of a given genome is therefore a poor

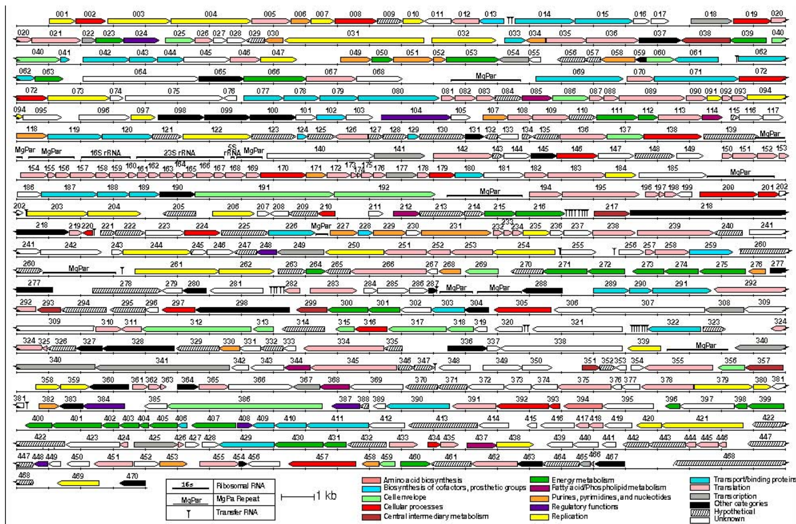


Figure 1.13 Small genome with 521 genes from Mycoplasma genitalium. Image taken from BioCyc.

indicator of how complex the organism might be. To give a better idea of the size and complexity of a small genome let us look more closely at a specific one, *E. coli*.

1.7 *E. coli*

The bacterium *E. coli* is probably one the best understood organisms and it is worth considering some of its features in detail. Much of the information provided here comes from the EcoCyc and RegulonDB online databases and their publications [47, 24].

E. coli is roughly a cylindrical body, with a length of about $2\mu m$ and diameter of about $0.8\mu m$. This gives us a volume of approximately 1×10^{-15} L. If one molecule is present in this volume this represents one molecule per 1×10^{-15} L. That is in 1 L we will have 1×10^{15} molecules which in terms of moles will be $1 \times 10^{15}/6.022 \times 10^{23} \simeq 1 \times 10^{-9}$ moles = 1 nM.

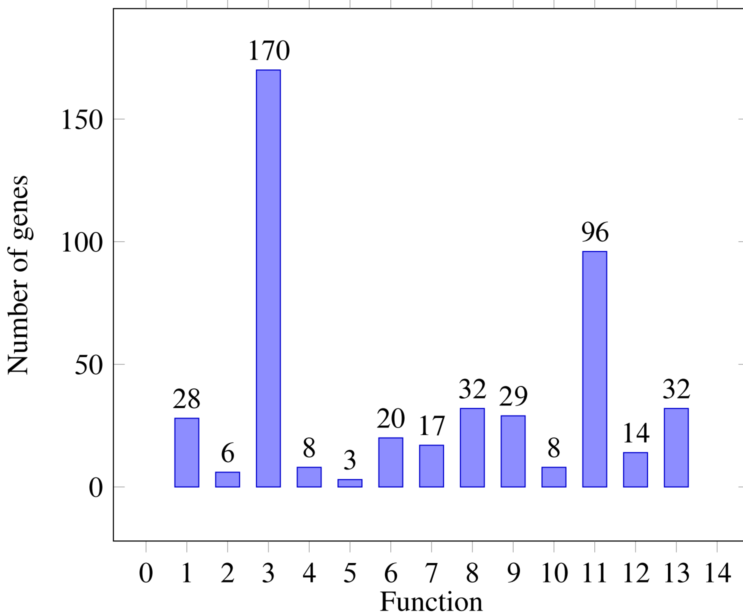


Figure 1.14 1: cell Envelope; 2: Regulatory; 3: Unknown; 4: Central Metabolism; 5: Cofactor Biosynthesis; 6: Purine/Pyrumdine metabolism; 7: Transcription; 8: Transport; 9: Replication/Repair; 10: Lipid Metabolism; 11: Translation: 12: Cellular Processes; 13: Energy Production

This means these dimensions offer a convenient translation between concentration and number of molecules in *E. coli*. 1 nM concentration roughly translates to one molecule per *E. coli* cell (See exercises at end of chapter). For example, ATP is present at a concentration of approximately 2 mM, this means there are roughly 2,000,000 molecules of ATP in a single *E. coli* cell. Another calculation we can do is estimate the maximum number of proteins that can be packed into a single *E. coli* cell. If we assume that of a protein is approximated by a cube of side length 5 nm then the volume of a single protein in units of L (dm^3) is $(5 \times 10^{-9} \times 10)^3 = 1.25 \times 10^{-22}$ L. In a volume of 1×10^{-15} L we can therefore fit a maximum of 8×10^6 proteins. In reality the number will be less than this because other compo-

Table 1.2 A comparison of genome sizes (base pairs) and estimated number of genes. Data from Taft and Mattick [81]

Organism	Genome Size	Est. Number of Genes
E. coli	4,639,221	4,316
Bacillus subtilis	4,214,810	4,100
Saccharomyces cerevisiae	12,100,000	6,000
Caenorhabditis elegans	97,000,000	19,049
Arabidopsis thaliana	115,409,949	25,000
Drosophila melanogaster	120,000,000	13,600
Mus musculus	2,500,000,000	37,000
Homo sapiens	3,000,000,000	30,000

nents such as water, metabolites and nucleic acids much also have space to occupy.

The *E. coli* genome is composed of 4,639,221 base pairs (490μm in diameter) encoding at least 4472 genes. Of this number, 4316 code for proteins, with the remainder coding for various RNA products such as tRNAs and rRNAs. The genes in *E. coli*, like other prokaryotes, do not have segmented genes (genes made of introns and exons); that is a gene in *E. coli* is contiguous sequence of DNA translated into the final protein without editing. In addition there is very little non-coding DNA in *E. coli* with almost 88% of the genome coding for proteins.

Almost one quarter of all proteins produced by gene expression in *E. coli* form multimers, that is proteins composed of multiple subunits. Many of these multimers are homomultimers, that is they are made up of the same subunits. Some of these proteins can also be covalently modified by phosphorylation, methylation and other means. There are estimated to be at least 171 transcription factors, that is proteins that directly control gene expression. This number gives some idea of the size of the gene regulatory network. The EcoCyc database reports at least 48 small molecules and ions that can regulate transcription factors.

Of the 4316 genes in *E. coli*, 3384 (76%) have been assigned a biochemical function. There are at least 991 genes that are involved in encoding

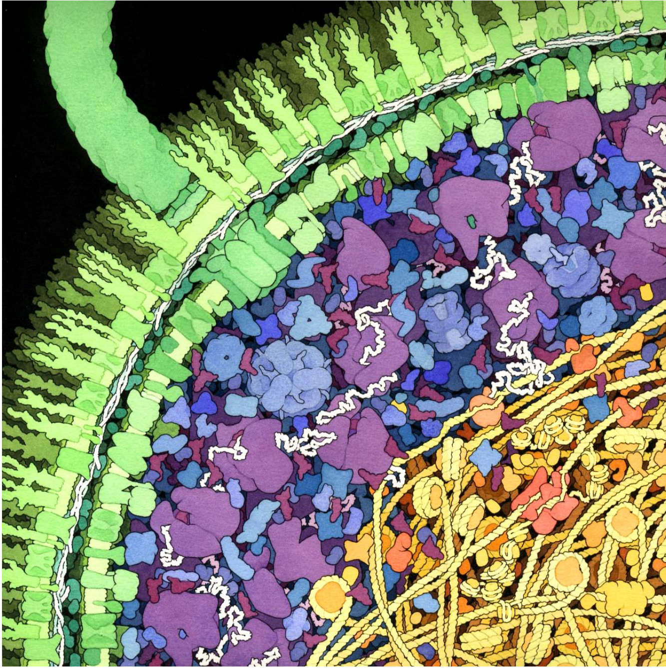


Figure 1.15 Artists impression (With permission Goodsell) of a cross-section through *E. Coli* illustrating the high density of proteins and other molecules in the the cytoplasm.

proteins involved directly in metabolism with a further 355 genes involved in transport. Other gene functions include DNA replication, recombination and repair, protein folding, transcription, translation and regulatory proteins. An inventory of small molecules has not been thoroughly made but EcoCyc records at least 1352 unique small organic molecules which is probably a significant underestimate.

These statistics suggest large numbers of interactions among many thousands of cellular components forming extensive **networks**.

Given the size of a single *E. Coli* cell, the concentration of protein in the cytoplasm and the average diameter of a protein (5 nm), it is estimated that the average spacing (center to center) between proteins is about 7 nm.

This means that the cytoplasm is quite dense. David Goodsell (http://mgl.scripps.edu/people/goodsell) is well known for his evocative illustrations of subcellular spaces. Figure 1.15 illustrates his rendition of a cross-section through *E. Coli*.

Property	Dimensions
Length	2 to 3 μm
Diameter	$\simeq 1\mu$m
Volume	1×10^{-15} L
Optimal generation time	20 to 30 mins
Translation rate	40 amino acids per sec
Transcription rate	70 nucleotides per sec
Number of ribosomes per cell	18,000
Average protein diameter	5 nm
Average concentration of protein	5-8 mM
Average number of proteins	3,600,000

Table 1.3 Basic Information on *E. coli*

There are two useful sites for obtaining basic operating information on *E. Coli*. The first site is the *E. Coli*. statistics site at Project CyberCell (http://gchelpdesk.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi). The other is a more generic and community based web site called BIONUMBRS (The Database of Useful Biological Numbers). Publications from the EcoCyc project also supply many useful statistics on *E. Coli* [47, 48].

The number of molecules in a typical *E. coli* varies with the molecule type. For example there are approximately 2,000,000 Na ions while only 300,000 tryptophan molecules. The larger the molecule the fewer their number, Table 1.4. For example transcription factors are only present in numbers ranging from 10s to 100s, whereas ions are present in the millions.

A major study by Bennett et al (Rabinowitz et al. Nature Chemical Biology (2009)) measured over 100 metabolites levels in the main metabolic pathways of glucose-fed, exponentially growing *E. coli*. The average concentration was found to be 0.22 mM. We can compare this with the av-

Molecule	Estimated Number
Ions	Millions
Small Molecules	10,000 - 100,000
Metabolic Enzymes	1000 - 10,000s
Signaling Molecules	100 - 1000s
Transcription factors	10s to 100s
DNA	1 - 10s

Table 1.4 Orders of magnitude for various molecule types.

erage K_m (concentration of substrate that gives half maximal activity) of approximately 0.1 mM as reported by the BRENDA database. This suggests that on average enzymes operate above their half maximal activity. However, a more detailed analysis revealed considerable variability among different metabolite types. For example, cofactors such as ATP and NAD+, were at concentrations significantly above their K_{ms}. In contrast substrate-enzyme pairs where the concentration was below the K_m were dominated by enzymes catalyzing nucleotide, nucleoside, nucleobase and amino acid degradation reactions. On the other hand the glycolytic pathway, the tricarboxylic acid cycle and the pentose-phosphate pathways showed substrate concentration that were similar to their K_m value.

We can also consider how fast processes occur in *E. coli*. As suggested earlier in the chapter, metabolic responses are the fastest followed by protein signaling networks and gene regulatory networks.

Table 1.7 lists some estimated response times for various biological processes.

The number of molecules and the rate of various processes gives some idea of the magnitude of systems we are dealing with. However, the economy of a typical cell, how ATP is distributed to different processes and how supply and demand are maintained is largely not understood since many of these processes are difficult to measure. Moreover there is no economic theory that describes the life of a cell.

Ions	Estimated Numbers
Na	3,000,000
Ca	2,300,000
Fe	7,000,000
Small Molecules	**Estimated Numbers**
Alanine	350,000
Pyruvate	370,000
ATP	2,000,000
ADP	70,000
NADP	240,000

Table 1.5 Some statistics on estimated numbers of small molecules in *E. coli*

Further Reading

General

1. Bray D (2011) Wetware: A Computer in Every Living Cell. Yale University Press. ISBN: 978-0300167849

2. Goodsell D S (2009) The machinery of life. Springer, 2nd edition. ISBN 978-0387849249

3. Phillips R, Kondev J and Theriot J (2010) Physical Biology of the Cell. Garland Science. ISBN 978-0-8153-4163-5

Specific

1. Nelson DL and Cox MM (2008) Wetware: Lehninger Principles of Biochemistry. W. H. Freeman; 5th edition. ISBN: 978-0716771081

2. Hartl DL (2008) Genetics: Analysis Of Genes And Genomes. Jones & Bartlett Learning; 7 edition. ISBN: 978-0763772154

Signaling Proteins	Estimated Numbers
LacI	10 to 50
CheA kinase	4,500
CheB	240
CheY	8,200
Chemoreceptors	15,000
Metabolic Enzymes	**Estimated Numbers**
Phosphofructokinase	1,550
Pyruvate Kinase	11,000
Enolase	55,800
Phosphoglycerate kinase	124,000
Malate Dehydrogenase	3,390
Citrate Synthase	1,360
Aconitase	1630

Table 1.6 Some statistics on estimated numbers of larger molecules in *E. coli*

3. Brown TA (2006) Genomes 3, Garland Science; 3 edition. ISBN: 978-0815341383

4. Alberts et al, (2002) General Principles of Cell Communication http://www.ncbi.nlm.nih.gov/books/NBK26813/

5. Hancock J (2010) Cell Signalling, Oxford University Press, 3rd edition, ISBN: 978-0199232109

6. Salway JG (2004) Metabolism at a Glance, Wiley-Blackwell; 3 edition ISBN: 978-1405107167

7. Gerhard M and Schomburg D (2012) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley; 2 edition, ISBN: 978-0470146842

Process	Rate
Cell Division Time	50 minutes
Rate of Replication	2000 bp/s
Protein Synthesis	1000 proteins/s
Lipid Synthesis	20,000 lipids/s
Ribosome Rates	25 amino acids per sec per ribosome
Number of ATP to make one cell	55 billion ATPs

Table 1.7 *E. coli* grown on minimal media plus Glucose. Data from Phillips et al (2010) and *E. coli* stats reference: http://ccdb.wishartlab.com

Exercises

In the following exercises, use the data given in the main text, and Tables 1.3, 1.4, 1.5, and 1.6.

1. How many *E. coli* cells laid end to end would fit across the full stop at the end of this sentence? Assume a diameter of the full stop to be 0.5 mm.

2. Estimate the volume of an *E. coli* cell.

3. Calculate the surface area of an *E. coli* cell. If a typical membrane protein is 5 nm in diameter, estimate the number of membrane proteins that can be laid out on the membrane if the center-center distance between each protein is 6 nm.

4. Show that a 1 nM concentration is roughly equivalent to 1 molecule in a volume of one *E. coli* cell.

5. Estimate the number of protein molecules a typical *E. coli* cell can make per second if one assumes that the average protein is 360 amino acids long. Assume that the number of proteins in a cell is 3,000,000. How long would it take to make 3,000,000 proteins?

6. If it takes 1,500 ATP molecules to make an average protein how long would it take before all the ATP is used up? Assume the ATP is not being replaced.

7. What are the visual symbols that are often used to represent activation and repression?

8. Draw a similar diagram to the glycolysis regulatory diagram (Figure 1.4) but for the lysine, threonine and methionine biosynthesis pathway from *E. coli*

2

Stoichiometric Networks

2.1 Mass-Balance Equations

Consider a simple network comprising two reactions, v_1 and v_2, with a common species, S. We will assume that the first reaction, v_1 produces S and the second reaction, v_2, consumes S (Figure 2.1).

Figure 2.1 Simple Two Step Pathway.

According to the law of conservation of mass, any observed change in the amount of species, S must be due to the difference between the inward rate, v_1 and outward rate, v_2. That is, the change in S will be given by the difference in the two rates, leading to the differential equation:

$$\frac{dS}{dt} = v_1 - v_2. \tag{2.1}$$

The above equation is called a **mass-balance equation**. Often S will be expressed in concentration (mol l^{-1}) but it is mass that is conserved not

concentration. We can reexpress equation 2.1 as:

$$\frac{dS_a}{dt}\frac{1}{V} = v_1 - v_2$$

where S_a is the amount of S in moles and V is the volume. Alternatively we can write:

$$\frac{dS_a}{dt} = V(v_1 - v_2)$$

This assumes that the reaction rates are expressed in mol l^{-1} t^{-1}. Often models assume a constant unit volume so that numerically:

$$\frac{dS}{dt} = \frac{dS_a}{dt}$$

and this will be the case is all the examples in this chapter. Although we will write the rate of change in terms of concentration, it is implied that we are dealing with a constant unit volume so that the change in concentration is the same as the change in amount. If movement is from one compartment to a compartment with a different volume then it is necessary to factor in the volume difference and express the rate of change in amounts.

For more complex systems such as the one shown in Figure 2.2 where there are multiple inflows and outflows, the mass-balance equation is given by:

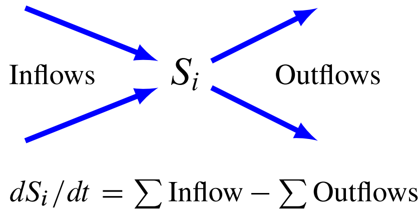


Figure 2.2 Mass Balance: The rate of change in species S_i is equal to the difference between the sum of the inflows and the sum of the outflows

$$\frac{dS_i}{dt} = \sum Inflows - \sum Outflows \tag{2.2}$$

For

an even more general representation, we can write the mass-balance equations by taking into account the stoichiometric coefficients. The rate at which a given reaction, v_j contributes to change in a species, S_i is given by the stoichiometric coefficient of the species, S_i with respect to the reaction, c_{ij}, **multiplied** by the reaction rate, v_j (See equation A.1). That is, a reaction j contributes, $c_{ij} v_j$ rate of change in species S_i. For a species, S_i with multiple reactions producing and consuming S_i, the mass-balance equation (assuming constant unit volume) is given by:

$$\frac{dS_i}{dt} = \sum_j c_{ij} v_j \tag{2.3}$$

where c_{ij} is the stoichiometric coefficient for species i with respect to reaction, j. For reactions that consume a species, the stoichiometric coefficient is often **negative** otherwise the stoichiometric coefficient is **positive** (See Appendix A). In considering the simple example in Figure 2.1, the stoichiometric coefficient for S with respect to v_1 is $+1$ and for v_2 is -1. That is

$$\frac{dS}{dt} = c_{s1} v_1 + c_{s2} v_2$$

or

$$\frac{dS}{dt} = (+1) v_1 + (-1) v_2 = v_1 - v_2$$

The way in which the construction of the mass-balance equation is described may seem overly formal, however the formality allows software to be written that can automatically convert network diagrams into mass-balance differential equations.

Example 2.1

Consider a linear chain of reactants from S_1 to S_5 shown in Figure 2.3. Write out the mass-balance equations for this simple system.

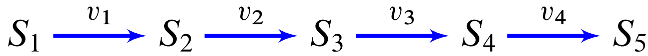


Figure 2.3 Simple Straight Chain Pathway.

$$\frac{dS_1}{dt} = -v_1 \qquad \frac{dS_2}{dt} = v_1 - v_2$$

$$\frac{dS_3}{dt} = v_2 - v_3 \qquad \frac{dS_4}{dt} = v_3 - v_4$$

$$\frac{dS_5}{dt} = v_4 \tag{2.4}$$

Each species in the network is assigned a mass-balance equation which accounts for the flows into and out of the species pool.

Example 2.2

Write out the mass-balance equation for the following branched system:

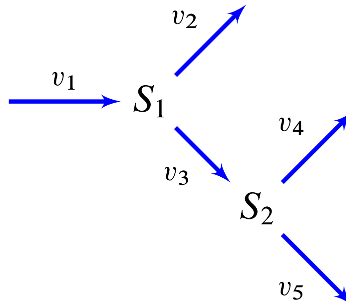


Figure 2.4 Multi-Branched Pathway.

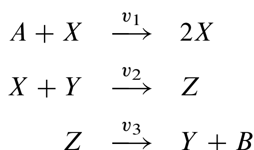
The mass-balance equations are given by:

$$\frac{dS_1}{dt} = v_1 - v_2 - v_3$$

$$\frac{dS_2}{dt} = v_3 - v_4 - v_5$$

Example 2.3

Write out the mass-balance equation for the more complex pathway:



$$\begin{aligned} A + X &\xrightarrow{v_1} 2X \\ X + Y &\xrightarrow{v_2} Z \\ Z &\xrightarrow{v_3} Y + B \end{aligned}$$

This example is more subtle because we must be careful to take into account the stoichiometry change between the reactant and product side in the first reaction (v_1). In reaction v_1, the stoichiometric coefficient for X is $+1$ because two X molecules are made for every one consumed. Taking this into account the rate of change of species X can be written as:

$$\frac{dX}{dt} = -v_1 + 2v_1 - v_2$$

or more simply as $v_1 - v_2$. The full set of mass-balance equations can therefore be written as:

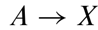
$$\frac{dA}{dt} = -v_1 \qquad \frac{dX}{dt} = v_1 - v_2$$

$$\frac{dY}{dt} = v_3 - v_2 \qquad \frac{dZ}{dt} = v_2 - v_3$$

$$\frac{dB}{dt} = v_3$$

The last example (2.3) illustrates a very important aspect of converting a network diagram into a set of differential equations. The process is potentially **lossy**. That is, it is not always possible to fully recover the original

network diagram from the set of derived differential equations. This is because in one or more of the reactions the stoichiometries may cancel out. In the example (2.3) the reaction, $A + X \longrightarrow 2X$ is not recoverable from the final set of differential equations. Instead if we reverse engineered the differential equations the first reaction would be:

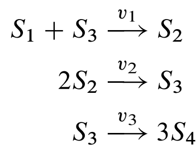


$$A \rightarrow X$$

which is not like the original. This is not perhaps a common occurrence although in protein signaling pathways it might be more common than other kinds of networks. What it means however is that sharing models by exchanging differential equations is **not** recommended. This is one reason why standard exchange formats such as SBML [41] store models explicitly as a set of reactions not as a set of differential equations. Many models are exchanged using Matlab which means that much of the biological information, particularly information on the underlining network, is lost. Exchanging models via computer languages such as Matlab is therefore not recommended.

Example 2.4

Write out the mass-balance equation for pathway:



$$\begin{aligned} S_1 + S_3 &\xrightarrow{v_1} S_2 \\ 2S_2 &\xrightarrow{v_2} S_3 \\ S_3 &\xrightarrow{v_3} 3S_4 \end{aligned}$$

In this example we have non-unity stoichiometries in the second and third reaction steps. The mass-balance equations are given by:

$$\frac{dS_1}{dt} = -v_1 \qquad \frac{dS_2}{dt} = v_1 - 2v_2$$

$$\frac{dS_3}{dt} = v_2 - v_3 \qquad \frac{dS_4}{dt} = 3v_3$$

From the previous examples we can see that it is fairly straight forward to derive the balance equations from a visual inspection of the network. Many software tools exist that will assist in this effort by converting network diagrams, either represented visually on a computer screen (for example, JDesigner) or by processing a text file that lists the reactions in the network (for example via Jarnac) into a set of differential equations (See Appendix **??**).

2.2 Stoichiometry Matrix

When describing multiple reactions in a network, it is convenient to represent the stoichiometries in a compact form called the **stoichiometry matrix**, traditionally denoted by $\mathbf{N}$, where the symbol $\mathbf{N}$ refers to number[1]. The stoichiometry matrix is a m row by n column matrix where m is the number of species and n the number of reactions:

$$\mathbf{N} = m \times n \text{ matrix}$$

The columns of the stoichiometry matrix correspond to the individual chemical reactions in the network, the rows to the molecular species, one row per species. Thus the intersection of a row and column in the matrix indicates whether a certain species takes part in a particular reaction or not, and, according to the sign of the element, whether there is a net loss or gain of substance, and by the magnitude, the relative quantity of substance that takes part in that reaction. That is the elements of the stoichiometry matrix **do not** concern themselves with the rate of reaction. This latter point is particular important when we will consider in a later chapter the various stoichiometric analyses that can be carried out purely on the stoichiometry without **any** reference to reaction rate laws.

> The stoichiometric matrix is not concerned with describing the reaction rates. Reaction rates are given by the rate laws which is a separate vector (See section 2.5).

[1] Some recent flux balance literature uses the symbol $\mathbf{S}$

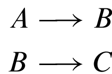
In general the stoichiometry matrix has the form:

$$\mathbf{N} = \begin{array}{c} \\ \uparrow \\ S_i \\ \downarrow \end{array} \overset{\longleftarrow \quad v_j \quad \longrightarrow}{\begin{bmatrix} c_{ij} & \cdots & \cdots \\ \vdots & & \\ \vdots & & \end{bmatrix}}$$

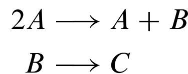
where c_{ij} is the stoichiometry coefficient for the ith species and j^{th} reaction. As was mentioned before the stoichiometry matrix is in general a lossy representation. That is, it is not always possible to revert back to the original biochemical network from which the matrix was derived. For example consider the simple stoichiometry matrix:

$$\mathbf{N} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$$

The most obvious network that this matrix could have been derived from is:



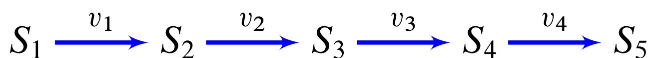
But equally plausible is this network:



If is not possible from the stoichiometry matrix alone to determine which was the original network.

Example 2.5

Write out the stoichiometry matrix for the simple chain of reactions which has five molecular species and four reactions as shown below. The four reactions are labeled, v_1 to v_4.



The stoichiometry matrix for this simple system is given by:

$$\mathbf{N} = \begin{array}{c} \\ \\ \end{array} \begin{matrix} v_1 & v_2 & v_3 & v_4 \end{matrix}$$
$$\mathbf{N} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix}$$

The rows and columns of the matrix have been labeled for convenience. Normally the labels are absent.

Example 2.6

Write out the stoichiometry matrix for the multibranched pathway shown in Figure 11.5

$$\mathbf{N} = \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 \end{matrix}$$
$$\mathbf{N} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 \end{bmatrix} \begin{matrix} S_1 \\ S_2 \end{matrix}$$

2.3 Reversiblity

Up to this point nothing has been said about whether a given reaction is reversible or not. When dealing with kinetic models, reversibility often manifests itself as a **negative reaction rate** in the rate law. For example the rate law for the simple mass-action reversible reaction, $A \rightleftharpoons B$ is often given by:

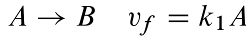
$$v = k_1 A - k_2 B$$

When this reaction goes in the reverse (right to left) direction, the reaction rate, v, will be negative. This may not be apparent from the stoichiometry

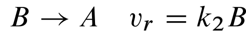
matrix, which in this case will be:

$$\mathbf{N} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Information on reversibility is therefore traditionally found in the rate law. Depending on the modeling problem, reversibility can be made more explicit in the stoichiometry matrix by specifying a **separate** reaction path for the reverse reaction. For example, in the previous example we might instead represent the system by two separate rate laws:



$$A \rightarrow B \quad v_f = k_1 A$$



$$B \rightarrow A \quad v_r = k_2 B$$

in which case the stoichiometry matrix now becomes:

$$\mathbf{N} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Splitting a reaction into separate forward and reverse steps might not always be possible however. For example an enzyme catalyzed reversible reaction such as $A \rightleftharpoons B$ **cannot** be represented using:

$$\frac{dB}{dt} = v_f - v_r$$

where the forward (v_f) and reverse (v_r) rates might be represented by irreversible Michaelis-Menten rate laws because the individual reactions are not independent but are connected by the shared enzyme pool. In such cases, the full enzyme mechanism in terms of elementary steps should be used.

To illustrate that we can apply the stoichiometry matrix to other kinds of networks, let us look at a simple signaling network and two simple gene regulatory networks.

2.4 Network Types

Signaling Networks

Figure 2.5 illustrates a simple protein signaling network, comprising two double phosphorylation cycles coupled by inhibition by protein C on the lower double cycle (D, E and F). In this model, all species are proteins and we assume that protein A and D are unphosphorylated, B and E singly phosphorylated and C and F doubly phosphorylated. C acts as a kinase and phosphorylates D and E. The reverse reactions, v_2, v_4, v_7 and v_8 are assumed to be catalyzed by phosphatases.

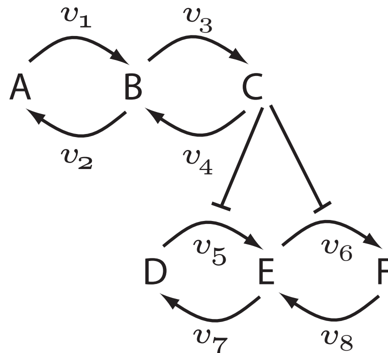


Figure 2.5 Simple Signaling Network. Protein C inhibits the activity of reactions v_5 and v_6.

There is no specified stoichiometric mechanism for the inhibition on v_5 and v_6. Therefore the stoichiometric matrix will contain no information

on this. The stoichiometric matrix for this system will look like:

$$
\mathbf{N} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array}
\begin{array}{c}
\begin{array}{cccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 \end{array} \\
\begin{bmatrix}
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1
\end{bmatrix}
\end{array}
\tag{2.5}
$$

The stoichiometric matrix can be seen to be composed of two separate blocks corresponding to the two cycle layers. It is important to note that whenever there are **regulatory** interactions in a pathway diagram, these **do not** appear in the stoichiometry matrix. Instead, such information will reside in the rate law that describes the regulation. If however the mechanism for the regulation is made explicit then details of the regulation will appear in the stoichiometry matrix. Figure 2.6 will shows a simple example of an inhibitor I regulating a reaction, S to P. On the left is displayed the implicit regulatory interaction. All we see is a blunt ended arrow indicating inhibition. In this case, details of the regulation will be found in the rate law governing the conversion of S to P. On the right is displayed an explicit mechanism, a simple competitive inhibition. In this case details of the inhibition mechanism will find its way into the stoichiometry matrix, although from an inspection of the stoichiometry matrix it is not obvious what kind of regulation it is.

Figure 2.7 shows a comparison of the implicit and explicit models in terms of the stoichiometry matrix. In each case the rate laws also change. In the implicit form, the rate law will be a Michaelis-Menten competitive inhibition model whereas in the explicit model, the rates laws (now multiplied in number) will be simple mass-action rate laws. The choice of what to use, an implicit or explicit model, will depend entirely on the type of question that the model is being used to answer. **There is no right or wrong way to do this**, the details of a model will depend on the type of question being asked.

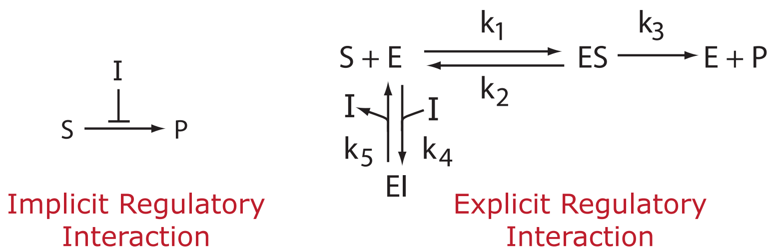


Figure 2.6 Example of implicit and explicit depiction of a regulatory interaction. The left-hand mechanism involving inhibitor I will not appear in the stoichiometry matrix whereas the explicit mechanism, right-hand figure, it will.

Gene Regulatory Networks

Consider a transcription factor P_1 that represses a gene with expression rate v_3 shown in Figure 2.8, left panel. In this model we have production of P_1 from reaction v_1 and degradation of P_1 via v_2. The construction of the stoichiometry matrix will depend on how we represent the regulated step, v_3. If regulation is implied, i.e. there is no explicit kinetic mechanism, then the regulation will not appear in the stoichiometry matrix. For the network on the left in Figure 2.8, the stoichiometry matrix will be given by:

$$\mathbf{N} = P_1 \begin{array}{c} \\ \left[\right. \end{array} \begin{array}{cc} v_1 & v_2 \\ 1 & -1 \end{array} \begin{array}{c} \\ \left. \right] \end{array} \tag{2.6}$$

The stoichiometry matrix has only one row indicating that there is only one species in the model, P_1 and there is no hint in the stoichiometry matrix that there is regulation. In this model, P_1 is not explicitly sequestered by the operator site that is upstream of the gene. We make the significant assumption that when P_1 regulates, it is itself is not affected in any way.

Consider now that the interaction between P_1 and v_3 is made mechanistically explicit. The right hand network in Figure 2.8 shows one possible way in which to represent the interaction of the transcription factor, P_1 with gene v_3. In the explicit model, the transcription factor, P_1 is as-

$$\mathbf{N} = \begin{array}{c} \\ S \\ P \\ I \end{array} \begin{array}{c} v_1 \\ \left[\begin{array}{c} -1 \\ 1 \\ 0 \end{array}\right] \end{array} \qquad \mathbf{N} = \begin{array}{c} \\ S \\ P \\ I \\ ES \\ EI \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \left[\begin{array}{ccccc} -1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right] \end{array}$$

Implicit Explicit

Figure 2.7 Stoichiometry matrices corresponding to the two models in Figure 2.6

sumed to bind to a repressor site preventing gene expression. In the ex-

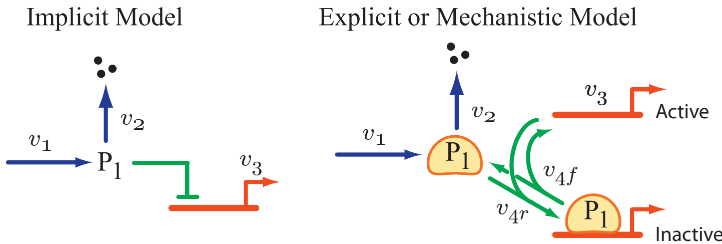


Figure 2.8 Two simple gene regulatory networks involving gene repression. On the left side is the implicit model where P_1 represses v_3, on the right side is the explicit model showing a more detailed mechanism for the regulation.

plicit model there are two new species, designated active gene and inactive gene. The stoichiometry matrix will therefore include two additional rows corresponding to these two new species. The stoichiometry matrix for the

explicit model is shown below:

$$
\mathbf{N} = \begin{array}{c} \\ P_1 \\ P_1(\text{Active}) \\ P_1(\text{InActive}) \end{array}
\begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_{4r} & v_{4f} \end{array} \\
\left[\begin{array}{cccc} 1 & -1 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{array} \right] \end{array} \tag{2.7}
$$

In this case, P_1 is actively sequestered on to the operator site and therefore appears in the stoichiometry matrix. Processes such as consumption, production or sequestration by some binding mechanism will appear as columns in the stoichiometry matrix. Regulation that is often depicted by arrow or blunt ends are modeled in the rate law itself and therefore do not appear in the stoichiometry matrix.

In conclusion, regulation does not appear explicitly in a stoichiometry matrix unless the regulation is represented in an explicit mechanistic scheme. The choice of implicit or explicit representations depends on the question being asked and the availability of suitable data.

2.5 The System Equation

Equation 2.3, which describes the mass balance equation, can be reexpressed in terms of the stoichiometry matrix to form the **system equation**.

$$
\frac{d\mathbf{s}}{dt} = \mathbf{N}\boldsymbol{v} \tag{2.8}
$$

where $\mathbf{N}$ is the $m \times n$ stoichiometry matrix and $\boldsymbol{v}$ is the n dimensional rate vector, whose ith component gives the rate of reaction i as a function of the species concentrations. $\mathbf{s}$ is the m vector of species.

Looking again at the simple chain of reactions in Figure 2.3, the system equation can be written down as:

$$
\frac{d\mathbf{s}}{dt} = \mathbf{N}\boldsymbol{v} = \left[\begin{array}{cccc} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{array} \right] \left[\begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \right] \tag{2.9}
$$

If the stoichiometry matrix is multiplied into the rate vector, the mass-balance equations show earlier (2.4) are recovered. To illustrate that the system equation might look like for a simple system, consider the model in Jarnac format:

```
p = defn cell
     A -> B; k1*A - k2*B;
     B -> C; k3*B - k4*C;
end;

p.k1 = 0.1; p.k2 = 0.02;
p.k3 = 0.3; p.k4 = 0.04;
p.A  = 10;  p.B  = 0; p.C = 0;
```

The system equation for this model will be given by:

$$\frac{d\mathbf{s}}{dt} = \mathbf{N}\boldsymbol{v} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 A - k_2 B \\ k_3 B - k_4 C \end{bmatrix} \tag{2.10}$$

All stoichiometric interactions are placed in the stoichiometry matrix.

The example shown in Figure 2.5 and Figure 2.8 illustrated non-stoichiometric interactions, namely two inhibition interactions from C to reactions v_5 and v_6 and repression on v_3 by P_1. As was noted, these interactions do no occur in the stoichiometry matrix. Instead they will be found in the rate vector, $\boldsymbol{v}$ in the form of a particular rate law.

The stoichiometry matrix represents the mass transfer connectivity of the network and contains information on the network's structural mass-transfer characteristics. These characteristics fall into two groups, relationships among the species and relationships among the reaction rates. These relationships will be considered in detail in a later chapter (Chapter **??** and **??**).

2.6 Jarnac

The modeling platform Jarnac provides facilities to extract the stoichiometry matrix from a model. The command for generating the stoichiometry matrix is p.sm assuming the model is stored in the variable p. The script and results of a run are given below:

```
p = defn cell
     J1: A -> B; k1*A - k2*B;
     J2: B -> C; k3*B - k4*C;
end;

p.k1 = 0.1; p.k2 = 0.02;
p.k3 = 0.3; p.k4 = 0.04;
p.A  = 10;  p.B  = 0; p.C = 0;

// Print out the stoichiometry matrix
println p.sm;
```

If this script is run, the output is as shown below:

```
       J1     J2
A{{    -1      0 }
B {     1     -1 }
C {     0      1 }}
```

Note that in Jarnac, matrices are labeled, this is useful for identifying the corresponding species and reactions in the stoichiometry matrix.

Further Reading

1. Sauro HM (2012) Enzyme Kinetics for Systems Biology. 2nd Edition, Ambrosius Publishing ISBN: 978-0982477335

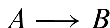
2. Stephanopoulos G, Aristidou A, and Nielsen J (1998) Metabolic engineering: principles and methodologies. Academic Press, ISBN: 978-0126662603

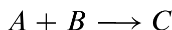
3. Palsson BO (2006) Systems Biology Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, ISBN: 978-0521859035

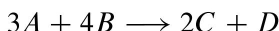
Exercises

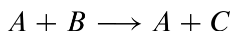
1. Explain the difference between the terms: Stoichiometric amount, Stoichiometric coefficient, rate of change (dX/dt) and reaction rate (v_i). Refer to Appendix A to answer this question.

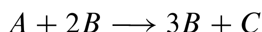
2. Determine the stoichiometric amount and stoichiometric coefficient for each species in the following reactions:





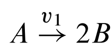


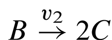


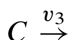


$$A \longrightarrow B$$
$$A + B \longrightarrow C$$
$$A \longrightarrow B + C$$
$$2A \longrightarrow B$$
$$3A + 4B \longrightarrow 2C + D$$
$$A + B \longrightarrow A + C$$
$$A + 2B \longrightarrow 3B + C$$

3. Derive the set of differential equations for the following model in terms of the rate of reaction, v_1, v_2 and v_3:

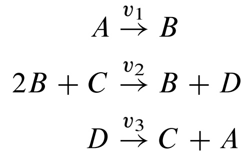






$$A \stackrel{v_1}{\rightarrow} 2B$$
$$B \stackrel{v_2}{\rightarrow} 2C$$
$$C \stackrel{v_3}{\rightarrow}$$

4. Derive the set of differential equations for the following model in terms of the rate of reaction, v_1, v_2 and v_3:



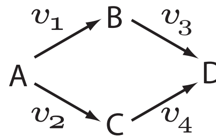
$$
\begin{aligned}
A &\stackrel{v_1}{\rightarrow} B \\
2B + C &\stackrel{v_2}{\rightarrow} B + D \\
D &\stackrel{v_3}{\rightarrow} C + A
\end{aligned}
$$

5. Write out the stoichiometry matrix for the networks in question 3 and 4

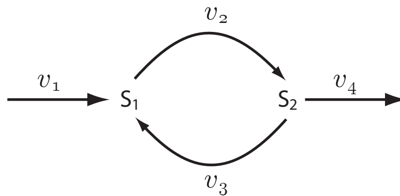
6. Enter the previous models, 3 and 4, into Jarnac and confirm that the Jarnac stoichiometry matrices are the same as those derived manually in the exercises.

7. Derive the stoichiometry matrix for each of the following networks. In addition write out the mass-balance equations in each case.

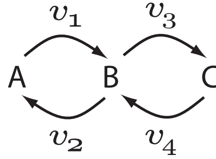
(a)



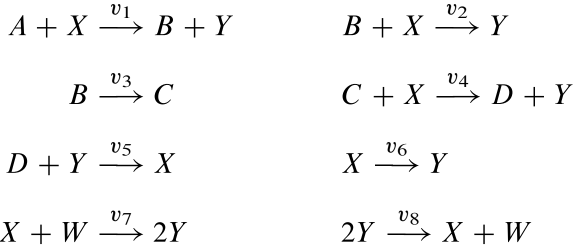
(b)



(c)



(d)



$$A + X \xrightarrow{v_1} B + Y \qquad B + X \xrightarrow{v_2} Y$$

$$B \xrightarrow{v_3} C \qquad C + X \xrightarrow{v_4} D + Y$$

$$D + Y \xrightarrow{v_5} X \qquad X \xrightarrow{v_6} Y$$

$$X + W \xrightarrow{v_7} 2Y \qquad 2Y \xrightarrow{v_8} X + W$$

8. For the irreversible enzyme catalyzed reaction, $A \rightarrow B$:

(a) Write out the stoichiometry matrix.

(b) Write out the stoichiometry matrix in terms of the elementary reactions that make up the enzyme mechanism.

9. A gene G_1 expresses a protein p_1 at a rate v_1. p_1 forms a tetramer (4 subunits), called p_1^4 at a rate v_2. The tetramer negatively regulates a gene G_2. p_1 degrades at a rate v_3. G_2 expresses a protein, p_2 at a rate v_9. p_2 is cleaved by an enzyme at a rate v_4 to form two protein domains, p_2^1 and p_2^2. p_2^1 degrades at a rate v_5. Gene G_3 expresses a protein, p_3 at a rate v_6. p_3 binds to p_2^2 forming an active complex, p_4 at a rate v_{10}, which can bind to gene G_1 and activate G_1. p_4 degrades at a rate v_7. Finally, p_2^1 can form a dead-end complex, p_5, with p_4 at a rate v_8.

10. (a) Draw the network represented in the description given above.

(b) Write out the differential equation for each protein species in the network in terms of $v_1, v_2, \ldots$.

(c) Write out the stoichiometric matrix for the network.

11. Write out the differential equations for the system depicted in equation 2.9.

12. Given the following stoichiometry matrix, write out the corresponding network diagram. Why might this process not fully recover the original network from which the stoichiometry matrix was derived?

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \\ G \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \left[\begin{array}{rrrrr} -1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 3 \\ 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{array} \right] \end{array} \tag{2.11}$$

13. Why is it better to store a model as a list of reactions rather than a set of differential equations?

3
How Systems Behave

3.1 System Behavior

Ultimately we are interested in what kinds of behavior systems can display, how that behavior is generated and with that understanding how systems can be manipulated and controlled. As we proceed through the book we will encounter many different kinds of behavior. At this stage however it is worth describing the states that are fundamental to all systems. These states fall into three groups: **(Thermodynamic) equilibrium**, **steady state**, and **transients**. In the literature the terms equilibrium and steady state are often used to mean the same thing but here they will be used to describe two very different states.

The simplest and arguably the least interesting is equilibrium, or more precisely thermodynamic equilibrium.

3.2 Equilibrium

Thermodynamic equilibrium, or simply equilibrium, refers to the state of a system when all forces are balanced. In chemistry, thermodynamic equi-

librium is when all forward and reverse rates are equal. This also means that the concentration of chemical species are also unchanging and all net flows are zero. Equilibrium is easily achieved in a closed system. For example, consider the simple chemical isomerization:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \tag{3.1}$$

Let the net forward rate of the reaction, v, be equal to $v = k_1 A - k_2 B$. The rates of change of A and B are given by:

$$\frac{dA}{dt} = -v \qquad \frac{dB}{dt} = v$$

At equilibrium dA/dt and dB/dt equal zero, that is $Ak_1 = Bk_2$, or $v = 0$. The analytical solution to the chemical isomerization can be derived as follows. Given that the system is closed we know that the total mass in the system, $A + B$ is constant. This constant is given by the sum of the initial concentrations of A and B which we will define as $A_o + B_o$. Note that $A_o + B_o = A(t) + B(t)$ is always true. We assume that the volume is constant and set to unit volume, this allows us to state that the sum of the concentrations is conserved. The differential equation for A is given by:

$$\frac{dA}{dt} = k_2 B - k_1 A$$

Before solving this equation, let us replace B by the term $A_o + B_o - A$. This yields:

$$\frac{dA}{dt} = k_2 A_o + k_2 B_o - k_2 A - k_1 A = k_2(A_o + B_o) - A(k_1 + k_2)$$

The easiest way to solve this equation is to use Mathematica or Maxima. The Mathematica command is `DSolve[A'[t] == k2 (Ao + Bo) - A[t] (k1 + k2), A[0] == Ao, A[t], t]`, where `A[0] == Ao` sets the initial condition for the concentration of A to be A_o. By implication, the initial condition for B_o is $(A_o + B_o) - A_o = B_o$. The result of applying the Mathematica command yields the following solution:

$$A(t) = \frac{(A_o + B_o)k_2}{k_1 + k_2} + \frac{e^{-(k_1+k_2)t}\, v_{\text{initial}}}{k_1 + k_2}$$

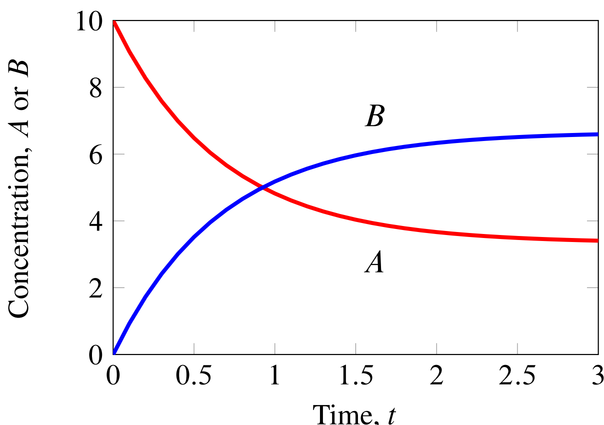


Figure 3.1 Time course for equilibration of the reversible reaction in model 3.1 where $k_1 = 1, k_2 = 0.5, A_o = 10, B_o = 0$. The ratio of the equilibrium concentration is given by k_1/k_2. Jarnac model: 3.1

The first term in the equation is a constant and equals the equilibrium concentration of A. The first term is however a function of the total mass in the system $(A_o + B_o)$ which means that the equilibrium solution is independent of the starting concentrations so long as the total remains the same. The second term is time dependent and describes the evolution of the system when the initial concentrations of A and B are not set to the equilibrium concentrations. The initial concentrations are set in the term v_{initial} which is the reaction rate, v, at $t = 0$. The second term also has an exponential component which approaches zero as time goes to infinity so that at infinite time we are left with the first term which equals the concentration of A when $dA/dt = dB/dt = 0$.

At equilibrium the reaction rate can be computed by substituting the equilibrium concentration of A and B into the reaction rate, $v = k_2 B - k_1 A$. We note that the equilibrium concentration of A is given by:

$$A_{eq} = \frac{(A_o + B_o)k_2}{k_1 + k_2}$$

and for B by subtracting A_{eq} from $A_o + B_o$. When the A_{eq} and B_{eq} relations are substituted into v, the result is:

$$v = 0$$

From this somewhat long winded analysis, it has been determined for the closed reversible system, at infinite time, the concentrations of A and B reach some constant values and that the net rate, v is zero. The system is therefore at thermodynamic equilibrium.

In biochemical models it is often assumed that when the forward and reverse rates for a particular reaction are very fast compared to the surrounding reactions that the reaction is said to be in **quasi-equilibrium**. That is, although the entire system may be out of equilibrium there may be parts of the system that can be approximated as though they were in equilibrium. This is often done to simplify the modeling process. Living organisms are not themselves at thermodynamic equilibrium, if they were then they would technically be dead. Living systems are open so that there is a continual flow of mass and energy across the system's boundaries.

3.3 Steady State

The **steady state**, also called the stationary state, is where the rates of change of all species, dS/dt are zero **but** at the same time the net rates are non-zero, that is $v_i \neq 0$. This situation can only occur in an open system, that is a system that can exchange matter with the surroundings. To convert the simple reversible model described in the last section into an open system we need only add a source reaction and a sink reaction as shown in the following scheme:

$$X_o \stackrel{v_o}{\rightarrow} A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \stackrel{k_3}{\rightarrow} \tag{3.2}$$

In this case simple mass-action kinetics is assumed for all reactions. It is also assumed that the source reaction, with rate v_o, is irreversible and originates from a boundary species, X_o, that is X_o is fixed. In addition it is assumed that sink reaction, with rate constant, k_3 is also irreversible.

For the purpose of making it easier to derive the time course solution, the reverse rate constant, k_2 will be assumed to equal zero and we will set the initial conditions for A and B to both equal zero. The mathematical solution for the system can again be obtained using Mathematica:

$$\begin{aligned} A(t) &= v_o \frac{1 - e^{-k_1 t}}{k_1} \\ B(t) &= v_o \frac{k_1 \left(1 - e^{-k_3 t}\right) + k_3 \left(e^{-k_1 t} - 1\right)}{k_3 \left(k_1 - k_3\right)} \end{aligned} \tag{3.3}$$

As t tends to infinity $A(t)$ tends to v_o/k_1 and $B(t)$ tends to v_o/k_3. In addition, the reaction rate through each of the three reaction steps is v_o. This can be confirmed by substituting the solutions for A and B into the reaction rate laws. Given that v_o is greater than zero and that A and B reach constant values given sufficient time, we conclude that this system eventually settles to a steady state rather than thermodynamic equilibrium. The system displays a continuous flow of mass from the sink to the source. This can only continue undisturbed so long as the source material, X_o never runs out and that the sink is continuously emptied. Figure 3.2 shows a simulation of this system.

> At steady state, the rate of mass transfer across a reaction is often called the flux, or J.

At steady state the net reaction rate is also called the **pathway flux**, often symbolized with the latter J.

Thermodynamic Equilibrium and Steady State

Thermodynamic equilibrium (or equilibrium for short) and the steady state are distinct states of a chemical system. If we consider a system where every part is in equilibrium then we can be sure of two things. That the species concentrations are unchanging and most importantly there are no net flows of mass or energy within the system or between the system and the environment. A system that is in equilibrium must have the following properties:

$$\frac{d\mathbf{s}}{dt} = 0$$

$$\text{for all } i: v_i = 0$$

where v_i is the net reaction rate for the ith reaction step. When a biological system is at equilibrium, we say it is **dead**. Thermodynamically we can also say that entropy production is at zero and has reached its **maximum value**.

The steady state has some similarities with the equilibrium state, species concentrations are still unchanging, however **there are net flows** of energy and mass within the system and with the environment. Systems at steady state must therefore be open and will necessarily continuously dissipate any gradients between the system and the external environment. This means that one or more v_is must be non-zero

The steady state is defined when all dS_i/dt are equal to zero while one or more reaction rates are non-zero:

$$\frac{d\mathbf{s}}{dt} = 0$$

$$v_i \neq 0$$

Thermodynamically, we can also say that the entropy production of the system at steady state is lower than the entropy production in the environment. In some of the literature the terms equilibrium and steady state are used interchangeably resulting in possible confusion. In this book, the word equilibrium will be used to refer to a system at thermodynamic equilibrium not at steady state.

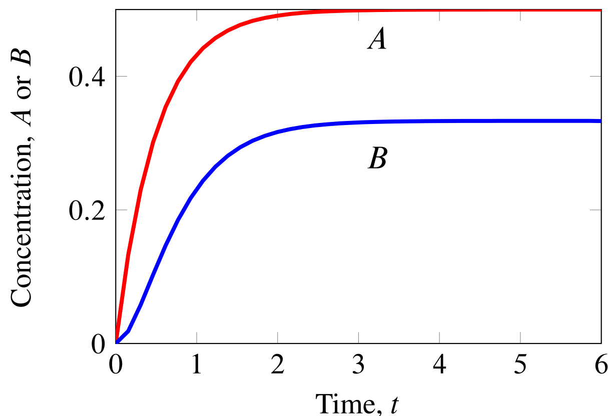


Figure 3.2 Time course for an open system reaching steady state in model 3.4 where $v_o = 1, k_1 = 2, k_2 = 0, k_3 = 3, A_o = 0, B_o = 0$. X_o is assumed to be fixed. The Jarnac model: 3.2

We can sometimes also calculate the steady state mathematically. In the last example we used the simplified model:

$$X_o \overset{v_o}{\rightarrow} A \overset{k_1}{\rightarrow} B \overset{k_3}{\rightarrow} \tag{3.4}$$

The differential equations for this system are:

$$\frac{dA}{dt} = v_o - k_1 A$$

$$\frac{dB}{dt} = k_1 A - k_3 B$$

If we set the rates of change to zero:

$$0 = v_o - k_1 A$$
$$0 = k_1 A - k_3 B$$

We have equations in two unknown, A and B. We can solve for A and B

to obtain:

$$A = v_o/k_1$$

$$B = v_o/k_3$$

Usually however we cannot solve the equations and so must revert to computer simulation or useing specialist software (such as Jarnac) to compute the steady state. The script below show a Jarnac model where we ask Jarnac to solve for the steady state using the command `p.ss.eval`.

```
p = defn model
     $Xo -> A;  vo;
      A -> B;   k1*A;
      B -> $X1; k3*B;
end;

// Set up the model initial conditions
p.Xo = 1;   p.X1 = 0;
p.k1 = 0.2; p.k3 = 0.3;
p.vo = 0.5;

// Evaluation the steady state
p.ss.eval;

println "Steady State values:", p.A, p.B;

// Output follows:
Steady State values:  2.5  1.66667
```

3.4 Transients

The final simple behavior that a system can show is a transient. A transient is usually the change that occurs in the species concentrations as the system moves from one state, often a steady state, to another. Equation 3.3 shows the solution to a simple system that describes the transient behavior of species A and B. Figure 3.2 illustrates the transient from an initial condition, in this case from a non-steady state condition to a steady state. A

periodic (such as an oscillation) or a chaotic system may be considered a transient, one that is unable to settle to a fixed steady state. In the case of a system showing periodic behavior, the transient repeats itself indefinitely at regular intervals called the period. In a chaotic system, the transient never repeats the exact same trajectory but will continue indefinitely.

3.5 Setting up a Model in Software

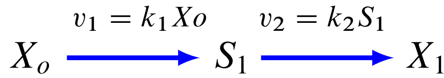
There are many software tools both commercial and free (including open source) that one can use to build models of cellular networks. In this book we will use Jarnac, a software tool written by the author. Jarnac is a script based tool where one enters a model as a text file, the model is then compiled, run and the results displayed. Jarnac is currently a windows based application. It is quite easy to set up models using Jarnac but it also has a fairly complete programming language built-in that allows advanced users to do some sophisticated analysis. A brief introduction on how to use Jarnac is given in Appendix **??**. For those who wish to use other tools, such as COPASI (`http://www.copasi.org`), CellDesigner (`http://celldesigner.org/` or even Matlab (`http://www.mathworks.com`, it is easy to convert Jarnac files into standard Systems Biology Markup Language (SBML) or Matlab scripts and then load the models into the simulation tool of choice.

3.6 Effect of Different Kinds of Perturbations

When we talk about model dynamics we mean how species levels and reaction rates change in time as the model evolves. There are a number of ways to elicit a dynamic response in a model, the two we will consider here are perturbations to species and perturbations to model parameters.

Effect of Perturbing Floating Species

Let us consider a two step pathway of the following form:



We assume that X_o and X_1 are fixed. If the initial concentration of S_1 is zero then we can run a simulation and allow the system to come to steady state. This is illustrated in Figure 3.3

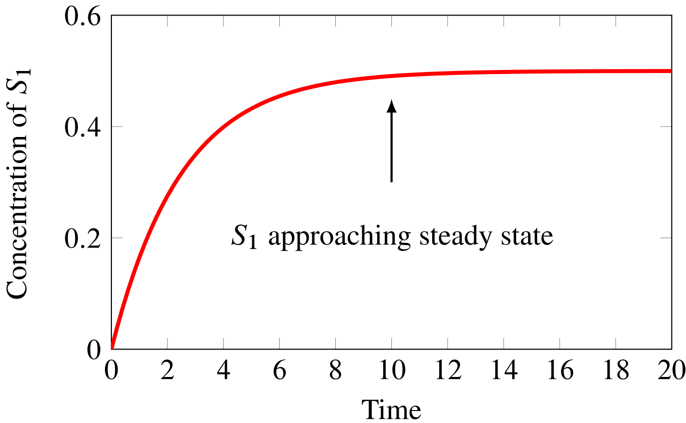


Figure 3.3 S_1 approaching steady state. Jarnac model: 3.3

Once at steady state we can consider applying perturbations to see what happens. For example, Figure 3.4 illustrates the effect of injecting 0.35 units of S_1 at $t = 20$ and watching the system evolve. The Jarnac script used to generate this graph is shown in the chapter Appendix. In practice this could be accomplished this by injecting 0.35 units of S_1 into the volume where the pathway operates. What we observe is that the concentration of S_1 initially jumps by the amount 0.35, then relaxes back to its steady state concentration before the perturbation was made (Figure 3.4). When we apply perturbations to species concentrations and the change relaxes back to the original state, we call the system **stable**. We will return to the topic of stability in the next section.

Figure 3.4 illustrates perturbing one of the floating molecular species by physically adding a specific amount of the substance to the pathway. In

many cases we will find that the system will recover from such perturbations as we see in Figure 3.4. We are not limited to single perturbations, Figure 3.5 shows multiple perturbations, both positive and negative. Not all systems show recovery like this but those that do not are called **unstable**. That is when we perturb a species concentration, instead of the perturbation relaxing back, it begins to diverge.

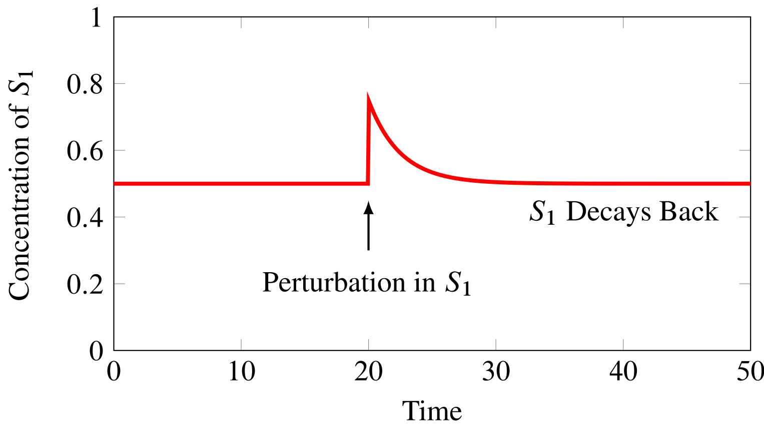


Figure 3.4 Stability of a simple biochemical pathway at steady state. The steady state concentration of the species S_1 is 0.5. A perturbation is made to S_1 by adding an additional 0.35 units of S_1 at time = 20. The system is considered stable because the perturbation relaxes back to the original steady state. Jarnac model: 3.4

.

Effect of Perturbing Model Parameters

In addition to perturbing floating species we can also perturb the model parameters. Such parameters include kinetic constants and boundary species. When changing a parameter we can do it in two ways, we can make a permanent change or we can make a change and at some time later return the parameter to its original value. Assuming that the steady state is stable, a temporary change will result in the steady state changing then recovering to the original state when the parameter is changed back. Figure 3.6

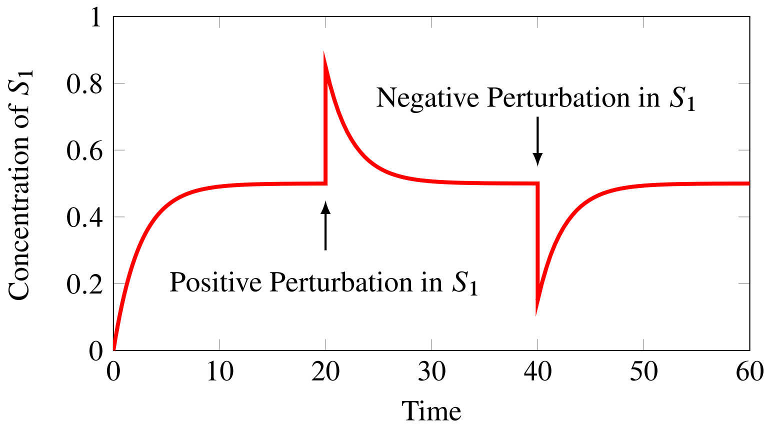


Figure 3.5 Multiple Perturbations. The steady state concentration of the species S_1 is 0.5 and a perturbation is made to S_1 by adding an additional 0.35 units of S_1 at time = 20 and removing 0.35 units at time = 40. In both cases the system relaxes back. Jarnac script: 3.5

shows the effect of perturbing the rate constant, k_1 and then restoring the parameter to its original value at some time later.

In some applications other types of perturbations are made. For example in studying the infusion of a drug where the concentration of the drug is a model parameter, one might use a slow linear increase in the drug concentration. Such a perturbation is called ramp. More sophisticated analyzes might require a sinusoidal change in a parameter, an impulse, a pulse or an exponential change. The main point to remember is that parameter changes will usually result in changes to the steady state concentrations and fluxes.

For completeness, Figure 3.7 shows what happens when we perturb both a parameter and a species concentration. As expected the species concentration does not recover to the original steady state.

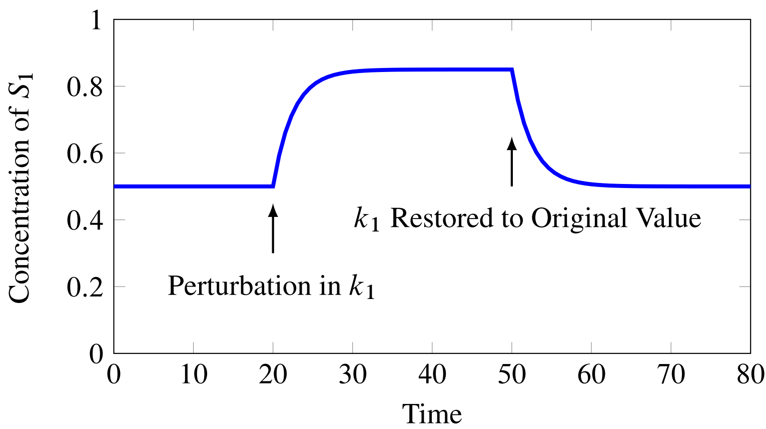


Figure 3.6 Effect of Perturbing Model Parameters. Jarnac script: 3.6

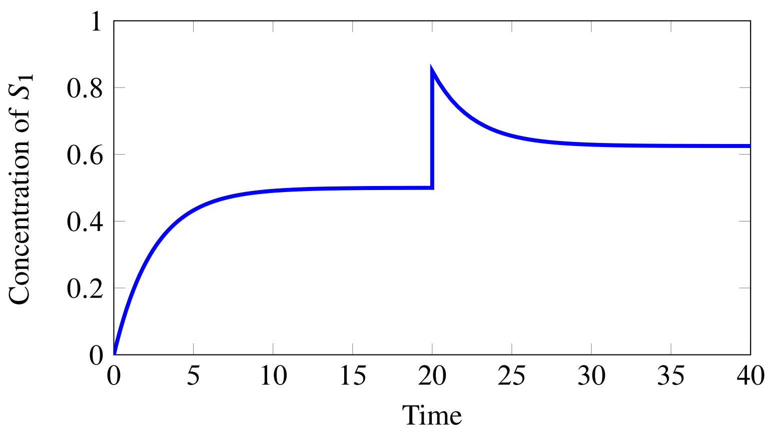


Figure 3.7 Effect of Perturbing Model Parameters and Species Concentration.

3.7 Sensitivity Analysis

Sensitivity analysis at steady state looks at how particular model variables are influenced by model parameters. There are at least two main reasons why it is interesting to examine sensitivities. The first is a practical one. Many kinetic parameters we use in building biochemical models can have a significant degree of uncertainty about them. By determining how much each parameter has an influence on the model's state we can decide whether we should improve our confidence in the particular parameter. A parameter that has considerable influence but at the same time has significant uncertainty is a parameter that should be determined more carefully by additional experimentation. On the other hand a parameter that has little influence but has significant uncertainly associated with it, is relatively unimportant. A sensitivity analysis can therefore be used to highlight parameters that need better precision.

The second reason for measuring sensitivities is to provide insight. The degree to which a parameter can influence a variable tells us something about how the network is responding to perturbations. Such a study can be used to answer questions about robustness and adaptation. We will delay further discussion of this important topic to the second half of the book where we will describe it much more detail.

How are sensitivities represented? Traditionally there are two way, one based on absolute sensitivities and the second based on relative sensitivities. Absolute sensitivities are simply given by the ratio of the absolute change in the variable to the absolute change in the parameter. That is:

$$S = \frac{\Delta V}{\Delta p}$$

where V is the variable and p the parameter. This equation shows finite changes to the parameter and variable. Unfortunately because most systems are nonlinear and therefore the value for the sensitivity will be a function of the size of the finite change. To make the sensitivity independent of the size of the change, the sensitivity is usually defined in terms of

infinitesimal changes:

$$S = \frac{dV}{dp}$$

Although absolute sensitivities are simple they have one drawback, the value can be influenced by the units used to measure the variable and parameter. Often in making experimental measurements we won't be able to measure the quantity using the most natural units, instead we may have measurements in terms of fluorescence, colony counts, staining on a gel and so on. Is is most likely that the variable and parameter units will be quite different and each laboratory may have its own way particular units is uses. Absolute sensitivities are therefore quite difficult to compare.

To get round the problem of units, many people will use relative sensitivities, These are simple scaled absolute sensitivities:

$$S = \frac{dV}{dp}\frac{p}{V}$$

The sensitivity is defined in terms of infinitesimal changes for the same reason cited before. The reader may also recall that elasticities are measured in this way. Relative sensitivities are immune from the units we use to measure quantities but also relative sensitivities correspond more closely to how many measurements are made, often in terms of relative or fold changes. In practice steady state relative sensitivities should be measured by taking a measurement at the operating steady state, making a perturbation (preferable a small one), waiting for the system to reach a new steady state then measuring the system again. It is important to be aware that the steady state sensitivities measure how a perturbation in a parameter moves the system from one steady state to another.

3.8 Robustness and Homeostasis

Biological organisms are continually subjected to perturbations. These perturbations can originate from external influences such as changes in temperature, light or the availability of nutrients. Perturbations can also arise internally due to the stochastic nature of molecular events or by

natural genetic variation. One of the most remarkable and characteristic properties of living systems is their ability to resist such perturbations and maintain very steady internal conditions. For example the human body can maintain a constant core temperature of 36.8°C ±0.7 even though external temperatures may vary widely. The ability of a biological system to maintain a steady internal environment is called **homeostasis**, a phrase introduced by Claude Bernard almost 150 years ago. Modern authors may also refer to this behavior as **robustness**, although this word is used in many other contexts.

There are a number of mechanisms that are used in biology to maintain homeostasis. Perhaps the most common is negative feedback. This is where the difference between the desired output and the actual output is used to modulate the process that determines the output. For example, if the output is lower than the desired output then the process will increase the output. Such systems are found at multiple levels in a living organism, including subcellular processes such as metabolism and multicellular processes that control for example the level of glucose in the blood stream. One way to measure the degree of robustness or homoeostasis in a system is to use sensitivity analysis. We investigate the use of negative feedback to maintain concentrations within a narrow range in a later chapter.

3.9 Stability

Figure 3.4 shows a simulation where a species concentration is disturbed and over time relaxes back to the original steady state. This is an example of a stable steady state.

The differential equation for the single floating species, S_1, is given by

$$\frac{dS_1}{dt} = k_1 Xo - k_2 S_1 \tag{3.5}$$

and as we saw before, with a steady state solution

$$S_1 = k_1 Xo / k_2 \tag{3.6}$$

The question we wish to ask here is whether the steady state is stable or not, that is whether perturbation to species recover or not? We can show that

the two step model is stable by using the following mathematical argument. The differential equation describing the two step model is given by,

$$\frac{dS_1}{dt} = k_1 X_o - k_2 S_1$$

If the system is at steady state, let us make a small perturbation to the steady state concentration of S_1, δS_1 and ask how δS_1 changes as a result of this perturbation, that is what is $d(\delta S_1)/dt$? The new rate of change equation is rewritten as follows:

$$\frac{d(S_1 + \delta S_1)}{dt} = k_1 X_o - k_2(S_1 + \delta S_1)$$

If we insert the steady state solution for S_1 (equation 3.6) into the above equation we are left with:

$$\frac{d\delta S_1}{dt} = -k_2 \delta S_1 \tag{3.7}$$

In other words the rate of change of the **disturbance itself**, δS_1 is negative, that is, the system attempts to reduce the disturbance so that the system returns back to the original steady state. Systems with this kind of behavior are called **stable**. If the rate of change in S_1 had been positive instead of negative however, the perturbation would have continued to diverge away from the original steady state and the system would them be considered **unstable**.

A biochemical pathway is dynamically stable at steady state if small perturbations in the floating species concentrations relax back to the original state.

To continue, let us divide both sides of equation 3.7 by δS_1 and taking the limit, we find that $\partial(dS_1/dt)/\partial S_1$ is equal to $-k_2$. The stability of this simple system can therefore be determined by inspecting the sign of $\partial(dS_1/dt)/\partial S_1$ which can be easily determined by taking the derivatives of the differential equations with respect to the species concentrations.

For larger systems the stability of a system can be determined by looking at all the terms $\partial(dS_i/dt)/\partial S_i$ which are given collectively by the expression:

$$\frac{d(d\mathbf{s}/dt)}{d\mathbf{s}} = \boldsymbol{J} \tag{3.8}$$

where $\boldsymbol{J}$ is called the **Jacobian matrix** containing elements of the form $\partial(dS_i/dt)/\partial S_i$. Equation 3.7 can be generalized to:

$$\frac{d(\delta\mathbf{s})}{dt} = \boldsymbol{J}\delta\mathbf{s} \tag{3.9}$$

where $\boldsymbol{J}$ is given by

$$\begin{bmatrix} \frac{\partial S_1/dt}{\partial S_1} & \cdots & \frac{\partial S_1/dt}{\partial S_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial S_m/dt}{\partial S_1} & \cdots & \frac{\partial S_m/dt}{\partial S_m} \end{bmatrix}$$

Equation 3.9 is an example of an unforced linear differential equation and has the general form:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x}$$

Solutions to such equations are well known and take the form:

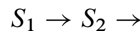
$$x_j(t) = c_1\boldsymbol{K}_1 e^{\lambda_1 t} + c_2\boldsymbol{K}_2 e^{\lambda_2 t} + \cdots c_n\boldsymbol{K}_n e^{\lambda_n t}$$

That is the solution to an unforced linear differential equations involves a sum of exponentials, $e^{\lambda_i t}$, constants c_i and vectors, $\boldsymbol{K}_i$. The exponents of the exponentials are given by the eigenvalues (See Appendix C) of the matrix, $\boldsymbol{A}$ and $\boldsymbol{K}_i$ the corresponding eigenvectors. The c_i terms are related to the initial conditions assigned to the problem. It is possible for the eigenvalues to be complex but in general if the real parts of the eigenvalues are negative then the exponents decay (stable) whereas if they are positive the exponents grow (unstable). We can therefore determine the stability properties of a given model by computing the eigenvalues of the Jacobian

matrix and looking for any positive eigenvalues. Note that the elements of the Jacobian matrix will often be a function of the species levels, it is therefore important that the Jacobian be evaluated at the steady state of interest.

Example 3.1

The following system:



$$S_1 \rightarrow S_2 \rightarrow$$

if governed by the following set of differential equations:

$$\frac{dS_1}{dt} = -2S_1$$

$$\frac{dS_2}{dt} = 2S_1 - 4S_2$$

The solution to this system can be derived using Mathematica or by using standard algebraic method for solving linear homogeneous systems. The solution can be found to be:

$$\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-2t} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-4t}$$

$$S_1 = c_1 e^{-2t}$$

$$S_2 = c_1 e^{-2t} + c_2 e^{-4t}$$

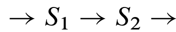
Since the exponents are all negative (-2, -2 and -4), the system is stable to perturbations in S_1 and S_2.

There are many software packages that will compute the eigenvalues of a matrix and there are a small number packages that can compute the Jacobian directly from the biochemical model. For example, the script below is taken from Jarnac, it defines the simple model, initializes the model values, computes the steady state and then prints out the eigenvalues of the Jacobian matrix. For a simple one variable model, the Jacobian matrix only has a single entry and the eigenvalue corresponds to that entry. The output from running the script is given below showing that the eigenvalue is -0.3. Since we have a negative eigenvalue, the pathway must be stable to perturbations in S_1.

```
p = defn model
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;
end;

// Set up the model initial conditions
p.Xo = 1;   p.X1 = 0;
p.k1 = 0.2; p.k2 = 0.3;

// Evaluation the steady state
p.ss.eval;
// print the eigenvalues of the Jacobian matrix
println eigenvalues (p.Jac);

// Output follows:
{    -0.3}
```

Example 3.2

The following system:



$$\rightarrow S_1 \rightarrow S_2 \rightarrow$$

if governed by the following set of differential equations:

$$\frac{dS_1}{dt} = 3 - 2S_1$$

$$\frac{dS_2}{dt} = 2S_1 - 4S_2$$

The Jacobian matrix is computed by differentiating the equations with respect to the steady state values of S_1 and S_2:

$$\boldsymbol{J} = \begin{bmatrix} -2 & 0 \\ 2 & -4 \end{bmatrix}$$

The eigenvalues for this matrix are: -2 and -4 respectively. Since both eigenvalues are negative the system is stable to small perturbations in S_1 and S_2.

The pattern of eigenvalues can tell us a lot about stability but also about the form of the transients that will occur when we perturb the state of the system. In the next section we will investigate this aspect.

Further Reading

1. Kipp E, Herwig R, Kowald A, Wierling C and Lehrach H (2005) Systems Biology in Practice, Wiley-VCH Verlag

2. Steuer R, Junker BH (2008) Computational Models of Metabolism: Stability and Regulation in Metabolic Networks, Advances in Chemical Physics, Volume 142, (ed S. A. Rice), John Wiley & Sons, Inc.

3. Stucki JW (1978) Stability analysis of biochemical systems–a practical guide. Prog Biophys Mol Biol. 33(2):99-187.

Exercises

1. Describe the difference between thermodynamic equilibrium and a steady state.

2. Write out the differential equations for the system $A \leftarrow B \leftarrow C$ where the reactions rates are given by:

$$
\begin{aligned}
v_1 &= k_1 A - k_2 B \\
v_2 &= k_3 B - k_4 C
\end{aligned}
$$

   Find the concentrations of A, B and C when the rates of change are zero $dA/dt = 0, dB/dt = 0, dC/dt = 0$. Show that this system is at thermodynamic equilibrium when the rates of change are zero.

3. What do we mean by the phrase quasi-equilibrium?

4. Find the mathematical expression for species A and B that describes the steady state for the network:

$$X_o \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A \overset{k_3}{\rightarrow} B \overset{k_4}{\rightarrow} \tag{3.10}$$

   Assume that X_o is fixed and that all reactions are governed by simple mass-action kinetics.

5. Explain what is meant by a stable and unstable steady state.

6. The steady state of a given pathway is stable. Explain the effect in general terms on the steady state if:

   a) A bolus of floating species is injected into the pathway

   b) A permanent change to a kinetic constant.

7. Determine whether the steady state for the system 3.10 is stable or not.

8. Why are scaled sensitivity sometime of more advantage that unscaled sensitivities?

9. Use a software tool of your choice to visualize the phase plot for the following system:

$$\frac{dx}{dt} = 2.55x - 4.4y$$

$$\frac{dy}{dt} = 5x + 2.15y$$

Appendix

See Appendix **??** for more details of Jarnac.

```
p = defn cell
       A -> B;  k1*A;
       B -> A;  k2*B;
```

```
end;

p.A = 10; p.k1 = 1;
p.B = 0;  p.k2 = 0.5;

m = p.sim.eval (0, 6, 100);
graph (m);
```

Listing 3.1 Script for Figure 3.1

```
p = defn cell
     $Xo -> S1;  vo;
      S1 -> S2;  k1*S1 - k2*S2;
      S2 -> $X1; k3*S2;
end;

p.vo = 1;
p.k1 = 2; p. k2 = 0;
p.k3 = 3;

m = p.sim.eval (0, 6, 100);
graph (m);
```

Listing 3.2 Script for Figure 3.2

```
p = defn newModel
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;
end;

p.k1 = 0.2;
p.k2 = 0.4;
p.Xo = 1;
p.S1 = 0.0;

m = p.sim.eval (0, 20, 100, [<p.time>, <p.S1>]);
graph (m);
```

Listing 3.3 Script for Figure 3.3

```
p = defn newModel
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;
end;

p.k1 = 0.2;
p.k2 = 0.4;
p.Xo = 1;
p.S1 = 0.5;

// Simulate the first part up to 20 time units
m1 = p.sim.eval (0, 20, 100, [<p.time>, <p.S1>]);

// Perturb the concentration of S1 by 0.35 units
p.S1 = p.S1 + 0.35;

// Continue simulating from last end point
m2 = p.sim.eval (20, 50, 100, [<p.time>, <p.S1>]);

// Merge and plot the two halves of the simulation
graph (augr(m1, m2));
```

Listing 3.4 Script for Figure 3.4

```
p = defn newModel
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;
end;

p.k1 = 0.2;
p.k2 = 0.4;
p.Xo = 1;
p.S1 = 0.0;

// Simulate the first part up to 20 time units
m1 = p.sim.eval (0, 20, 100, [<p.time>, <p.S1>]);

// Perturb the concentration of S1 by 0.35 units
p.S1 = p.S1 + 0.35;
```

```
// Continue simulating from last end point
m2 = p.sim.eval (20, 40, 50, [<p.time>, <p.S1>]);
// Merge the data sets
m3 = augr(m1, m2);
// Do a negative perturbation in S1
p.S1 = p.S1 - 0.35;

// Continue simulating from last end point
m4 = p.sim.eval (40, 60, 50, [<p.time>, <p.S1>]);

// Merge and plot the final two halves of the simulation
graph (augr(m3, m4));
```

Listing 3.5 Script for Figure 3.5

```
p = defn newModel
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;
end;

p.k1 = 0.2;
p.k2 = 0.4;
p.Xo = 1;
p.S1 = 0.5;

// Simulate the first part up to 20 time units

m1 = p.sim.eval (0, 20, 5, [<p.time>, <p.S1>]);

// Perturb the parameter k1
p.k1 = p.k1*1.7;

// Simulate from the last point
m2 = p.sim.eval (20, 50, 40, [<p.time>, <p.S1>]);

// Restore the parameter back to ordinal value
p.k1 = 0.2;

// Carry out final run of the simulation
```

```
m3 = p.sim.eval (50, 80, 40, [<p.time>, <p.S1>]);

// Merge all data sets and plot
m4 = augr(augr(m1, m2), m3);
graph (m4);
```

Listing 3.6 Script for Figure 3.6

4
Flux Constraints

4.1 Flux Constraints

The study of metabolism, that is the chemical reactions that are involved in breaking down nutrients and building up more complex molecules, was one of the earliest topics of study in biochemistry. Glycolysis, which concerns the breakdown of glucose in to pyruvate, was one of the first metabolic pathways to be investigated during the early part of the 20th century. In the period since, numerous other pathways have been uncovered. One of the most widely studied organisms, *E. coli*, has been shown at last count to have at least 918 enzymes catalyzing a wide range of metabolic functions [47]. In any particular pathway, enzymes catalyze the conversion of substances from one form to another. The rate of conversion is often called the **flux** which is simply another word for a reaction rate but refers specifically to the reaction rate when the enzyme is embedded in a pathway. Figure 4.1 shows a simplified metabolic map from Corynebacterium glutamicum [**?**]. The numbers next to the reaction steps indicate the flux through each step and shows how the flow of mass through the different metabolic pathway are distributed.

Stoichiometry has a significant effect on the possible space of flux distributions and in chapter the focus will be stoichiometry and flux balance.

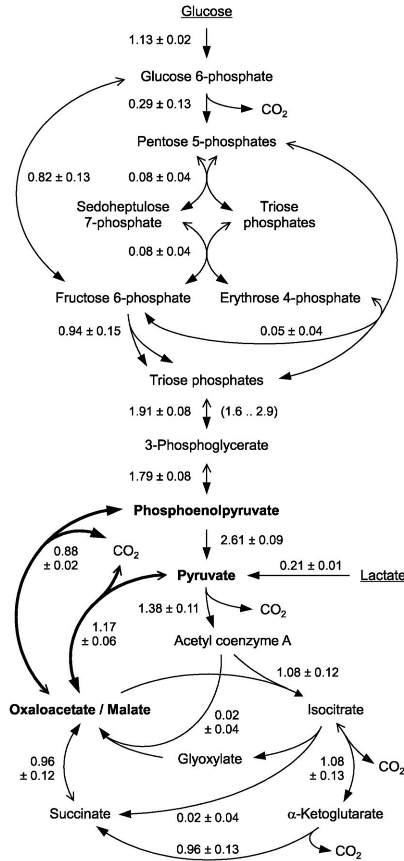


Figure 4.1 Metabolic Map of Corynebacterium glutamicum central metabolism adapted from [**?**].

4.2 Flux Balance Laws

While the rows of the stoichiometry matrix, **N**, indicate possible conservation relationships among the molecular species, the columns provide information on the constraints among the reaction rates at steady state. The steady state of a system is defined when the rates of change are zero, that

is when:

$$\mathbf{N}\boldsymbol{v} = 0$$

Note that although the rates of change may be zero at steady state, the net flows (or fluxes) through individual reactions will not be zero. By illustration, let us look at the very simple branched pathway shown in Figure 4.2. The stoichiometry matrix for this pathway is: $\mathbf{N} = [1 - 1 - 1]$ and the balance equation at steady state is given by:

$$\begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0$$

Or, more simply, $v_1 - v_2 - v_3 = 0$.

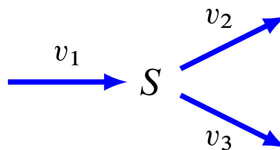


Figure 4.2 Simple branched pathway.

A common need by metabolic engineers is to know the flux distribution throughout a reaction network. One approach to obtain this information is to measure every individual flux in the network. This can be done, at least in principle, by measuring the consumption or turnover rates of all the metabolites in the network. The easiest rates to measure are on the reaction steps that connect directly to the external environment, such steps might be involved in nutrient and oxygen consumption, carbon dioxide, ethanol or biomass production, quantities that can be measured experimentally. However, the internal fluxes that are deep inside the metabolic networks are much more difficult to measure, although the use of ^{13}C-labeled substrates has made such measurements more accessible.

In practice it is extremely difficult to measure every reaction rate directly, instead the balance equations can be exploited to reduce the number of necessary flux measurements. To illustrate, the balance equation for the

simple branched pathway shows us that only two rates actually need be measured because the third can be computed. For example, if v_2 and v_1 were measured, the third rate, v_3, could be calculated from the balance equation $v_3 = v_2 - v_1$, taking note that the pathway must be in steady state. For an experimentalist this is a great benefit because it reduces the number of measurements that need to be made.

One of the practical aims of flux balance analysis is to devise methods that allow all the fluxes in a pathway to be determined with the minimum effort. To devise such methods however, a number of questions need to be answered. For example, what are the minimum number of fluxes that should to be measured experimentally to fully determine all fluxes in a pathway? In the simple branch pathway (Figure 4.2) a minimum of two fluxes were required. Alternatively it may not be possible to measure even the minimum number, in such cases can a best estimate for the flux distribution in a pathway be computed? The following sections will consider approaches to answering all these questions, particularly for arbitrary networks where systematic approaches are required.

Box 5.1 The Null Space

Given a matrix equation of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ where $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{x}$ is a column vector of n elements, the solution, that is all the vectors $\boldsymbol{x}$ that satisfy this equation, is called the **null space** of $\boldsymbol{A}$.

The minimum number of vectors required to fully describe the null space is called the **dimension** and is equal to the rank of the matrix rank($\boldsymbol{A}$) minus the number of columns, n. These vectors form what is called a **basis** for the space and linear combinations of these vectors can generate any other vector in the null space. In order to form a basis, the vectors must also be linearly independent.

Many tools can compute the basis for the null space, for example `null (A, 'r')` will compute the basis in Matlab, while `NullSpace[A]` can be used to compute the basis in Mathematica.

Box 5.2 Linear Dependence and Independence

One of the most important ideas in linear algebra is the concept of linear dependence and independence. Take three vectors, say $[1,-1,2]$, $[3,0,-1]$ and $[9,-3,4]$. If we look at these vectors carefully it should be apparent that the third vector can be generated from a combination of the first two, that is $[9,-3,4] = 3[1,-1,2]+2[3,0,-1]$. Mathematically we say that these vectors are *linearly dependent*.

In contrast, the following vectors, $[1,-1,0]$, $[0,1,-1]$ and $[0,0,1]$, are independent because there is no combination of these vectors that can generate even one of them. Mathematically we say that these vectors are *linearly independent*.

4.3 Determined Systems

Consider the more complicated pathway shown in Figure 4.3.

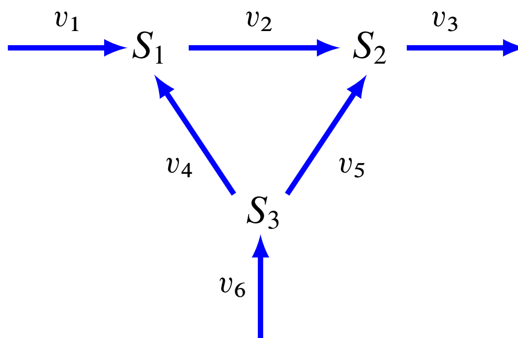


Figure 4.3 Complex branched pathway.

The stoichiometry matrix for this pathway is:

$$\mathbf{N} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ \left[\begin{array}{cccccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array} \right] \end{array} \tag{4.1}$$

which corresponds to the following three balance equations:

$$v_1 - v_2 + v_4 = 0$$
$$v_2 - v_3 + v_5 = 0$$
$$v_6 - v_4 - v_5 = 0$$

The first question to ask is what is the minimum number of fluxes that need to be measured so that the reminder of the fluxes can be estimated? Since there are three equations and six unknowns, at least three of the fluxes must to be measured so that number of unknowns can be reduced to three. However, of the six, which of the three fluxes should be measured? For example, measuring v_1, v_2 and v_4, will not help because it is not possible to compute the others from these fluxes.

Box 5.3 Rank of a Matrix

Closely related to linear independence (Box 5.2) is the concept of Rank. Consider the three vectors described in Box 5.2, $[1, -1, 2]$, $[3, 0, 1]$ and $[9, -3, 4]$ and stack them one atop each other to form a matrix:

$$\left[\begin{array}{ccc} 1 & -1 & 2 \\ 3 & 0 & 1 \\ 9 & -3 & 4 \end{array} \right]$$

then the Rank is simply the number of linear independent vectors that make up the matrix. In this case the Rank is 2, because there are only two linear independent row vectors in the matrix.

In order to answer this question let us divide the fluxes into two groups, call one the measured fluxes ($\boldsymbol{J_M}$) and the other the computed fluxes ($\boldsymbol{J_C}$). The computed fluxes will be calculated from some combination of the mea-

Box 5.4 Elementary Matrices

Elementary matrix operations such as row exchange, row scaling or row replacement can be represented by simple matrices called elementary matrices, called Type I, II and III respectively. Elementary matrices can be constructed from the identity matrix. For example a scaling operation can be represented by an identity matrix by replacing one of the elements of the main diagonal by the scaling factor. The following matrix represents a type II matrix which will scale the second row of a given matrix by the factor k:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Type I elementary matrices will exchange two given rows in a given matrix and are constructed from an identity matrix where rows in an identity matrix are exchanged that correspond to the rows exchanged in the target matrix. The following type I matrix will exchange rows 2 and 3 in a target matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Type III elementary matrices will add/subtract a given row in a target matrix to another row in the same matrix. Type III matrices are constructed from an identity matrix where a single off diagonal element is set to the multiplication factor and the specific location represents the two rows to combine. If an elementary matrix adds a row i to a row j multiplied by a factor α, then the identity matrix with entry i, j is set to α. In the following example, the type III elementary matrix will subtract five times the 2nd row from the 3rd row.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -5 & 0 \end{bmatrix}$$

A particularly important property of elementary matrices is that they can all be inverted. In addition, pre-multiplying by an elementary matrix will modify the rows of a target matrix while post-multiplying will operate on the columns.

Box 5.5 Echelon Forms

There are two kinds of matrices that one frequently encounters in the study of linear equations. These are the **row echelon** and **reduced echelon forms**. Both matrices are generated when solving sets of linear equations. The row echelon form is derived using forward elimination and the reduced echelon form by Gauss-Jordan Elimination.

A **row echelon matrix** is defined as follows:

1. All rows that consist entirely of zeros are at the bottom of the matrix.
2. In each non-zero row, the first non-zero entry is a 1, called the leading one.
3. The leading 1 in each row is to the right of all leading 1's above it. This means there will be zeros below each leading 1.

The following three matrices are examples of row echelon forms:

$$\begin{bmatrix} 1 & 4 & 3 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 5 & 3 & 0 \\ 0 & 1 & 7 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The **reduced echelon form** has one additional characteristic:

4. Each column that contains a leading one has zeros above and below it. The following three matrices are examples of reduced echelon forms:

$$\begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Sometimes the columns of a reduced echelon can be ordered such that each leading one is immediately to the right of the leading one above it. This will ensure that the leading 1's form an identity matrix at the front of the matrix. The reduced echelon form will therefore have the following general block structure:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{A} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

It is always possible to reduce any matrix to its echelon or reduced echelon form by an appropriate choice of elementary operations. The function `rref()` implemented in many math applications will generate a reduced row echelon.

sured fluxes. Consider the system equation at steady state:

$$\mathbf{N}\boldsymbol{v} = 0$$

Let us apply row reduction to the system equation until $\mathbf{N}$ is in reduced echelon form (See Box 5.5). Since the right-hand side is zero, it remains unchanged in the process. These operations lead to:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{M} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{v} = 0 \tag{4.2}$$

The process is likely to result in column as well as row exchanges and as a result the linearly independent columns will move to the left partition forming the identity matrix and the linearly dependent columns will be found in the partition corresponding to $\boldsymbol{M}$. Let us partition the $\boldsymbol{v}$ vector to correspond to the partitioning in the echelon matrix, so that:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{M} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{bmatrix} = 0$$

which when multiplied out gives $\boldsymbol{v_1} = -\boldsymbol{M}\ \boldsymbol{v_2}$. This implies that the flux terms in the $\boldsymbol{v}_1$ partition correspond to the computed fluxes, $\boldsymbol{J_C}$, and $\boldsymbol{v}_2$ to the measured fluxes, $\boldsymbol{J_M}$, that is $\boldsymbol{J_C} = -\boldsymbol{M}\ \boldsymbol{J_M}$.

This relation describes a set of computed fluxes, $\boldsymbol{J_C}$, as a function of a set of measured fluxes, $\boldsymbol{J_M}$ via a transformation matrix, $\boldsymbol{M}$. To follow conventional notation, the term $-\boldsymbol{M}$ will be renamed to $\boldsymbol{K_0}$ (that is $\boldsymbol{M} = -\boldsymbol{K_0}$) so that

$$\boldsymbol{J_C} = \boldsymbol{K_0}\ \boldsymbol{J_M}. \tag{4.3}$$

and equation 4.2 can be reexpressed as:

$$\begin{bmatrix} \boldsymbol{I} & -\boldsymbol{K_0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{J_C} \\ \boldsymbol{J_M} \end{bmatrix} = 0 \tag{4.4}$$

Returning to the example shown in Figure 4.3, let us apply a series of elementary operations to the stoichiometry matrix to reduce the stoichiometry to its reduced echelon form (Equation 4.4):

1. Start with the stoichiometry matrix.

$$\begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}$$

1. Multiply the 3rd row by -1.

$$\begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

2. Add the 2nd row to the 1st row.

$$\begin{bmatrix} 1 & 0 & -1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

3. Add the 3rd row times -1 to the 1st row.

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

4. And finally, exchange the 3rd and 4th columns.

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{bmatrix}$$

These operations lead to the following reduced echelon matrix (leading ones are shown in a darker gray):

$$\text{Reduced Echelon} = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_4 & v_3 & v_5 & v_6 \end{matrix} \\ \begin{bmatrix} \mathbf{1} & 0 & 0 & -1 & 0 & 1 \\ 0 & \mathbf{1} & 0 & -1 & 1 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 1 & -1 \end{bmatrix} \end{array} \tag{4.5}$$

Note that during the reduction, the third and forth columns were exchanged. The partition that holds the identity matrix marks the computed fluxes and the right-hand partition which holds the $\boldsymbol{K_0}$ matrix marks the measured fluxes. Thus the **computed fluxes** correspond to the **independent columns** and the **measured fluxes** to the **dependent columns**. If we extract the $\boldsymbol{K_0}$ partition, equation 4.3 can be used to relate the computed to the measured fluxes as follows:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_4 \end{bmatrix} = \ominus \begin{bmatrix} -1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_3 \\ v_5 \\ v_6 \end{bmatrix} \tag{4.6}$$

Or

$$\begin{aligned} v_1 &= v_3 - v_6 \\ v_2 &= v_3 - v_5 \\ v_4 &= v_6 - v_5 \end{aligned}$$

This shows that in principle only v_3, v_5 and v_6 need be measured from which all remaining rates can be calculated. A visual inspection of the pathway in Figure 4.3, will reveal this to be true, thus, v_4 can be computed from v_5 and v_6; v_2 can be computed from v_5 and v_3; and lastly, v_1 can be computed from v_2 and v_4.

Software tools such as PySCeS [**?**] can be used to automatically compute the $\boldsymbol{K_0}$ matrix along with an appropriately reordered stoichiometry matrix. In summary, the method outlined above enables us to derive the minimum set of fluxes to measure in order to determine all fluxes in an arbitrary pathway.

Linear Algebra of Determined Systems

An alternative but related approach to derive the computed from the measured fluxes is as follows. Let us assume we can reorder the columns of the stoichiometry matrix so that all the dependent columns are moved to the left-side of the matrix and the independent columns are moved to the right-side of the matrix. Note this is the opposite order to the columns in equations 4.5 and 4.2. Furthermore, let us also assume that the rows have also been reordered so that the independent rows are moved to the top and the dependent rows to the bottom of the matrix. We will consider the meaning of the rows in a subsequent chapter. These prerequisites means that the stoichiometry matrix has a partitioned structure shown in Figure 4.4.

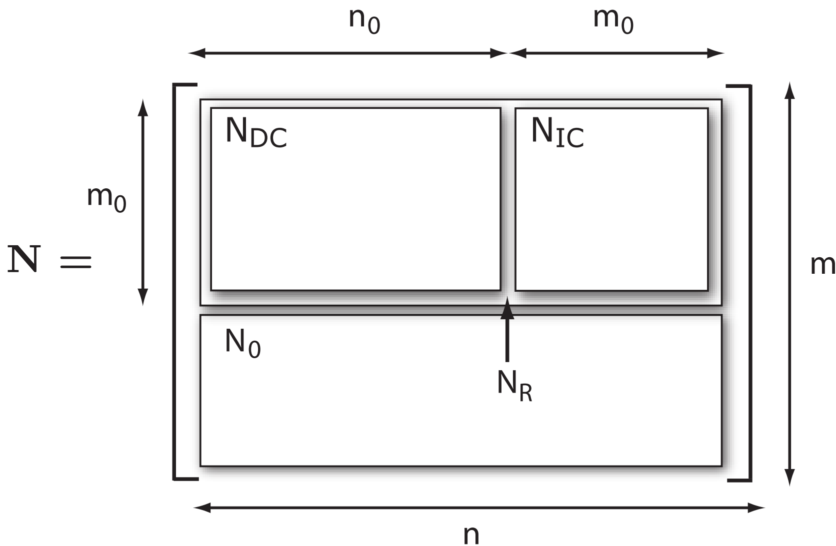


Figure 4.4 Partitioned Stoichiometry Matrix: n = number of reactions; m = number of species; $\boldsymbol{N_{DC}}$ = partition of linearly dependent columns; $\boldsymbol{N_{IC}}$ = partition of linearly independent columns; $\boldsymbol{N_R}$ = reduced stoichiometry matrix; $\boldsymbol{N_0}$ partition of linearly dependent rows.

We don't prove it here but it is possible to show that that at steady state:

$$\left[\ \boldsymbol{N_R}\ \right]\left[\begin{array}{c}\boldsymbol{J_M}\\ \boldsymbol{J_C}\end{array}\right]=0$$

$\boldsymbol{N_R}$ can be partitioned as shown in Figure 4.4:

$$\left[\ \boldsymbol{N_{DC}}\quad \boldsymbol{N_{IC}}\ \right]\left[\begin{array}{c}\boldsymbol{J_M}\\ \boldsymbol{J_C}\end{array}\right]=0$$

where $\boldsymbol{N_{DC}}$ represents the set of linearly dependent columns and $\boldsymbol{N_{IC}}$ the set of linearly independent columns. To reemphasize again, the order of the computed and measured fluxes are swapped compared to that shown in equation 4.4.

Multiplying out this equation gives $\boldsymbol{N_{DC}}\ \boldsymbol{J_M} + \boldsymbol{N_{IC}}\ \boldsymbol{J_C} = 0$. This equation can be rearranged and both sides multiplied by the inverse of $\boldsymbol{N_{IC}}$ to obtain:

$$\boldsymbol{J_C} = -(\boldsymbol{N_{IC}})^{-1}\ \boldsymbol{N_{DC}}\ \boldsymbol{J_M} \tag{4.7}$$

This result gives us a relationship between the computed and measured fluxes. The term $-(\boldsymbol{N_{IC}})^{-1}\ \boldsymbol{N_{DC}}$ can be replaced by, $\boldsymbol{K_0}$, so that $\boldsymbol{J_C} = \boldsymbol{K_0}\ \boldsymbol{J_M}$. This equation is identical to equation 4.3 but offers an alternative approach to computing $\boldsymbol{K_0}$ and is the method often cited in the literature [?, ?]. The inverse of $\boldsymbol{N_{IC}}$ is guaranteed to exist because the matrix is square and all rows and columns are guaranteed by construction to be linearly independent.

The equation, $\boldsymbol{K_0} = -(\boldsymbol{N_{IC}})^{-1}\ \boldsymbol{N_{DC}}$ can be rearranged into the following form:

$$\left[\ \boldsymbol{N_{DC}}\quad \boldsymbol{N_{IC}}\ \right]\left[\begin{array}{c}\boldsymbol{I}\\ \boldsymbol{K_0}\end{array}\right]=\boldsymbol{0} \tag{4.8}$$

or more simply:

$$\boldsymbol{N_R}\ \boldsymbol{K} = \boldsymbol{0} \tag{4.9}$$

This shows that the $\boldsymbol{K_0}$ matrix is related to the null space of the reordered stoichiometry matrix.

Examples

The following examples illustrate the application of equation 4.7.

a) Consider the branched pathway shown in Figure 4.3. The columns of the stoichiometry matrix can be reordered so that the linearly dependent columns ($\boldsymbol{N_{DC}}$) are first, followed by the linearly independent columns ($\boldsymbol{N_{IC}}$). Row reduction to the reduced echelon form (equation 4.4) can be used to determine which are the linearly independent and dependent columns (equation 4.5). In the stoichiometry matrix below, the partitions have been exchanged so that the linearly independent columns are first, followed by the linearly dependent columns:

$$\mathbf{N} = \begin{array}{c} \begin{matrix} v_3 & v_5 & v_6 & v_1 & v_2 & v_4 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 1 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

From the reordered matrix, the $\boldsymbol{N_{DC}}$ and $\boldsymbol{N_{IC}}$ partitions can be extracted from which the dependency relations can be derived by applying equation 4.7.

$$\boldsymbol{K_0} = -\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

The derived $\boldsymbol{K_0}$ corresponds to the same result found in equation 4.6.

b) A more complex example of a pathway is shown in Figure 4.5. The stoichiometry matrix for this network is given by:

$$\mathbf{N} = \begin{array}{c c} & \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} & \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

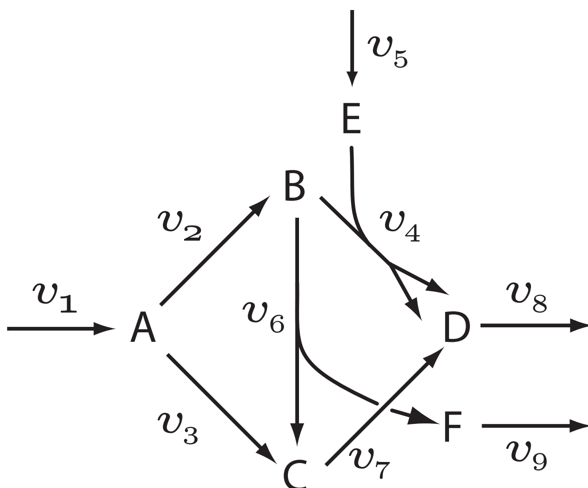


Figure 4.5 Complex Network incorporating two input fluxes and two output fluxes, coupled internally by multiple branches and one reaction that exhibits non-unity stoichiometry (v_4).

and the balance equations by:

$$\begin{aligned} v_1 - v_2 - v_3 &= 0 & \qquad v_2 - v_4 - v_6 &= 0 \\ v_3 + v_6 - v_7 &= 0 & 2v_4 + v_7 - v_8 &= 0 \\ v_5 - v_4 &= 0 & v_6 - v_9 &= 0 \end{aligned}$$

Let us reorder the columns of the stoichiometry matrix so that the linearly dependent columns are on the left and linearly independent columns are on the right (Figure 4.4). Note that all the rows are linearly independent so that there is no $\boldsymbol{N_0}$ partition in the reordered matrix. Reordering can be accomplished by carrying out a row reduction on the matrix to reduced echelon form (equation 4.2) and recording the column changes in the stoichiometry matrix. Note that the partitions must be exchanged to match the structure shown in equation 4.8. The simplest reordering is given by the

following stoichiometry matrix:

$$
\mathbf{N} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array}
\begin{array}{c}
\begin{array}{ccc|cccccc} v_7 & v_8 & v_9 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\
\left[\begin{array}{ccc|cccccc}
0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 \\
-1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
$$

The $\boldsymbol{K_0}$ matrix can be computed from the null space (4.9) of this reordered matrix:

$$
\boldsymbol{K} = \begin{array}{c} v_7 \\ v_8 \\ v_9 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array}
\left[\begin{array}{ccc}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\ \hline
0.5 & 0.5 & 0 \\
-0.5 & 0.5 & 1 \\
1 & 0 & -1 \\
-0.5 & 0.5 & 0 \\
-0.5 & 0.5 & 0 \\
0 & 0 & 1
\end{array}\right]
\qquad
\boldsymbol{K_0} = \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array}
\left[\begin{array}{ccc}
0.5 & 0.5 & 0 \\
-0.5 & 0.5 & 1 \\
1 & 0 & -1 \\
-0.5 & 0.5 & 0 \\
-0.5 & 0.5 & 0 \\
0 & 0 & 1
\end{array}\right]
$$

From the $\boldsymbol{K_0}$ matrix the relation between the measured and computed fluxes can be determined. From the reordering of the stoichiometry matrix it should be apparent that the measured fluxes are v_7, v_8, and v_9, that is a minimum of three fluxes must be measured in order to fully determine the remainder. Of the three measured fluxes, v_7 is the most problematic because it is an internal flux which experimentally would not be easy to determine. It is however possible to derive other combinations of measured and computed fluxes. Most notable is the following list of independent fluxes, v_5, v_8 and v_9. All three are edge fluxes which in principle are easier to measure.

The stoichiometry matrix can be reordered as follows:

$$\mathbf{N} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccc|cccccc} v_5 & v_8 & v_9 & v_6 & v_2 & v_3 & v_4 & v_1 & v_7 \end{array} \\ \left[\begin{array}{ccc|cccccc} 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

which yields the following $\boldsymbol{K_0}$ matrix from the null space:

$$\boldsymbol{K_0} = \begin{array}{c} v_6 \\ v_2 \\ v_3 \\ v_4 \\ v_1 \\ v_7 \end{array} \left[\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 1 \\ -2 & 1 & -1 \\ 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 0 \end{array}\right]$$

In turn this gives the dependency equations using equation 4.3:

$$\begin{aligned} v_6 &= v_9 \\ v_2 &= v_5 + v_9 \\ v_3 &= v_8 - v_9 - 2\, v_5 \\ v_4 &= v_5 \\ v_1 &= v_8 - v_5 \\ v_7 &= v_8 - 2\, v_5 \end{aligned}$$

In summary, measuring only v_5, v_8 and v_9 allows us to completely determine all the fluxes in the network. Unfortunately in real systems the internal structure of the network will be much more complex and will include many more degrees of freedom. This means that in many cases there will be insufficient information to fully determine the internal fluxes. Such cases are called **underdetermined systems** and alternative strategies must be used to gain access to the unknown fluxes. Two common strategies to the study of underdetermined systems include **flux balance analysis** and

metabolic flux analysis. Flux balance analysis relies on linear programming while metabolic flux analysis uses ^{13}C-labeled substrates to estimate fluxes.

4.4 Flux Balance Analysis

The previous section described how one can determine the set of computed and measured fluxes and how to calculate one set from the other. It assumed that it was possible to measure all the measured fluxes. However it is often the case that experimentally it is very difficult to measure all the required measured fluxes. In this situation, the problem becomes underdetermined and alternative strategies are required to determine the fluxes in a pathway. One method is to use linear programming. By its nature, linear programming only gives an estimate of the fluxes and predictions based on linear programming should be supported by additional measurements, however the approach has proved to be popular in the metabolic community [**?**].

Linear programming is an optimization method that requires two inputs, a linear objective function that is generally a sum of terms that contains weighted measurable elements from a metabolic model and a set of linear constraints. The maximizing linear programming problem can be expressed by the relations shown in equation 4.10.

There are a number of algorithms that can be used to solve linear programming problems, but by far the most popular is the simplex method – not to be confused with the simplex method developed by Nelder and Mead for solving nonlinear optimization problems. The simplex method can be motivated by a simple example.

Consider a pharmaceutical company that manufactures two drugs, say x and y, from two genetically engineered organisms, A and B. Let us assume that organism A can produce at maximum 4 kg of drug x per day and organism B a maximum of 2 kg of y per day. Let us also assume that the factory can only process a total of 5 kg of any drug per day due to packaging equipment limitations. If the company can make a profit of \$100 per kg for drug x and a profit of \$150 per kg for drug y, what is

$$
\begin{aligned}
\text{Maximize: } Z &= c_1 x_1 + c_2 x_2 + \cdots = \boldsymbol{c}^{\mathsf{T}} \boldsymbol{x} \\
\text{Subject to:}& \\
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\le b_1 \\
a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\le b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_m \\
\text{Or: } \boldsymbol{A}\boldsymbol{x} &\le \boldsymbol{b} \\
\text{where all: } x_i &\ge 0
\end{aligned}
\tag{4.10}
$$

the optimal rate at which each drug should be manufactured in order to maximize profit?

This problem is sufficiently small that it can be easily solved manually. To maximize profit, it would be prudent to first produce the maximum amount of most profitable drug first, y, then to use what ever spare capacity remains in the packaging department to manufacture drug x. This would mean producing 2 kg per day of drug y, which leaves 3 kg capacity left in the packaging department to produce 3 kg per day of drug x. Therefore the total profit for this scenario is $2 \times 150 + 3 \times 100 = \$600$.

The problem of drug manufacture allocation can be easily expressed as a linear programming problem. For example, the objective function for the problem is to maximize profit, that is to maximize:

$$\text{Maximise: } Z = \$100 \times x + \$150 \times y$$

The constraints on the problem can also be easily expressed. For example, the quantity of drug manufactured cannot be negative, that is:

$$x \ge 0 \text{ and } y \ge 0$$

In addition, the problem states that a maximum of 4 kg of x can be manu-

factured per day and a maximum of 2 kg of y per day, that is:

$$x \leq 4 \text{ and } y \leq 2$$

Finally, the packaging department can only process a maximum of 5 kg per day, that is:

$$x + y \leq 5$$

This problem can be reexpressed in graphical form as shown in Figure 4.6. The figure plots all the linear constraints that define the problem, including, $x \leq 4$, $y \leq 2$ and $x + y \leq 5$. The limits of the object function is indicated by the hashed line. Points where two or more constraints intersect are called cornerpoints or vertices. Figure 4.7 illustrates the feasible solu-

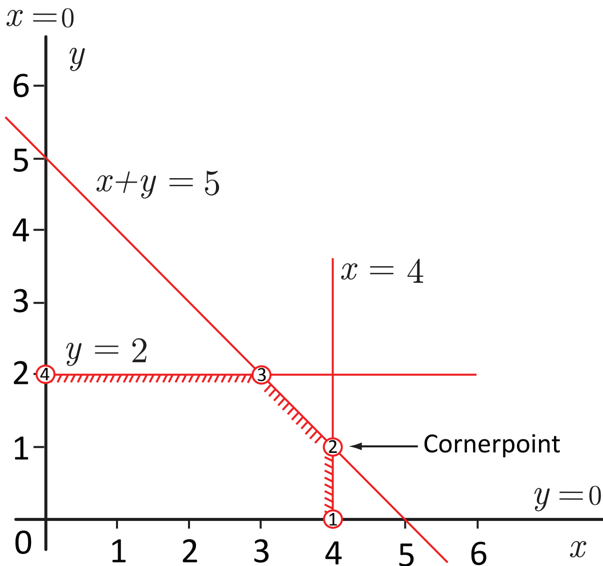


Figure 4.6 Linear Programming: Constants displayed as edges on a graph for the drug manufacturing problem.

tion bounded by the constraints and the maximum value of the objective function. The simplex method works by traversing the cornerpoints one by

one. The method first starts at one of the cornerpoints, say cornerpoint ① and then attempts to move to an adjacent cornerpoint which yields a better value for the objective function. If the method is unable to move to a better objective function it stops and reports the last cornerpoint as the optimal solution. For example, the value of the objective function at cornerpoint ① is 400 dollars. An adjacent cornerpoint is cornerpoint ②. The value of the objective function at this point is $550. Since the objective function at the new cornerpoint is larger, the method moves to this cornerpoint. From the second cornerpoint the method moves to the next adjacent cornerpoint, cornerpoint ③. The value of the objective function at cornerpoint ③ is $600. This again is larger than the value at cornerpoint ②. Once again, the method moves to the next adjacent cornerpoint, cornerpoint ④. The value of the objective function at ④ is $300 which is less that the value at ③. Since there are no other cornerpoints to traverse, the method stops and assigns the optimal value at $600 on cornerpoint ③. When a single point is located it represents a unique solution. However, it is possible for optimum solutions to lie on a line that joins two cornerpoints, that is two cornerpoints yield the same value for the objective function. In higher dimensions, optima may lie on hyperplanes connecting multiple cornerpoints. In such situations the solution is termed degenerate because there are now an infinite number of optimal solutions and other non-quantifiable criterion may be used to judge the 'best' solution. For example, a degenerate solution may indicate that two different combinations of drug x and y are equally profitable. However, one of the drugs may have toxicity issues in which case the optimum with the lowest level of this drug is better.

Another important aspect that Figure 4.7 illustrates what would happen to the optimal solution if the constraints change. This question leads to the idea of sensitivity and what are called **shadow prices**. A shadow price is the change in the optimal solution if a constraint is changed by one unit. For example what would happen to the optimal solution if the manufacture of drug y were to be increased from 2 to 3 kgs per day? Sensitivity analysis can answer these questions and provide additional information on interpreting the optimal solutions and to gauge how robust the solutions are to the constraints and/or objective function.

The drug manufacturing example was a relatively simple problem and

could be solved without recourse to the simplex method. For larger problems, particulary as the number of variables increases, the number of cornerpoints rises considerably. In addition, rather than being a simple two dimensional problem real problems are invariable hyper-dimensional. Linear programming is therefore rarely done by hand, instead software is employed to find solutions. Given the popularity of linear programming in general, there is a very wide range of software tools available, including well known tools such as Excel, Matlab and Mathematica or more specialized tools such as LINDO http://www.lindo.com or CPLEX http://www.ilog.com. However there is also a wide range of equally good open source alternatives. Probably the most notable of these include the GNU Linear Programming Kit (GLPK) or better still the lp_solve library by Peter Notebaert. lp_solve is notable for a number of reasons, its licence is less restrictive (LGPL) and there are language bindings that allows lp_solve to be easily called from many different computer languages, including for example, Java, Delphi, C#, Matlab, Excel, Python and SciLab. Both GLPK and lp_solve have a very active community forums. One enterprizing individual (Henri Gourvest) has written an excellent graphical front end to lp_solve, called the LPSolve IDE. This front-end makes it very easy to specify the objective function and constraints and solve the linear programming problem with the press of a single button. Further discussion of LPSolve IDE will be given in the next section.

Objective Functions

The choice of objective function is critical for the linear programming approach to be effective and there has been much discussion in the literature on what a suitable objective function might be for biological systems. For example, one of the earliest reported efforts to use linear programming in metabolic modeling was by Fell and Small [?]. These authors investigated fat synthesis in adipose tissue, and used a variety of objective functions, include minimizing the amount of glucose used per triacylglycerol formed or maximizing the generation of NADH from the pentose pathway. The authors subsequently used the model to study how the efficiency of conversion was affected by the availability of ATP.

One of the early attempts to determine the flux distribution in *E. coli* was conducted by Palsson's group [**?**, **?**, **?**]. An objective function used in this work involved maximizing the production of biomass, the assumption being that growing single celled organisms have been selected for growth (Unlike cells in multicellular organism where the objective function is more obscure). In order to relate biomass to a metabolic map, the authors obtained data [**?**, **?**] that described how 1 gram of *E. coli* biomass was derived from various metabolic precursors and cofactors (See Table 4.1). The objective function used to optimize the flux distribution was then defined as the sum of all the fluxes that produce each of the precursors, weighted by the amount of precursor required. Thus, a suitable objective function

Metabolite	Demand (mmol)
ATP	41.2570
NADH	-3.5470
NADPH	18.2250
G6P	0.2050
F6P	0.0709
R5P	0.8977
E4P	0.3610
T3P	0.1290
3PG	1.4960
PEP	0.5191
PYR	2.8328
AcCoA	3.7478
OAA	1.7867
AKG	1.0789

Table 4.1 Number of mmoles of precursors and cofactors that are required to yield 1 gram of biomass of *E. coli* [**?**, **?**]

may be written as:

$$Z = 41.257\, v_{ATP} - 3.547\, v_{NADH} + 18.225\, v_{NADH} + 0.205\, v_{G6P} + \ldots$$

The use of this objective function yielded results which overdetermined the experimentally determined glucose yield. This suggested that the stoichiometry model was missing an important component. In order to correct the discrepancy, the authors introduced ATP maintenance into the calculation since cells will use energy not just to achieve growth but also to maintain other non-growth functions such as maintenance of transmembrane gradients and cellular motility. The addition of ATP maintenance into the calculation yielded better estimates for glucose yield.

Another but quite different example of an objective function relates to the flux balance analysis of the mycolic acid pathway in Mycobacterium tuberculosis. In the work by Raman et al [**?**], the authors selected an objective function based on maximizing the different proportions of mycolates that make up the cell wall. Given that cell wall composition is important to the structural integrity of the cell wall, optimal production of mycolates would appear to be an appropriate optimum for the organism to achieve. With the objective function set, linear programming then requires a set of linear constraints that will restrict the limits of the objective function and allow one to find the maximum.

Flux Balance Constraints

In addition to an objective function, linear programming also requires a set of constraints to limit the scope of the solution space. Of these, the most important group are the steady state constraints on the pathway, that is $\mathbf{N}\boldsymbol{v} = 0$. There is one restriction on the steady state constraints, all rates must be positive. This means that reversible reactions must be split into their separate forward and reverse reactions.

In addition to the steady state constraints, other constraints can be added to the mix. The most common of these include constraints on the values of the external fluxes. Such fluxes, which might include nutrient uptake or oxygen consumption, will most likely be known and will contribute an important source of constraints on the model.

Other constraints include thermodynamic and capacity constraints. Capacity constraints impose upper bounds on a flux ($0 \leq v_i \leq b_i$). Such limits can be set by the Vmax of the enzyme catalyzing the reaction. Sometimes lower bounds may be set so that in general capacity constraints are set with the inequality ($a_i \leq v_i \leq b_i$). In addition some reaction steps under specific growth conditions may be absent all together due to catabolite repression, the rates through such reactions can be constrained to zero.

Thermodynamic constraints are more difficult to set and require the use of plausible ranges for metabolite levels. Thermodynamic constraints attempt to impose flux directions that are consistent with changes in the Gibb's free energy across each reaction which naturally require knowledge of metabolite levels (ref).

Finally, there will sometimes be available internal fluxes that have been measured. This means that such reactions have specific rates and can be added to the list of model constraints.

Through a judicious use of constraints it is possible to reduce the solution space and thus improve the reliability of the optimized solution.

In summary, a linear programming problem for estimating the fluxes in a metabolic pathway takes the form:

$$\begin{aligned} \text{Maximize: } & Z = c_i v_i + c_j v_j + \cdots \\ \text{Subject to: } & \mathbf{N}\boldsymbol{v} = 0 \\ \text{where: } & \boldsymbol{v} \geq 0 \end{aligned} \tag{4.11}$$

Example

Consider again the network shown in Figure 4.5. Let us assume that only v_5 and v_1 have been measured. Clearly there is insufficient information to compute the remaining fluxes in the pathway without recourse to linear programming. To solve the problem using linear programming, an objective function and a set of constraints will be required. For illustration, the model will be optimized for maximum production of biomass and for the sake of argument let us assume that fluxes v_8 and v_9 contribute to biomass. The objective can then be some weighted sum of the fluxes that contribute

to biomass, that is $Z = c_1 v_8 + c_2 v_9$.

As for the constraints, the most important are the steady state conditions on each of the nodes in the network. In this case the steady state constraints include:

$$\begin{aligned} v_1 - v_2 - v_3 &= 0 \\ v_2 - v_6 - v_4 &= 0 \\ v_3 + v_6 - v_7 &= 0 \\ 2v_4 - v_8 + v_7 &= 0 \\ v_5 - v_4 &= 0 \\ v_6 - v_9 &= 0 \end{aligned}$$

Two other constraints include the measured fluxes on v_1 and v_5. For illustration assume that $v_1 = 10$ flux units and $v_5 = 6$ flux units. This sets up the problem. Figure 4.8 shows a screen-shot of the LPSolve IDE software where the problem has been setup. The following code illustrate the problem expressed in the script language used by LPSolve.

```
/* Objective function */
max: 0.5*v9 + 0.75*v8;

/* Steady State Constraints */
v1 - v2 - v3 = 0;   /* A */
v2 - v4 - v6 = 0;   /* B */
v3 + v6 - v7 = 0;   /* C */
v6 - v9 = 0;        /* F */
v5 - v4 = 0;        /* E */
2 v4 - v8 + v7 = 0; /* D */

/* Known Flux Constraints */
v1 = 10;
v5 = 6;
v3 >= 1;
```

4.5 Isotopic Flux Measurements

In the previous section linear programming and its application to flux balance analysis was described as a method for estimating fluxes in undetermined systems. The method carried with it a number of assumptions, one in particularly was the choice of objective function which can in some systems be difficult to describe or justify. In addition, flux balance analysis has difficulties in estimating the fluxes in certain cases without more information, in particular the flux in parallel pathways, metabolic cycles such as futile cycles, and cofactor linked cycles cannot always be resolved by the method (See Figure 4.9). For this reason, other more experimentally based approaches have been devised to try and gather data on fluxes more directly. The most important approach by far is the use of isotopic tracer techniques, often referred to as **metabolic flux analysis** or MFA. The method proceeds in two phases, one experimental and another computational. The computational analysis is very important as the data analysis is complex owing to the size of the data sets and the resulting combinatory expansion of the system equations. Let us first consider the experimental phase.

Isotopes are atoms that have the same number of protons but differ in the number of neutrons. For example, carbon has three naturally occurring isotopes, the common and stable ^{12}C (6 protons and 6 neutrons), the stable and relatively uncommon (1%) ^{13}C (6 protons and 7 neutrons) and trace amounts of radioactive ^{14}C (6 protons and 8 neutrons), Table 4.2. In practice a given substrate, such as glucose will be labeled, that is one or more of the atoms in the glucose molecule will be replaced by a different isotope. For example, the ^{12}C on position one might be replaced with an atom of ^{13}C. In this case the glucose is referred to as [1-^{13}C]glucose to distinguish it from natural glucose. The main advantage to using isotopes is that they can be measured, that is in a mixture of labeled and unlabeled glucose it is possible to distinguish between the two molecules. The way labeled molecules are identified depends on whether radioactive or stable isotopes are used. Radioactive isotopes can clearly be identified by their decay emissions, for example β decay in ^{14}C and ^{3}H by using scintillation counters. The advantage to using radioactive isotopes is their great

Table 4.2 Isotopes commonly used in biological research.

Common Isotope	Rare Stable Isotope	Radioactive Isotope
^{1}H	^{2}H (0.02%)	^{3}H
^{12}C	^{13}C (1.1%)	^{14}C
^{14}N	^{15}N (0.37%)	^{13}N
^{16}O	^{18}O (0.04%)	^{11}O

sensitivity. However they are also difficult to handle due to the radiation hazard.

Stable isotopes can be identified by measuring the difference in mass between labeled and unlabeled molecules using mass spectroscopy combined with gas chromatography (GC/MS). Gas chromatography is used to separate the initial mixture of compounds based on differential equilibration between a gas and solid phase. Once separated, each compound is fed into the mass spectrometer where each compound is broken into fragments by an electron beam. The fragments, now charged, are first accelerated in an electric field that travel through a magnetic field on a circular path. The path that an individual fragment actually takes will depend on its charge and mass. The end result is a MS spectrum which records the relative proportion of the different fragments that were detected. If similar fragments contain different isotopes then different peaks will emerge in the spectrum and the proportion of the different labeled compounds can be determined. The introduction of high performance GC/MS in the last 10 years or so has revolutionized metabolic flux analysis and is now probably the preferred choice for estimating fluxes.

The basis for MFA is that when a labeled substrate is fed to an organism, the labeled atoms distribute themselves throughout the chemical composition of the organism. In microbial studies, commonly used substrates include specifically labeled glucose such as [1-^{13}C]glucose, uniformly labeled glucose ([U-^{13}C]glucose) or labeled amino acids. Once administered, the labeled molecules are metabolized by the organism and through various metabolic processes the atoms in the labeled substrate are rearranged by separation and recombination of molecular fragments. In ad-

dition some labeled isotope is either lost as metabolic waste, for example, CO_2 or incorporated into biomass. Assuming no further changes take place and the substrate is constantly applied, the distribution of the isotopes will reach what is called isotopic steady state. This can occur quite rapidly in about an hour. Once in isotopic steady state, GC/MS or NMR is used to determine how the label has been distributed in the various metabolites of interest. This is the raw data that is used to determine the fluxes through the various pathways.

In order to understand the process of generating fluxes from the isotopic data a number of terms must first be defined and understood.

Isotopomer One of the most important concepts in MFA is the **isotopomer**. Consider a molecule of alanine which has three carbon atoms; there are eight different ways to label a three carbon alanine molecule, Figure 4.10. As label enters the metabolic pathways from an external source there is the potential for the label to partition itself into every possible isotopomer. In general for a molecule with n potentially labeled atoms there will be 2^n different isotopomers, for example alanine with three atoms has $2^3 = 8$ possible isotopomers. Most often it is the relative mole fraction of isotopomers for a given molecular type that is considered and the vector of that holds the fractional contribution of each isotopomer is usually called the **isotopomer distribution vector**, or IDV. **Mass Distribution Vector** Another useful concept is the mass distribution vector, often abbreviated to MDV in the literature. An element from the mass distribution vector gives the proportion of mass in a group of isotopomers of the same mass. For n potentially labeled atoms in a molecule there will be $n + 1$ elements in the MDV. The $+1$ element corresponds to the fully unlabeled molecule.

Figure 4.11 illustrates the relationship between the IDV and MDV measures. The key reason for considering these two different descriptions is that the MDV are measurable while the IDV are on the whole more difficult to obtain experimentally, although a careful study of the fragmentation patterns from the mass spectrometry can sometimes give information on the IDV itself. In addition NMR can also be used to gain some information on the relative distribution of specific isotopomers, but the MDVs are the primary experiential data. Figure 4.12 shows a simple hypothetical network that illustrates three ways to view a such a network, as a stoichiomet-

ric network, as an atom transition network and as a isotopomer network. The stoichiometric network, a), is the simplest and most familiar, with six species and five connecting reactions. If we assume that the species, A, B, E, and F contain two atoms that could be potentially labeled, and species C and D contain one atom each that could be potentially labeled, then b) in Figure 4.12 shows the species with their atomic structure explicitly given, hence the atom transition network. Panel c in Figure 4.12 shows the isotopomer network. For example, molecule A has two carbon atoms that can be labeled therefore there are four possible isotopomers which are indicated by four blocks. The point of the figure is that it shows the increase in complexity when we consider the individual isotopomers, in this case in increase from six species to twenty. For a complex pathway the number of distinct species can explode to many thousands. In the computational phase each of these distinct species is modeled resulting in a large number of differential equations.

A number of assumptions are invoked in order for the subsequent analysis to be valid. The most important is that the system is at steady state, that is the fluxes and the isotopic distribution are steady. Some of the fluxes in the system can be measured directly, for example most of the external fluxes such as substrate uptake and product and biomass formation are known. What is left are the intracellular fluxes and it is these that will be estimated from the isotopic data.

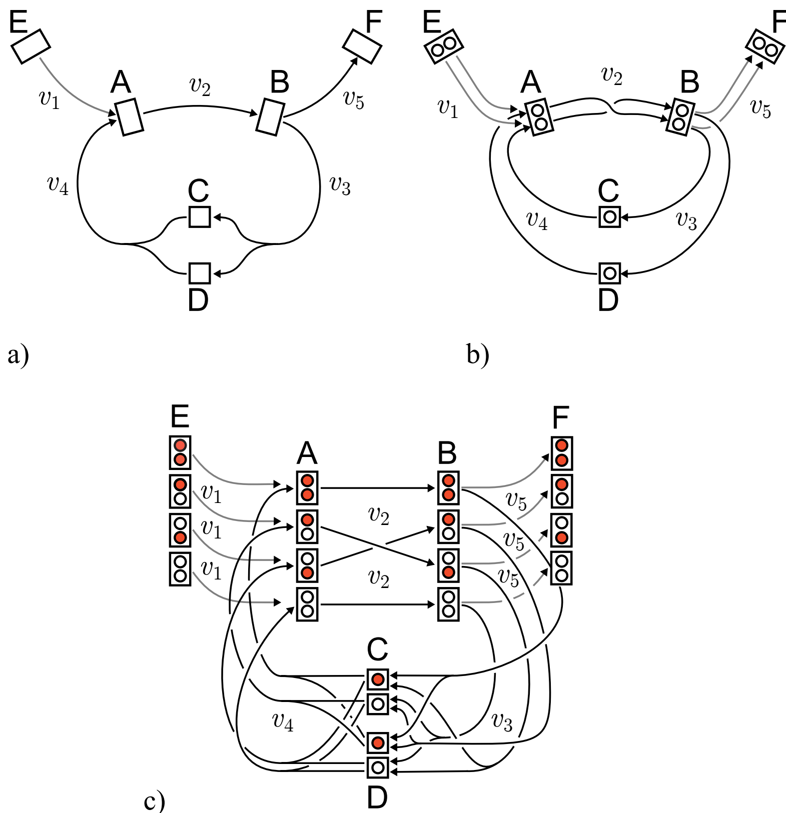


Figure 4.12 Label distribution in a simple network: a) Stoichiometry network, b) Atom transition network, c) Isotopomer network. Figure adapted with permission from Weitzel et al. [**?**], BioMed Central

The second phase in MFA is the computational effort. This is a fairly sophisticated and computationally intensive procedure. Here we describe the basic approach but many refinements have been introduced in recent years [**?**, **?**, **?**].

The essential idea behind the computational phase is the construction of a set of differential equations that describe the time evolution of the iso-

topomer distribution vector. These equations include two kinds of terms, fluxes and elements from the isotopomer distribution vector. The equations are used to predict the steady state levels of the various isotopomers, or more precisely the fractional distribution of the isotopomers at steady state. The nature of these equations will be described more fully later, for now let us designate the isotopomer distribution vector with the symbol **p** so that the set of differential equations can be written as:

$$\frac{d\mathbf{p}}{dt} = f(\mathbf{p}, \boldsymbol{J})$$

At steady state the left-hand side is zero and the isotopomer can be written, at least in principle, as a function of the fluxes, $\boldsymbol{J}$.

$$\mathbf{p} = g(\boldsymbol{J})$$

We say in principle because the equations will tend to be non-linear, rendering an analytical solution difficult if not impossible to obtain, instead numerical methods are used to find the solution, **p**. Once a solution has been found, the vector $\boldsymbol{p}$ is compared to the real measurements and a difference computed. The procedure now makes small adjustments to the flux values and the steady state equations is solved again to obtain a new $\boldsymbol{p}$ vector. If the difference between the new values and the measured values is small then the flux values are accepted and the procedure repeated otherwise the fluxes are adjusted again. The actual strategy for adjusting the fluxes will be described later but what we have is an iterative procedure where the flux values are adjusted until the measured values of the isotopomers match the computed values. The procedure just outlined is of course a classic optimization problem and many strategies exist for adjusting the flux values at each iteration including gradient search methods such as Levenberg-Marquardt or better still evolutionary algorithms [**?**, **?**] that are less likely to fail to converge.

In practice the measured values for the isotopomer distribution are not usually available, instead the model values are converted to the mass distributions and it is these that are compared to the measured mass distributions. One can imagine that in a large network, particularly where the metabolites have many potentially labeled atoms (say six or more carbon atoms)

then the number of isotopomers can become very large with a corresponding increase in the number of model differential equations. Large models can have thousands of differential equations that need to be solved at each iteration. The computational cost is therefore relatively high although with the availability of cheap and powerful personal computers the issue is not so significant as it used to be.

One question remains which relates to the exact nature of the model equations that are used to predict the isotopomers. Of all the steps required during the computational phase, generating the model equations is probably the most tedious and error prone, especially given the large number of equations that need to be deployed. With this in mind a number of authors have devised specialized software that can automate this phase and much else. Here a brief description of the equations themselves will be given. What may not be obvious is that the model equations do not assume any kinetics from the reaction steps themselves, that is there are no rates that depend on Michaelis-Menten rate laws or other more complicated functions. Instead linear equations are devised that assume that the rate of reaction between two particular label molecules is a linear function of the isotopomer concentrations. This is possible because the underlying metabolic state is assumed to be at steady state. In addition, the individual rates are simply scaled terms containing the fluxes.

Consider the system depicted in Figure 4.13. The overall reaction is given as A $\rightarrow$ B $\rightarrow$ 2 C in the upper panel. In the lower panel we see the individual species represented by their groups of isotopomers. For simplicity the species are assumed to only contain two potentially labeled carbon atoms. The first reaction, v_1 swaps the carbon atoms and the second reaction, v_2, dissociates the species into two one carbon units, C and D. The fractional distribution of isotopomers in the A species is given by A_1 and A_2, and in the B species by B_1 and B_2. Note that in each case the following is also true, $A_1 + \mathrm{A}_2 = 1$ and $\mathrm{B}_1 + \mathrm{B}_2 = 1$.

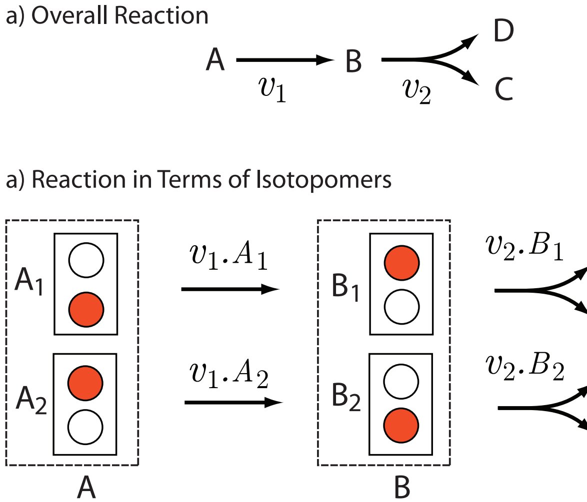


Figure 4.13 v_1 swaps the two atoms in the molecule, v_2 dissociates the two atoms. A_i and B_i are the proportion of mass in each group of isotopomers. For example $v_1 A_1$ equals the proportion of reaction velocity that is associated with the A_1 isotopomer.

At steady state the flux from species A to B and from species B to C plus D is v_1 and v_2 respectively, these are the fluxes we would like to know. However the isotopomer computational model considers each isotopomer reaction transition as a separate reaction such that the rate from from one isotopomer to another is proportional to the fraction of isotopomer.

For example, the rate of reaction from isotopomer A_1 to B_1 is the fraction of the overall rate, $v_1 A_1$. Likewise for the other isotopomers. For this system, the rate of change of the fraction B_1 and B_2 is then give by:

$$\frac{dB_1}{dt} = v_1 A_1 - v_2 B_1$$

$$\frac{dB_2}{dt} = v_1 A_2 - v_2 B_2$$

Note that these equations compute the rate of change on the fraction of isotopomers, not the absolute amount of isotopomers. This approach eliminates the need for a complex kinetic model whose construction would be extremely difficult to construct and suspect at best.

The computational effort required to estimate the fluxes are as formidable as the experimental effort and for this reason a number of authors have devised software for the automatic construction and solution to the equations. One of the earliest and most comprehensive is the software tool by Wiechert [**?**], 13C-FLUX[1] who was one of the pioneers in developing the current state of MFA [**?**, **?**, **?**]. Other tools of note include FluxSimulator from Binsl ([**?**]) and FiatFlux from [**?**].

There are many other details of MFA that have not been mentioned and the area is still under rapid development with an ever increasing number of researchers turning to use the approach to estimate fluxes [**?**, **?**, **?**, **?**].

Further Reading

1. Wiechert W. (2001) 13C metabolic flux analysis. Metabolic Engineering Jul;3(3):195-206.

2. Stephanopoulos, Gregory (1998). "Chapter 9: Methods for the Experimental Determination of Metabolic Fluxes by Isotope Labeling". Metabolic engineering: principles and methodologies. San Diego: Academic Press. pp. 356Ű404. ISBN 0-12-666260-6.

[1] see http://www.uni-siegen.de/fb11/simtec/software/13cflux/

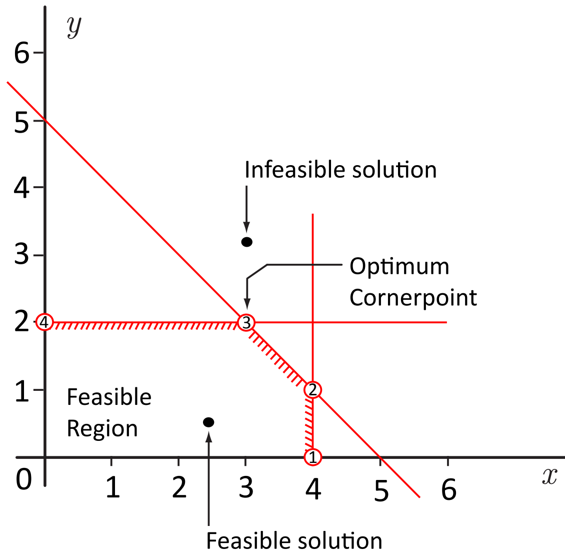


Figure 4.7 Linear Programming: Area within the confinement of the constraints is marked as the feasible region. All potential solutions to the problem reside in this region. Linear Programming attempts to locate the optimum solution within this region given an objective function. The simplex method moves from cornerpoint (vertex) to cornerpoint searching for the maximum value of the objective function. In this problem, the third cornerpoint indicates the optimal solution.

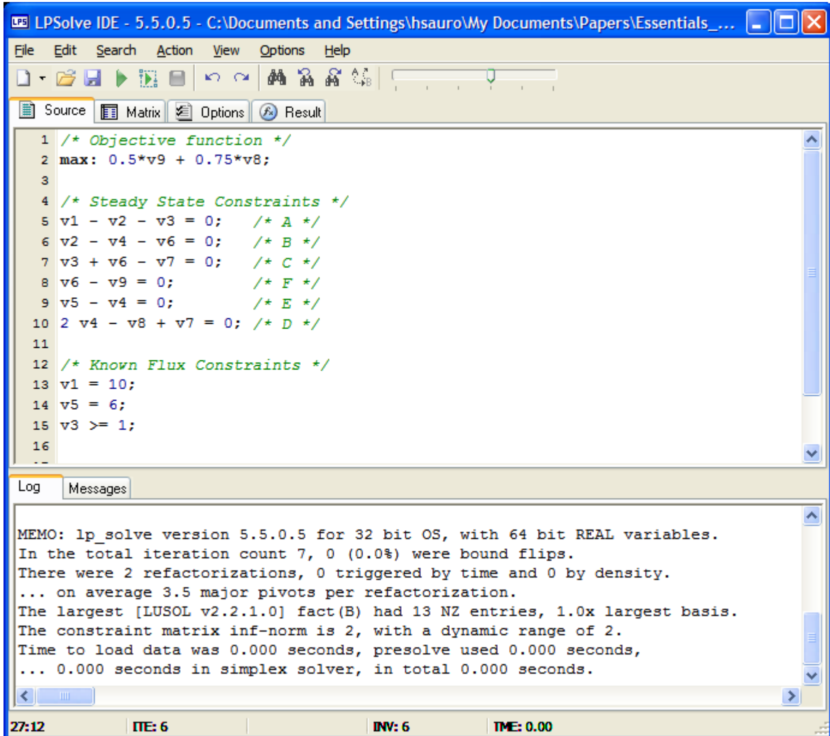


Figure 4.8 LPSolve IDE used to model a simple metabolic model problem. Available from http://lpsolve.sourceforge.net/5.5/IDE.htm

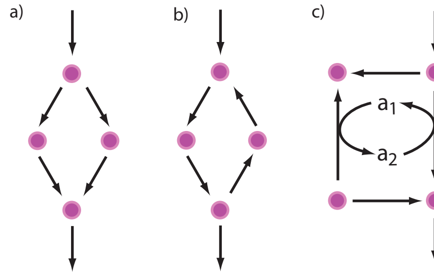


Figure 4.9 Typical situations where linear programming based flux balance analysis cannot resolve fluxes: a) Parallel pathways; b) Metabolic Cycles; c) Pathways with closed cofactor cycles.

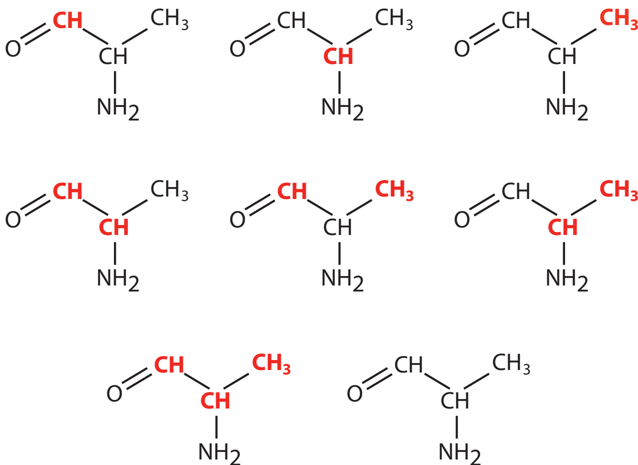


Figure 4.10 Alanine is a three carbon amino acid. If Alanine were labeled with ^{13}C, there would be eight possible different labeling patterns. These different labeled forms are call Isotopomers. For a molecules with n potentially labeled atoms, there will 2^n possible isotopomers.

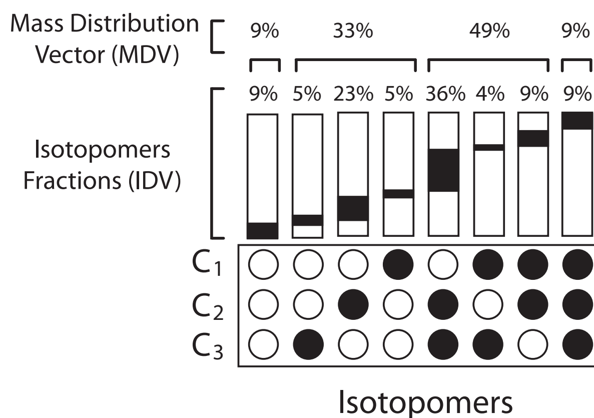


Figure 4.11 This figure illustrates the relationship between the isotopomer fraction (IDC) and the mass distribution vector (MDV). The example uses a three carbon molecule of which there are eight possible isotopomers. For each labeled molecule there is a fraction that is labeled, for example the unlabeled molecule is 9% of the total fraction. To compute the mass distributions, we collect all isotopomers having the same number of labeled atoms, for example, the 2nd, 3rd and 4th isotopomers have one labeled atom each, therefore this group constitutes a particular element in the MDV, in this case 33%

Part II

Metabolic Control Analysis

Preface to Part II

Metabolic control analysis (MCA) is a general approach to understanding how perturbations propagate through a biochemical network. Although called metabolic control analysis, the approach applies equally well to gene regulatory and protein signaling networks. Historically, during the inception of MCA, the focus was on metabolic systems and the name metabolic control analysis seems to have stuck. Ideally it should be called biochemical control analysis but it's difficult to change something after it has been used for so long. Savageau who developed Biochemical Systems Theory in the US, roughly an equivalent approach to MCA, was forward thinking enough to realize that the approach was quite general and not just restricted to metabolic systems. In this book I will use the notation that the Kacser/Heinrich groups in Europe developed rather than the notation developed by Savageau. There are two reasons for this, one is my own familiarity with the European notation; secondly, and this is no criticism of Savageau's work, the European model has been developed much more. I think this has partly to do with cultural differences. Having living the US for over 12 years, I have found that the American culture is very practical and invention orientated and theory is not as important. Savageau's work, which was largely theoretical, therefore languished in somewhat obscurity in the US whereas MCA was of more interest in Europe.

5 Elasticities

5.1 Introduction

Enzymes catalyze virtually all the chemical transformations of metabolism. They coordinate all the primary activities of a cell, ranging from energy transformations and storage, through to maintenance of cellular structure and integrity. They manage directly the expression and maintenance of the host DNA, including replication. Enzymes clearly serve an essential and fundamental role in the activity of a cell and for this reason we can regard them as the fundamental units of life. If we are to understand how cellular systems work an appreciation of the properties of these fundamental units is obviously essential. This first section will focus attention on the properties of the isolated enzymes and only later will intact pathways be considered.

In traditional enzyme kinetics the emphasis has been on mechanism, culminating in the derivation of an algebraic rate law describing the rate of reaction in terms of concentrations and kinetic constants.

The approach that MCA takes is quite different, empathizing instead the response of enzymes to changes in their environment. Where enzyme kinetics essentially stops at the rate law, MCA begins it journey. MCA is

concerned with responses, so instead of asking how a reaction rate depends on a concentration or kinetic constant, MCA considers how the rate of a reaction responds to changes in concentrations and kinetic constants.

Since enzymes are the functional units of metabolism, it is important to understand how an enzyme responds to changes in its environment and an important part of MCA is the consideration of this question. In MCA the measure that describes this response is called the elasticity coefficient. Elasticity coefficients are so important to MCA that the remainder of the chapter is devoted to their discussion.

Elasticities describe how sensitive a reaction rate is to changes in reactant, product and effector concentrations. They represent the degree to which changes are transmitted from the immediate environment to the reaction rate. From a systems perspective they are critical components in understanding how a disturbance, such as the introduction of a drug applied at one or more points in a cellular pathway, propagates to the rest of the system. It is the magnitude and signs of the elasticities that determines how far and at what strength the disturbance travels. Elasticities are therefore central in helping us understand how networks function. In this chapter we will focus on describing the properties of elasticities, how they can be computed and used to describe changes at a reaction step.

To study the properties of an individual enzyme, the usual experimental procedure is to purify the enzyme and study it in vitro. Once purified and isolated, the environment of the enzyme can be controlled and in principle the concentrations of all the participating molecules manipulated at will. Individual substrates, effectors etc., can be selectively changed and any change in rate recorded. In this manner, the response of the reaction rate to changes in factors that might affect the reaction rate can be studied. It is important that only one factor at a time is manipulated so that relative effectiveness of each can be assessed.

Consider an experiment where we wish to investigate the response of the rate of reaction to changes in substrate concentration. For notational convenience let us denote the concentration of substrate by the symbol S_j and the rate of reaction by v_i. The experiment would proceed in two steps. The first step would involve measuring the rate of reaction, v_i, at some substrate concentration of interest, say S_j. In the second part of the experiment the

concentration of substrate would be increased by an amount given by δS_j, and the experiment repeated at the new concentration of $S_j + \delta S_j$. The increase in S_j is likely to cause a change in the rate of reaction from v_i to v_{inew}. The difference between the two rates, $v_{inew} - v_i$ is the change in rate as a result of the change, δS_j and we can denote this change in rate by δv_i. Depending on the particular enzyme, the effectors, the substrates and products, the change we observe in the rate might be large or small. In order to judge the relative effectiveness of any particular modifier we can form the ratio

$$\frac{\delta v_i}{\delta S_j}.$$

This will give us the change in v_i per unit change in S_j. By measuring this ratio for each factor that might affect the rate we can gauge which ones have more of an effect or less of an effect.

There are however, two problems with this ratio. The first is that its value depends on the size of the change we make to S_j, this is particularly true if the response of v_i to changes in S_j is non-linear (as most enzyme rate responses are). The second problem is that the ratio depends on the units we choose to measure the rate and concentration. A possible solution to the later problem would be for all experimenters to employ a standard set of units but this would be almost impossible to achieve in practice. A much easier way around this problem is to eliminate the units altogether by scaling the ratio with the rate and concentration. We can eliminate the concentration units by dividing the change, δS_j, by the concentration of S_j, i.e. $\delta S_j / S_j$. Likewise we can eliminate the reaction units by dividing by v_i. Therefore, rather than measure $\delta v_i / \delta S_j$ it would be more sensible to measure:

$$\frac{\delta v_i}{\delta S_j} \frac{S_j}{v_i}.$$

This still leaves us with the first problem which is that the value of the ratio varies with the amount of change we make to S_j. We could all decide on a standard change to make in S_j, say doubling S_j and measuring the change in v_i. There is however a much more elegant and ultimately more profitable approach to standardizing the change in S_j.

Assume that the substrate concentration has been set to a value S_b. At this

concentration, the enzyme will show a reaction rate of v_b. If we make a change δS_1 to S_b, then this will cause a change in rate δv_1. The ratio, $\delta v_1/\delta S_1$ is the slope of the line. If we now make the change in S smaller, say δS_2, the ratio will be slighly different because the response is non-linear. If we were to continue making δS smaller and smaller, the slope given by $\delta v/\delta S$ would slowly approach a limiting slope. This line is the tangent to the curve at the point S_b. Those familiar with the calculus will recognize that in reducing δS to a smaller and smaller increment, the ratio, $\delta v/\delta S$, has reached a limiting value called the derivative:

$$\frac{\delta v}{\delta S} \underset{\text{as } \delta S \to 0}{\longrightarrow} \frac{dv}{dS}. \tag{5.1}$$

The ratio, $\delta v/\delta S$ tends to the differential dv/dS as δS tends to zero. The differential has a precise meaning, it is the slope of the curve at the point S and significantly for us, it has a *unique* value at this point.

As before, we can scale dv/dS to eliminate the measuring units so that we end up with

$$\frac{dv}{dS}\frac{S}{v}.$$

This expression represents the scaled slope of the response curve at S, and is called the *elasticity coefficient* of the rate of reaction v with respect to the concentration of metabolite S. It measures how responsive a reaction rate is to changes in the concentration of a modifier, in this case the concentration of substrate, S. We could have changed the concentration of the product, P, or the concentration of an effector. In either case we would be able to measure an elasticity. This means there will be as many elasticity coefficients for a particular enzyme as there are modifiers that might affect its reaction rate. Thus, not only will an enzyme be characterized by a substrate elasticity but also by a product elasticity and any effector elasticities. In addition, other factors which might affect the reaction rate, such as pH, ionic strength and so on, will also have associated elasticity coefficients. Any particular enzyme will thus be fully characterized when all its elasticities have been measured or computed.

As will be more fully explained in later sections the value of a particular elasticity depends on the concentrations of all the modifiers that the enzyme may interact with. This behavior may not be too obvious at this point but is crucial to a proper understanding of elasticities. In practice it means that if an enzyme is purified with the intention of measuring its elasticities then the concentrations of the substrates and products, the pH, ionic strength and so on should be faithfully recreated in order to mimic the *in vivo* conditions. If this is not done, then the measured values for the elasticities will not reflect the elasticities *in vivo* and their usefulness will be lost. As will be revealed in the next chapter, the elasticities are the building blocks with which we can begin to understand the properties of intact pathways.

5.2 Elasticity Coefficients

The **elasticity coefficient** is defined by the following expression:

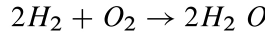
$$\varepsilon^v_{S_i} = \left(\frac{\partial v}{\partial S_i} \frac{S_i}{v} \right)_{S_j, S_k, \ldots} = \frac{\partial \ln v}{\partial \ln S_i} \approx v\% / S_i\% \tag{5.2}$$

The symbol for an elasticity is the Greek epsilon, ε. It measures how responsive a reaction rate is to changes in the concentration of a modifier, in this case the concentration of modifier S_i. Any modifier can be changed to see how they affect the reaction rate, we could have changed the concentration of the product, effector or anything else that might affect the reaction rate. This means there will be as many elasticity coefficients for a particular reaction step as there are modifiers that might affect it. Thus, not only will a reaction step be characterized by a substrate elasticity but also by a product elasticity, any effector elasticities, the enzyme concentration, the pH, ionic strength and any other elasticity coefficients. Any particular enzymic step will thus be fully characterized when all its elasticities have been measured or computed.

When writing down the elasticity symbol, ε, a subscript is often used to indicate the modulating factor (S), and a superscript to indicate the effect

that is being measured (v). Those familiar with quantitative economics will have come across a similar concept. The subscripts, $S_j, S_k, \ldots$ in the definition (5.2) indicate that any species or factor that could also influence the reaction rate must be held constant at their current value when species S_i is changed. This is also implied in the use of the **partial derivative** symbol, ∂, rather than the derivative symbol, d. In normal usage, these subscripts are often left out as the partial derivative symbol is usually sufficient to indicate what is meant. Given that the elasticity is defined in terms of a derivative it is possible to computed an elasticity for a give rate law by differentiation (See Example 5.1).

The elasticity is closely related to the **kinetic order**, sometimes called the reaction order. For simple mass-action chemical kinetics, the kinetic order is the power to which a species is raised in the kinetic rate law. Reactions with zero-order, first-order and second-order are commonly found in chemistry, and in each case the kinetic order is zero, one and two, respectively. For a reaction such as:



$$2H_2 + O_2 \rightarrow 2H_2\,O$$

where the irreversible mass-action rate law is given by:

$$v = k\ H_2^2 \cdot O_2$$

the kinetic order with respect to hydrogen is two and oxygen one. In this case the kinetic order also corresponds to the stoichiometric amount of each molecule although this may not always be true.

Example 5.1 shows the elasticities for zero, first, second, and nth order reactions. From the example we see that the elasticity reduces to the expected kinetic order for simple mass-action kinetics.

From the definition it is apparent that elasticities are dimensionless quantities. In biochemical systems theory, elasticities are also called the **apparent kinetic order**.

Example 5.1

Determine the elasticities for the following mass-action rate laws:

1. $v = k$
   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = 0$
2. $v = kA$
   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = \dfrac{A\,k}{kA} = 1$
3. $v = kA^2$
   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = \dfrac{2kAA}{kA^2} = 2$
4. $v = kA^n$
   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = \dfrac{nkA^{n-1}A}{kA^n} = n$

Operational Interpretation

The definition of the elasticity (5.2) also gives us a useful operational interpretation.

Operational Definition: The elasticity is the fractional change in reaction rate in response to a fractional change in a given reactant or product while keeping all other reactants and products constant.

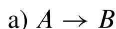
That is, the elasticity measures how responsive a reaction is to changes in its immediate environment. Since the elasticity is expressed in terms of fractional changes, it is also possible to get an approximate value for the elasticity by considering percentage changes. For example, if we increase the substrate concentration of a particular reaction by 2% and the reaction rate increases by 1.5%, then the elasticity is given by $1.5/2 = 0.75$. The elasticity is however only strictly defined (See equation (5.2)) for infinitesimal changes and not finite percentage changes. So long as the changes are small, the finite approximation is a good estimate for the true elasticity.

For a given reaction, there will be as many elasticity coefficients as there are reactants, products and other effectors of the reaction. For species that cause reaction rates to increase, the elasticity is **positive**, while for species

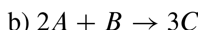
that cause the reaction rate to decrease, the elasticity is **negative**. Therefore, reactants generally have positive elasticities and products generally have negative elasticities (Figure 5.2).

Example 5.2

How many elasticities are there for the following mass-action reactions:


a) $A \rightarrow B$

There are two elasticities, ε^v_A which will be positive and ε^v_B which will be negative.


b) $2A + B \rightarrow 3C$

There are three elasticities, ε^v_A which will be positive, ε^v_B which will also be positive and ε^v_C which will be negative.

There are different ways to calculate an elasticity including numerical, algebraic, and experimental. The numerical and algebraic methods rely on knowing the reaction rate law. We saw in example (5.1) how elasticities were computed algebraically. Numerically the elasticity can by estimated by making a small change (say 5%) to the chosen reactant concentration and measuring the change in the reaction rate. For example, assume that the reference reaction rate is v_o, and the reference reactant concentration, S_o. If we increase the reactant concentration by ΔS_o and observe the new reaction rate at v_1, then the elasticity can be estimated by using Newton's difference quotient:

$$\varepsilon^v_S \simeq \frac{v_1 - v_o}{\Delta S_o}\frac{S_o}{v_o} = \frac{v_1 - v_o}{v_o} \bigg/ \frac{S_1 - S_o}{S_o}$$

Newton's quotient method relies on making one perturbation to S_o. A much better estimate for the elasticity can be obtained by doing two separate perturbations in S_o. One perturbation to increase S_o and another to decrease S_o. In each case the new reaction rate is recorded; this is called the three-point estimation method. For example if v_1 is the reaction rate when we increase S_o, and v_2 is the reaction rate when we decrease S_o, then we can use the following three-point formula to estimate the elasticity:

$$\varepsilon^v_S \simeq \frac{1}{2}\frac{v_1 - v_2}{S_1 - S_o}\left(\frac{S_o}{v_o}\right)$$

Example 5.3

Estimate the elasticity using Newton's difference quotient and the three-point estimation method. Compare the results with the exact value derived algebraically:

Let $v = S/(0.5 + S)$. Assume S is 0.6.

a) Algebraic Evaluation

Differentiation and scaling the rate law gives the elasticity as $0.5/(0.5 + S)$. At a value of 0.6 for S, the exact value for the elasticity is: 0.4546

b) Difference Quotient

Let us use a step size of 5%. Therefore $h = 0.05 \times 0.6 = 0.03$ from which $S_1 = 0.63$. $S_o = 0.6$. From these values we can compute v_1 and v_o. $v_o = 0.6/(0.5 + 0.6) = 0.5454$, $v_1 = 0.63/(0.5 + 0.63) = 0.5575$. From these values the estimated elasticity is given by: $\varepsilon^v_S = ((0.5575 - 0.5454)/0.5454)$ / $((0.63 - 0.6)/0.6) = 0.443$

Compared to the exact value the error is 0.0116, or **2.55 % error**

c) Three-Point Estimation

In addition to calculating v_1 in the last example, we must also compute v_2. To do this we subtract h from S_o to give $v_2 = 0.533$. The Three-Point estimation formula gives us: $\varepsilon^v_S = 0.5\, \frac{0.5575 - 0.5327}{0.03} \frac{0.6}{0.5454} = 0.4549$

Compared to the exact value the error is only 0.0033, or **0.7 % error**, a significant improvement over the difference quotient method.

The degree of error in the difference quotient method will depend on the value of S, which in turn determines the degree of curvature (or nonlinearity) at the chosen point. The more curvature there is the more inaccurate the estimate. The value in this example was chosen where the curvature is high, therefore the error was larger.

In the examples shown in (5.1) the elasticities were constant values. However for more complex rate law expressions this need not be the case (See Example (5.4)) and the elasticity will change in response to changes in the reactant and product concentrations. Consequently when measuring the elasticity numerically or experimentally one has to choose a particular operating point.

Example 5.4

Determine the elasticities for the following rate laws:

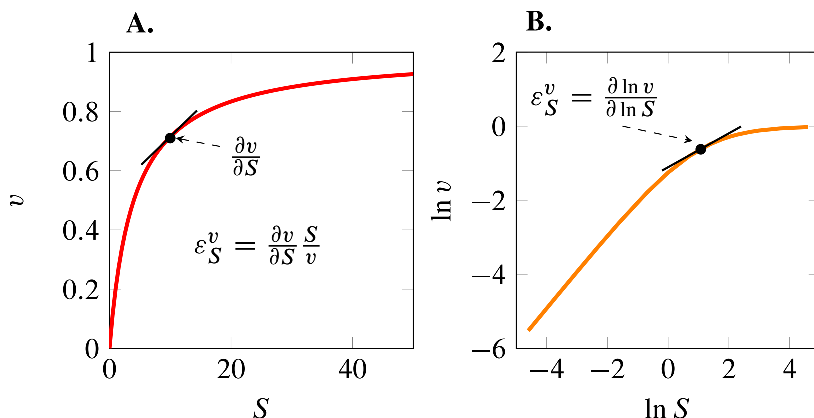


Figure 5.1 **A.** The slope of the reaction rate versus the reactant concentration scaled by both the reactant concentration and reaction rate yields the elasticity, ε^v_S. **B.** If the log of the reaction rate and log of the reactant concentration are plotted, the elasticity can be read directly from the slope of the curve. Curves are generated by assuming $v = S/(2 + S)$.

1. $v = k(A + 1)$

   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = k\dfrac{A}{k(A+1)} = \dfrac{A}{(A+1)}$

2. $v = k/(A + 1)$

   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = -\dfrac{k}{(1+A)^2}\dfrac{A}{k/(A+1)} = -\dfrac{A}{A+1}$

3. $v = A/(A + 1)$

   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = \dfrac{1}{(A+1)^2}\dfrac{A}{A/(A+1)} = \dfrac{1}{A+1}$

4. $v = kA(A + 1)$

   Elasticity: $\varepsilon^v_A = \dfrac{\partial v}{\partial A}\dfrac{A}{v} = k(1+2A)\dfrac{A}{kA(A+1)} = 1 + \dfrac{A}{A+1}$

The examples illustrate that for more complex rate laws, the elasticity becomes a function of the reactant concentrations.

Experimentally, we can measure an elasticity using the following experiment. Consider a simple reaction such as $A \rightarrow B$ and let us measure the elasticity of reaction A. We must first select an operating point for A and B. This choice will depend on the system under study. For example, perhaps we are interested in the value of the substrate elasticity for an enzyme catalyzed reaction when the substrate and product concentration are at their K_m levels. Once the operating point has been chosen, the reaction is started and the rate of reaction is measured. It is important that during the measurement only a small amount of substrate is consumed and product produced. We now begin the experiment again but this time the substrate concentration is increased by a small amount and the product concentration is reset to its value in the first experiment. The reaction is started and the new reaction rate measured. The fractional change in reaction rate and substrate is recorded and the ratio computed to give the substrate elasticity. In principle the same kind of experiment could be performed on the product, this time keeping the substrate concentration constant.

Simple protocol for estimating the substrate elasticity

1. Set substrate and product concentrations to their operating points.
2. Record the reaction rate at the operating point.
3. Restore all concentrations to their original starting points.
4. Increase the concentration of substrate by a small amount.
5. Record the new reaction rate.
6. Compute the elasticity by dividing the fractional change in reaction rate by the fractional change in substrate concentration.
7. At all times, maintain other substrate, product and effector concentrations at the operating point.

The algebraic definition of the elasticity automatically suggests ways to estimate their values. Here we have seen a number of methods, including

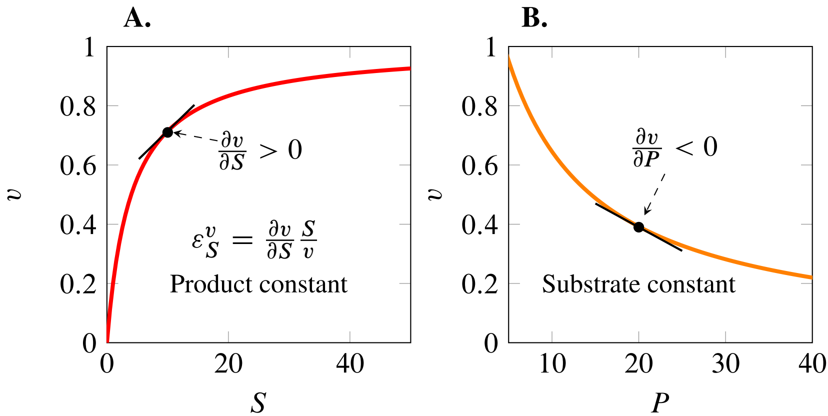


Figure 5.2 **A.** Reaction rate versus reactant. Increases in the reactant cause an increase in the rate. A positive slope yields a positive elasticity. **B.** Reaction rate versus product (assuming a positive rate from reactant to product). Increases in product result in a decrease in reaction rate; a negative slope yields a negative elasticity. Curves generated by assuming $v = S/(2 + S)$ and $2/(1 + (0.1 + 0.2P))$, respectively.

algebraic differentiation of the rate law (if the rate laws is available), plotting kinetic data on a log/log plot, and numerical computation of values by simulation. Both the algebraic differentiation and the numerical simulation require some kind of qualitative model. The log/log plot can in principle be used to empirically determine an elasticity from a an enzyme kinetics experiment. For example, the following table

Log Form

The definition of the elasticity in equation (5.2) shows the elasticity expressed using a log notation:

$$\varepsilon^v_S = \frac{\partial \ln v}{\partial \ln S}$$

This notation is frequently used in the literature. The right panel of Fig-

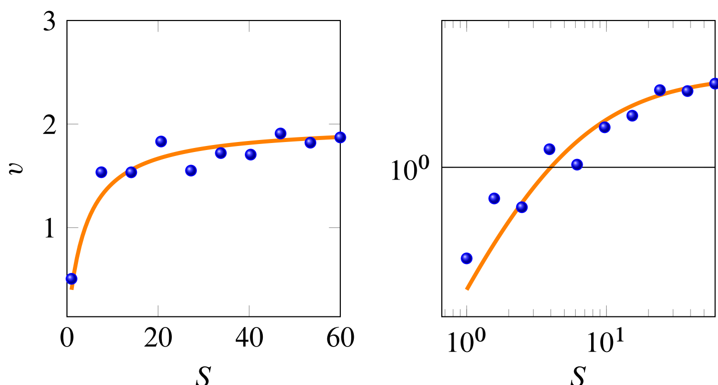


Figure 5.3 **A.** Left Side: Plot of reaction rate versus substrate concentration. Measurements include errors. Right Side: Same data but plotted in log space. The elasticity can be read directly from the slope of the curve. Curves are generated by assuming $v = 2S/(4 + S)$.

ure 5.3 shows one application of this notation, namely that a log-log plot of reaction rate versus reactant concentration yields a curve where the elasticity can be read directly from the slope. The origin of this notation will be explained here.

If we examine the growth pattern of a micro-organism, we will often find that it follows a pattern of the kind, $y = a^x$. What this means is that the number of microorganisms increases by a fixed proportion per unit time. Often such data is plotted on a semi-logarithmic scale rather than the usual linear scale as it helps to emphasize the fact that the relative growth under these conditions is the same throughout the growth phase. To explain this statement, a numerical example will be useful. Of the two sequences of numbers

$$\begin{array}{llllll} 100, & 150, & 200, & 250, & 300, & \dots \\ 100, & 150, & 225, & 337.5, & 506.25, & \dots \end{array}$$

the first shows a regular increase of 50 units and the second a regular increase of 50 per cent. from one number to the next. On a linear scale, the

points representing the first sequence appear as equal distances from each other and those representing the second sequence at increasing distances along the axis. If instead, we take the logarithms of these numbers as in the following sequence:

$$\begin{array}{cccccc} 2, & 2.176, & 2.301, & 2.398, & 2.477, & \ldots \\ 2, & 2.176, & 2.352, & 2.528, & 2.704, & \ldots \end{array}$$

then on the logarithmic scale, it is the second sequence that gives points at equal distances from each other while the first sequence shows points at decreasing distances along the axis. It would seem, therefore, that equal distances between points on a linear scale indicate equal *absolute* changes in the variable and equal distances between points on a logarithmic scale indicate equal *proportional* changes in the variable. Before taking the logarithm, the second sequence increased by 50% each time, in log form however, it increased by a constant absolute amount of 1.176.

More formally we can describe this effect as follows. Consider a variable y to be some function $f(x)$, that is $y = f(x)$. If x increases from x to $(x+h)$ then the change in the value of y will be given by $f(x+h)-f(x)$. The **proportional** change however, is given by:

$$\frac{f(x+h) - f(x)}{f(x)}$$

The **rate of proportional change** at the point x is given by the above expression divided by the step change in the x value, namely h:

Rate of proportional change =

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{hf(x)} = \frac{1}{f(x)} \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \frac{1}{y} \frac{dy}{dx}$$

From calculus we know that $d \ln y/dx = 1/y dy/dx$, therefore the rate of proportional change equals:

$$\frac{d \ln y}{dx}$$

and serves as a measure of the rate of *proportional* change of the function y. Just as dy/dx measures the gradient of the curve, $y = f(x)$ plotted

on a linear scale, $d \ln y/dx$ measures the slope of the curve when plotted on a semi-logarithmic scale, that is the rate of proportional change. For example, a value of 0.05 means that the curve increases at 5% per unit x.

We can apply the same argument to the case when we plot a function on both x and y logarithmic scales. In such a case, the following result is true:

$$\frac{d \ln y}{d \ln x} = \frac{x}{y}\frac{dy}{dx}$$

This shows use the relationship between the log form and non-log form of the elasticity. In approximate terms, we can say that if we make a x % change in the concentration of a molecular species then the elasticity tells us the percentage change, v %, in the reaction rate. For this reason one will sometimes find the elasticity expressed as a ratio of percentage changes:

$$\varepsilon^v_{S_i} \approx \frac{\%\text{ change in } v}{\%\text{ change in } S_i} \tag{5.3}$$

For example, if the concentration of a substrate is increased from 1.5 mM to 1.95 mM then the percentage increase in substrate concentration is 30 %. If at the same time, the reaction rate of the enzyme increases from 55 $\mu M g^{-1}\text{min}^{-1}$ to 12 $\mu M g^{-1}\text{min}^{-1}$, then the percentage increase in rate must be 24 %. Therefore the elasticity can be estimated approximately from the ratio 24/30, which is equal to 0.8; that is, the enzyme rate changes almost in proportion to a change in substrate. If the enzyme were acting *in vivo* and a disturbance upstream caused the concentration of substrate to rise then this enzyme would respond by increasing its rate almost in proportion to the change in substrate concentration.

5.3 Mass-action Kinetics

Computing the elasticities for mass-action kinetics is straight forward. For a reaction such as $v = kS$, we showed earlier (5.1) that $\varepsilon^v_S = 1$. For a generalized irreversible mass-action law such as:

$$v = k \prod S_i^{n_i}$$

the elasticity for species S_i is n_i. For simple mass-action kinetic reactions, the kinetic order and elasticity are therefore identical and independent of species concentration.

For a simple irreversible mass-action reaction rate law such as:

$$v = k_1 S - k_2 P \tag{5.4}$$

The elasticities for the substrate and product are given by:

$$\varepsilon^v_S = \frac{k_1 S}{k_1 S - k_2 P} = \frac{v_f}{v} \tag{5.5}$$

$$\varepsilon^v_P = -\frac{k_2 P}{k_1 S - k_2 P} = -\frac{v_r}{v} \tag{5.6}$$

In the above equations v_f is the forward rate, v_r is the reverse rate, and v is the net rate. Note that ε^v_S is **positive** and ε^v_P **negative**. In general, the elasticity for an effector that results in an increase in reaction rate will be positive and conversely if the effector results in a decrease in the reaction rate.

If we divide top and bottom by k_1 and S in equation (5.5), and k_2 and P in equation (5.6), and noting that the ratio $k_1/k_2 = K_{eq}$ (See equation (A.2)), $P/S = \Gamma$ and $\Gamma/K_{eq} = \rho$ we can express the elasticities in the form:

$$\begin{aligned}
\varepsilon^v_S &= \frac{1}{1 - \Gamma/K_{eq}} = \frac{1}{1-\rho} \\
\varepsilon^v_P &= -\frac{\Gamma/K_{eq}}{1 - \Gamma/K_{eq}} = -\frac{\rho}{1-\rho}
\end{aligned} \tag{5.7}$$

These expressions can vary over a wide range of values. Far from equilibrium ($\rho \simeq 0$) ε^v_S will lie close to 1.0, while ε^v_P will be close to -0.0. When operating close to equilibrium however ($\rho \approx 1$), the same elasticities will tend to $+\infty$ and $-\infty$, respectively. This behavior is depicted in Figure 5.4.

Of interest is the following relation (See equation (A.5)):

$$\frac{\varepsilon^v_P}{\varepsilon^v_S} = \frac{v_r}{v_f} = -\rho$$

which connects the ratio of the elasticities to the disequilibrium ratio. It also follows from the above equations that the sum of the elasticities for mass-action kinetic rate laws is always one:

$$\varepsilon^v_S + \varepsilon^v_P = 1 \tag{5.8}$$

This means that if one of the elasticities is known, the other can be easily determined by subtraction.

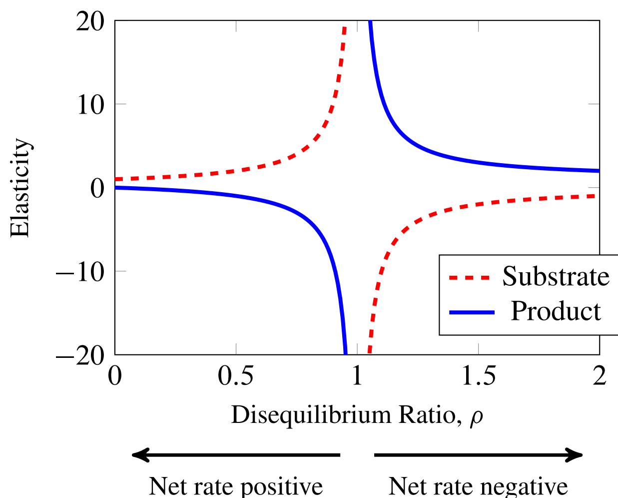


Figure 5.4 Elasticities as a function of the disequilibrium ratio, ρ.

Equation (5.8) is significant for another reason. The absolute magnitude of ε^v_S will **always** be larger than the absolute value for ε^v_P when dealing with mass-action kinetics. That is:

ρ	$\varepsilon^v_S = 1/(1-\rho)$	$\varepsilon^v_P = -\rho/(1-\rho)$
0.9	10	-9
0.5	2	-1
0.2	1.2	-0.25
0.1	1.111	-0.111

Table 5.1 Selected values for the elasticities and the disequilibrium ratio, ρ: $\|\varepsilon^v_S\| > \|\varepsilon^v_P\|$

$$\|\varepsilon^v_S\| > \|\varepsilon^v_P\|$$

For small elasticity values the relative difference between the elasticities can be significant. This means that changes in substrate concentrations will have a much greater effect on the reaction velocity than changes in product concentrations. As will be discussed more fully in a separate volume, 'Control Theory for Bioengineers', the propagation of signals along a pathway is **determined** by the elasticity values. Given that substrate elasticities are larger than product elasticities, signal propagation tends to amplify when traveling downstream compared to signals traveling upstream which tend to be attenuated. For the general reversible mass-action rate law:

$$v = k_1 \prod S_i^{n_i} - k_2 \prod P_i^{m_i} \tag{5.9}$$

The elasticities can be shown to equal:

$$\begin{aligned} \varepsilon^v_{S_i} &= \frac{n_i}{1-\rho} \\ \varepsilon^v_{P_i} &= -\frac{m_i \rho}{1-\rho} \end{aligned} \tag{5.10}$$

5.4 Local Equations

The elasticity coefficient is of central importance to metabolic control analysis. Just as the Michaelian constants are essential to describing the rate of a enzyme-catalysed reaction, so the elasticities are essential to describing the behavior of whole pathways. Before we can discuss in any great detail how this is achieved, since that is the topic of the next chapter, some direct uses of the elasticities will be given here.

It will be recalled that the elasticity coefficient is given by

$$\varepsilon^v_S = \frac{\partial v}{\partial S}\frac{S}{v}$$

This definition can be rearranged and an approximate equation written in the form

$$\frac{\delta v}{v} \approx \varepsilon^v_S \frac{\delta S}{S}$$

This relation is approximate because the changes considered are finite, and the definition of an elasticity applies strictly to infinitesimal changes. The equation describes how, given a fractional change in some effector S, the resulting fractional change in rate can be computed. For example, if the elasticity of a enzyme reaction towards an effector S is 0.8, then given a fractional change in S of 0.05 (a 5% change in S), the fractional change in rate is given by

$$0.8 \times 0.05 = 0.04$$

in others words, a 5% change in S leads to a 4% change in reaction rate. But what of the following situation:

The diagram (Figure 5.5) shows a fragment from a larger pathway. The central reaction step has three effectors which could potentially affect its rate v, these are S, P and an inhibitor, I. Let us consider a disturbance[1] somewhere in the pathway but *not* originating at the reaction step under consideration. This disturbance will ultimately cause changes in each of

[1] This could be one of a number of causes, a change in enzyme expression, nutrient supply change, hormonal change etc.

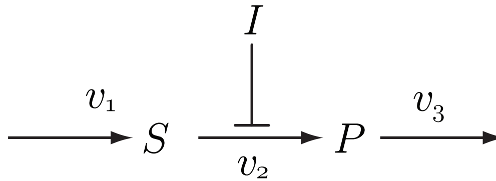


Figure 5.5 Species I inhibits reaction v_2 in addition to potential affects from S and P.

the effectors by amounts δS, δP and δI. These changes will also be accompanied by a change to the reaction rate by an amount, δv. There are two immediate questions we can ask, what is the relationship between the change in the effectors and the change in rate and what is the contribution that each change in effector makes to the final change in rate?

The answers to these questions are straight-forward to obtain. Provided that the changes are small, then the fractional change in rate, $\delta v/v$ is defined by the sum of the individual contributions:

$$\frac{\delta v}{v} \approx \varepsilon^v_S \frac{\delta S}{S} + \varepsilon^v_P \frac{\delta P}{P} + \varepsilon^v_I \frac{\delta I}{I}$$

For example, let us assume the following values for the elasticities, $\varepsilon^v_S = 0.4$; $\varepsilon^v_P = -0.5$; $\varepsilon^v_I = -0.2$, and assuming the following changes in effectors, $\delta S/S = 0.05$; $\delta P/P = 0.03$; $\delta I/I = 0.01$, then the fractional change in rate through the step is given by:

$$\frac{\delta v}{v} \approx 0.4 \times 0.05 + (-0.5) \times 0.03 + (-0.2) \times 0.01 = 0.003$$

The rate has only changed by 0.3 %, much of the potential increase that could have been obtained by the change in S has been reduced by the strong product inhibition. To answer the question, what contribution does each effector make to the final change in rate is given simply by examining the individual changes. Thus out of the total absolute change in rate, the change brought about by S contributed 54 % while the change in product and inhibitor contributed -41 and -5 % respectively. Clearly the change in inhibitor was not an important factor.

In general, for a reaction step embedded in a pathway and acted upon by m modifiers, the change in rate due to changes in all modifiers is given by the relation

$$\frac{dv}{v} = \sum_{j=1}^{m} \varepsilon^{v}_{S_j} \frac{dS_j}{S_j}$$

where the symbol, $\sum$ means 'sum of'. In the equation, the small, but finite changes have been replaced by differentials so that the relation is exact. If the concentration of enzyme is also changed then we may also add the enzyme elasticity to the sum, as in:

$$\frac{dv}{v} = \sum_{j=1}^{m} \varepsilon^{v}_{S_j} \frac{dS_j}{S_j} + \varepsilon^{v}_{E} \frac{dE}{E}$$

These relations are probably the most important mathematical relations used in MCA and we will come across their application in subsequent chapters. They are a modification of standard total derivative relation. This states that if y is some function f of m variables, x_i,

$$y = f(x_1, x_2, \ldots, x_m),$$

then the total derivative of y is given according to the definition:

$$dy = \frac{\partial y}{\partial x_1} dx_1 + \frac{\partial y}{\partial x_2} dx_2 + \ldots + \frac{\partial y}{\partial x_m} dx_m$$

5.5 General Elasticity Rules

Just as there are rules for differential calculus, there are similar rules for computing elasticities. These rules can be used to simplify the derivation of elasticities for complex rate law expressions. Table 5.2 shows some common elasticity rules, where a designates a constant and x the variable. For example the first rule says that the elasticity of a constant is zero.

1.	$\varepsilon(a) = 0$
2.	$\varepsilon(x) = 1$
3.	$\varepsilon(f(x) \pm g(x)) = \varepsilon(f(x))\, \frac{f(x)}{f(x)+g(x)} \pm \varepsilon(g(x))\, \frac{g(x)}{f(x)+g(x)}$
4.	$\varepsilon(x^a) = a$
5.	$\varepsilon(f(x)^a) = a\varepsilon(f(x))$
6.	$\varepsilon(f(x)\, g(x)) = \varepsilon(f(x)) + \varepsilon(g(x))$
7.	$\varepsilon(f(x)/g(x)) = \varepsilon(f(x)) - \varepsilon(g(x))$

Table 5.2 Transformation rules for determining the elasticity of a function, a = constant, x = variable.

We can illustrate the use of these rules with a simple example. Consider the reversible mass-action rate law (5.4):

$$v = k_1\, S - k_2\, P$$

To determine the elasticity we first apply rule 3 to give:

$$\varepsilon^v_S = \varepsilon_S(k_1\, S)\frac{k_1\, S}{k_1\, S - k_2\, P} - \varepsilon_S(k_2\, P)\frac{-k_2\, P}{k_1\, S - k_2\, P}$$

where $\varepsilon_S(f)$ means the elasticity of expression f with respect to variable S.

Now transform the elasticity terms by applying additional rules. Let us apply rule 6 to the expression $\varepsilon_S(k_1\, S)$ to give:

$$\varepsilon_S(k_1\, S) = \varepsilon_S(k_1) + \varepsilon_S(S)$$

We can now apply rule 1 to the first term on the right and rule 2 to the second term on the right to give:

$$\varepsilon_S(k_1\, S) = 0 + 1$$

Since we're evaluating the elasticity of S, P in this situation is a constant, therefore:

$$\varepsilon_S(k_2\, P) = \varepsilon_S(k_2) + \varepsilon_S(P) = 0 + 0$$

Combining these results yields:

$$\varepsilon^v_S = \frac{k_1 \ S}{k_1 \ S - k_2 \ P}$$

which corresponds to the first equation in (5.5).

Now consider a simple enzyme kinetic rate equation. One of the most famous is the Michaelis-Menten equation:

$$v = \frac{V_m \ S}{K_m + S}$$

where V_m is the maximal velocity and M_m the substrate concentration at half maximal velocity.

The elasticity for this equation can be derived by first using the quotient rule (rule 7) which gives:

$$\varepsilon^v_S = \varepsilon(V_m \ S) - \varepsilon(K_m + S)$$

The rules can now be applied to each of the sub-elasticity terms. For example we can apply rule 6 to the first term, $\varepsilon(V_m \ S)$, and rule 3 to the second term, $\varepsilon(K_m + S)$, to yield:

$$\varepsilon^v_S = (\varepsilon(V_m) \ + \varepsilon(S)) - \left(\varepsilon(K_m) \frac{K_m}{K_m + S} + \varepsilon(S) \frac{S}{K_m + S} \right)$$

Applying rules 1 and 2 allows us to simplify ($\varepsilon(V_m) = 0; \varepsilon(\text{Km}) = 0; \varepsilon(S) = 1$) the equation to:

$$\varepsilon^v_S = 1 - \left(\frac{S}{K_m + S} \right)$$

or

$$\varepsilon^v_S = \frac{K_m}{K_m + S}$$

Example 5.5

Determine the elasticity expression for the rate laws using log-log rules:

1. $v = k(A+1)$
   Begin with the product rule 6:

$$\varepsilon_A^v = \varepsilon(k) + \varepsilon(A+1) = \varepsilon(A+1)$$

   Next use the summation rule 3 and rule 2:

$$\begin{aligned}\varepsilon_A^v &= \varepsilon(A+1) = \varepsilon(A)\frac{A}{A+1} + \varepsilon(1)\frac{1}{A+1} \\ &= \frac{A}{A+1} + 0 = \frac{A}{A+1}\end{aligned}$$

2. $v = k/(A+1)$
   Begin with the quotient rule 6 followed by Rule 3 and 2:

$$\begin{aligned}\varepsilon_A^v &= \varepsilon(k) - \varepsilon(A+1) = -\frac{1}{A+1} \\ &= -\frac{A}{A+1}\end{aligned}$$

3. $v = A(A+1)$
   Begin with the quotient rule 6:

$$\varepsilon_A^v = \varepsilon(A) + \varepsilon A + 1$$

   Next use Rule 2, 3 and 2:

$$\varepsilon_A^v = 1 + \frac{1}{A+1}$$

To make matters even simpler we can define the elasticity rules using an algebraic manipulation tool such as Mathematica (`http://www.wolfram.com/`) to automatically derive the elasticities [89]. To do this we must first enter the rules in Table 5.2 into Mathematica. The script shown in Figure 5.6 shows the same rules (with a few additional ones) in Mathematica format.

The notation `f[x_,y_] := g()` means define a function that takes two arguments, `x_` and `y_`. The underscore character in the argument terms is essential. Note also the symbol ‘`:`’ in the assignment operator.

```
(* Define elasticity evaluation rules *)
el[x_, x_]      := 1
el[k_, x_]      := 0
el[Log[u_,x_]   := el[Log[u],x] = el[u,x]/Log[u]
el[Sin[u_],x_]  := el[Sin[u],x] = u el[u,x]Cos[u]/Sin[u]
el[Cos[u_],x_]  := el[Sin[u],x] = -u el[u,x]Sin[u]/Cos[u]
el[u_*v_,x_]    := el[u*v,x] = el[u,x] + el[v,x]
el[u_/v_,x_]    := el[u/v,x] = el[u,x] - el[v,x]
el[u_+v_,x_]    := el[u+v,x] = el[u,x]u/(u+v) + el[v,x]v/(u+v)
el[u_-v_,x_]    := el[u-v,x] = el[u,x]u/(u-v) - el[v,x]v/(u-v)
el[u_^v_,x_]    := el[u^v,x]  = v (el[u,x] + el[v,x] Log[u])
```

Figure 5.6 Elasticity rules expressed as a Mathematica script.

Typing `el[k1 S - k2 P, S]` into Mathematica will result in the output:

```
k1 S/(-k2 P + k1 S)
```

5.6 Summary

The elasticity coefficient is a measure of how sensitive the rate of a reaction is to changes in its environment. The factors that are of usual interest in MCA are the concentrations of modifiers, that is, the substrates, products and effectors and the concentration of enzyme.

There will be as many elasticities as there are modifiers of the reaction. The elasticity is strictly defined in terms of a partial derivative which means that it measures the change in rate when one modifier is changed. For example the substrate elasticity is measured when all other modifiers are held constant except for the substrate concentration. Algebraically this is achieved by partial differentiation and experimentally by clamping the appropriate modifier concentrations.

The elasticity coefficient can be written in various equivalent forms each

reflecting a different emphasis:

$$\varepsilon^v_{S_j} = \left(\frac{\partial v_i / v_i}{\partial S_j / S_j} \right)_{S_k, S_l, \ldots} = \frac{S_j}{v_i} \left(\frac{\partial v_i}{\partial S_j} \right)_{S_k, S_l, \ldots} = \left(\frac{\partial \ln v_i}{\partial \ln S_j} \right)_{S_k, S_l, \ldots}$$

The first form is the ratio of fractional changes, the second form the scaled slope on a linear plot, and the third form the slope of a log/log plot.

The notation, $S_k, S_l \ldots$ means that these modifiers are held constant during the measurement of the partial derivative. An approximate form of the elasticity is given by

$$\varepsilon^v_S \approx \frac{\%\text{ change in } v}{\%\text{ change in } S}$$

which can be used to estimate an elasticity is changes in reaction rate and modifier are known. Elasticities have a number of important properties:

- The elasticity coefficient is **not** a constant but depends on the concentrations of all modifiers that might affect the rate of the reaction. An elasticity is not like a K_m or K_i; the Michaelian constants are characteristic for a particular enzyme and modifier, reflecting the enzymes' kinetic mechanism and interaction energy with the modifier. Kinetic constants do not in general depend on the concentrations of the modifiers; elasticities **do**.

- In general, an elasticity is a function of both the kinetic characteristics of an enzyme and the concentration of all the various modifiers that might interact with the enzyme.

- For the standard irreversible Michaelian mechanism, the elasticity of a substrate at saturating levels is zero and when the substrate is below its K_m, the elasticity is unity. When the substrate concentration is equal to the K_m, the elasticity has a value of 0.5.

Given a change in concentration of a modifier, it is possible to use the value of the elasticity coefficient to predict approximately the change in rate, thus

$$\frac{\delta v}{v} \approx \varepsilon^v_S \frac{\delta S}{S}$$

If more than one modifier is changing at a time, then the approximate change in rate is given by the sum of the individual contributions:

$$\frac{\delta v}{v} \approx \sum_j \varepsilon^v_{S_j} \frac{\delta S_j}{S_j}$$

It is very important to appreciate that the elasticities used in the above equation must be measured at the prevailing state of the modifiers. It makes no sense to use an elasticity that has previously been measured at a substrate concentration of 2mM and then to use the same elasticity value *value* at a substrate concentration of 20mM. The value of an elasticity is dependent on the state of all modifiers.

Further Reading

1. Fell D A (1996) Understanding the Control of Metabolism. Portland Press, ISBN: 185578047X

2. Heinrich R and Schuster S (1996) The Regulation Of Cellular Systems. Springer; 1st edition, ISBN: 0412032619

3. Sauro H M (2013) Control Theory for Bioengineering. www.sys-bio.org

4. Sauro HM (2012) Enzyme Kinetics for Systems Biology. 2nd Edition, Ambrosius Publishing ISBN: 978-0982477335

Exercises

1. What is the relevance of elasticity coefficients in understanding network dynamics?

2. State the operational interpretation of an elasticity.

3. Why is the elasticity coefficient expressed in terms of a partial derivative? What does it mean in terms of an experimental operation?

4. An experiment indicates that a given molecule X has an elasticity of -0.5 with respect to the rate of a reaction. State two key aspects that this elasticity describes.

5. What is the elasticity with respect to the species A for the rate law $v = kA^3$?

6. Work out algebraically the elasticity for the rate law, $v = k_1 * S + k_2$. Describe its properties at high and low levels of S.

7. Derive the elasticity expression with respect to x for the following:

   a) $v = x^2 + 1$
   b) $v = x^2 + x$
   c) $v = x/(x^2 + 1)$

8. Describe one technique for numerically estimating an elasticity.

9. Given a change in $\delta v/v$ equal to 0.04, and that $\varepsilon^v_S = 0.1$, what was the change in $\delta S/S$? If the concentration of S was 2.5mM, what was the absolute change in S?

10. Given that the concentration of S is 3mM, and that the elasticity of an enzyme with a rate law, $VmaxS/(Km + S)$ is 0.6, what is the Km of the enzyme? What would be the elasticity at 8mM? Why does the elasticity change?

11. What does the term Γ/K_{eq} measure?

12. Describe the value of disequilibrium ratio as a reaction nears equilibrium.

13. For a mass-action reversible reaction, describe what happens to the substrate and product elasticities as the reaction approaches equilibrium.

14. Derive the two equations in (5.7).

15. Describe the significance of equation (5.8).

16. Using the elasticity rules in Table 5.2, derive the elasticity for the following equation indicating all intermediate steps.

$$v = S^n/(K_m + S^n)$$

6

Introduction to Biochemical Control

6.1 Control, Responses and Regulation

In the vernacular, the word control usually means the ability to influence, command or to restrain a situation or process.[1] In this chapter the term control will be used to describe how much influence a given reaction step in a network has on the system. To make matters simpler the system will be considered at steady state so that control will refer to much influence a given reaction step has on the steady state, that is how fluxes and concentrations are influenced. Most reaction steps in a cell are controlled by proteins and the question then becomes how much influence a given protein has on the system's steady state. Experimentally such control can be measured by changing the concentration of an enzyme or changing its activity via an inhibitor and measuring the effect on the steady state flux

[1] In engineering, control theory refers to the body of knowledge concerned with the design and study of systems that can perform specific tasks or achieve a particular objective.

and species concentrations. Concentrations of proteins can be changed in various ways, for example by using irreversible inhibitors, changing the promoter consensus sequence on the gene that codes for the protein or by employing antisense RNA to reduce the expression level. Measurement approaches will be described in more detail in the next chapter.

In the biological literature, the amount of control that a particular reaction step has on a flux or species concentrations is called the **control coefficient**.

In addition to investigating how individual reaction steps control the fluxes and concentrations in a network, we are also interested in how external factors influence the network. Examples of external factors will include the level of nutrients, hormones, and of particular interest to human health, therapeutic drugs. In these situations, rather than use the word control, we will use the term response. Thus a biological cell will have a response to a particular infusion of a drug. The degree of influence an external factor has on a biological system will be described using **response coefficients**.

Finally, the degree to which an external factor or an enzyme has on a cellular network will depend on the network's regulatory mechanisms. The degree of regulation at a given reaction step will be described using the **regulatory coefficient**.

6.2 Control Coefficients

Control coefficients are used to describe how much influence (i.e. control) a given reaction step has on the steady state flux or species concentration level. It is common to measure this influence by changing the concentration of the enzyme that catalyzes the reaction. To describe control coefficients in more detail let us consider a thought experiment.

The following discussion will be centered on the simple linear pathway shown in Figure 6.1. Let us walk through a thought experiment where we will assume to begin with that the species pools, S_1 to S_4 in the pathway

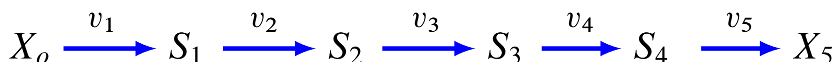


Figure 6.1 Five step linear pathway

are empty, and that X_o and X_5 are fixed species forming the boundary.

To make matters simpler, we can also assume that the right-hand boundary pool, X_5 is set to zero. In order to have a net flux through the pathway, the external metabolite, X_o must have a positive value, perhaps 1 mM. This is the situation at time zero. Let us allow the pathway to evolve in time. The first thing that happens is that the reaction catalyzed by the first enzyme begins to convert X_o into product S_1. Since we assume that X_o is fixed, the concentration of X_o is unaffected by this rate of consumption. However, the product S_1 is a floating species and as time goes on, its concentration will rise. As the concentration of S_1 increases, two things will happen, first the second enzyme will begin to convert S_1 into S_2 and S_1 will begin to inhibit its own production rate by the first enzyme on account of product inhibition. The first reaction will therefore begin to rise at a **slower** rate.

Since the second enzyme is now generating S_2, S_2 starts to increase. S_2 in turn it stimulates the third enzyme to begin making S_3 but it also begins to inhibit the second enzyme. And so on down the chain, all concentrations begin to rise and all enzymes begin to operate by showing a positive rate. The concentrations of the floating species and the reaction rates cannot however go on rising forever. We have already seen that as the species concentrations rise they begin to inhibit the enzymes that produce them. The net effect of these many interactions is that the concentrations slowly settle to a constant value such that the rates at which they are being made is exactly balanced by the rates at which they are being consumed. The rate of the first enzyme must balance the rate of the second enzyme, that is $v_1 = v_2$, but the second and third rates will also be in balance, so that $v_2 = v_3$. This must mean that the rate through the first enzyme must be the same as the rate through the third enzyme, $v_1 = v_3$. In fact all rates of flow across each enzyme will equal each other, that is

$$v_1 = v_2 = v_3 = v_4 = v_5$$

This state is the steady state, the concentrations of all the metabolites settle to some values and no longer evolve in time, and the rate through each step is the **same**.

The fact that the rate across each enzyme is the same also means that there is a constant flow of material through the pathway, which we call the **flux**, symbolized by J. At steady state there are no 'slow' rates or 'fast' rates, they are all the **same**.

In a linear pathway, the rates of reaction are equal to each other at steady state.

With the pathway at steady state we can now consider some additional thought experiments such as the effect of perturbations on the steady state. Let us change the concentration of one of the enzymes and see what happens to the steady state concentrations and the flux, J. Let's try this by doubling the concentration of enzyme, E_2, that catalyzes the second step. The immediate effect is to increase the rate, v_2, through the step. This in turn results in more S_2 being produced and more S_1 consumed, S_2 will therefore rise and S_1 fall. The rise in S_2 will cause a cascade of changes down the pathway towards X_5. The fall in S_1 has a different effect. Assuming that the first enzyme is product inhibited by S_1, the fall in S_1 will cause a **rise** in the rate through v_1. The net effect of all these changes is that the net flux through the pathway will **increase**, all species concentration downstream of v_2 will **increase** and S_1 will **decrease**. Figure 6.2 illustrates a simulation that shows the change in flux through E_2 as the pathway approaches steady state, followed by the effect of a perturbation in E_2 at $t = 0.2$.

We can get an idea of how effective the change in enzyme concentration was by taking the ratio of the change in the species concentration or the flux to the change we made in the enzyme concentration, that is we could measure:

$$\frac{\Delta J}{\Delta E_2}, \quad \frac{\Delta S_1}{\Delta E_2}, \ldots \frac{\Delta S_4}{\Delta E_2}$$

where Δ means 'a change in'. However, because enzyme kinetic rate laws are usually nonlinear, the degree of influence we measure will depend on the size of the ΔE. Therefore instead of making large changes to the

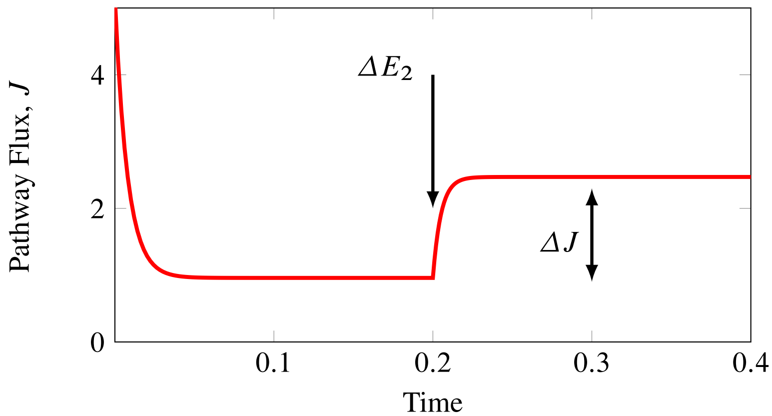


Figure 6.2 Effect of a perturbation in E_2 at $t = 0.2$ on the flux through the pathway, Figure 6.1. Note the initial transition to steady state between $t = 0$ and $t = 0.2$. At $t = 0.2$ a change ΔE_2 is made to E_2 resulting in a change ΔJ in the steady state pathway flux/ Parameters are given in the Jarnac Script: 6.2 at the end of the chapter.

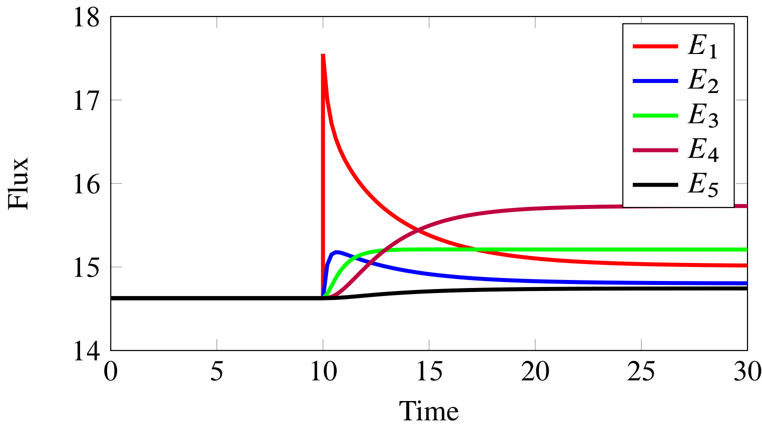


Figure 6.3 Effect of perturbing each enzyme by 20% in the linear pathway, Figure 6.1. Note that each enzyme affects the system differently both in the transient response and in the final steady state response. Parameters given in Jarnac script: 6.3

enzyme concentration we should make small changes, for example:

$$\frac{\delta J}{\delta E_2}, \quad \frac{\delta S_1}{\delta E_2}, \ldots \frac{\delta S_4}{\delta E_2}$$

where δ means 'a small change'. We can be more precise mathematically if we make the changes infinitesimally small, our measurement of influence then becomes:

$$\frac{dJ}{dE_2}, \quad \frac{dS_1}{dE_2}, \ldots \frac{dS_4}{dE_2}$$

Finally, if we want to make the measurement useful to experimentalists we can remove the units by scaling the derivatives, such that:

$$\frac{dJ}{dE_2}\frac{E_2}{J}, \quad \frac{dS_1}{dE_2}\frac{E_2}{S_1}, \ldots \frac{dS_4}{dE_2}\frac{E_2}{S_4}$$

Obviously in an experiment we cannot make infinitesimals changes but we can make changes sufficiently small (but still measurable) that we can

approximate the derivatives. The scaled derivatives are called **control coefficients** and we will define both the flux and concentration control coefficients as follows:

$$C^J_{E_i} = \frac{dJ}{dE_i}\frac{E_i}{J} = \frac{d\ln J}{d\ln E_i} \approx J\%/E_i\% \tag{6.1}$$

$$C^{S_j}_{E_i} = \frac{dS_j}{dE_i}\frac{E_i}{S_j} = \frac{d\ln S_j}{d\ln E_i} \approx S_j\%/E_i\% \tag{6.2}$$

Example 6.1

A given enzyme catalyzed reaction in a metabolic pathway has a flux control coefficient equal to 0.2:

$$C^J_E = 0.2$$

What does this mean?

A flux control coefficient of 0.2 means that increasing the enzyme activity of the step by 1% will increase the steady state flux through the pathway by 0.2%.

In the expression above, J is the flux through the pathway and E_i the enzyme concentration of the ith step. Operationally, an individual C_{E_i} is measured by making a small change in E_i, waiting for the system to reach a new steady state and than taking the ratio of the change. Before moving on to another step, the level of E_i must be restored by to its original value.

From a practical standpoint we see that the control coefficients can also be approximated by the ratio of **percentage changes** which is a useful interpretation for measurement purposes. The other point to note is that like elasticities we can express the control coefficients in log form (See section 5.2).

The flux control coefficient measures the fractional change in flux brought about by a given fractional change in enzyme concentration and the con-

centration control coefficients measure the fractional change in species concentration given a fractional change in enzyme concentration.

Control coefficients are useful because they tell us how much influence each enzyme or protein has in a biochemical reaction network. For example we could increase the copy number of a given gene to change the concentration of the expressed protein and find out how much influence the gene has on its own gene product and on other processes. These changes might include the concentration of other gene products, production of an important commodity metabolite or even the growth rate of the organism. From an engineering perspective, knowing the degree to how processes are influenced is very important. If is important to note however, that knowing the values of the control coefficients does not tell us why certain enzymes or proteins have more influence than others. To answer the 'why' question we must consider the theorems associated with how control is distributed and the relationship of the control coefficients to the elasticities of the network.

6.3 Distribution of Control

Flux control coefficients are a useful measure to judge the degree to which a particular step influences the steady state flux. Even more interesting is that there are numerous relationships between the various coefficients.

Consider the simple two step pathway:

$$X_o \xrightarrow{v_1} S \xrightarrow{v_2} X_1$$

There is a simple graphical technique we can use to study how the enzyme concentrations, E_1 and E_2 control the steady state concentration S, and the steady state flux, J through the pathway. In this system, the steady state flux, J will be numerically equal to the reaction rates v_1 and v_2,

$$J = v_1 = v_2$$

It is important to recall that for many enzyme catalyzed reactions the rate, v is proportional to the concentration of enzyme, E, $v \propto E$.

Let us plot both reaction rates, v_1 and v_2 against the substrate concentration S, Figure 10.9.

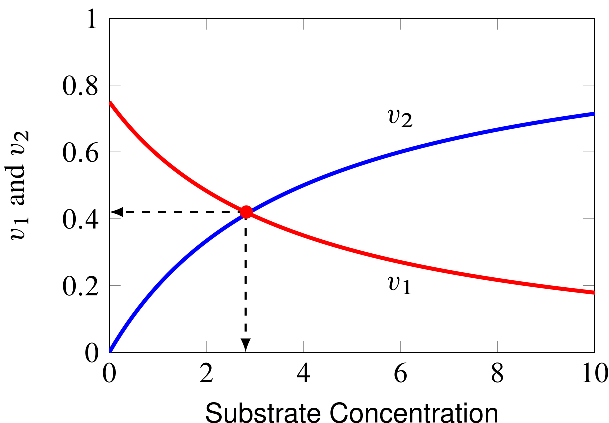


Figure 6.4 Plot of v_1 and v_2 versus the concentration of S for a simple two step pathway. The intersection of the two curve marks the point when $v_1 = v_2$, that is steady state. A perpendicular dropped from this point gives the steady state concentration of S.

Note the response of v_1 to changes in S. v_1 falls as S increases due to product inhibition by S. The intersection point of the two curves marks the point when $v_1 = v_2$, that is the steady state. A line dropped perpendicular from the intersection point marks the steady state concentration of S

Let us now increase the concentration of E_2 by 30% by adding more enzyme (Figure 6.5). Because the reaction rate is proportional to E_2, the curve is scaled upwards although its general shape stays the same. Note how the intersection point moves to the left, indicating that the steady state concentration of S **decreases** relative to the reference state. This is understandable because with a higher v_2, more S is consumed therefore S decreases.

In the next experiment, let us restore E_2 back to its original level and instead increase the amount of E_1 by 30% (Figure 6.6). Again, changing E_1 scales the v_1 curve but because of the negative curvature, the v_1 curve

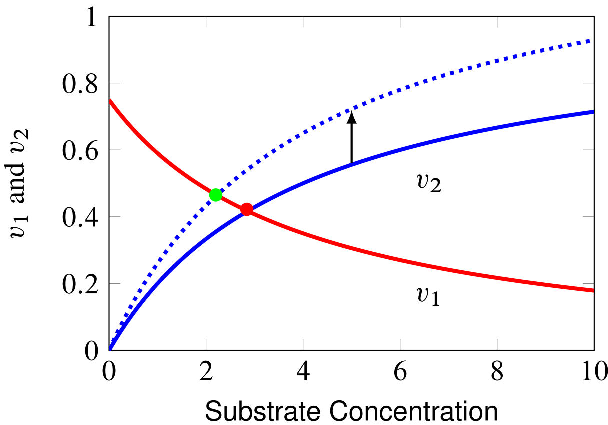


Figure 6.5 v_2 has been increased by 30% (dotted line) by increasing the enzyme concentration on v_2. This results in a displacement of the steady state curve to the left, leading to a decrease in the steady state concentration of S.

shifts right. This moves the intersection point to the right, indicating that the steady state concentration of S **increases** relative to the reference state.

Let us now change the activity of **both** E_1 and E_2 by 30% (Figure 6.7). Note that the curves for v_1 and v_2 are both scaled upwards, this moves the intersection point vertically upwards and therefore **doesn't** change the steady state concentration of S. This happens because both curves move vertically by the same fraction so that the intersection point can only move vertically.

This experiment highlights an important result, when all enzyme concentrations are increased by the same fraction, the flux increases by that same fraction but the species or metabolite levels remain **unchanged**. We can summarize this with the following statement:

If all E_i are increased by a factor α then the steady state change in J and S_i is:

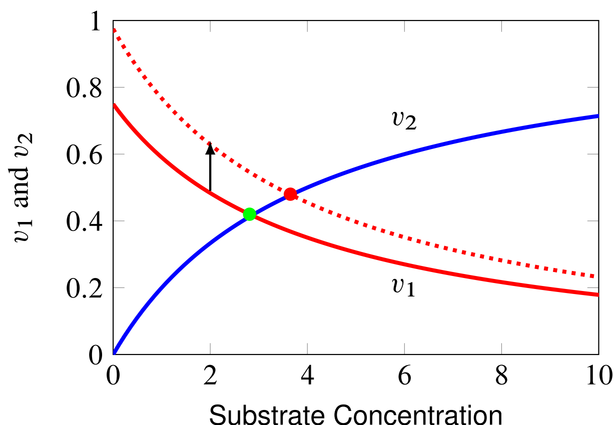


Figure 6.6 v_1 has been increased by 30% (dotted line) by increasing the enzyme concentration on v_1. This results in a displacement of the steady state curve to the right, leading to an increase in the steady state concentration of S.

$$\delta J = \alpha\ J \quad \text{and} \quad \delta S_j = 0$$

This is such an important result that it will be repeated again: increasing the activities of both enzymes by the **same fraction** will increase the flux through the pathway by the same fraction but will not change the concentration of the pathway species, S.

If we had a pathway with n steps and increased every enzyme concentration by the some factor then we would observe the same thing, the species concentrations would remain unchanged but the flux would increase. This observation is in fact true no matter how complex the pathway topology and doesn't just apply to linear chains of reaction steps.

Another way to understand why J increases by αJ is as follows. Since $\delta S = 0$, the only change that could possibly effect the flux is the change in enzyme concentration, since the enzyme concentration has increased by a given proportion (30%), the flux must also have increased by the same

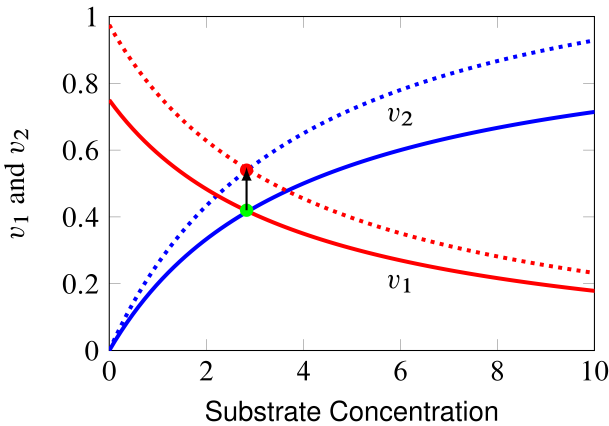
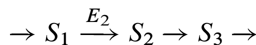


Figure 6.7 In this experiment, both E_1 and E_2 are increased by 30 % (dotted lines). Because both rates are increased by the same amount, the rate of change of S does not change. This means that there is no resulting change to the steady state concentration of S. The net flux through the pathway has however increased by 30 %.

proportion since the rate is proportional to the enzyme concentration (i.e $v_i \propto E_i$, hence $J \rightarrow \alpha J$.

Example 6.2

In the following pathway, an increase in E_2 by 20% results in a 5% increase in the steady state flux, estimate the value of $C^J_{E_2}$



$$\rightarrow S_1 \xrightarrow{E_2} S_2 \rightarrow S_3 \rightarrow$$

$C^J_{E_2}$ is the ratio of the fractional change in flux divided by the fraction change in the enzyme concentration, therefore an estimate for the control coefficient is given by:

$$C^J_{E_2} = \frac{0.05}{0.2} = 0.25$$

Prediction of Flux and Concentration Changes

We can also rearrange equations 6.1, 6.2 control coefficients into the following form:

$$\frac{dJ}{J} = C^J_{E_i} \frac{dE_i}{E_i}$$

$$\frac{dS_j}{S_j} = C^{S_j}_{E_i} \frac{dE_i}{E_i}$$

These simple relations allow us to compute the change in flux or concentration given a change in enzyme concentration. These relations only hold true if the changes in enzyme concentration are infinitesimal. For practical purposes the relationships will approximately hold provided the changes in E_i are small. Of more interest is if we make changes to multiple enzymatic steps, the overall change will be the sum of the individual changes. The technical reason for this is that small changes in E_i mean that the response is linear so that multiple responses can be summed to obtain the total response. In general, if we make changes to n reaction steps, then the overall change in flux and species concentrations is given by:

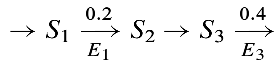
$$\frac{dJ}{J} = \sum_{i=1}^{n} C^J_{E_i} \frac{dE_i}{E_i} \quad (6.3)$$

$$\frac{dS}{S} = \sum_{i=1}^{n} C^S_{E_i} \frac{dE_i}{E_i} \quad (6.4)$$

Example 6.3

In the following pathway, the numbers refer to the flux control coefficients for the

respective reaction steps:



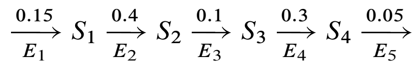
what is the percentage change in flux if we increase E_1 by 10% and E_3 by 20%?

To calculate this we use equation 6.3:

$$\frac{\delta J}{J} = 0.1\ 0.2 + 0.2\ 0.4 = 0.1 \text{ or } 10\%$$

Example 6.4

In the following pathway, the numbers refer to the flux control coefficients for the respective reaction steps:



Given no other information, if you could increase enzyme concentrations by 20% which two steps would you engineer to increase the flux the most?

Engineering the second and fourth enzymes will have the most effect on the steady state flux since they have the highest flux control coefficients. If we increased the second and fourth step by 20% the percentage change in flux will be:

$$\frac{\delta J}{J} = 0.2\ 0.4 + 0.2\ 0.3 = 0.14 \text{ or } 14\%$$

Summation Theorems

In this section we will introduce the concept of **operational proofs**. These proofs rely on carrying out thought experiments on a system and then casting the experiments in algebraic form from which theorems can be derived. Although perhaps not as rigorous as a purely algebraic approach, operational proofs offer insight in to the underlying biology and dynamics of the system and are therefore very useful exercises.

Consider the following two step pathway:

where X_o and X_1 are fixed. Let the pathway be at steady state and imagine increasing the concentration of enzyme, E_1 that catalyzes the first step by an amount, δE_1. The effect of this is to increase the steady state levels of S and flux, J. Let us now increase the level of E_2 by δE_2 **such that** the change in S is restored to the value it had in the original steady state. The net effect is that $\delta S = 0$. There are two ways to look at this thought experiment, from the perspective of the **system** and from the perspective of **local changes**. For the system we can compute the overall change in flux or species concentration by adding the two control coefficient terms, thus:

$$\begin{aligned} \frac{\delta J}{J} &= C^J_{E_1} \frac{\delta E_1}{E_1} + C^J_{E_2} \frac{\delta E_2}{E_2} \\ \frac{\delta S}{S} &= C^S_{E_1} \frac{\delta E_1}{E_1} + C^S_{E_2} \frac{\delta E_2}{E_2} = 0 \end{aligned} \tag{6.5}$$

We can also look at that is happening locally at every reaction step. Since the thought experiment guarantees that $\delta S = 0$ the local equations (See section 5.4) are quite simple:

$$\frac{\delta E_1}{E_1} = \frac{\delta v_1}{v_1}$$

$$\frac{\delta E_2}{E_2} = \frac{\delta v_2}{v_2}$$

Because the pathway is linear, at steady state, $v_1 = v_2 = J$. We can substitute these expressions into 6.5 therefore we can rewrite the system equations as:

$$\frac{\delta J}{J} = C^J_{E_1} \frac{\delta v_1}{v_1} + C^J_{E_2} \frac{\delta v_1}{v_1}$$

$$\frac{\delta S}{S} = C^S_{E_1} \frac{\delta v_1}{v_1} + C^S_{E_2} \frac{\delta v_1}{v_1} = 0$$

Since $\delta J/J = \delta v_1/v_1$, we can rewrite the above equations as:

$$\alpha = C^J_{E_1}\alpha + C^J_{E_2}\alpha$$

$$0 = C^S_{E_1}\alpha + C^S_{E_2}\alpha$$

where $\alpha = \delta J/J$. We then conclude that:

$$1 = C^J_{E_1} + C^J_{E_2}$$
$$0 = C^S_{E_1} + C^S_{E_2}$$

A number of assumptions were made in this derivation, these include the fact that $v_i \propto E_i$ and that changes in E_i has no effect on any other enzyme concentration. The first assumption is called the linearity assumption and the second the independence assumption. These assumptions can be relaxed if we describe the theorems using canonical control coefficients (See section at end of chapter).

To illustrate a more complex example, consider a branched pathway (Figure 6.8):

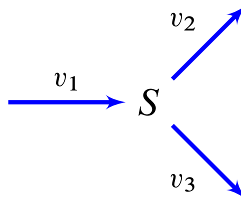


Figure 6.8 Simple Branched Pathway.

At steady state the following statement is true:

$$v_1 = v_2 + v_3$$

As before let us make a positive perturbation, δE_1 in the reaction step v_1. This will cause the steady state level of S and all reactions rates downstream to increase. Unlike the simple linear chain, there are multiple ways

to perturb E_2 and E_3 in order to restore S back to its original level in the branched pathway. Instead we can impose the constraint that what ever relative perturbation we make in v_1 we will make the same relative perturbation in v_2. Given this constraint what is the relative perturbation in v_3 that achieves, $\delta S = 0$?

To answer this question we must be aware that to satisfy the steady state condition the sum of the changes in δv_2 and δv_3 must equal the change δv_1, that is:

$$\delta v_1 = \delta v_2 + \delta v_3$$

We now divide both sides by v_1 and adjusting the denominators of v_2 and v_3 to obtain:

$$\frac{\delta v_1}{v_1} = \frac{\delta v_2}{v_2}\frac{v_2}{v_1} + \frac{\delta v_3}{v_3}\frac{v_3}{v_1}$$

Since we imposed the condition $\delta v_1 = \delta v_2$, we rewrite the above equation as:

$$\frac{\delta v_1}{v_1} - \frac{\delta v_1}{v_1}\frac{v_2}{v_1} = \frac{\delta v_3}{v_3}\frac{v_3}{v_1}$$

That is:

$$\frac{\delta v_1}{v_1}\left(1 - \frac{v_2}{v_1}\right) = \frac{\delta v_3}{v_3}\frac{v_3}{v_1}$$

v_2/v_1 is the fraction of flux going down the upper branch, which we will term, α. v_3/v_1 is the fraction of flux going down the lower arm, that is $1 - \alpha$, therefore:

$$\frac{\delta v_1}{v_1}(1 - \alpha) = \frac{\delta v_3}{v_3}(1 - \alpha)$$

$$\frac{\delta v_1}{v_1} = \frac{\delta v_3}{v_3}$$

This simple analysis shows that one way to achieve $\delta S = 0$ is for all fractional changes in δv_i to be equal, that is $\delta v_1/v_1 = \delta v_2/v_2 = \delta v_3/v_3$. As before, because $\delta S = 0$, it is also true that $\delta E_i/E_i = \delta v_i/v_i$ so that

(we assume that $v_i \approx E_i$):

$$\frac{\delta J}{J} = C^J_{E_1}\frac{\delta v_1}{v_1} + C^J_{E_2}\frac{\delta v_1}{v_1} + C^J_{E_3}\frac{\delta v_1}{v_1}$$

$$\frac{\delta S}{S} = C^S_{E_1}\frac{\delta v_1}{v_1} + C^S_{E_2}\frac{\delta v_1}{v_1} + C^S_{E_3}\frac{\delta v_1}{v_1} = 0$$

and therefore:

$$1 = C^J_{E_1} + C^J_{E_2} + C^J_{E_3}$$
$$0 = C^S_{E_1} + C^S_{E_2} + C^S_{E_3}$$

Equations 6.6 summarizes the two summation theorems.

$$\sum_{i=1}^{n} C^J_{E_i} = 1$$
$$\sum_{i=1}^{n} C^{S_j}_{E_i} = 0 \tag{6.6}$$

In both theorems, n is the number of reaction steps in the pathway. The flux summation theorem suggests that there is a finite amount of 'control' in the system and that control is shared between all steps. In addition, it states that if one step gains control then one or more other steps must lose control.

A more sophisticated analysis involving the stoichiometry matrix, implicit differentiation together with the application of linear algebra techniques reveals that the summation theorems apply to networks of arbitrary complexity. If the control coefficients are expressed in terms of changes in enzyme concentrations then there is the implicit assumption that $v_i \approx E_i$ and than changes in a particular E_i has no effect on other enzyme concentrations. Later on we will see that even these assumptions can be relaxed by using an alternative definition for the control coefficient.

Example

Consider the simple pathway shown in Figure 6.9. A simple mathematical model for this pathway is given in the Jarnac script 6.1.

```
p = defn cell
   J1: $Xo -> S1;
       Vm1/Km1*(Xo - S1/Keq1)/(1 + Xo/Km1 + S1/Km2);
   J2:  S1 -> S2;
       Vm2/Km2*(S1 - S2/Keq2)/(1 + S1/Km3 + S2/Km4);
   J3:  S2 -> $X1;
       Vm3*S2/(Km5 + S2);
end;

p.Xo = 2;  p.X1 = 0;
p.Keq1 = 1.2; p.Keq2 = 2.5;
p.Vm1 = 3.4; p.Vm2 = 8.2; p.Vm3 = 2.3;
p.Km1 = 0.6; p.Km2 = 0.78;
p.Km3 = 0.9; p.Km4 = 1.2;
p.Km5 = 0.5;
p.S1 = 0; p.S2 = 0;

p.ss.eval;
println " C1 = ", p.cc (<p.J1>, p.Vm1),
        " C2 = ", p.cc (<p.J1>, p.Vm2),
        " C3 = ", p.cc (<p.J1>, p.Vm3);
```

Listing 6.1 Three step pathway model

Table 6.1 shows the values for the flux control coefficients. We can see that flux control is distributed across all three reaction steps. Almost 50% of control is located on the last step. This show it is possible that in a linear pathway the committed step (i.e./ the firs step) is necessarily the step with the most control. By varying the values of the various kinetic parameters it is possible to obtain almost any pattern of control.

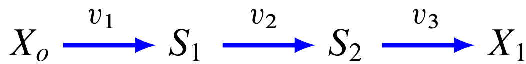


Figure 6.9 Three step linear pathway

$C^J_{Vm_1}$	$= 0.3677$
$C^J_{Vm_2}$	$= 0.1349$
$C^J_{Vm_2}$	$= 0.4989$

Table 6.1 Flux Control Coefficients

To summarize:

1. Control is shared throughout a pathway, that is the degree to which flux is limited in a pathway is shared. It is unlikely that only a single step in pathway limits flux.

2. If one step gains control, one of more other steps must loose control.

3. Control coefficients are system properties, they can only be computed or measured in the intact system. Inspection of a single enzymatic step will not reveal its degree of control (or influence) on the pathway.

Rate-limiting Steps

In much of the literature and many contemporary textbooks, one will often find a brief discussion of an idea called the rate-limiting step. In text books we will find statements such as:

"It is of course a truism to say that every metabolic pathway can and must have only one rate-liming step", Denton and Pogson, 1976.

"In order to exert control on the flux of metabolites through a metabolic pathway, it is necessary to regulate its rate-limiting step" Voet and Voet Biochemistry (p522), 3rd edition, 2004

These statements do not match the discussion given in the previous sections where control, or rate-limitingness was considered to be shared among all steps and not confined to just one. Other than a few books such as the 4th edition of Lehninger Biochemistry, most textbooks will have similarly expressed statements. The 4th edition of Lehninger states more correctly:

"Metabolic control analysis shows that control of the rate of metabolite flux through a pathway is distributed among several of the enzyme in that path".

There are also a variety of related terms such as rate-determining, pacemaker, bottleneck, master reaction or key enzyme. What is common is that authors will argue that a metabolic pathway will contain a single step that has overall influence over the pathways' flux. One of the earliest references to the concept of the rate-limiting step is a quote from Blackman [6]:

'When a process is conditioned as to its rapidity by a number of separate factors, the rate of the process is limited by the pace of the slowest factors."

This sentence started a century long love-affair with the idea of the rate-limiting step in biochemistry, a concept that has lasted to this very day. From the 1930s to the 1950s there were however a number of published papers which were highly critical of the concept, most notably Burton [12], Morales [60] and Hearon [31] in particular. Unfortunately much of this work did not find its way into the rapidly expanding fields of biochemistry and molecular biology after the second world war and instead the intuitive idea first pronounced by Blackman still remains today one of the corner stones in understanding cellular regulation. The concept drives much of metabolic engineering and drug targeting of metabolism. What is most surprising however is that a simple quantitative analysis shows that it cannot be true, and there is ample experimental evidence [32, 11] to support the alternative notion, that of shared control. The concept of the rate-limiting step is both inconsistent with logic and more importantly experimental ev-

idence.

The confusion over the existence of rate-limiting steps stems from a failure to realize that rates in cellular networks are governed by the law of mass-action, that is, if a concentration changes, then so does its rate of reaction. Some researchers try to draw analogies between cellular pathways and human experiences such as traffic congestion on roads or customer lines at shopping store checkouts. In each of these analogies, the rate of traffic and the rate of customer checkouts does not depend on how many cars are in the traffic line or how many customers are waiting. Such situations do warrant the use of the phrase rate-limiting step. Traffic congestion and the customer line are rate-limiting because the only way to increase the flow is to either widen the road or increase the number of cash tills, that is there is a single factor that determines the rate of flow. In reaction networks the flow is governed by many factors including substrate/product/effector concentrations as well as the capacity of the reaction (Vmax). Unless a reaction step is saturated (unlikely), it is possible to increase the flow by increasing the reactant concentration or decreasing the product concentration. In biological pathways, rate-limiting steps are therefore the exception rather than the rule. It is highly unlikely for a single reaction to be fully rate limiting because it can be influenced by many factors. Many hundreds of measurements of control coefficients have born out this prediction.

Most biochemistry and molecular biology literature interpret the rate-limiting step to be the single step in a pathway which limits the flux. In terms of our control coefficients we can interpret the rate-limiting step as the step with a flux control coefficient of unity. This means, by the summation theorem, that all other steps (at least in a linear chain) must have flux control coefficients of zero. However, when we consider branched and cyclic systems it is possible to have flux control coefficient much greater than one (other control coefficient must then be negative to satisfy the summation theorem). In these cases what adjective should we use, hyper-rate-limiting steps? In the final analysis, it is better to try and assign a value to the rate-limitingness of a particular step in a pathway rather than designate a given reaction step as either a rate-limiting step or not.

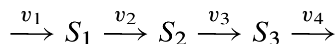
6.4 Connectivity Theorems

In the previous section, we saw that there are summation theorems relating the control coefficients. Here we will introduce an additional set of theorems that relate the control coefficients to the substrate, product and effector elasticities. There theorems are called the **connectivity theorems** and represent the most important result in metabolic control analysis. The theorem relates the control coefficients to the elasticities, that is it relates system wide properties to local properties.

In the derivation of the summation theorems, certain operations were performed on the pathway such that the flux changed value but the concentrations of the species were unchanged, thus $dJ/J \neq 0$ and $dS/S = 0$.

The constraints on the flux and concentration variables in the summation theorems suggest a complementary set of operations. That is can we perform one or more operations to the enzymes such that the opposite is true, $dJ/J = 0$ and $dS/S \neq 0$. The short answer is such a set of operations which preserve the flux but change the species concentrations leads to another set of theorems, called the **connectivity theorems**.

Consider the following pathway fragment:



$$\xrightarrow{v_1} S_1 \xrightarrow{v_2} S_2 \xrightarrow{v_3} S_3 \xrightarrow{v_4}$$

Let us make a change to the rate through v_2 by increasing the concentration of enzyme catalyzing E_2. Let us assume we increase E_2 by an amount, δE_2. This will result in a change in the steady state of the pathway. The concentration of S_2, S_3, and the flux through the pathway will rise and the concentration of S_1 will decrease because it is upstream of the disturbance.

The condition we now wish to impose is to make a second change to the pathway such that we restore the flux back to what it was before we made original change. Since the flux increased when we changed E_2 we need to decrease the flux and we can easily do this by decreasing one of the other enzyme levels. If we decrease the concentration of E_3 this will reduce the flux. Decreasing E_3 will also cause the concentration of S_2 to increase further. However, S_1 and S_3 will change in the opposite direction compared to when E_2 was increased.

In fact when E_3 is changed sufficiently so that the flux is restored to its original value, the concentrations of S_1 and S_3 will also be restored to their original values and it is only S_2 that will be different. This is true because the flux through v_1 is now the same as it was originally (since we've restored the flux), and coupled to the fact that E_1 has not been manipulated in any way must mean that the concentrations of S_1 and all species upstream of S_1 must be the same as they were before the modulations were made. The same arguments apply to S_3 and all species downstream of v_4.

We have thus accomplished the following: E_2 has been increased by δE_2, this results in a change δJ to the flux. We now **decrease** the concentration of E_3 such that the flux is restored to its original value. In the process, S_2 has changed by δS_2 and neither S_1 nor S_3 have changed. In fact *no other* species in the entire system has changed other than S_2.

The ability to perform such a manipulation is general and even if a particular species had many rates coming in and many rates leaving we would still be able to perform the necessary manipulations on all the adjacent enzymes such that only that species changed in concentration and the flux was unaltered.

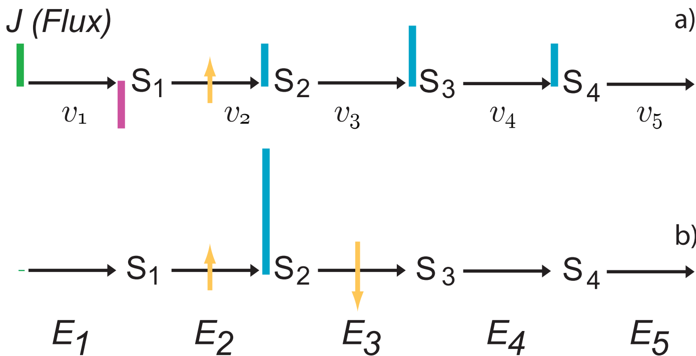


Figure 6.10 Connectivity Theorem: a) A change is made to E_2, this causes changes to species concentrations upstream and downstream including a change in flux. b) Another change is made to E_3 to oppose the change in flux, this results in only three net changes, a change in E_2, E_3 and S_2, no other changes occur.

Flux Connectivity Theorem

Considering the thought experiment described in the last section we can now write down two sets of equations which apply simultaneously to the pathway, a **local** equation and a **system** equation. The system equation will describe the effect of the enzyme changes on the flux. Since the net change in flux is zero and the fact that we only changed E_2 and E_3, we can write the change in the system flux using the following system equation:

$$\frac{\delta J}{J} = 0 = C^J_{E_2} \frac{\delta E_2}{E_2} + C^J_{E_3} \frac{\delta E_3}{E_3}$$

To determine the local equations we concentrate on what is happening at a particular reaction step. For example, as a result of making changes to E_2 and E_3, the change in rate at v_2 is given by

$$0 = \frac{\delta v_2}{v_2} = \frac{\delta E_2}{E_2} + \varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2}$$

and at v_3

$$0 = \frac{\delta v_3}{v_3} = \frac{\delta E_3}{E_3} + \varepsilon^{v_3}_{S_2} \frac{\delta S_2}{S_2}$$

Note that $\delta E_2/E_2$ will not necessarily equal $\delta E_3/E_3$. No other changes took place so that these are the only local equations to consider. We can rearrange the local equations so that:

$$0 = \frac{\delta E_2}{E_2} = -\varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2} \tag{6.7}$$

$$0 = \frac{\delta E_3}{E_3} = -\varepsilon^{v_3}_{S_2} \frac{\delta S_2}{S_2} \tag{6.8}$$

We can now insert $\delta E_2/E_2$ and $\delta E_3/E_3$ from the local equations into the system equations and obtain:

$$0 = \frac{\delta J}{J} = -\left(C^J_{E_2} \varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2} + C^J_{E_3} \varepsilon^{v_3}_{S_2} \frac{\delta S_2}{S_2} \right)$$

and therefore:

$$0 = \frac{\delta S_2}{S_2} \left(C^J_{E_2} \varepsilon^{v_2}_{S_2} + C^J_{E_3} \varepsilon^{v_3}_{S_2} \right)$$

Since we know that $\delta S_2 / S_2$ is not equal to zero it must be true that

$$0 = C^J_{E_2} \varepsilon^{v_2}_{S_2} + C^J_{E_3} \varepsilon^{v_3}_{S_2}$$

This derivation can be applied to a species that interacts with any number of steps. In general the number of the terms will equal the number of interactions a species makes. For example, in the pathway fragment in Figure 6.11.

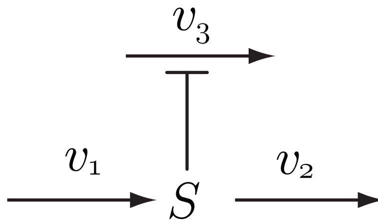


Figure 6.11 Pathway Fragment for Connectivity Theorem.

where S interacts with it production rate, v_1, a consumption rate, v_2, and an inhibitory interaction with v_3, the connectivity may be written as

$$C^J_{E_1} \varepsilon^{v_1}_S + C^J_{E_2} \varepsilon^{v_2}_S + C^J_{E_3} \varepsilon^{v_3}_S = 0$$

For a species, S that interacts with r other steps, the flux connectivity theorem is written as:

$$0 = \sum_{i=1}^{r} C^J_{E_i} \varepsilon^{v_i}_S$$

Concentration Connectivity Theorem

To derive the flux connectivity theorem we had to use the system equation that was related to the flux. It is however possible to use a different set

of systems equations, those with respect to the species concentrations. In the case of the species there will be two distinct systems equations. One of these will describe the effect that our modulation has on the common species (S_2 in the example), and a second describing the effect on any other species (S_1, S_3, etc.) in the pathway. Consider first the system equation involving the common species; for the pathway under consideration this equation is given by:

$$\frac{\delta S_2}{S_2} = C^{S_2}_{E_2} \frac{\delta E_2}{E_2} + C^{S_2}_{E_3} \frac{\delta E_3}{E_3}$$

We must remember that the change in the common species, $\delta S_2/S_2$, is non-zero (Figure 6.10). Therefore substituting in the local equations given previously (Equation 6.8) leads to:

$$\frac{\delta S_2}{S_2} = -C^{S_2}_{E_2} \varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2} - C^{S_2}_{E_3} \varepsilon^{v_3}_{S_2} \frac{\delta S_2}{S_2}$$

Since $\delta S_2/S_2 \neq 0$, we can cancel the term dS_2/S_2 which leads to the first concentration connectivity theorem:

$$-1 = C^{S_2}_{E_2} \varepsilon^{v_2}_{S_2} + C^{S_2}_{E_3} \varepsilon^{v_3}_{S_2}$$

A second theorem can be derived by considering the effect of our modulation on a distant species, for example S_3. In this case, the system equation now with respect to S_3, becomes:

$$0 = \frac{\delta S_3}{S_3} = C^{S_3}_{E_2} \frac{\delta E_2}{E_2} + C^{S_3}_{E_3} \frac{\delta E_3}{E_3}$$

Note that the equation equals zero because our operations ensure that species other than the common species do not change in concentration.

Substituting once again the local equations into the above system equation leads to:

$$0 = \frac{\delta S_3}{S_3} = -C^{S_3}_{E_2} \varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2} - C^{S_3}_{E_3} \varepsilon^{v_3}_{S_2} \frac{\delta S_2}{S_2}$$

or

$$0 = -\frac{\delta S_2}{S_2} \left(C^{S_3}_{E_2} \varepsilon^{v_2}_{S_2} + C^{S_3}_{E_3} \varepsilon^{v_3}_{S_2} \right)$$

However, we know that dS_2/S_2 is not zero, therefore it must be the case that:

$$0 = C^{S_3}_{E_2} \varepsilon^{v_2}_{S_2} + C^{S_3}_{E_3} \varepsilon^{v_3}_{S_2}$$

That completes the proof for the concentration connectivity theorems. As with the flux connectivity theorems, the concentration connectivity theorems can be generalized to any number of steps that a species might interact with.

To summarize, the connectivity theorems are:

Flux Connectivity Theorem with respect to a common metabolite, S_k where r is the number of interactions it makes with neighboring reaction steps.

$$\sum_{i=1}^{r} C^{J}_{E_i} \varepsilon^{v_i}_{S_k} = 0 \tag{6.9}$$

Concentration Connectivity Theorem with respect to the common metabolite S_k where r is the number of interactions it makes with neighboring reaction steps.

$$\sum_{i=1}^{r} C^{S_k}_{E_i} \varepsilon^{v_i}_{S_k} = -1 \tag{6.10}$$

Concentration Connectivity Theorem with respect to the common metabolite S_k and a distant metabolite, S_m where r is the number of interactions it makes with neighboring reaction steps.

$$\sum_{i=1}^{r} C^{S_m}_{E_i} \varepsilon^{v_i}_{S_k} = 0$$

Interpretation The connectivity theorems are important for a number of reasons. The first and foremost is that the theorems link local effects in terms of the elasticities to global effects in terms of the control coefficients. Consider for example the following linear pathway.

$$\xrightarrow{v_1} S_1 \xrightarrow{v_2}$$

The flux connectivity can be written in the form:

$$\frac{C^J_{E_1}}{C^J_{E_2}} = -\frac{\varepsilon^{v_2}_{S_1}}{\varepsilon^{v_1}_{S_1}}$$

That is the ratio of two adjacent flux control coefficients is inversely proportional to the ratio of the corresponding elasticities. This means that high flux control coefficients tend to be associated with small elasticities and small flux control coefficients with large elasticities. This is explained in terms of changes to species opposing changes in rates by species moving in a direction opposite to the rate change. Since species with high elasticities are able to oppose rate changes more effectively than small elasticities then it follows that large elasticities are associated with small flux control coefficients and *vice versa*.

The classic example of this is the case of a reaction operating near equilibrium where the elasticities are very high relative to adjacent elasticities on neighboring enzymes. In such situations the flux control coefficients of near equilibrium enzymes are *likely* to be small. However, one must bear in mind that it is the ratio of elasticities which is important and not their absolute values. Simply examining the elasticity of a single reaction may lead to an incorrect conclusion. Even more so, one must also consider all the ratios of the elasticities along a pathway because even though one elasticity ratio may suggest a high or low flux control coefficient on a

particular enzyme, it is a consideration of the other ratios coupled to the flux summation theorem that will give an absolute value to a particular flux control coefficient. Control coefficients are truly system wide properties.

The examination of a single enzyme will not give an absolute indication of the ability of that enzyme to control the flux or species concentrations.

6.5 Response Coefficients

Control coefficients measure the response of a pathway to changes in enzyme activities. What about the effect of external factors such as inhibitors, pharmaceutical drugs or boundary species? Such effects are measured by another coefficient called the **response coefficient**. The flux response coefficient is defined by:

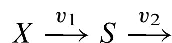
$$R_X^J = \frac{dJ}{dX}\frac{X}{J}$$

and the concentration response coefficient by:

$$R_X^S = \frac{dS}{dX}\frac{X}{S}$$

The response coefficient measures how sensitive a pathway is to changes in external factors other than enzyme activities. What is the relationship of the response coefficients with respect to the control coefficients and elasticities?

Like many of the proofs in this chapter we can carry out a thought experiment to investigate the response coefficients more closely. Consider the pathway fragment below:



$$X \xrightarrow{v_1} S \xrightarrow{v_2}$$

where X is the fixed boundary species. Let us increase the activity of v_1 by increasing the concentration of E_1. This will cause the steady state flux

and concentration of S and in fact all downstream species beyond v_2 to increase. Let us now decrease the concentration of X such that we restore the flux and steady state concentration of S back to its original value. From this thought experiment we can write the operations in terms of the local response equation and a system response equation as follows:

$$\frac{\delta v_1}{v_1} = \varepsilon^{v_1}_X \frac{\delta X}{X} + \varepsilon^{v_1}_{E_1} \frac{\delta E_1}{E_1} = 0$$

$$\frac{\delta J}{J} = R^J_X \frac{\delta X}{X} + C^J_{E_1} \frac{\delta E_1}{E_1} = 0$$

Note that the right-hand sides are zero because the thought experiment guarantees that the flux has not changed due to these operations. We can eliminate the $\delta E_1/E_1$ term in the system response equation by substituting the term from the local response equation. In addition if we assume that the reaction rate for an enzyme catalyzed reaction is proportional to the enzyme concentration, then we know that $\varepsilon^{v_1}_{E_1} = 1$.

Proof that $\varepsilon^v_E = 1$. If

$$v = \frac{E k_{cat} S}{S + K_m} = E\ f(S)$$

then

$$\varepsilon^v_E = \frac{\partial v}{\partial E}\frac{E}{v} = f(S)\frac{E}{E f(s)} = 1$$

Therefore:

$$0 = R^J_X \frac{\delta X}{X} - C^J_{E_1} \varepsilon^{v_1}_X \frac{\delta X}{X}$$

Since $\delta X/X \neq 0$ we cancel $\delta X/X$ yielding:

$$R^J_X = C^J_{E_1} \varepsilon^{v_1}_X$$

This gives use the relationship we seek. It can be generalized for multiple ex-

ternal factors acting simultaneously by summing up individual responses:

$$R^J_X = \sum_{i=1}^{n} C^J_{E_i} \varepsilon^{v_i}_X$$

Likewise the response of a species, S to an external factor is given by:

$$R^S_X = \sum_{i=1}^{n} C^S_{E_i} \varepsilon^{v_i}_X$$

The response coefficient carries an important message, which is that the response of some external factor, X, is a function of two things, the effect the factor has on the step it acts upon and the effect that the step itself has on changing the system. This means than an effective external factor, such as a pharmaceutical drug, must not only be able to bind and inhibit the enzyme being targeted, but the step itself must be able to transmit the effect to the rest of the pathway and ultimately affect the phenotype.

The ability of an external factor to influence a given species or flux depends on:

1. The ability of the external factor to influence its immediate target.

2. The ability of the target to influence the network it is connected to.

6.6 Canonical Control Coefficients

Let us write the response coefficient equation in a different way. Consider:

$$C^J_{E_i} = \frac{R^J_{E_i}}{\varepsilon^{v_i}_{E_i}}$$

Expand the terms and replace E_i with a general parameter p to yield:

$$C^J_{v_i} = \frac{\frac{dJ}{dp}\frac{p}{J}}{\frac{\partial v_i}{\partial p}\frac{p}{v_i}} = \frac{dJ}{dv_i}\frac{v_i}{J}$$

What we have here is a **parameterless control coefficient**, $C^J_{v_i}$, also called a canonical control coefficient. It describes the effect of a change in the reaction rate on the steady state level of the pathway flux. This may seem like an odd definition because isn't the change in reaction rate the same as the change in steady state pathway flux? In this case not so. We have to be clear what the derivative, dv_i, in the denominator actually means. Operationally the change indicated by dv_i refers to a change in v_i by some unspecified means under the conditions where the reactants, products and any other effectors remain constant. dv_i in this context is also sometimes referred to as the local rate, i.e the change in the reaction rate we could impose if the reaction were not connected to the rest of the network. The way dv_i is changed is unspecified but it must be done via a parameter of the system, not by a variable quantity such as one of the reactants or products. The identity of the parameters will depend on the constraints put on the system but common parameters are the concentration of expressed enzyme, the catalytic constant of the enzyme or an external inhibitor.

For example, let us suppose that the change we make to alter v_i is via a change in the enzyme concentration, E_i. Such a change will cause an immediate change dv_i in the reaction rate. We now let the system go to its new steady state. The dv_i will cause changes in the immediate environment of the reaction, causing the enzyme's substrate to decrease and its product to increase. These changes in turn will propagate throughout the system. Once the system has settled to the new steady state an inspection of v_i will reveal that the final change in rate does not equal the original dv_i (because the local environment has now changed). We refer to the final change in v_i as dJ, that is the change in flux through the system. Taking the ratio of dJ and dv_i and scaling we obtain the canonical control coefficient. In many situations, the enzyme elasticity, $\varepsilon^{v_i}_{E_i} = 1$, that is:

$$C^J_{E_i} = C^J_{v_i}$$

In other words the control coefficient we measure, C^J_E is identical to the canonical control coefficient. Strictly speaking the summation and connectivity theorems apply to the canonical control coefficients but because $\varepsilon^{v_i}_{E_i} = 1$ we can often safely express the theorems using the control coefficients with respect to enzyme concentration.

6.7 Computing Control Equations

In previous sections we have seen how the control coefficients can give useful information on the sensitivity of a network to parameter changes. In addition we have seen that relationships exist between the control coefficients and the elasticities. In this section we will look at ways to express the control coefficients in terms of the elasticities of which there are a number. The simplest way to derive the control equations is to combine the summation and connectivity theorems. For example, consider a two step pathway such as:

where X_o and X_1 are fixed species. There is one flux connectivity theorem with respect to every species in a pathway so that in the above example there will only be one connectivity theorem centered around S:

$$C^J_{E_1}\varepsilon^{v_1}_S + C^J_{E_2}\varepsilon^{v_2}_S = 0$$

In addition, there will be a flux summation theorem:

$$C^J_{E_1} + C^J_{E_2} = 1$$

These two equations can be combined to give expressions that relate the control coefficients in terms of the elasticities, thus:

$$C^J_{E_1} = \frac{\varepsilon^2_S}{\varepsilon^2_S - \varepsilon^1_S}$$

$$C^J_{E_2} = -\frac{\varepsilon^1_S}{\varepsilon^2_S - \varepsilon^1_S}$$

These equations, possibly the most important result of the theory, allows us to understand how system responses depend on local properties. This topic will covered in the next chapter.

For more complex pathways such as branches and moiety conserved cycles, additional theorems are required to solve the equations. There is however an alterative approach to deriving the control equations based on a purely algebraic method. We will describe this in the next section.

Pure Algebraic Method

The pure algebraic method relies on the use of implicit differentiation of the system equation. Consider the simplest two step pathway already considered:

$$X_o \xrightarrow{v_1} S \xrightarrow{v_2} X_1$$

At steady state the rate of change of S is given by:

$$\frac{dS}{dt} = v_1 - v_2 = 0$$

Assuming that v_1 can be changed by perturbations to the concentration of catalyzing enzyme, E_1, we can write the rate of change as:

$$\frac{dS}{dt} = v_1(S(E_1), E_1) - v_2(S(E_1)) = 0$$

Here we can stated explicitly how each reaction rate, v_i is a function of both the steady state species concentration and the perturbing parameter, E_1. Note that v_1 is both a function of S and E_1 and S in turn is a function of E_1. v_2 is only a function of S and not directly a function of E_1. We can implicitly differentiate this equation to yield:

$$0 = \frac{\partial v_1}{\partial S}\frac{dS}{dE_1} + \frac{\partial v_1}{\partial E_1} - \frac{\partial v_2}{\partial S}\frac{dS}{dE_1}$$

We can scale each of the derivatives by multiplying by the appropriates factors, that is:

$$0 = \frac{\partial v_1}{\partial S}\frac{S}{v_1}\frac{dS}{dE_1}\frac{E_1}{S} + \frac{\partial v_1}{\partial E_1}\frac{E_1}{v_1} - \frac{\partial v_2}{\partial S}\frac{S}{v_1}\frac{dS}{dE_1}\frac{E_1}{S}$$

which can be simplified to:

$$0 = C^S_{E_1} \varepsilon^1_S + \varepsilon^1_{E_1} - C^S_{E_1} \varepsilon^2_S$$

Solving for $C^S_{E_1}$ and assuming v_1 is first-order with respect to E_1 so that $\varepsilon^1_{E_1} = 1$ yields:

$$C^S_{E_1} = \frac{-1}{\varepsilon^1_S - \varepsilon^2_S}$$

We can derive $C^S_{E_2}$ in the same way by implicitly differentiating:

$$\frac{dS}{dt} = v_1(S(E_2)) - v_2(S(E_2), E_2) = 0$$

The flux control coefficients can be computed in a similar way. For example to find $C^J_{E_1}$ we can implicitly differentiate:

$$J = v_1(S(E_1), E_1)$$

$$\frac{dJ}{dE_1} = \frac{\partial v_1}{\partial S} \frac{dS}{dE_1} + \frac{\partial v_1}{\partial E_1}$$

Scaling yields:

$$C^J_{E_1} = C^S_{E_1} \varepsilon^1_S + 1$$

Substituting $C^S_{E_1}$ gives:

$$C^J_{E_1} = -\frac{1}{\varepsilon^1_S - \varepsilon^2_S} \varepsilon^1_S + 1 = -\frac{\varepsilon^1_S}{\varepsilon^1_S - \varepsilon^2_S}$$

6.8 Regulatory Coefficients

One of the questions that comes up sometimes is the distinction between the phrase 'control' and 'regulation' [37, 69, 38]. In this book control is used to indicate how much influence a given reaction step in a network has on the system. This allows us to quantify the degree of control in a system and gives us an operational definition of the word. It is therefore clear

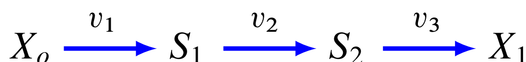


Figure 6.12 Three step linear pathway

what is meant. The word regulation however doesn't have such a clear interpretation. One way to think of regulation is that it describes how a given system achieves control. For example, if a reaction step has a flux control coefficient of 0.4 we can ask how that level of control is achieved. However this question a little vague and therefore difficult to quantify. Another way to look at regulation is to investigate how the internal effectors change in response to an external perturbation since it is the change in the internal state that ultimately decides the final response. Let us consider the following thought experiment on the three step pathway shown in Figure **??**.

Let us consider a positive perturbation in the enzyme activity on the third step, that is δE_3. This will cause S_2 and S_1 to fall and the steady state flux to rise. If we focus on the second steps, v_2, we can see that the change in flux at this step was caused by changes in both S_1 and S_2. Each change contributed to the final change in flux at this step. What we can do is quantify this contribution. To do this we first write down the change in flux at v_2 by considering the local changes:

$$\frac{\delta J}{J} = \varepsilon^{v_2}_{S_1} \frac{\delta S_1}{S_1} + \varepsilon^{v_2}_{S_2} \frac{\delta S_2}{S_2}$$

Each term on the right indicates the contribution the term makes to the overall flux change. We will call this the partition equation. The change in the concentrations S_1 and S_2 can be computed from a knowledge of the concentration control coefficients, that is:

$$\frac{\delta S_1}{S_1} = C^{S_1}_{E_3} \frac{\delta E_3}{E_3} \qquad \frac{\delta S_2}{S_2} = C^{S_2}_{E_3} \frac{\delta E_3}{E_3}$$

We can now substitute these terms into the partition equation to obtain:

$$\frac{\delta J}{J} = \varepsilon^{v_2}_{S_1} C^{S_1}_{E_3} \frac{\delta E_3}{E_3} + \varepsilon^{v_2}_{S_2} C^{S_2}_{E_3} \frac{\delta E_3}{E_3}$$

Dividing both sides by $\delta E_3/E_3$, we obtain:

$$C^J_{E_3} = \varepsilon^{v_2}_{S_1} C^{S_1}_{E_3} + \varepsilon^{v_2}_{S_2} C^{S_2}_{E_3}$$

Finally dividing both sides by $C^J_{E_3}$ we get:

$$1 = \varepsilon^{v_2}_{S_1} \frac{C^{S_1}_{E_3}}{C^J_{E_3}} + \varepsilon^{v_2}_{S_2} \frac{C^{S_2}_{E_3}}{C^J_{E_3}}$$

$$^{v_3}p^{v_2}_{S_1} = \varepsilon^{v_2}_{S_1} \frac{C^{S_1}_{E_3}}{C^J_{E_3}} \qquad ^{v_3}p^{v_2}_{S_2} = \varepsilon^{v_2}_{S_2} \frac{C^{S_2}_{E_3}}{C^J_{E_3}}$$

We will call the two terms, $^{v_3}p^{v_2}_{S_1}$ and $^{v_3}p^{v_2}_{S_2}$ on the right-hand side of the equation the partitioned regulatory coefficients. These terms quantify all the factors that contribute to a response around a given reaction step as a result of a perturbation. In this case there are two partitioned coefficients, one for S_1 and a second for S_2. The first quantifies how much S_1 contributes to the net change in v_2 and the second term how much S_2 contributes, both relative to a perturbation in E_3. By way of example consider the simulation shown in listing 6.1. We can compute the partitioned coefficients for this system using the Jarnac commands:

```
p1 = p.ee (<p.J2>, p.S1) * p.cc(<p.S1>, p.Vm3) / p.cc(<p.J1>, p.Vm3);
p2 = p.ee (<p.J2>, p.S2) * p.cc(<p.S2>, p.Vm3) / p.cc(<p.J1>, p.Vm3);
println p1, p2, p1+p2;
```

```
p1 = -2.7260 p2 = 3.7259 sum = 0.9999
```

$$^{v_3}p^{v_2}_{S_1} = -2.7260 \quad ^{v_3}p^{v_2}_{S_2} = 3.7259$$

What do these numbers means? The first thing to note is that a positive number means that the effector in question contributes by increasing the reaction rate through the step. Conversely, a negative term means that the effector contributes to decreasing the reaction rate. The magnitude of the terms express the relative contribution of each. For example, of the total change that occurred in v_2 due to the original perturbation, 57% originated from S_2 and the remaining, 42% from S_1. S_2 therefore contributed the

most and contributed to the overall positive change in reaction rate. This highlights the significant contribution made by product inhibition during a perturbation, an effect often ignored. Without product inhibition there would be no effect on the reaction rate. It is interesting to note that because the summation is to one, product inhibition must always necessarily contribute more in these simple systems.

The flux control coefficient for the flux through v_3 with respect to E_3 is 0.498. This is the amount of control that E_3 has on the flux. This degree of control is mediated at v_2 by changes in S_1 and S_2 and the partitioned coefficients measure the degree that each effector at v_2 mediates this control. ${}^{v_3}p^{v_2}_{S_1}$ and ${}^{v_3}p^{v_2}_{S_2}$ measure the amount of regulation that occurs at v_2 and the relative contribution of each in order to achieve a flux control coefficient of 0.498.

For the first step, v_1, S_1 contributes 100% to the regulatory response at v_1 due to the perturbation at v_3. This is expected given that S_1 is the only effector at v_1. For v_3 we have the additional effector in the perturbation itself in E_3. The partition equation is given by:

$$1 = \frac{\varepsilon^{v_3}_{E_3}}{C^J_{E_3}} + \varepsilon^{v_3}_{S_2} \frac{C^{S_2}_{E_3}}{C^J_{E_3}}$$

$$1 =^{v_3} p^{v_3}_{E_3} +^{v_3} p^{v_3}_{S_2}$$

We can see from this that even though there is only on internal regulator, S_2, we must also consider the contribution from the external perturbation in E_3. Running a simulation on the previous three step model reveals the values for the partition coefficients:

$${}^{v_3}p^{v_3}_{E_3} = 2.0 \qquad {}^{v_3}p^{v_3}_{S_2} = -1$$

From this we can see that E_3 is twice as important at regulating the flux at v_3 as is S_2.

In Chapter 10 we will look at negative feedback in more details and consider again the use of regulation coefficients in understanding regulation.

Further Reading

1. Rafael Moreno-Sanchez, Emma Saavedra, Sara Rodrguez-Enriquez, and Viridiana Olin-Sandoval (2008) Metabolic Control Analysis: A Tool for Designing Strategies to Manipulate Metabolic Pathways, Journal of Biomedicine and Biotechnology, Volume 2008, Article ID 597913, doi:10.1155/2008/597913

Exercises

1. At steady state all reactions rate are equal in a linear chain of reactions. Explain this statement.

2. The control coefficients are defined in terms of infinitesimal relative changes. An alternative would be to define them using large finite changes, that is $\Delta J/\Delta E$ which could be more easily measured. What is the main disadvantage to defining control coefficients in terms of large finite changes?

3. List three properties of control coefficients.

4. In a given reaction step E_i, the enzyme concentration is increased by 15%. The steady state change in flux was found to be 5% and the change in a species, S_j changed by -3%. Estimate the values for the flux control coefficient, $C^J_{E_i}$ and the concentration control coefficient, $C^{S_j}_{E_i}$.

5. In last question you were asked to find estimates for the control coefficients. Why were you ask to estimate the control coefficients and not their precise values?

6. A given reaction step has a flux control coefficient of 0.6. If the enzyme concentration is increased by 40% what is the approximate change in the steady state flux?

7. Two reactions have flux control coefficients of 0.2 and 0.3 respectively. The concentration of the first enzyme is changed by 10% and

the second enzyme by 30%, What is the approximate change in the steady state flux if both changes are made?

8. What assumption(s) are made in the derivation of the summation theorems?

9. In a linear pathway, the concentration control coefficient for a give species, S is found to be negative with respect to one enzyme but positive with respect to every other enzymatic step. Where is the species, S located in the pathway, explain your answer.

10. The last four steps in a five step pathway are found to have concentration control coefficients for a species, S, of -0.1, -0.2, -0.5 and -0.05. What is the concentration control coefficient with respect top the first step?

11. Locate five biochemistry and molecular cell biology text books and describe how the books describe regulation in pathway with respect to control of flux. If they mention rate-limiting steps or rate-determining steps, describe how they justify these statements if at all.

12. Why are rate-limiting steps unlikely to be found in natural pathways?

13. A given species S that has a single production step and a single consumption step, the elasticity of the production step with respect to S was found to be -1.6 and for the consumption step 0.12. a) Explain why the production step elasticity is negative. b) From the information what can you say about the flux control coefficients of the production and consumption steps?

14. Explain why examination of a single enzyme in a pathway will not necessarily give a good indication of how flux limiting the enzyme is.

15. The response coefficient relationship has two important lessons for those looking to develop new therapeutic drugs, what are they?

16. What is a canonical control coefficient?

17. Given the model described by the Jarnac script 6.2, ignore everything in the script after line 14, find the flux control coefficient for each step and confirm numerically that the flux summation theorem holds. Use perturbations to estimate the flux control coefficients.

6.A Jarnac Scripts

```
p = defn cell
   J0: $Xo -> S1; (E1/0.3)*(0.5*Xo-100*S1)/(1+Xo+S1);
   J1: S1 -> S2; E2*(3*S1-0.2*S2)/(1+S1+S2);
   J2: S2 -> S3; E3*(500*S2-10*S3)/(1+S2+S3);
   J3: S3 -> S4; E4*(200*S3-2*S4)/(1+S3+S4);
   J4: S4 -> $X1; E5*(200*S4-2*X1)/(1+S4+X1);
end;

p.Xo = 10;  p.X1 = 0;
p.E1 = 3.4; p.E2 = 8.2;
p.E3 = 2.3; p.E4 = 1.8;
p.E5 = 4.5;
p.S1 = 0; p.S2 = 0; p.S3 = 0; p.S4 = 0;

m1 = p.sim.eval (0, 0.2, 100, [<p.time>, <p.J0>]);
setColumnName (m1, 1, "E2");
p.E2 = p.E2*4;
m2 = p.sim.eval (0.2, 0.4, 200, [<p.time>, <p.J0>]);
alldata = augr(m1, m2);

setAxes ({0, 0.4, 0, 5});
graph (alldata);
str = exportpgfplot (alldata);
copyToClipboard (str);
```

Listing 6.2 Script for Figure 6.2

```
p = defn cell
   J0: $Xo -> S1; E1*(10*Xo-2*S1)/(1+Xo+S1);
   J1:  S1 -> S2; E2*(10*S1-2*S2)/(1+S1+S2);
   J2:  S2 -> S3; E3*(10*S2-2*S3)/(1+S2+S3);
   J3:  S3 -> S4; E4*(10*S3-2*S4)/(1+S3+S4);
   J34: S4 -> $X1; E5*(10*S4-2*X1)/(1+S4+X1);
end;

p.Xo = 10;  p.X1 = 0;
p.E1 = 3.4; p.E2 = 8.2;
```

```
p.E3 = 2.3; p.E4 = 1.8;
p.E5 = 4.5;
p.S1 = 8.359; p.S2 = 17.68; p.S3 = 6.938; p.S4 = 0.4816;
savedState = p.pv;

//  -------------- E1 --------------
p.sim.eval (0, 80, 500);
m1 = p.sim.eval (0, 10, 100, [<p.Time>, <p.J0>]);
p.E1 = p.E1*1.2;
m2 = p.sim.eval (10, 30, 100, [<p.Time>, <p.J0>]);
alldata = augr (m1, m2); // Combine the two segments
setColumnName (alldata, 2, "E1");

//  -------------- E2 --------------
p.pv = savedState; p.sim.eval (0, 80, 500);
m1 = p.sim.eval (0, 10, 100, [<p.J0>]);
setColumnName (m1, 1, "E2");
p.E2 = p.E2*1.2;
m2 = p.sim.eval (10, 30, 100, [<p.J0>]);
alldata = aug (alldata, augr(m1, m2));

//  -------------- E3 --------------
p.pv = savedState; p.sim.eval (0, 80, 500);
m1 = p.sim.eval (0, 10, 100, [<p.J0>]);
setColumnName (m1, 1, "E3");
p.E3 = p.E3*1.2;
m2 = p.sim.eval (10, 30, 100, [<p.J0>]);
alldata = aug (alldata, augr (m1, m2));

//  -------------- E4 --------------
p.pv = savedState; p.sim.eval (0, 80, 500);
m1 = p.sim.eval (0, 10, 100, [<p.J0>]);
setColumnName (m1, 1, "E4");
p.E4 = p.E4 *1.2;
m2 = p.sim.eval (10, 30, 100, [<p.J0>]);
alldata = aug (alldata, augr (m1, m2));

//  -------------- E5 --------------
p.pv = savedState; p.sim.eval (0, 80, 500);
m1 = p.sim.eval (0, 10, 100, [<p.J0>]);
setColumnName (m1, 1, "E5");
```

```
p.E5 = p.E5 *1.2;
m2 = p.sim.eval (10, 30, 100, [<p.J0>]);
alldata = aug (alldata, augr (m1, m2));
p.pv = savedState;

setAxes ({0, 30, 14, 18});
graph (alldata);
str = exportpgfplot (alldata);
copyToClipboard (str);
```

Listing 6.3 Script for Figure 6.3

7

Understanding Metabolism

7.1 Introduction

This book is about the control of biochemical systems with a focus on metabolic pathways. The ability to control reaction rates and concentrations in a changing environment is one of the characteristics of living systems. Cells must monitor prevailing conditions and make appropriate decisions. Cells make sure, for example, that adequate phosphate and redox potentials are available at all times. They also have to ensure that major transitions from one state to another (for example cell division) avoid any disruption to subsystems that are essential to cell viability. These activities presumably require a great deal of coordination and control and indeed over sixty years of research has uncovered a myriad number of feedback and feedfoward control loops together with many less obvious means of control.

It is worth examining some of the history of how we came to understand control in biological cells. The first thing to note is that understanding control in any complex system is difficult. It was difficult in the past and it is difficult now. Man's propensity to grasp the many factors involved in a complex system is limited. As a result, reasoning about complex systems cannot be done by intuition alone but requires expertise and the application

of approaches from mathematics, engineering and computer science.

7.2 Early Quantitative Efforts

During the early part of the 20th century it became apparent that chemical processes in biological cells were a result of sequences of separate chemical transformations. The first such sequence of steps discovered, later to be called a 'pathway', was yeast glycolysis. Subsequently, many other pathways were discovered including the Calvin and Krebs cycle and the many pathways involved in amino acid biosynthesis and degradation. As early as the 1930s, various individuals began taking a theoretical interest in the dynamic properties of such pathways. Much of the early work focused on the question of limiting factors. This may have originated from a statement by Blackmam [6] in 1905 who stated as an axiom: "when a process is conditioned as to its rapidity by a number of separate factors, the rate of the process is limited by the pace of the slowest factor". This implied that the understanding of a complex system could be accomplished by identifying the limiting factor; and so was born the idea of the rate-limiting step, the pacemaker, the bottleneck, or master reaction.

The Pacemaker

Although the idea of a pacemaker reaction in a pathway was extremely attractive, there were opponents to the idea even as early as the 1930s. Burton [12] was probably one of the first to point out that: "In the steady state of reaction chains the principle of the master reaction has no application". Hearon [31] made a more general mathematical analysis and developed strict rules for the prediction of mastery in a linear sequence of enzyme-catalysed reaction. Webb [87] gave a severe criticism of the concept of the pacemaker and of its blind application to solving problems of regulation in metabolism. Waley [85] made a simple but clear analysis of simple linear chains that showed that rate-limitingness was a shared commodity in a chain of reactions. Later authors from the biochemical community, such as Higgins [35] but particularly Heinrich and Rapoport [33] supported the same conclusion with more advanced analysis. In parallel with this work

other communities were coming to the same conclusion. Most notably Sewell Wright, a geneticist, wrote a treatise on 'Physiological and Evolutionary Theories of Dominance' [90] where he discussed the limiting factors in relation to hypothesized networks controlled by 'genes'. This work was taken up by Kacser and Burns [46] in Edinburgh and was developed in to a major theory of control in pathways. Heinrich and Rapoport [33] simultaneously accomplished the same feat but from a more biochemical perspective. Finally Savageau [75] in the United States, an engineer by training, developed the same approach and reached similar conclusions.

7.3 Prevailing Ideas

Nevertheless, although there was considerable theoretical and some experimental work that suggested that the concept of the pacemaker was erroneous, the biochemical community, for what ever reason, ignored these results. Instead the biochemistry community, which had largely morphed into molecular biology in the 1970s developed its own framework for understanding the operating principles of cellular networks. This framework was derived largely through an intuitive approach, based neither on experimental evidence or mathematical reasoning. This ultimately led to a number of unfortunate miss-understandings in how cellular networks operate, misunderstandings that still prevail today.

One of the chief concepts in the traditional control framework is the pacemaker or rate-limiting step. The rate-limiting step is thought to be located near the start of a pathway and because it is rate-limiting, the pathway is controlled by this one key step. In addition, it is proposed that rate-limiting steps are likely to be the site for allosteric regulation. There were a number of criteria that are used to identify possible rate-limiting step though there was no real definitive test. These criteria included:

- The rate-limiting step is the slowest step in the pathway.
- The rate-limiting step has the lowest substrate-affinity (highest K_m), this means that the reaction velocity is the lowest when saturating substrate concentrations are present at all enzymes.

- The rate-limiting step will be the regulated step.
- The rate-limiting step is an irreversible reaction.
- The rate-limiting step is usually the first step in the pathway.
- The rate-limiting step is far from equilibrium.

No single criterion could positively identify a rate limiting step but the cross-over theorem is one that was considered important. The technique worked as follows. A metabolic pathway is perturbed by adding an inhibitor of one of the enzyme catalyzed steps and the metabolite concentrations before and after the inhibited step are measured. If the inhibited step is rate limiting then those metabolites upstream would increase and those downstream decrease. The technique was originally developed by Britton Chance [13] in the 1950s as a means to study the electron transport chain in mitochondria. The advantage here was that many of the intermediates had characteristic absorption spectra. The method was used to identify the sites where electron transfer was being coupled to ATP production. Although applicable to the electron transport chain (Fell, 1996), its subsequent use to identify sites of regulation in metabolic pathways has been considered on theoretical grounds to be untrustworthy (Heinrich et al, 1974).

Regulatory Enzymes and Feedback Regulation

A key concept in traditional metabolic control theory is that feedback regulation by an end product will necessarily act on the rate-limiting step. Control may also be exerted by inducing or repressing the synthesis of the rate-limiting enzyme. Rate-limiting enzymes could therefore be identified simply by locating regulated steps. For example, a classic rate limiting step in glycolysis was phosphofructokinase since it was regulated by many effectors. Such an assertion makes perfect sense if metabolism it seen as a series of connected pipes and tanks with valves[1] that turn on and off the flow of water. However metabolism is not like this, in particular the valves

[1] Taps or faucets depending on where you live in the world.

are part of the system and cannot be independently controlled of the pathway. This makes understanding how pathways are controlled much more subtle. Measurements (**ref**) have shown for example that phosphofructokinase is in fact not rate-limiting even though it is heavily regulated. Since the development of recombinant technology in the late 70s, the ability to control enzyme levels has become relatively easy. There are many experiments reported where over expression of a regulated step resulted in no change in the pathway flux even though such steps were considered rate limiting (**tryp reference, pfk and find others**). Even in the face of considerable experimental evidence the idea that regulated steps are rate limiting continues to persist.

7.4 A Modern Understanding of Metabolism

Although the metabolic parts list is almost complete as witnessed by the development of genomic scale metabolic reconstruction (**ref**) our understanding of how metabolism operates is still primitive and incomplete. The last four decades however has seen some progress as described in the earlier chapters of this book. The traditional concepts of metabolic control described in the previous sections are logically untenable and many experiential measurements support this notion. In addition there are now many examples where metabolic engineers have discovered that the traditional approach is next to useless in predicting how to engineer a metabolic pathway. In this final chapter I will summarize some of the more modern operating principles that can help us understand and ultimately successfully engineer metabolic pathways.

7.5 Operating Principles

- Democracy

  A living cell is a molecular democracy with distributed decision making.

  Possibly the most important key concept to understand is that the

components of a cell do not act in isolation. This seems a very obvious thing to say but how may times do we hear about a key enzyme, a master protein, a hub, a hot spot or other metaphor in an attempt to reduce a complex system to one component. In practice we often try to simplify a complex problem down to a single entity. One example of this is related to circadian rhythms or any oscillatory system in a cell. The temptation is to find the 'oscillatorphore', that single protein responsible for the oscillation. No such protein exists of course. The oscillator is the result of a collection of components acting together. Likewise in metabolism, the behavior we observe is the result of all enzymes acting in unison. The idea that a complex pathway can be distilled down to a single enzyme is too simplistic.

- Context

  The behavior of a part or set of parts only makes sense when related to its functional context.

  The influence that a cellular component has on the phenotype is always modified according to the context in which we find the component. This is related to the first principle but adds the qualifying point that the influence of a component can change according to context, and context can change according to the state of the organism.

- Operation

  A living cell does not operate like a digital computer, with a program and a sequential operation. There have been many times in the past when we have tried to compare a digital computer to a living cell sugegsting that cells can be programmed just like a digital computer. This is no surprise since throughout history we have often tried to compare natural systems to what ever is the current major technological development at the time. Biological cells are however quite different from digital computers. Biological cells work in parallel, digital computer work in a sequential mode; biological cell process information in many ways, including analog and sometimes digital. The major problem in comparing a digital computer to a biological cell is that it puts the mind into the wrong mode of thinking and

colors subsequent analysis of a biological system which can lead to incorrect conclusions.

- Rate-Limiting Steps

  There is no such thing as a rate-limiting step, only degrees of limitingness which can be quantified.

- Systemic Property

  Whether an enzyme limits flux or not is a systemic property and cannot be determined from looking at the enzyme alone.

- Front-Loading

  In an unregulated pathway with only mass-action or Michaelis-Menten kinetics, sensitivity of the flux to enzyme changes is biased toward the front of the pathway.

- Feedback

  Steps regulated by feedback control (or within the feedback loop) are insensitive to changes in enzyme activity (eg by gene expression changes of addition of inhibitors).

- Homeostasis

  Metabolic feedback and the rarer feed-forward loops ensure homeostasis of metabolites far from equilibrium through the control of supply and demand. They are not involved in flux control.

- Flux Control

  Flux control is achieved by targeted up or down regulation of a set of enzymes or entire pathways either by gene expression changes or much more rapid kinase/phosphatase action on enzymes

- Control is Dynamic

  The ability of a give step to control a flux or concentration is dynamic and changes according to the state of the cell.

7.6 Characteristics of a bootleneck?

What are they?

7.7 'Excess' Enzymes

Many the misunderstandings of how metabolism operates can be blamed on the use of verbal logic, intuition and inappropriate analogies (such as connected tanks of water or traffic flow). A key area of confusion is the observation that some enzymes appear to be in excess. That is the Vmax of the enzyme exceeds by a wide margin the range of pathway fluxes that will flow through the enzyme. Without further analysis, it seems that evolution has made a mistake and is maintaining high levels of enzyme concentrations without any apparent fitness advantage. This however cannot be the case because any excess enzyme is likely to reduce the fitness of an organism and therefore by selection any excess enzyme will eventually disappear, and yet the excess enzyme observation is still discussed in the literature.

paradox

7.8 Why are regulated enzymes regulated?

Text

Further Reading

1. Sauro HM (2012) Enzyme Kinetics for Systems Biology. Ambrosius Publishing. ISBN: 978-0982477311

Exercises

1. A question

8

Measuring

8.1 Introduction

Up to now we've largely discussed theory. However, without the ability to make measurements and test predictions on real biological systems, the approach outlined in the previous chapters will find little use in practice. In this chapter we will outline various methods that have been developed to measure control coefficients and some examples of applications to real metabolic pathways.

Measuring Control Coefficients

Various ways have been used in the past to measure control coefficients experimentally. All revolve around the need to change either the concentration or activity of an enzyme or protein. The different methods can be categorized into six general approaches:

1. Use of classical genetics to manipulate gene expressions
2. Titration of enzymes with specific inhibitors
3. Computer modeling
4. Double modulation method
5. *in vitro* reconstitution and enzyme titration
6. Gene engineering to change enzyme levels *in vivo*

Of the six approaches, we will only consider the experimental based methods, this means excluding computer modeling form our consideration.

8.2 Using Classical Genetics

There is an interesting story related to the origins or metabolic control antilysis from the Kacser lab[1]. Early on it was noted that a mutation in one of the amino acid biosynthesis enzymes in the fungus *Neurospora crassa* yielded a loss of 95% enzyme activity and yet resulted is hardly any change to the observed phenotype. The question that arose was how could this be? Instinctively one might try to answer this question by studying the enzyme in question, perhaps studying its kinetics in details understanding its catalytic activity by determining its protein structure. However this would not ultimately explain the effect of enzyme loss on the phenotype. In an insightful move, the Kacser lab decided that the answer must lie within the context in which the enzyme operated. That is the network within which the enzyme operated. It was this observation that began the study of how networks behaved and how phenotype was related to phenotype, not in terms of individual proteins or enzymes but in terms of networks. Similar conclusions were made by the other two groups who also developed MCA independently, namely the Heinrich and Rapoport in Berlin and and Savageau ion Michigan, USA.

The Kacser lab investigated the amino acid biosynthesis pathway by changing the copy number of a gene using classical genetics [23]. This assumed that the activity or concentration of the enzyme or protein in question was proportional to the copy number. This work was carried out using the

[1] personal communication

fungus *Neurospora crassa* where arginine biosynthesis was studied. This fungus forms multinucleated mycelia that generate polyploid spores. By mixing different ratios of spores containing genes encoding wild and mutant enzymes it was possible to generate mycelia with different activities of the arginine pathway enzymes. From such experiments it was determined that four enzymes, acetyl-ornithine aminotransferase, ornithine transcarbamoylase, arginine succinate synthetase, and arginine-succinate lyase all had flux control in the range 0.02 to 0.2. This indicated that none of the enzyme exerted significant control of arginine synthesis.

Another study by the same group was to investigate the flux control of alcohol dehydrogenase (ADH) in *Drosophila melanogaster*. ADH is present in three alleles that encode isoforms with different maximal activities. When mixing the various combinations including mutations of the isoforms it was possible to change the total activity of ADH and measure the ethanol production. From this study is was concluded the ADH has a flux control coefficient of zero zero.

8.3 Genetic Engineering

Neurospora crassa and *Drosophila melanogaster* were special cases where gene dosage could be engineered by classical genetic methods. However most systems are no so easily manipulated. As a result alterative methods based on inhibitors were developed. In particular oxidative phosphorylation is susceptible to a large repertoire of inhibitors. Control coefficients were estimated by titrating in a given inhibitor and measuring the effect on a flux or species concentration [30]. By taking into account how the inhibitor acted it is possible to obtain estimates for the control coefficient of the inhibited step.

With modern developments in genetic engineering and molecular biology, it is now relatively straightforward to estimate control coefficients. For example, inducible or repressor operator sites can be added to a gene of interest and the effect of up regulating or down regulating the wild-type activity measured. In addition, protein levels can be knocked down by using RNA antisense. This method was used to estimate the flux control

coefficient for ribulose-bisphosphate-carboxylase (Rubisco) which is responsible for fixing carbon dioxide in plants [80]. The traditional view has been that Rubisco catalyzes a rate-limiting step, that is a step with a high flux control coefficient. However, studies that used DNA antisense in tobacco plants to reduce the level of Rubisco showed that during high illumination, flux control was estimated in the range 0.69 to 0.83 and in moderate illumination or high carbon dioxide levels, the flux control fell dramatically to 0.05 to 0.2. This study highlights two important points that were highlighted in the theoretical analysis, the first is that control is not fixed but depends on external conditions and secondly that rate limitingness cannot be determined by simple inspection of the pathway but must be actively measured.

8.4 Titration by Inhibitors

The first approach devised to estimate control coefficients that the use of classical genetic to manipulate gene copy number. This however is difficult to do and requires considerable expertise in genetic manipulation. In the late 1970s and early 1980s it was realized that inhibitors could also be used to change protein and enzyme activity. One of the most well studied pathways is oxidative phosphorylation due in part to the wide variety of available inhibitors. Inhibitors include irreversible, for example cyanide that can bind to cytochrome c oxidase (Site 3), noncompetitive inhibitors such as Rotenone that can bind to NADH-CoQ-oxidoreductase (Site I) or Antimycin that can bind to CoA-cytochrome c oxidoreductase and competitive such as malate that binds to dicarboxylate transporter.

By titrating the inhibitor and measuring the steady state response it is possible to estimate the control coefficient at the inhibited site by extrapolating the response curve back to zero inhibitor. Each inhibitor type however must be treated differently. In the last Chapter the response coefficient was introduced:

$$R^J_X = \frac{dJ}{dX}\frac{X}{J}$$

The response coefficient describe the effect of an external signal, X on a pathway. The response coefficient can be expressed in terms of the control

coefficient and the external signal elasticity using the relationship:

$$R^J_X = C^J_E \varepsilon^v_X$$

Expanding this relationship yields:

$$\frac{dJ}{dX}\frac{X}{J} = \frac{dJ}{dE}\frac{E}{J}\frac{\partial v}{\partial X}\frac{X}{v}$$

We can cancel the X terms on both sides and rearranging so that C^J_E is on one side we obtain:

$$C^J_E = \frac{dJ}{dXJ} \Big/ \frac{\partial v}{\partial X v}$$

The term $\frac{dJ}{dXJ}$ can be derived form the initial slop of the inhibition curve and $\frac{\partial v}{\partial X v}$ from the inhibition characteristics of the inhibited enzyme. Since we seek the value of C^J_E when there is no inhibitor present we should measure both terms when $X = 0$. For a non-competitive inhibitor it is possible to show that the control coefficient at $X = 0$ is given by:

$$C^J_X = -\frac{K_i}{J}\frac{dJ}{dX}$$

where K_i is the inhibition constant. For a competitive inhibitor titration, the control coefficient can be computed from the equation:

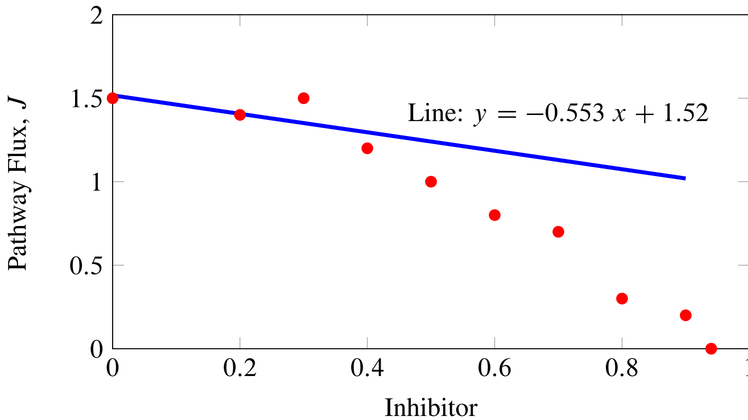
$$C^J_E = -\frac{K_i(S_j/K_m + S_{j+1}/K_{m+1} + 1}{J}\frac{dJ}{dX}$$

Example 8.1

The following data (constructed from a simulated pathway with added noise) was collected from a pathway and measures the flux through the pathway at various concentration of an irreversible inhibitor. Use the data to estimate the flux control coefficient through the pathway.

Inhibitor Concentration	Pathway Flux
0	1.5
0.2	1.4
0.3	1.5
0.4	1.2
0.5	1.0
0.6	0.8
0.7	0.7
0.8	0.3
0.9	0.2
0.94	0.0

A plot of the data is shown below



The curve that follows the points is not known in general which means it is difficult to select a suitable nonlinear function to fit the points. Instead we will take the first four data points and plot a straight line through them. This is shown by the blue continuous line on the plot. The slope of the fitted line was found to be -0.553. This corresponds to the dJ/dX term in equation 8.1. The flux at zero inhibitor is 1.5 and the I_{max} is 0.94, the concentration of inhibitor that yields zero flux. Taking these together and inserting the values into equation 8.1, we obtain the flux control coefficient:

$$C^J_E = -0.553 \frac{0.94}{1.5} = 0.35$$

The simulation model that was used to obtain the data gave a control coefficient of 0.367 which is close to the estimated value. The key to obtaining a reasonable estimate is to secure sufficient points at low inhibitor concentration in order to compute a best line fit through the first few points. Attempts to fit polynomials, logistic curves or hyperbolic curves will yield poor estimates.

Irreversible Inhibitors

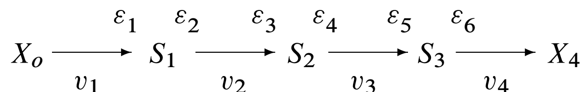
Irreversible inhibitors are a special case, in the sense that the amount of inhibitor required to completely inhibit the enzyme should be equal to the amount of enzyme. This assumes that one molecule of inhibitor binds to a single enzyme and completely inhibits the enzyme's activity. That is $X_{max} = E$. The equation to compute the control coefficient using an irreversible enzyme is then given by:

$$C^J_E = -\frac{X_{max}}{J}\frac{dJ}{dX} \tag{8.1}$$

The negative sign is included because the slope of dJ/dX is negative.

8.5 Double Modulation Technique

The double modulation method, first proposed by Kacser and Burns in 1979 is an elegant method for estimating elasticities *in vivo*. Consider the pathway:



Let us focus on reaction v_3 flanked by species S_2 and S_3 respectively. Let us make a perturbation in the upstream source metabolite, X_o. Changes will propagate through the pathway resulting in changes in S_2 and S_3. If the changes are sufficiently small we can write down the change in flux

using the following relation:

$$\frac{\delta J^1}{J^1} = \varepsilon^{v_3}_{S_2} \frac{\delta S^1_2}{S^1_2} + \varepsilon^{v_3}_{S^1_3} \frac{\delta S^1_3}{S^1_3} \tag{8.2}$$

We can now carry out a separate experiment where we perturb the downstream sink pool X_4. Again, we will observe propagations in the pathway resulting in changes to S_2 and S_3. Note that because the disturbance is from another source, the changes in S_2, S_3 and flux J will be different. That is:

$$\frac{\delta J^2}{J^2} = \varepsilon^{v_3}_{S_2} \frac{\delta S^2_2}{S^2_2} + \varepsilon^{v_3}_{S_3} \frac{\delta S^2_3}{S^2_3}$$

We now have two equations in two unknowns, $\varepsilon^{v_3}_{S_2}$ and $\varepsilon^{v_3}_{S_3}$. Assuming we can measure the changes in metabolite and fluxes we can use the two equations to solve for the elasticities. In principle, if we measured all the metabolite changes in the entire pathway we could estimate all the elasticities. Once we have the elasticities the control coefficients can be estimated using the methods described in the last chapter. Another point work making, it doesn't matter what changes are made to illicite the perturbation, we used X_o and X_4 but change sin enzyme levels or addition of inhibitors are equally valid ways to perturb the system. The method has been generalized by Acerenza and Cornish-Bowden [1].

Example 8.2

Using the same system from the previous example, the following perturbation data were obtained by carrying out two perturbations, one upstream and one downstream of the reaction under observation.

Before any perturbations were made the following reference flux and concentrations were recorded:

$$J = 1.5$$
$$S_1 = 0.74$$
$$S_2 = 0.92$$

A perturbation that involved increasing the input pool by 30% was applied the

following new steady state flux and concentrations were recorded:

$$J = 1.6$$
$$S_1 = 0.91$$
$$S_2 = 1.15$$

It was noted that it was not possible to perturb the output metabolite form the pathway, instead an inhibitor was applied to one of the downstream steps. The degree of inhibition at the inhibited site is not known (and is irrelevant) but the new steady state flux and concentrations were recorded:

$$J = 1.28$$
$$S_1 = 0.905$$
$$S_2 = 1.32$$

Form this data estimate the two elasticities, $varepsilon_1$ and ε_2 with respect to S_1 and S_2 respectively. Formulate two equations, corresponding to each perturbation, of the form:

$$\frac{\delta J}{J} = \varepsilon_1 \frac{\delta S_1}{S_1} + \varepsilon_1 \frac{\delta S_2}{S_2}$$

From the data these two equations are:

$$\frac{0.1}{1.5} = \varepsilon_1 \frac{0.17}{0.74} + \varepsilon_2 \frac{0.23}{0.92}$$

$$\frac{-0.22}{1.5} = \varepsilon_1 \frac{-0.165}{0.74} + \varepsilon_2 \frac{0.4}{0.92}$$

Evaluating the ratios we obtain:

$$0.066 = \varepsilon_1\, 0.23 + \varepsilon_2\, 0.25$$

$$-0.14 = -\varepsilon_1\, 0.23 + \varepsilon_2\, 0.43$$

Solving for ε_1 and ε_2 yields:

$$\varepsilon_1 = 1.531 \quad \varepsilon_2 = -1.144$$

The values form the model are $\varepsilon_1 = 1.69 \quad \varepsilon_2 = -1.3$. The discrepancy in the values is a result of two factors. The first is that all values were rounded down to two decimal places, secondly and more importantly we made relatively

large perturbations. The double modulation method depends on making small enough changes such that the relationship describe by equation 8.2 remains true. If the perturbations are too high equation 8.2 is only an approximation. Ideally one might make multiple perturbation at different strengths, plotting the resulting changes and extrapolating the plotted response back to the zero axis.

Determining Elasticities from Multiple Measurements of Flux Rates and Metabolite Concentrations. Christoph Giersch, uropean Journal of Biochemistry Volume 227, Issue 1-2, pages 194Ű201, January 1995

8.6 Reconstitution Methods

Further Reading

1. Kacser H and Burns J. (1979) Molecular Democracy: Who Shares the Controls? Biochem Soc Trans, 7, 1149-1160

2. Rafael Moreno-Sanchez, Emma Saavedra, Sara Rodrguez-Enriquez, and Viridiana Olin-Sandoval (2008) Metabolic Control Analysis: A Tool for Designing Strategies to Manipulate Metabolic Pathways, Journal of Biomedicine and Biotechnology, Volume 2008, Article ID 597913, doi:10.1155/2008/597913

Exercises

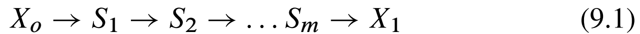
1. Estimate the other two flux control coefficients using the data below:

9

Linear Pathways

9.1 Basic Properties

Linear pathways represent the simplest network motif and are a good starting point to begin to gain insight into how cellular networks operate. The simplest linear pathway is one where the kinetics are mass-action. Consider the following linear pathway:



$$X_o \to S_1 \to S_2 \to \ldots S_m \to X_1 \tag{9.1}$$

This pathway has m floating species and n reactions ($n = m + 1$). X_o and X_1 are are fixed species representing the source and sink pools respectively. We can assume that each reaction obeys the following simple reversible mass-action kinetic law:

$$v_i = k_i S_{i-1} - k_{-i} S_i \tag{9.2}$$

where k_i and k_{i-1} are the forward and reverse rate constants respectively. S_{i-1} is the substrate and S_i the product. Recall that the equilibrium con-

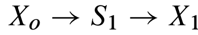
stant for such as simple reaction is given by

$$K_{eq} = q = \frac{k_i}{k_{-i}} = \frac{S_i}{S_{i-1}}$$

which means we can replace the reverse rate constant and rewrite the rate law as

$$v_i = k_i \left(S_{i-1} - \frac{S_i}{q_i} \right) \tag{9.3}$$

This model is simple enough that we can derive the analytical equation for the steady state flux through the pathway. One way to do this is to first start with a two step pathway:



$$X_o \rightarrow S_1 \rightarrow X_1$$

where the rates for the two steps are given by:

$$v_1 = k_1 \left(X_0 - \frac{S_1}{q_1} \right) \qquad v_2 = k_2 \left(S_1 - \frac{X_1}{q_2} \right)$$

By setting $v_1 = v_2$ we can solve for the steady state concentration of S_1 and then insert this solution into one of the rate laws. This leads to the steady state flux:

$$J = \frac{X_o q_1 q_2 - X_1}{\frac{1}{k_2} q_1 q_2 + \frac{1}{k_1} q_2}$$

We might also note that the steady state solution for S_1 is given by:

$$S_1 = \frac{q_1}{q_2} \frac{k_2 X_1 + k_1 q_2 X_o}{k_1 + k_2 q_1}$$

We can also derive the flux equation for a three step pathway and by comparing the two solutions we can deduce that the flux for a pathway of arbitrary length will be given by:

$$J = \frac{X_o \prod_{i=1}^{n} q_i - X_1}{\sum_{i=1}^{n} \frac{1}{k_i} \left(\prod_{j=i}^{n} q_j \right)} \tag{9.4}$$

where n is the number of steps in the linear chain of reactions. For example if the pathway has four steps then the steady state flux is given by

$$J = \frac{X_o\, q_1 q_2 q_3 q_4 - X_1}{\dfrac{1}{k_1} q_1 q_2 q_3 q_4 + \dfrac{1}{k_2} q_2 q_3 q_4 + \dfrac{1}{k_3} q_3 q_4 + \dfrac{1}{k_4} q_4}$$

and so on. The first thing to note about the flux relationship is that the flux is a function of all kinetic and thermodynamic parameters. There is no single parameter that determines the flux completely. This means that for a pathway with randomly assigned parameters it is extremely unlikely to have the first step as the rate limiting step. It would require a very unlikely set of parameter values for that to occur.

From the flux expression we can also compute the corresponding flux control coefficients. For this we need to differentiate the flux equation with respect to an enzyme activity-like parameter. One way to do this is to add an e_i term to each rate law, such as:

$$v_i = e_i k_i \left(S_{i-1} - \frac{S_i}{q_i} \right)$$

We can eliminate the e_i terms afterwards by setting them to one. The result of this yields the following expression for the flux control coefficient of the ith step:

$$C_i^J = \frac{\dfrac{1}{k_i} \displaystyle\prod_{j=i}^{n} q_j}{\displaystyle\sum_{j=1}^{n} \frac{1}{k_j} \prod_{k=j}^{n} q_k} \tag{9.5}$$

Note that the sum, $\sum C_i^J = 1$ in accordance to the flux summation theorem. The equation also indicates that, at least in this case, the control coefficients are less than one but greater than zero, $0 \leq C_i^J \leq 1$.

For a three step pathway the flux control coefficients for each step will be

given by:

$$
\begin{aligned}
D &= \frac{1}{k_1} q_1 q_2 q_3 + \frac{1}{k_2} q_2 q_3 + \frac{1}{k_3} q_3 \\
C_1^J &= \frac{1}{k_1} q_1 q_2 q_3 / D \\
C_2^J &= \frac{1}{k_2} q_2 q_3 / D \\
C_3^J &= \frac{1}{k_3} q_3 / D
\end{aligned}
$$

Each term in a numerator can also be found in the denominator.

9.2 Product Insensitive Steps and Fast Reactions

From the flux control coefficient equation we can make some general statements. Let us assume for example that each equilibrium constant, q_i is greater than one, $q_i > 1$ and also that all forward rate constants are equal to each other and all reverse rate constants are equal to each other. This also means that all equilibrium constants are the same. If we now take the ratio of two adjacent steps, for example the ith and $i + 1^{\text{th}}$ step, then we find:

$$
\frac{C_i^J}{C_{i+1}^J} = \frac{1/k_i \prod_{j=i}^{n} q_j}{1/k_{i+1} \prod_{j=i+1}^{n} q_j} = \frac{k_{i+1}}{k_i} q_i
$$

Since $q_i = k_i / k_{-i}$

$$
\frac{C_i^J}{C_{i+1}^J} = \frac{k_{i+1}}{k_i} \frac{k_i}{k_{-i}} = \frac{k_{i+1}}{k_{-i}}
$$

Since all the forward rate constants are equal and all the reverse rate constants are equal, the ratio k_{i+1}/k_{-i} must equal the equilibrium constant, q, therefore

$$
\frac{C_i^J}{C_{i+1}^J} = q
$$

That is the ratio of two adjacent control coefficients is equal to the equilibrium constant. Given that we assumed that $q > 1$, then it must be true that $C_i^J > C_{i+1}^J$, that is **earlier steps** will have more flux control. This pattern applies across the entire pathway such that steps near the beginning of a pathway will have more control than steps near the end. We will call this effect **front loading** and gives some credence to the traditional idea that the first or committed step is the most important step in a pathway. However, front loading only applies to unregulated pathways, the moment we add regulation to the pathway this picture changes. We will consider front loading again in a later section.

Another way to look at a linear pathway is via the mass-action ratio:

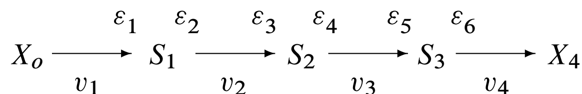
$$\Gamma = \frac{S_i}{S_{i-1}}$$

where the species concentrations are measured at steady state. We define the disequilibrium ratio, ρ to be equal to:

$$\rho = \frac{\Gamma}{K_{eq}}$$

If a step is near equilibrium, then $\rho \simeq 1$ whereas if a step is far from equilibrium then $\rho \ll 1$.

Consider the following linear pathway, where X_o and X_1 are fixed species:



The elasticities have been labeled 1 to 6, for example ε_1 represents $\varepsilon_{S_1}^{v_1}$, ε_2 represents $\varepsilon_{S_1}^{v_2}$ etc. Considering the connectivity theorem for each metabolite, the ratios of all the flux control coefficients can be shown to be:

$$C_1^J : C_2^J : C_3^J : C_4^J = 1 : -\frac{\varepsilon_1}{\varepsilon_2} : -\frac{\varepsilon_1}{\varepsilon_2}\left(-\frac{\varepsilon_3}{\varepsilon_4}\right) : -\frac{\varepsilon_1}{\varepsilon_2}\left(-\frac{\varepsilon_3}{\varepsilon_4}\right)\left(-\frac{\varepsilon_5}{\varepsilon_6}\right)$$

or for a pathway of arbitrary length, the nth term will equal:

$$\prod_{i=1}^{n-1} \left(-\frac{\varepsilon_i}{\varepsilon_{i+1}} \right)$$

If we assume that the enzymes are operating below saturation so that they are governed by the rate law, $v_i = \text{Vm}_i/\text{Km}_i(S_{i-1} - S_i/\text{Keq}_i)$, then we can replace the substrate elasticities by $1/(1-\rho_i)$ and the product elasticities by $-\rho_i/(1-\rho_i)$. If we do these substitutions, the ratios of flux control coefficients become:

$$C_1^J : C_2^J : C_3^J : C_4^J = (1-\rho_1) : \rho_1(1-\rho_2) : \rho_1\rho_2(1-\rho_3) : \rho_1\rho_2\rho_3(1-\rho_4) \quad (9.6)$$

or for an arbitrary length pathway, the nth term is equal to:

$$\left(\prod_{i=1}^{n-1} \rho_i \right) (1-\rho_n) \quad (9.7)$$

This is an important result, because by just knowing the equilibrium constants and the concentrations of the intermediate pools it is possible to obtain an idea of the relative strengths of the flux control coefficients across the pathway.

Irreversible Steps: We can draw some interesting conclusions from relation 9.7. Let us make one of the steps irreversible, say step i, so that the disequilibrium ratio for that step is zero, ($\rho_i = 0$). We can see that since ρ_i appears as a multiplier in the terms down-stream of the irreversible step, all the flux control coefficients for steps beyond will be zero. Thus steps beyond an irreversible reaction have no control over the flux (This also assumes no product inhibition). However, steps up-stream of the irreversible step may still have control. Therefore, provided the irreversible step is not the first step of the pathway, an irreversible step will not necessarily carry a control coefficient of one.

In a linear pathway governed by **linear kinetics** and without the presence of regulatory interactions, all steps downstream of an irreversible step, ($\rho_i = 0$), have no flux control.

Although this result was derived assuming linear kinetics, the result is more general and applies equally to steps governed by non-linear Michaelis-Menten kinetic laws or steps that show cooperativity. The more general result will be shown in a later section.

It is fairly easy to understand why steps beyond an irreversible step have not control. Imagine a perturbation in an enzyme activity at a step downstream of an irreversible step. This perturbation will result in changes in metabolites concentration upstream of the perturbed step. However, a perturbation in the concentration of the product of the irreversible step will by definition have no effect on the reaction rate of the irreversible step. This also means that reactions rates of all reaction steps upstream of the irreversible also remain unchanged. This means that it is impossible for downstream perturbations to change the overall flux through the pathway. In the extreme case where the first step is irreversible, the only step that has any influence on the pathway flux is the first step. All other steps have no influence. This means that the flux control coefficient for the first step will one and all downstream steps zero.

Steps close to Equilibrium: If any of the steps is near equilibrium then the disequilibrium ratio for that step will be nearly equal to one. i.e. for step i close to equilibrium, $\rho_i \approx 1$. Under these conditions, the term, $(1 - \rho_i)$ will equal approximately zero and therefore the flux control coefficient for that step will also be near zero. In addition, steps other than step i, act as if step i is not part of the pathway and the pathway appears shortened.

In a linear pathway governed by linear kinetics and without regulation, any step that is very close to equilibrium will have a flux control coefficient close to zero.

It is possible to show that the disequilibrium ratio, ρ is equal to the ratio of

the reverse and forward rates for a given reaction:

$$\rho = \frac{v_r}{v_f}$$

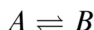
Since the forward rate will always be greater than the reverse rate for a pathway showing a positive net rate, the disequilibrium ratio will always be less than one:

$$\rho \leq 1$$

Because ρ is always less than one, the tendency is for flux control to be higher near the front of the pathway since downstream steps have greater multiples of ρ values that are less than one (See later Section on front loading 9.4).

Relaxation Times

For a simple decay process the half-life is given by $\ln 2/k_1$ where k_1 is the rate constant for the process. The term $1/k_1$ is often called the **relaxation time** and gives an idea of how fast the process changes. For a reversible system such as:



$$A \rightleftharpoons B$$

where the initial concentration of $A = A_o$ and for B is zero, the change in the concentration of A as a function of time is given by:

$$A = A_o e^{-t(k_1 + k_{-1})}$$

where k_1 and k_{-1} are the forward and reverse rate contacts respectively. The term $(k_1 + k_{-1})$ is analogous to the half-life for the simple decay process and by analogy, the reciprocal of $(k_1 + k_{-1})$ is called the **relaxation time**, usually denoted by τ:

$$\tau = \frac{1}{(k_1 + k_{-1})}$$

Returning to the linear pathway 9.1, let us assume that all the equilibrium constants are equal to one, $q_i = 1$. This means that the forward and reverse

rate constants for each reaction are equal. Applying these assumptions to the flux control equation 9.5 we find that:

$$C_i^J = \frac{\dfrac{1}{k_i}}{\displaystyle\sum_{j=1}^{n} \frac{1}{k_j}}$$

and noting that since $k_i = k_{-1}$, then $\tau = 1/(k_i + k_{i-1}) = 1/(2k_i)$ we finally obtain:

$$C_i^J = \frac{\tau_i}{\tau_1 + \ldots + \tau_n}$$

This relation shows how a given flux control coefficient depends on the relaxation time of the particular step relative the sum of all the relaxation times. That is the higher the relaxation time the larger the flux control. This result relates to the previous section where steps close to equilibrium tended to have small flux control coefficients. Steps close to equilibrium will necessarily have small relaxation times.

Although the results shown in this section and the pervious tell us that steps close to equilibrium will tend to have small flux control coefficients we must be careful in this assertion. In all the equations that predict the values for the flux control coefficients, the one common theme is that no step can be considered in isolation. Thus although a step may be close to equilibrium, this observation must be considered in the context of all the others.

9.3 NonLinear Kinetics

The previous examples used linear mass-action kinetics for the individual steps. What happens if we replace linear mass-action kinetics with non-linear enzymatic rate laws? In such situations we are unable to generate analytical solutions for the flux, as in equation 9.4 and since we cannot derive flux expression we also cannot generate sensitivity equations such as 9.5. Instead we must use the method describe in section 6.7, that is derive the control coefficients in terms of the elasticities. One way to derive

the control coefficients is to use the summation and connectivity theorems. Section 6.7 derived the equations for a two step pathway:

$$C^J_{E_1} = \frac{\varepsilon^2_S}{\varepsilon^2_S - \varepsilon^1_S} \quad C^J_{E_2} = -\frac{\varepsilon^1_S}{\varepsilon^2_S - \varepsilon^1_S}$$

$$C^S_{E_1} = \frac{1}{\varepsilon^2_S - \varepsilon^1_S} \quad C^S_{E_2} = -\frac{1}{\varepsilon^2_S - \varepsilon^1_S}$$

Using these equations we can look at some simple extreme behaviors. For example, let us assume that the first step is completely insensitive to its product, S, then $\varepsilon^1_S = 0$. In this case, the control coefficients reduce to:

$$C^J_{E_1} = 1$$

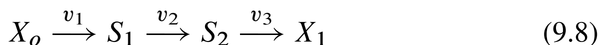
$$C^J_{E_2} = 0$$

That is all the control (or sensitivity) is on the first step. This situation represents the classic rate-limiting step. The flux through the pathway is completely dependent on the first step. Under these conditions, no other step in the pathway can affect the flux. The effect is however dependent on the complete insensitivity of the first step to its product. Such a situation is likely to be rare in real pathways. In fact the classic rate limiting step has almost never been observed experimentally. Instead, a range of "limitingness" is observed, with some steps having more "limitingness" (control) than others. We can shift control off the first step by increasing the product inhibition.

What happens if the first step is near equilibrium? In this situation, the ε^1_1 will approach $-\infty$ (See Figure 5.4) so that the first step hardly has any flux control and all the control is on the second step.

Control Coefficients for a Three Step Pathway

What about a three step pathway:



$$X_o \xrightarrow{v_1} S_1 \xrightarrow{v_2} S_2 \xrightarrow{v_3} X_1 \tag{9.8}$$

The flux control coefficient summation theorem is given by:

$$C^J_{E_1} + C^J_{E_2} + C^J_{E_3} = 1$$

Given that we have two species concentrations, S_1 and S_2, we have two connectivity theorems:

$$C^J_{E_1} \varepsilon^{v_1}_{S_1} + C^J_{E_2} \varepsilon^{v_2}_{S_1} = 0$$

$$C^J_{E_2} \varepsilon^{v_2}_{S_2} + C^J_{E_3} \varepsilon^{v_3}_{S_2} = 0$$

These three equations can be combined to give expressions that relate the flux control coefficients in terms of the elasticities, thus:

$$C^J_{E_1} = \frac{\varepsilon^2_1 \varepsilon^3_2}{\varepsilon^2_1 \varepsilon^3_2 - \varepsilon^1_1 \varepsilon^3_2 + \varepsilon^1_1 \varepsilon^2_2}$$

$$C^J_{E_2} = -\frac{\varepsilon^1_1 \varepsilon^3_2}{\varepsilon^2_1 \varepsilon^3_2 - \varepsilon^1_1 \varepsilon^3_2 + \varepsilon^1_1 \varepsilon^2_2}$$

$$C^J_{E_3} = \frac{\varepsilon^1_1 \varepsilon^2_2}{\varepsilon^2_1 \varepsilon^3_2 - \varepsilon^1_1 \varepsilon^3_2 + \varepsilon^1_1 \varepsilon^2_2}$$

The first thing to note from these equations is that if the first step is product insensitive, that is $\varepsilon^1_1 = 0$ then $C^J_{E_1} = 1$ and $C^J_{E_2}$ and $C^J_{E_3}$ are zero. As with the two step example, if any of the steps is close to equilibrium (compared to the other two), its flux control coefficient will be close to zero. For example, if the second step is close to equilibrium, that is $\varepsilon^2_1 \rightarrow \infty$ and $\varepsilon^2_2 \rightarrow -\infty$, the $C^J_{E_2} \rightarrow 0$.

If the first step of the pathway is product insensitive, then flux control resides exclusively on the first step. All other steps have no influence on the flux through the pathway.

Let us consider some realistic values for the elasticities. Let us assume that each enzyme experiences a small amount of product inhibition, let us say that each product elasticity is equal to -0.1, that is ε^1_1 and ε^2_2. Let us also assume that the substrate levels are roughly at the K_m for each

enzyme. This means that each substrate elasticity will be 0.5, this includes ε_1^2 and ε_2^3. With these values we can estimate the flux control coefficients, shown in Table 9.1. Flux control is clearly biased towards the start of the pathway but some control is found in steps downstream of the first step. Flux control that is biased towards the front of the pathway is called front-loading, a topic we will discuss later in the chapter.

Step	Flux Control Coefficient
J_1	0.9
J_2	0.09
J_3	0.009

Table 9.1 Distribution of flux control assuming weak product inhibition and substrate levels at the enzyme's K_m.

What happens if all three steps are close to equilibrium? At first glance it might seem that no step has flux control because we know from the previously results that steps close to equilibrium have little ability to control flux. However, every system must obey the flux summation theorem where all flux control coefficients sum to one. The division of control in a pathway were all steps are close to equilibrium is instead decided by the relative degree of equilibrium between the each step.

It is possible for steps close to equilibrium to have significant flux control depending on the context of the reaction.

Concentration Control Coefficients

To compute the concentration control coefficients we need a different set of theorems. There are two sets of concentration control coefficients, one with respect to S_1 and another with respect to S_2. For example if we were to consider the control coefficients with respect to S_2 we would use the following summation theorem:

$$C_{E_1}^{S_2} + C_{E_2}^{S_2} + C_{E_3}^{S_2} = 0$$

and the two connectivity theorems:

$$C^{S_2}_{E_2}\varepsilon^2_2 + C^{S_2}_{E_3}\varepsilon^3_2 = -1$$

$$C^{S_2}_{E_1}\varepsilon^1_1 + C^{S_2}_{E_2}\varepsilon^3_1 = 0$$

Solving for $C^{S_2}_{E_1}$, $C^{S_2}_{E_2}$ and $C^{S_2}_{E_3}$ yields:

$$C^{S_2}_{E_1} = \frac{\varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^{S_2}_{E_2} = \frac{-\varepsilon^1_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^{S_2}_{E_3} = \frac{\varepsilon^1_1 - \varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

Note that the denominator is positive and the numerators for $C^{S_2}_{E_1}$ and $C^{S_2}_{E_2}$ are positive indicating that increases in E_1 or E_2 result in increases in S_2. In contrast the numerator for $C^{S_2}_{E_3}$ is net negative indicating that increases in E_3 result in decreases in S_2. We can apply similar reasoning to derive the concentration control coefficients with respect to S_1:

$$C^{S_1}_{E_1} = \frac{\varepsilon^3_2 - \varepsilon^2_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^{S_1}_{E_2} = \frac{-\varepsilon^3_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^{S_1}_{E_3} = \frac{\varepsilon^2_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

In a linear pathway where increases in substrate increase the reaction rate and increases in product decrease the reaction rate, increases in activity of a given enzyme, will result in all species down stream of the enzyme to increase and all species upstream of the enzyme to decrease.

Let us estimate the concentration control coefficients using some realistic values for the elasticities. Let us assume that each enzyme experiences a small amount of product inhibition, let us say that each product elasticity is equal to 0.1, that is ε_1^1 and ε_2^2. Let us also assume that the substrate levels are roughly at the K_m for each enzyme. This means that each substrate elasticity will be 0.5, this includes ε_1^2 and ε_2^3. Table 9.2 shows the results of the calculations.

Step	$C_i^{S_1}$	$C_i^{S_2}$
E_1	1.982	1.802
E_2	-1.802	0.180
E_3	-0.180	-1.982

Table 9.2 Distribution of concentration control assuming weak product inhibition and substrate levels at the enzyme's K_m.

Note how increases in enzymes downstream of a metabolite result in the metabolite decreasing in concentration (negative coefficient) while changes in enzymes upstream of a metabolite result in increases in the metabolite.

Another important observation is that reactions close to equilibrium have little influence over the species concentrations. Consider the middle reaction v_2. If v_2 is close to equilibrium then its substrate elasticity, $\varepsilon_1^2 \gg 0$ and the product elasticity, $\varepsilon_2^2 \ll 0$. Under these conditions we see that $C_{E_2}^{S_1}$ and $C_{E_2}^{S_2}$ tend to zero because the term ε_1^2 or ε_2^2 only appear in the denominator.

Reaction steps which are close to equilibrium have little influence over species concentrations in a linear pathway.

Heinrich and Shuster in their book The Regulation of Cellular System [34] also showed it is possible to express the concentration control coefficients in terms of the flux control coefficients in a linear pathway. They showed by application of the connectivity and summation theorems that:

For steps at or before i, that is $1 \leq j \leq i$:

$$C_j^{S_i} = \frac{C_j^J}{C_{i+1}^J \varepsilon_i^{i+1}} \sum_{k=i+1}^{n+1} C_k^J$$

For steps downstream of i, that is $i + 1 \leq j \leq n + 1$:

$$C_j^{S_i} = \frac{C_j^J}{C_i^J \varepsilon^i + i} \sum_{k=1}^{J} C_k^J$$

What both equations tell is us that the value for a concentration control coefficient at a step i is proportional to the flux control coefficient at step i. Therefore if flux control at a particular step is small then the ability of the same step to control concentration is also diminished. Given that the denominator contains elasticity terms, a low flux control coefficient isn't a sufficient criterion for low concentration flux control.

Matrix Method

An examination of the theorems that were used to derive both the flux and concentration control coefficients reveals that we can recast the theorems in matrix form as follows:

$$\begin{bmatrix} C_1^J & C_2^J & C_3^J \\ C_1^{S_1} & C_2^{S_1} & C_3^{S_1} \\ C_1^{S_2} & C_2^{S_2} & C_3^{S_2} \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_1^1 & 0 \\ 1 & -\varepsilon_1^2 & -\varepsilon_2^2 \\ 1 & 0 & -\varepsilon_2^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9.9}$$

For example, the first row of the first matrix multiplied by the first column of the elasticity matrix yields the flux summation theorem. It is a simple matter to extend the matrix to any size linear pathway by following the pattern. For example a four step pathway is represented by:

$$\begin{bmatrix} C_1^J & C_2^J & C_3^J & C_4^J \\ C_1^{S_1} & C_2^{S_1} & C_3^{S_1} & C_4^{S_1} \\ C_1^{S_2} & C_2^{S_2} & C_3^{S_2} & C_4^{S_2} \\ C_1^{S_3} & C_2^{S_3} & C_3^{S_3} & C_4^{S_3} \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_1^1 & 0 & 0 \\ 1 & -\varepsilon_1^2 & -\varepsilon_2^2 & 0 \\ 1 & 0 & -\varepsilon_2^3 & -\varepsilon_3^3 \\ 1 & 0 & 0 & -\varepsilon_3^4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

One advantage of writing the equations in matrix form is that it makes it easy to evaluate the control coefficients by inverting the elasticity matrix. For example, if a three step linear pathway has the following elasticities:

$$\begin{bmatrix} 1 & 0.6 & 0 \\ 1 & -1.2 & 0.2 \\ 1 & 0 & -0.5 \end{bmatrix}$$

Then the control coefficient matrix is given by:

$$\begin{bmatrix} 1 & 0.6 & 0 \\ 1 & -1.2 & 0.2 \\ 1 & 0 & -0.5 \end{bmatrix}^{-1} = \begin{bmatrix} 0.588 & 0.294 & 0.118 \\ 0.686 & -0.49 & -0.196 \\ 1.176 & 0.588 & -1.765 \end{bmatrix}$$

Note how the top row of values sum to one reflecting the flux summation theorem and the second and third rows sum to zero corresponding to the concentration summation theorem.

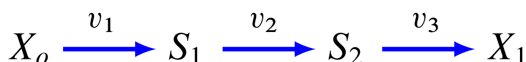
9.4 Front Loading

In a linear pathway with linear reversible kinetics on each step, given two adjacent flux control coefficients, the upstream coefficient will always be equal or larger than the downstream coefficient, that is for the ith step the following is true:

$$C_i^J \geq C_{i+1}^J$$

This means that in a linear pathway control will be concentrated upstream. To understand why this should be the case we must consider the elasticities and control equations for a linear pathway.

Using the flux summation and connectivity theorems it is straight forward to derive the flux control equations. For example for the three step pathway:



one can derive the following flux control coefficient equations:

$$C^J_{E_1} = \varepsilon^2_1 \varepsilon^3_2 / D$$

$$C^J_{E_2} = -\varepsilon^1_1 \varepsilon^3_2 / D$$

$$C^J_{E_3} = \varepsilon^1_1 \varepsilon^2_2 / D$$

where D the denominator is given by:

$$D = \varepsilon^2_1 \varepsilon^3_2 - \varepsilon^1_1 \varepsilon^3_2 + \varepsilon^1_1 \varepsilon^2_2$$

It is possible to do this for pathways with additional steps from which a clear pattern emerges in the equations. For a pathway with n steps where n is even, we have the following equations:

$$\begin{aligned}
C^J_1 &= \varepsilon^2_1 \, \varepsilon^3_2 \, \varepsilon^4_3 \, \varepsilon^5_4 \cdots \varepsilon^{n+1}_n / D \\
&\vdots \\
C^J_m &= \prod_{k=m}^{n} \varepsilon^{k+1}_k \prod_{k=m-1}^{1} \varepsilon^k_k / D \\
&\vdots \\
C^J_n &= \varepsilon^1_1 \, \varepsilon^2_2 \, \varepsilon^3_3 \, \varepsilon^4_4 \cdots \varepsilon^{n+1}_{n+1} / D
\end{aligned}$$

If we look carefully at C^J_1 we see that the numerator is the product of all the substrate elasticities. This implies that a perturbation in E_1 'hops' from one enzyme to the next until it reaches the end of the pathway. Conversely, the control coefficient of the last enzyme includes all the product elasticities, that is the perturbation 'hops' from one enzyme to the next until it reaches the beginning of the pathway.

If we looked at any intermediate enzyme step we would find two groups of elasticities, one group representing the perturbation traveling downstream via the substrate elasticities and the other representing the perturbation traveling upstream via product elasticities.

We must now recall that given a reversible mass-action rate law, such as $k_1 S - k_2 P$, the elasticities are given by:

$$\varepsilon^v_S = \frac{1}{1-\rho}$$

$$\varepsilon^v_P = -\frac{\rho}{1-\rho}$$

From these equations it follows that $\varepsilon^v_S + \varepsilon^v_P = 1$, that is:

$$\| \varepsilon^v_S \| \geq \| \varepsilon^v_P \|$$

That is the absolute value of the substrate elasticity is always greater than the product elasticity. Given that an upstream enzyme will have more substrate elasticities than product elasticities, it follows that the numerator will be larger when compared to an enzyme further downstream which will have more of the small value product elasticities. What this means is that perturbations at a downstream enzyme will be attenuated compared to a similar perturbation at an upstream step. Hence the control coefficients upstream will on average be larger.

The origins of the asymmetry between the substrate and product elasticities is a thermodynamic one. If the thermodynamic gradient were to be reversed so that the pathway flux traveled 'upstream', the elasticity values exchange so that now the front loading occurs downstream, although 'downstream' is now 'upstream' because the flux has reversed.

In a linear pathway governed by linear kinetics and without regulation, flux control is biased towards the start of the pathway, an effect called **front loading**.

9.5 Optimal Allocation of Protein

Protein synthesis constitutes a significant drain on resources in a cell. For example, protein synthesis consumes approximately 7.5 ATP equivalents

per peptide bond compared to one glucose molecule yielding roughly 36 molecules of ATP. If the average number of peptide bonds in a protein is 300, then it takes roughly 62 molecules of glucose to make just one protein molecule, not including the cost of the amino acids. In some cultured mammalian cells, protein synthesis consumes 35% to 50% of all ATP production. In addition to the energetic cost, proteins also occupy a significant proportion of cell volume at around 20 to 30% of the cell. This high level approaches the solubility limit of proteins and also limits the diffusion of other smaller molecules. These and other issues effectively put an upper limit on the total amount of protein in a cell. It would seem logical to assume that the distribution of a fixed amount of protein is not evenly distributed because some processes may requires higher levels of protein compared to others suggesting competition for protein between different processes. Such distributions are likely to be under evolutionary selection so that there exists an optimal allocation of the fixed amount of protein to all process in the cell. The optimal allocation is also likely to shift as environmental conditions change.

In this section we will consider what is the optimal allocation of a fixed amount of protein in a metabolic pathway such that the steady state pathway flux is maximized.

Let us consider a very simple two step metabolic scheme shown below:

$$X_o \rightarrow S_1 \rightarrow X_1$$

Assume that the first step is catalyzed by an enzyme E_1 and the second step by an enzyme E_2. Let us reduce the amount of enzyme E_1 by a small amount, δE_1, such that the pathway flux is reduced by an amount δJ. We can now increase the level of E_2 by δE_2 so that the pathway flux is returned to the original state. The net change in protein is therefore $\delta E_1 + \delta E_2$.

Let us also assume that the levels of E_1 and E_2 had previously being adjusted so that for a given flux, the total $E_1 + E_2$ was at a minimum, that is the distribution of protein was optimal. In other words it would not be possible to reduce the total amount of protein and at the same time adjust the protein distribution such that the flux is unchanged. Then it must be true that:

$$\delta E_1 + \delta E_2 = 0$$

Given these changes in E_i and the fact that the flux does not change, we can write the following:

$$C^J_{E_1} \frac{\delta E_1}{E_1} + C^J_{E_2} \frac{\delta E_2}{E_2} = \frac{\delta J}{J} = 0$$

Submitting $\delta E_1 + \delta E_2 = 0$ into the above relation yields:

$$C^J_{E_1} \frac{1}{E_1} = C^J_{E_2} \frac{1}{E_2}$$

We can now invoke the flux summation theorem to eliminate one of the control coefficients to yield:

$$C^J_{E_1} \frac{1}{E_1} = \left(1 - C^J_{E_1}\right) \frac{1}{E_2}$$

Rearranging this to solve for $C^J_{E_1}$ yields:

$$C^J_{E_1} = \frac{E_1}{E_1 + E_2}$$

This result can be generalized to any length pathway so that for a given total amount of protein and a given flux, the optimal allocation of protein at a particular step, i, is given by:

$$C^J_{E_i} = \frac{E_i}{\sum E_i}$$

Further Reading

1. Heinrich and Rapoport

Exercises

1. Show that summing all the C^J_i coefficients in equation 9.5 equals one.

2. Prove equation 9.6 in the main text.

3. In general if a given enzymatic step is very close to equilibrium, what can be say about the flux control coefficient of that step?

4. In a four step pathway each step is catalyzed by a reversible Michaelis-Menten rate law. In addition each step is close to equilibrium. Does this mean that no step in the pathway can control flux? Explain your answer.

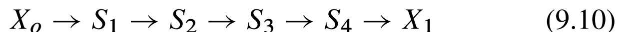
5. An unregulated linear pathway is made of up eight enzymatic reaction steps, all steps are product sensitive except for the fifth step. What can you say about the distribution of flux control in this pathway?

6. Show that the ratio of flux control coefficients in a linear pathway, such as 9.1, where each reaction is governed by equation 9.3 is given by:

$$C^J_1 : C^J_2 : C^J_3 : \ldots =$$
$$(X_o - S_1/q_1) : (S_1 - S_2/q_2)/q_1 : (S_2 - S_3/q_3)/(q_1 q_2) : \ldots$$

7. What is front-loading?

8. Metabolic engineers wish to increase the production of an important commodity that is synthesized by a five step metabolic pathway (9.10).



$$X_o \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow X_1 \qquad (9.10)$$

The pathway has no known negative feedback loops. In order to obtain a rough idea of the distribution of control in this pathway, the engineers obtain values for all the standard ΔG^os and obtain estimates for the concentrations of all the metabolite pools in the pathway. ΔG^os were obtained at 25^oC.

The table below shows the data they obtained:

Step	ΔG^o	Metabolite	Concentration
1	-12 kJ mol^{-1}	X_o	0.9 mM
2	-2 kJ mol^{-1}	S_1	0.2 mM
3	+1 kJ mol^{-1}	S_2	0.05 mM
4	-5 kJ mol^{-1}	S_3	0.45 mM
5	-4 kJ mol^{-1}	S_4	0.15 mM
		X_1	0.01 mM

From the data they collected, what advice would you give concerning which step(s) are worth increasing in activity in order to increase the flux through the pathway?

9. Using a four step linear pathway where each reaction uses reversible mass-action kinetics (Equation 9.3), generate 10,000 variations of this pathway. Do this by setting the equilibrium constants to fixed values of $q_1 = 2; q_2 = 4; q_3 = 8; q_4 = 16$ and then randomizing the forward rate constant between 0 and 1.0. For each pathway variant compute the flux control coefficients. This can be done by modulating the rate constant for each step and observing the effect on the pathway flux or by inserting the relevant values into equation 9.5. From the 10,000 variants, compute the distribution of flux control coefficients in the pathway. Explain the distribution of control coefficients you observe.

10. Derive the concentration control coefficient equations for a three step pathway.

11. It is known that in a given linear pathway the distribution of protein across the enzymes is optimized for flux. In this situation, what it the easiest way to estimate all the flux control coefficients?

10 Negative Feedback

10.1 Historical Background

Feedback is widespread in biochemical networks and physiological systems in general. Some form of feedback permeates almost every known biological process. On the face of it, feedback is a simple process that involves sending a portion of the output to the input. If the portion sent back reduces the input then the feedback is called negative feedback otherwise it is called positive feedback.

Water Clocks

The concept of feedback control goes back at least as far as the Ancient Greeks. Of some concern to the ancient Greeks was the need for accurate time keeping. In about 270 BC the Greek Ktesibios invented a float regulator for a water clock. The role of the regulator was to keep the water level in a tank at a constant depth. This constant depth yielded a constant flow

of water through a tube at the bottom of the tank which filled a second tank at a constant rate. The level of water in the second tank thus depended on time elapsed.

Philon of Byzantium in 250 BC is known to have kept a constant level of oil in a lamp using a float regulator and in the first century AD Heron of Alexandria experimented with float regulators for water clocks. Philon and particularly Heron (13 AD) have left us with an extensive book (Pneumatica) detailing many amusing water devices that employed negative feedback.

Governors

It wasn't until the industrial revolution that feedback control, or devices for automatic control, became economically important. Probably the most famous modern device that employed negative feedback was the governor. Thomas Mead in 1787 took out a patent on a device that could regulate the speed of windmill sails. His idea was to measure the speed of the mill by the centrifugal motion of a revolving pendulum and use this to regulate the position of the sail. Very shortly afterwards in early 1788, James Watt is told of this device in a letter from his partner, Matthew Boulton. Watt recognizes the utility of the governor as a device to regulate the new steam engines that were rapidly becoming an important source of new power for the industrial revolution.

The device employed two pivoted rotating flyballs which were flung outward by centrifugal force. As the speed of rotation increased, the flyweights swung further out and up, operating a steam flow throttling valve which slowed the engine down. Thus, a constant speed was achieved automatically. So popular was this innovation that by 1868 it is estimated that 75,000 governors (A History of Control Engineering, 1800-1930 By Stuart Bennett, 1979) were in operation in England. Many similar devices were subsequently invented to control a wide range of processes, including water wheels, telescope drives and temperature and pressure control.

The description of the governor illustrates the operational characteristics of **negative feedback**. The output of the device, in this case the steam engine speed, is Ťfed backŤ to control the rate of steam entering the steam

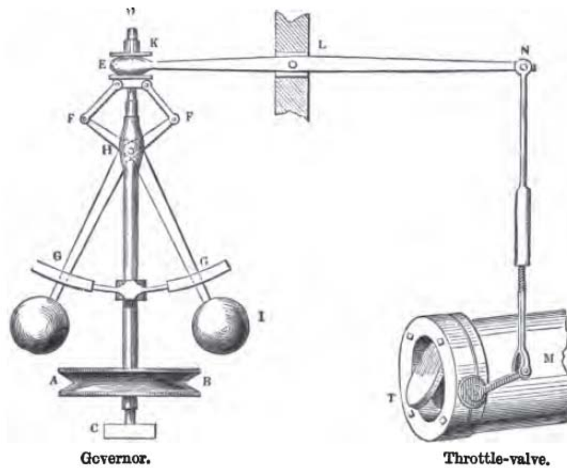


Figure 10.1 A typical governor from J. Farley, A Treatise on the Steam Engine: Historical, Practical, and Descriptive (London: Longman, Rees, Orme, Brown, and Green, 1827, p436

engine and thus influence the engine speed.

During this period devices for automatic control were designed through trial and error and little theory existed to understand the limits and behavior of feedback control systems. One of the difficulties with feedback control is the potential for instability. As the governor became more widespread, improvements were made in manufacturing mechanical devices which reduced friction. As a result engineers began to notice a phenomena they termed hunting. This was where after a change in engine load, the governor would begin to 'hunt' in an oscillatory fashion for the new stream rate that would satisfy the load. This effect caused considerable problems with maintaining a stable engine speed and resulted in James Maxwell and independently Vyshnegradskii, undertaking the first theoretical analysis of a negative feedback system.

Until the 20th century, feedback control was generally used as a means to achieve automatic control, that is to ensure that a variable, such as a temperature or a pressure was maintained at some set value. However, an entirely new application for feedback control was about to emerge with the

advent of electronics in the early part of the 20th century.

Feedback Amplifiers

Amplification is one of the most fundamental tasks one can demand of an electrical circuit. One of the challenges facing engineers in the 1920's was how to design amplifiers whose performance was robust with respect to the internal parameters of the system and which could overcome inherent nonlinearities in the implementation. This problem was especially critical to the effort to implement long distance telephone lines across the USA.

These difficulties were overcome by the introduction of the feedback amplifier, designed in 1927 by Harold S. Black (Mindell, 2000), who was an engineer for Western Electric (the forerunner of Bell Labs). The basic idea was to introduce a negative feedback loop from the output of the amplifier to its input. At first sight, the addition of negative feedback to an amplifier might seem counterproductive. Indeed Black had to contend with just such opinions when introducing the concept. His director at Western Electric dissuaded him from following up on the idea and his patent applications were at first dismissed. In his own words, "our patent application was treated in the same manner as one for a perpetual motion machine" (Black, 1977).

While Black's detractors were correct in insisting that the negative feedback would reduce the gain of the amplifier, they failed to appreciate his key insight that the reduction in gain is accompanied by increased robustness of the amplifier and improved fidelity of signal transfer.

Unlike the steam engine governor which is used to stabilize some system variable, negative feedback in amplifiers is used to accurately track an external signal. These two applications highlight the two main ways in which negative feedback can be used, namely as a **regulator** or as a **servomechanism**.

As a regulator, negative feedback is used to maintain a controlled output at some constant desired level, whereas a servomechanism will slavishly track a reference input. We can see both applications at work in the eye. On the one hand there is the need to control the level of light entering the

pupil. The diameter of the pupil is controlled by two antagonistic muscles. If the external light intensity increases, the muscles respond by reducing the pupil diameter, whereas the muscles increase the pupil diameter if the light intensity falls. The pupil reflex serves as an example of negative feedback using in a regulator mode. In contrast tracking an object involves maintaining the eyeball fixed on the object. In this mode the eye functions as a servomechanism.

Both regulator and servomechanism are implemented using the same operational mechanism. Figure 10.2 shows a generic negative feedback circuit. On the left of the figure can be found the input, sometimes terms the desired value or more often the set point. If the circuit is used as a servomechanism then the output tracks the set point. As the set point changes the output follows. If the circuit is used as a regulator or homeostatic device then the set point is held constant and the output is maintained at or near the set point even in the face of disturbances.

The central mechanism in the feedback circuit is the generation of the error signal, that is the difference between the desired output (set point) and the actual output. The error is fed into a controller (often something that simply amplifies the error) which is used to increase or decrease the process. For example, if a disturbance on the process block reduces the output, then the feedback operates by generating a positive error, this in turn increases the process and restores the original drop in the output.

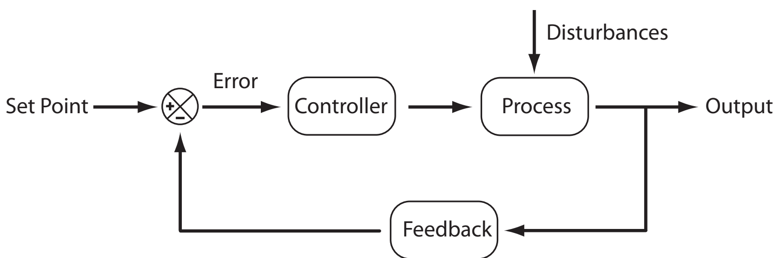


Figure 10.2 Generic structure of a negative feedback system.

10.2 Simple Quantitative Analysis

The figure of the generic negative feedback circuit (Figure 10.2) is highly stylized which makes it difficult to identify the various component in a real biological system. In addition, biological system are invariably more complex with multiply nested feedback loops and multiple inputs and outputs. If it remarkable that even after 50 or 60 years of research, the role of many of the feedback systems in biochemical networks take is still highly speculative.

In the remainder of this section we will consider some basic properties of negative feedback systems. The simplest way to think about feedback quantitatively is by reference to Figure 10.3.

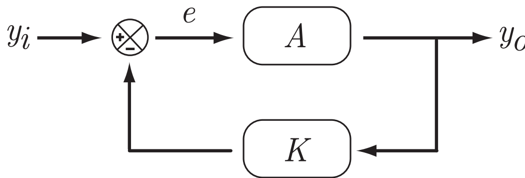


Figure 10.3 Generic structure of a negative feedback system.

We will assume some very simple rules that govern the flow of information in this feedback system. For example, the output signal, y_o will be given by the process A multiplied by the error, e. The feedback signal will be assumed to be proportional to y_o, that is Ky_o. Finally, the error signal, e will be given by the difference between the set point, y_i and the feedback signal, Ky_o (Figure 10.4).

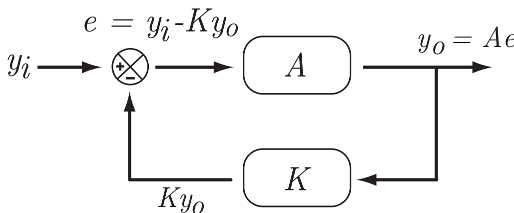


Figure 10.4 Generic structure of a negative feedback system.

From these simple relations it is straightforward to show that:

$$y_o = \frac{Ay_i}{1 + AK} \qquad \text{or more simply} \quad y_o = Gy_i \tag{10.1}$$

G is called the gain of the feedback loop, often called the **closed loop gain**. *Gain* is a term that is commonly used in control theory and refers to the scalar change between an input and output. Thus a gain of 2 simply means that a given output will be twice the input. In addition to the close loop gain, engineers also define two other gain factors, the **open loop gain** and the **loop gain**. The open loop gain is simply the gain from process, A, alone. It is the gain one would achieve if the feedback loop were absent. The loop gain is the gain from the feedback and process A combined, AK. The loop gain is a significant quantity when discussing the stability of feedback circuits. Figure 10.4 illustrates the different types of gain in a feedback circuit.

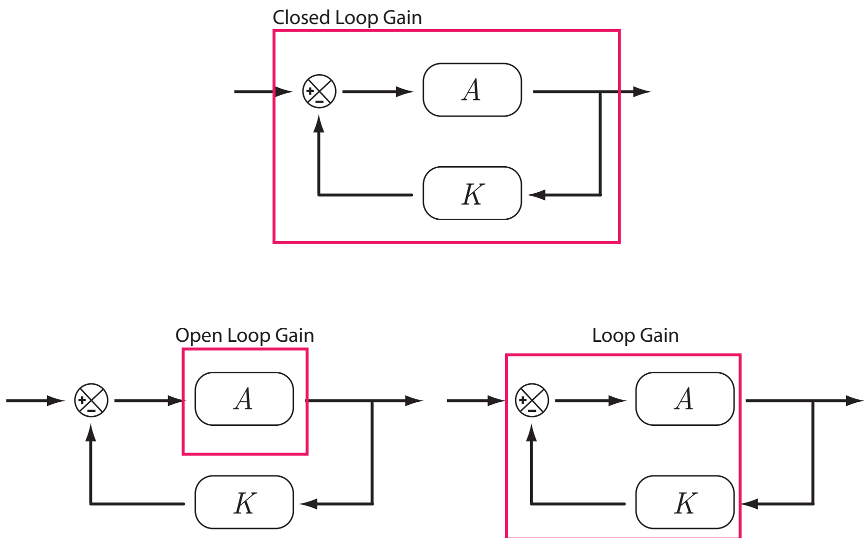


Figure 10.5 Generic structure of a negative feedback system.

We can use equation 10.1 to discover some of the basic properties of a negative feedback circuit. The first thing to note is that as the loop gain, AK,

increases, the system behavior becomes more dependent on the feedback loop and less dependent on the rest of the system:

$$\text{when} \quad AK \gg 1 \quad \text{then} \quad G \simeq \frac{A}{AK} = \frac{1}{K}$$

This apparently innocent effect has significant repercussions on other aspects of the circuit. To begin with, as the system becomes less important on A, so does variation in the properties of A. Feedback makes the performance of the system independent of any variation in A. Such variation might include noise or variation as a result of the manufacturing process or in the case of biological systems, genetic variation. To be more precise we can compute the sensitivity of the gain G with respect to variation in A.

$$\frac{\partial G}{\partial A} = \frac{\partial}{\partial A} \frac{A}{1 + AK} = \frac{1}{(1 + AK)^2}.$$

If we consider the relative sensitivity we find:

$$\frac{\partial G}{\partial A} \frac{A}{G} = \frac{1}{1 + AK}$$

In addition to resistance to parameter variation, feedback also confers a resistance to disturbances in the output. Suppose that a nonzero disturbance d affects the output. The system behavior is then described by

$$y = Ae - d \qquad e = u - Ky.$$

Eliminating e, we find

$$y = \frac{Au - d}{1 + AK}.$$

The sensitivity of the output to the disturbance is then

$$\frac{\partial y}{\partial d} = -\frac{1}{1 + AK}.$$

The sensitivity decreases as the loop gain AK is increased. In practical terms, this means that the imposition of a load on the output, for example a current drain in an electronic circuit, protein sequestration on a signaling network or increased demand for an amino acid will have less of an effect on the circuit as the feedback strength increases. In electronics this property essentially modularizes the network into functional modules.

Last but not least, feedback also improves the fidelity of the response. That is, for a given change in the input, a system with feedback is more likely to faithfully reproduce the input at the output that a circuit without feedback. An ability to faithfully reproduce signals is critical in electronics communications and in fact it was this need that was the inspiration for the development of negative feedback in the early electronics industry.

Consider now the case where the amplifier A is nonlinear. For example a cascade pathway exhibiting a sigmoid response. Then the behavior of the system G (now also nonlinear) is described by

$$G(y_i) = y_o = A(e) \qquad e = y_i - Ky_o = y_i - KG(y_i).$$

Differentiating we find

$$G'(y_i) = A'(y_i)\frac{de}{dy_i} \qquad \frac{de}{dy_i} = 1 - KG'(y_i).$$

Eliminating $\frac{de}{dy_i}$, we find

$$G'(y_i) = \frac{A'(y_i)}{1 + A'(y_i)K}.$$

We find then, that if $A'(y_i)K$ is large ($A'(y_i)K \gg 1$), then

$$G'(y_i) \approx \frac{1}{K},$$

so, in particular, G is approximately linear. In this case, the feedback compensates for the nonlinearities $A(\cdot)$ and the system response is not distorted. (Another feature of this analysis is that the slope of $G(\cdot)$ is less

than that of $A(\cdot)$, i.e. the response is "stretched out". For instance, if $A(\cdot)$ is saturated by inputs above and below a certain "active range", then $G(\cdot)$ will exhibit the same saturation, but with a broader active range.)

A natural objection to the implementation of feedback as described above is that the system sensitivity is not actually reduced, but rather is shifted so that the response is more sensitive to the feedback K and less sensitive to the amplifier A. However, in each of the cases described above, we see that it is the nature of the loop gain AK (and not just the feedback K) which determines the extent to which the feedback affects the nature of the system. This suggests an obvious strategy. By designing a system which has a small "clean" feedback gain and a large "sloppy" amplifier, one ensures that the loop gain is large and the behavior of the system is satisfactory. Engineers employ precisely this strategy in the design of electrical feedback amplifiers, regularly making use of amplifiers with gains several orders of magnitude larger than the feedback gain (and the gain of the resulting system).

1. Amplification of signal.
2. Robustness to internal component variation.
3. High fidelity of signal transfer.
4. Low output impedance so that the load does not affect the performance of the circuit.

These are the main advantages of negative feedback but as we will see in a later section on frequency response, feedback can confer additional useful features.

10.3 Negative Feedback in Biochemical Systems

It was Umbarger (Umbarger, 1956) and Yates and Pardee (Yates & Pardee, 1956) who discovered feedback inhibition in the isoleucine biosynthesis

pathway and the inhibition of aspartate transcarbamylase in *E. coli*. It wasn't long afterwards that some researchers began to investigate such feedback systems mathematically. Probably the most extensive mathematical analysis of biochemical feedback was conducted by Savageau (Savageau, 1972; Savageau, 1974; Savageau, 1976) and Burns and Kacser (Burns, 1971; Kacser & Burns, 1973) and Othmer and Tyson (Othmer, 1976; Tyson & Othmer, 1978) in the 1970s and Dibrov et. al. in the early 1980s (Dibrov et al., 1982). More recently, Cinquin and Demongeot have published an interesting review on the roles of feedback in biological systems (Cinquin & Demongeot, 2002).

In the last section we considered a simple analysis of negative feedback and its behavioral effects. The treatment was however very generic and the question we wish to address here is how can we apply the same kind of analysis to biochemical feedback systems? This question is harder to answer that it seems. To begin with, biochemical systems are governed by nonlinear rate laws not the simple linear rules we used in the previous analysis. Secondly how do we map the generic diagram (Figure 10.2) onto a biochemical feedback circuit (Figure 10.7), on the surface they look similar but the initial impression is misleading.

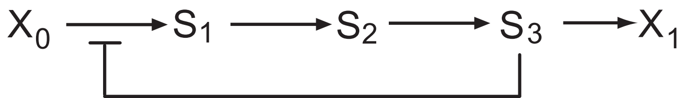


Figure 10.6 Simple four step pathway with negative feedback.

In order to be clear we need to identify the input (set point), output, the feedback loop and the process block in the biochemical network Figure 10.7).

In naturally evolved systems it is sometimes difficult to identify the various parts in a negative feedback circuit. The most difficult element to identify is the set point. In biochemical networks the set point is embedded in the regulated enzyme, often in the form of the half saturation constant of the allosteric regulator.

The pathway in Figure 10.8 shows a simple negative feedback, a common motif in many metabolic pathways. Here we see a downstream species, S_2

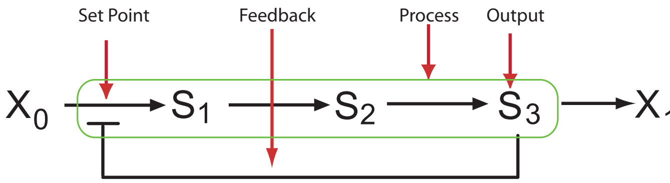


Figure 10.7 Simple four step pathway with negative feedback.

controlling the first step in the pathway, v_1. Control is often achieved using allosteric enzymes which have distinct binding sites for the controlling species that are separate from the main active site.

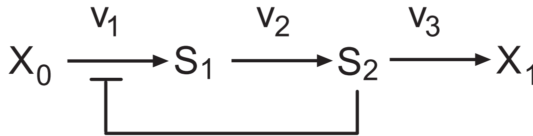


Figure 10.8 Simple three step pathway with negative feedback.

When comparing the block diagram to the biological pathway in Figure 10.8 it may not be apparent how the two representations can be matched. It is however possible to pair each component in the block diagram to an equivalent component in the biological pathway. Thus the output, y_o in the block diagram corresponds to the concentration of S_2. The negative feedback component k corresponds to the interaction of S_2 with the allosteric enzyme in the first step. The set point, y_i is more problematic but it is most likely embedded in the kinetic characteristics of the allosteric enzyme. Finally the controller A is represented by the steps v_1 and v_2. The load on the system is represented by the last step, v_3 and other disturbances can be assigned to v_1, v_2 and the input concentration, X_o.

Graphical understanding of feedback

If is possible to appreciate the effect of negative feedback using a graphical approach. Figure 10.9 shows to plots, one with strong and the other with

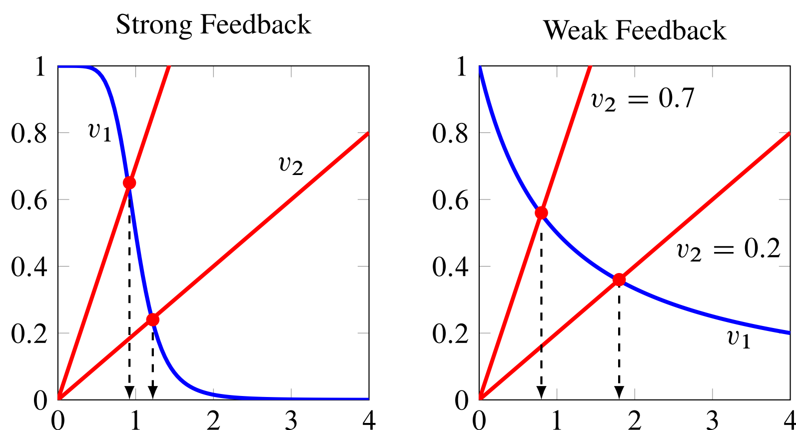


Figure 10.9 Plot of v_1 and v_2 versus the concentration of S for a simple two step pathway with negative feedback. Two perturbations in k_2 that determines v_2 are shown. In the left panel where the feedback is string, changes in k_2 have hardly any effect on S. On the right panel, the same change in k_2 results in a much larger change in S. This illustrates the homeostatic property of negative feedback. Left Panel: $v_1 = 1/(1 + S^4)$, Right Panel: $v_1 = 1/(1 + S)$

weak feedback. The plots show the reaction rates v_1 and v_2 as a function of the intermediate species, S for a simple two step pathway where S can negatively feedback on to the first step. We assume that the second step follows first-order kinetics so that the v_2 is a straight line. The feedback response curve shows a decline from high to low as S increases. For strong feedback the decline is sheep (left plot). If we now change the rate constant for the second step, this changes the slope of v_2, this is equivalent to a perturbation in the system. In the case of weak feedback, changes in v_2 result in significant changes to S, this is because the feedback response is shallow. In contrast, when we have strong feedback (left panel), where the slope is very steep, any changes in v_2 results in only small changes in S. In this way we can see how strong negative feedback can buffer changes in v_2.

Control Analysis

Just as we did earlier, we can derive the flux and concentration control coefficients in terms of the elasticities. For convenience, we will write out the theorems in matrix form (See equation 9.9), note the presence of the **feedback term**, ε^1_2 in the matrix.

$$\begin{bmatrix} C^J_1 & C^J_2 & C^J_3 \\ C^{S_1}_1 & C^{S_1}_2 & C^{S_1}_3 \\ C^{S_2}_1 & C^{S_2}_2 & C^{S_2}_3 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon^1_1 & -\varepsilon^1_2 \\ 1 & -\varepsilon^2_1 & -\varepsilon^2_2 \\ 1 & 0 & -\varepsilon^3_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Rearranging the matrix equation yields:

$$\begin{bmatrix} C^J_1 & C^J_2 & C^J_3 \\ C^{S_1}_1 & C^{S_1}_2 & C^{S_1}_3 \\ C^{S_2}_1 & C^{S_2}_2 & C^{S_2}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon^1_1 & -\varepsilon^1_2 \\ 1 & -\varepsilon^2_1 & -\varepsilon^2_2 \\ 1 & 0 & -\varepsilon^3_2 \end{bmatrix}^{-1}$$

Inverting the elasticity matrix yields the following equations for the control coefficients with and without feedback to illustrate the difference in the results.

With Feedback

$$C^J_{E_1} = \frac{\varepsilon^2_1\varepsilon^3_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

$$C^J_{E_2} = \frac{-\varepsilon^1_1\varepsilon^3_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

$$C^J_{E_3} = \frac{\varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

Without Feedback

$$C^J_{E_1} = \frac{\varepsilon^2_1\varepsilon^3_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^J_{E_2} = \frac{-\varepsilon^1_1\varepsilon^3_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

$$C^J_{E_3} = \frac{\varepsilon^1_1\varepsilon^2_2}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2}$$

We can see that the addition of feedback adds a new term to the denominator and to the numerator for $C^J_{E_3}$. If we make the feedback elasticity, ε^1_2 larger we can see that the demand flux control coefficient tends to unity. That is, all control moves out of the feedback loop. In terms of the steam engine analogy, it is equivalent to being able to change the demand on the steam engine without loss of power. This can be seen more clearly if we look at the concentration control coefficients.

$$C^J_{E_3} \to 1$$

$$C^{S_2}_{E_3} \to 0$$

In the list below we give the three control coefficients with respect to S_2. Looking at $C^{S_2}_{E_3}$ we see that as the feedback strength is increased (ε^1_2) in magnitude, the control coefficient tends to zero. This means that the feedback locks the concentration of S_2 in to a very narrow range in response to changes in demand.

$$C^{S_2}_{E_1} = \frac{\varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

$$C^{S_2}_{E_2} = -\frac{\varepsilon^1_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

$$C^{S_2}_{E_3} = \frac{\varepsilon^1_1 - \varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2 - \varepsilon^2_1\varepsilon^1_2}$$

Classically, allosteric enzymes have been considered as flux controllers. The analysis here suggests the opposite picture. Allosteric enzymes, when part of a negative feedback loop are very poor controllers of flux. Instead allow distal steps to be good controllers of a pathway flux. For demand driven systems this is a logical arrangement.

The nagging suspicion remain however that intuitively that the regulated step must have some kind of ability to enable the system to operate in the

way it does. There are at least two answers to this. The first is that the feedback elasticity will be strongly negative. For example, if the regulated step were determined by a modified Hill like equation such as:

$$v = \frac{V_{\max}\, Xo}{S^n + Xo + K_m}$$

where S is the feedback signal, then the elasticity of the reaction rate with respect to the signal is given by:

$$\varepsilon^v_S = \frac{n Xo K_m}{K_m + (S/K_m)^n}$$

We see that at low signal, S, the elasticity is proportional to n, the Hill coefficient. The reaction is therefore very sensitive to changes in the signal molecule.

The second way to answer the question is to consider the pathway with and without the negative feedback loop. We can compare for example the flux control coefficient on the first step with and without negative feedback. When we remove the negative feedback the pathway will change to a new state where the concentrations of S_1 and S_2 are higher. To make the comparison fair we should adjust the level of enzyme in the first step to restore the levels of S_1 and S_2 to the values they had before the feedback was removed. When we do this we will also automatically restore the pathway flux to its original value. In practice this means that all the elasticities except for the feedback elasticity are exactly in the same in both systems. To keep the analysis simple, let us assume that the product inhibition elasticities, ε^1_1 and ε^2_2 are both zero. We now take the ratio of the flux control coefficient on the first step, C^J_1 without feedback to the same coefficient with feedback, which will be denoted by ${}^nC^J_1$:

$$\frac{C^J_1}{{}^nC^J_1} = \frac{\varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_2\varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2} = 1 - \frac{\varepsilon^1_2\varepsilon^2_1}{\varepsilon^2_1\varepsilon^3_2} = 1 - \frac{\varepsilon^1_2}{\varepsilon^3_2}$$

Given that ε^1_2 is negative, the term, $-\varepsilon^1_2/\varepsilon^3_2$ is positive. That is:

$$C^J_1 >^n C^J_1$$

This tells us that for a negative feedback loop to be effective, the flux control coefficient in the unregulated pathway must be higher than the flux control coefficient with feedback. In a sense the first step in the pathway must have some degree of rate limitingness if a negative feedback is to be effective.

10.4 Robustness and Supply/Demand

One way to look at metabolic systems is to divide them into two separate but connected blocks, the supply block and the demand block. Negative feedback makes this division very straight forward. Consider a factory that make cars. Should the rate of car production be controlled by the demand or the supply of cars? Economically the most efficient strategy is to let demand decide how many cars to make, this ensure that excess cars do not build up and thereby waste resources. In certain metabolism situations the same reasoning can be used. For example the production of amino acids would best be determined by the demand from protein synthesis. This means that if protein synthesis slows, amino acid production should also slow. A metabolic system that supplies amino acids should be able to supply amino acids effectively at both high and low demand. One way to do this is to maintain the amino acid level at a relatively constant level independent of the demand block. Let us assume for the moment that there is no feedback regulation in the supply/demand pathway (Figure 10.10). If demand rises, this will result in the intermediate metabolite, P, falling. As P falls the ability to supply the increased demand becomes more and more difficulty. If demand falls, the flux through the pathway will fall. This will cause the intermediate metabolite, P to rise. Since the equilibrium constant across the supply block is likely to be large, the concentration of P could raise to toxic high levels as the supply block approach equilibrium at low fluxes.

Figure 10.10 A system divided into supply and demand blocks.

The solution to avoid both problems at high and low demand is to use negative feedback (Figure 10.11). With negative feedback, high demand will result in a decrease in the intermediate metabolite, P, which in turn will release repression in the supply block to restore some the loss in P. Alternatively, at low demand, the increase in P will suppress its own production preventing excessive production of P.

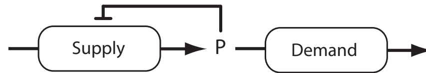


Figure 10.11 A system divided into supply and demand blocks with negative feedback.

10.5 Instability

Although negative feedback offers considerable advantages to a system, too much feedback and delays can result instability in the form of sustained oscillations. We will not discuss stability of negative feedback systems in any detail in this book (See Control Theory for Bioengineers), but there are no a number of examples of metabolic systems or metabolism systems interacting with protein networks where negative feedback appears to cause instability and the emergence of sustained oscillations (Westerhoff and Hasty).

Further Reading

1. Sauro HM (2013) Control Theory for Bioengineers. Web site

Exercises

1. A regulated step via a negative feedback look has a flux control coefficient of 0.9. Would you consider this system to be a well regulated pathway? Explain your answer.

2. A system is driven by the supply. What could go wrong with such a system when compared to one that is driven by demand?

11

Branched and Cyclic Systems

11.1 Branched Pathways

Branching structures are probably one of the most common patterns in biochemical networks. Even a pathway such as glycolysis, often depicted as a straight chain in textbooks is in fact a highly branched pathway.

At any given branch node, where a node is a molecular species, there will be conservation of mass. Given a node species, x_i, with b branches entering the node and d branches leaving, the net rate of change in concentration of x_i is:

$$\sum_{i=1}^{b} v_i - \sum_{j=1}^{d} v_j = \frac{dx_i}{dt}$$

At steady state when $dx_i/dt = 0$, it must also be true that:

$$\sum_{i=1}^{b} v_i = \sum_{j=1}^{d} v_j$$

$C^{J_1}_{E_1}$	$C^{J_2}_{E_1}$	$C^{J_3}_{E_1}$	$C^{S}_{E_1}$
$C^{J_1}_{E_2}$	$C^{J_2}_{E_2}$	$C^{J_3}_{E_2}$	$C^{S}_{E_2}$
$C^{J_1}_{E_3}$	$C^{J_2}_{E_3}$	$C^{J_3}_{E_3}$	$C^{S}_{E_3}$

Table 11.1 Set of control coefficients for a simple branch

In this section we will investigate the control of flux through a branched system in response to changes in enzyme activity. Let us consider the simple branched pathway depicted in Figure 11.1.

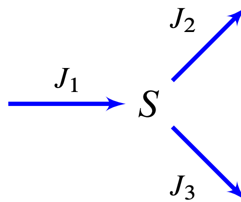


Figure 11.1 Simple branched pathway. This pathway has three different fluxes, J_1, J_2, and J_3 which at steady state are constrained by $J_1 = J_2 + J_3$.

In the figure, J_1, J_2 and J_3 are the steady state fluxes. By the law of conservation of mass, at steady state, the fluxes in each limb will be governed by the relationship:

$$J_1 = J_2 + J_3$$

Given three different fluxes and one intermediate, there will be four sets of control coefficients, one set concerned with changes in the intermediate, S, and three sets corresponding to each of the three fluxes (Table 11.1).

For the branched system we can write a summation and a connectivity theorem with respect to each flux. For example, with respect to J_1 we can write:

$$C^{J_1}_{E_1} + C^{J_1}_{E_2} + C^{J_1}_{E_3} = 1$$

and

$$C^{J_1}_{E_1} \varepsilon^{v_1}_{S} + C^{J_1}_{E_2} \varepsilon^{v_2}_{S} + C^{J_1}_{E_3} \varepsilon^{v_3}_{S} = 0$$

This gives us two equations but three unknown flux control coefficients. To solve for the $C^{J_1}_{E_i}$ we need another equation.

Let the fraction of flux through J_2 be given by $\alpha = J_2/J_1$ and the fraction of flux through J_3 be $1 - \alpha = J_3/J_1$. Let us carry out the following thought experiment.

1. Increase the concentration of E_2 by δE_2, this will cause a decrease in S, an increase in J_1 (relief of product inhibition) and a decrease in J_3.

2. Restore the change in J_1 by decreasing E_3 such that S is restored to its pre-perturbation state. That at the end the thought experiment $\delta S = 0$.

3. Since we have not changed E_1, it must be the case that $\delta J_1 = 0$.

From this experiment we can write down the system and local equations. The system equation is given by:

$$C^{J_1}_{E_2} \frac{\delta E_2}{E_2} + C^{J_1}_{E_3} \frac{\delta E_3}{E_3} = \frac{\delta J_1}{J_1} = 0$$

Note that the system equation only has two terms because we did not change E_1. The local equations are quite simple because $\delta S = 0$ and as before we assume that $\varepsilon^{v}_{E_i} = 1$.

$$\frac{\delta v_2}{v_2} = \frac{\delta E_2}{E_2} \quad \text{and} \quad \frac{\delta v_3}{v_3} = \frac{\delta E_3}{E_3}$$

By substitution, the system equation can be written as:

$$C^{J_1}_{E_2} \frac{\delta v_2}{v_2} + C^{J_1}_{E_3} \frac{\delta v_3}{v_3} = 0$$

Since $\delta J_1 = 0$, it must be the case the net change in flux downstream of S must also be zero, that is $\delta v_2 + \delta v_3 = 0$, or $\delta v_2 = -\delta v_3$. We can therefore

eliminate the δv_3 term:

$$C^{J_1}_{E_2}\frac{\delta v_2}{v_2} - C^{J_1}_{E_3}\frac{\delta v_2}{v_2}\frac{v_2}{v_3} = 0$$

Canceling terms we obtain:

$$C^{J_1}_{E_2} - C^{J_1}_{E_3}\frac{v_2}{v_3} = 1$$

We can substitute the absolute rates, v_2 and v_3 with the fractional fluxes, α and $1-\alpha$ to give:

$$C^{J_1}_{E_2} - C^{J_1}_{E_3}\frac{\alpha}{1-\alpha} = 0$$

and finally:

$$C^{J_1}_{E_2}(1-\alpha) - C^{J_1}_{E_3}\alpha = 0$$

This result is called the **flux branch point theorem**. We can derive similar theorems with respect to J_2 and J_3. In each case we carry out the same thought experiment such that the reference flux, J_2 or J_3 is unchanged. The two additional theorems are given below with respect to J_2 and J_3.

$$C^{J_2}_{E_1}(1-\alpha) + C^{J_2}_{E_3} = 0$$

$$C^{J_3}_{E_1}\alpha + C^{J_3}_{E_2} = 0$$

We can also derive using the same thought experiment branch point theorems with respect to the species concentration, S. This time the systems equation is:

$$C^{S}_{E_2}\frac{\delta E_2}{E_2} + C^{S}_{E_3}\frac{\delta E_3}{E_3} = \frac{\delta S}{S} = 0$$

Substituting in the same local equations as before and noting that $\delta v_2 = -\delta v_3$ we obtain after some rearrangement:

$$C^{S}_{E_2}(1-\alpha) + C^{S}_{E_3}\alpha = 0$$

This result is known as the **concentration branch point theorem** and as can be seen it is very similar to the flux branch point theorem. There are also a set of variants that correspond to the flux branch theorems for J_2 and J_3:

$$C^S_{E_1}(1-\alpha) + C^S_{E_3} = 0$$

$$C^S_{E_1}\alpha + C^S_{E_2} = 0$$

We can write out the theorems in matrix form (See equation 9.9) using the theorems expressed in terms of J_2, this includes one summation, one connectivity and one branch theorem:

$$\begin{bmatrix} C^{J_2}_1 & C^{J_2}_2 & C^{J_2}_3 \\ C^S_1 & C^S_2 & C^S_3 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon^1_1 & 0 \\ 1 & -\varepsilon^2_1 & 1-\alpha \\ 1 & -\varepsilon^3_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

We can solve for the control coefficient matrix by rearranging:

$$\begin{bmatrix} C^{J_2}_1 & C^{J_2}_2 & C^{J_2}_3 \\ C^S_1 & C^S_2 & C^S_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon^1_1 & 0 \\ 1 & -\varepsilon^2_1 & 1-\alpha \\ 1 & -\varepsilon^3_1 & 1 \end{bmatrix}^{-1}$$

Inverting the second matrix we can derive $C^{J_2}_{E_2}$ and $C^{J_2}_{E_3}$ [20]. In the following we have simplified the notation by setting $\varepsilon_1 = \varepsilon^1_S, \varepsilon_2 = \varepsilon^2_S$, and $\varepsilon_3 = \varepsilon^3_S$. The denominator, $\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1$ is **positive**, therefore the following equalities hold given that $\varepsilon_1 < 0$, $\varepsilon_2 > 0$ and $\varepsilon_3 > 0$:

$$C^{J_2}_{E_1} = \frac{\varepsilon_2}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} > 0$$

$$C^{J_2}_{E_2} = \frac{\varepsilon_3(1-\alpha) - \varepsilon_1}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} > 0$$

$$C^{J_2}_{E_3} = \frac{-\varepsilon_2(1-\alpha)}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} < 0$$

And for the concentration control coefficients:

$$C^S_{E_1} = \frac{1}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} > 0$$

$$C^S_{E_2} = \frac{-\alpha}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} < 0$$

$$C^S_{E_3} = \frac{-(1-\alpha)}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} < 0$$

Referring to the concentration control coefficient first we note that C^S_1 is positive while the two branch coefficients, C^S_2 and C^S_3 are negative. This is as expected. The degree to which each of the output branches affects the concentration is in proportion to the amount of flux carried by the branch. This means that a branch that only carries a small amount of flux will have little effect on the branch species concentration.

Both flux control coefficients, $C^{J_2}_1$ and $C^{J_2}_2$ are positive which we would expect. The flux control coefficient, $C^{J_2}_3$ however is negative, indicating that changes in the activity of E_3 decreases the flux in the other limb, J_2. This means there is **competition** in each output branch for flux. If one branch becomes more active then it can "steal" flux from the other branch. The amount stolen will depend on the various kinetic properties of the branch enzymes. To answer what determines the competition between the output branches we must look at the control equations in more detail, in particular we must look at how the distribution of control is affected by different flux distributions and the kinetics of the branch enzymes. In the following analysis J_2 will be the flux we observe as a result of perturbations to the enzymes in the branched pathway.

Most Flux Through J_3

The first situation to consider is the case when the bulk of flux moves along J_3 and only a small amount goes through the upper limb J_2, that is $\alpha \rightarrow 0$ and $1-\alpha \rightarrow 1$ (See Figure 11.2(b)). Let us examine how the small amount of flux through J_2 is influenced by the two branch limbs, E_2 and E_3.

As $\alpha \to 0$ and $1-\alpha \to 1$, then:

$$C^{J_2}_{E_2} \quad \to \quad \frac{\varepsilon_1 - \varepsilon_3}{\varepsilon_1 - \varepsilon_3} = 1$$

$$C^{J_2}_{E_3} \quad \to \quad \frac{\varepsilon_2}{\varepsilon_1 - \varepsilon_3}$$

The first thing to note is that E_2 tends to acquire proportional influence over its own flux, J_2. Since J_2 only carries a very small amount of flux, any changes in E_2 will have little effect on S, hence the flux through E_2 is almost entirely governed by the activity of E_2. Because of the flux summation theorem and the fact that $C^{J_2}_{E_2} = 1$ it means that the remaining two coefficients must be equal and opposite in value. Since $C^{J_2}_{E_3}$ is negative, $C^{J_2}_{E_1}$ must be positive.

Unlike a linear pathway, the values for $C^{J_2}_{E_2}$ and $C^{J_2}_{E_1}$ are not bounded between zero and one and depending on the values of the elasticities it is possible for the control coefficients in a branched system to greatly exceed one [45, 51].

It is also possible to arrange the kinetic constants so that every step in the branch with respect to J_2 has a control coefficient of unity (one of which must be -1 in order to satisfy the summation theorem). We could therefore claim that **every step** in the pathway is a rate limiting step with respect to J_2. This clearly shows us again that rate limitation is not a simple concept as is traditionally supposed.

> In a branched pathway it is possible to arrange the kinetic constants of the enzymes such that the feed branch has a flux control coefficient of +1, one of the output branch a coefficient of -1 and the other output branch a coefficient of +1. That is, **every step** in the pathway is equally rate limiting.

It is also possible to arrange the kinetic constants in the pathway such that the flux coefficients for E_1 and E_3 are much greater than one. This effect has been termed ultrasensitivity [51]. The Jarnac script 11.1 in the chapter Appendix illustrates a branched pathway with control coefficients over 8.0.

$C^{J_2}_{E_1}$	8.34
$C^{J_2}_{E_2}$	0.99
$C^{J_2}_{E_3}$	-8.51

Table 11.2 Results showing high flux control coefficients in a simple branch model, see script 11.1

Table 11.2 shows the results from the Jarnac script simulation.

The explanation for these high control coefficients is straight forward. Any changes in the two limbs that carry the high flux will have an adverse effect on the very small flux that is carried by J_2. Imagine a small stream coming off a large river. Any flooding in the large river is likely to have a huge impact on the small stream.

> In a branched pathway it is possible to arrange the kinetic constants of the enzymes such the flux control coefficients in the feed and output branch can greatly **exceed one**.

Other than an asymmetric distribution of flux the ability to achieve high flux sensitivity at a branch point also depends on the relative values of the elasticities. For example increasing the value ε_2 relative to ε_3 increases the sensitivity of the branch point. This could be achieved in a number of ways:

1. E_2 can show positive cooperativity with respect to the branch species. That is any changes in E_3 become amplified through E_2.

2. v_3 is operating in a more saturated regime compared to v_2. This will make ε_3 smaller than ε_2 and amounts to ensuring that the K_m for v_2 is higher than the K_m of v_3.

3. Product inhibition on v_1 is very small.

Most Flux Through J_2

Let us now consider the other extreme, that is when most of the flux is through J_2, in other words $\alpha \rightarrow 1$ and $1 - \alpha \rightarrow 0$ (See Figure 11.2(a)). Under these conditions the control coefficients yield:

$$\begin{aligned} C^{J_2}_{E_2} &\rightarrow \frac{\varepsilon_1}{\varepsilon_1 - \varepsilon_2} \\ C^{J_2}_{E_3} &\rightarrow 0 \end{aligned}$$

In this situation the pathway has effectively become a simple linear chain. The influence of E_3 on J_2 is negligible. By analogy, changing the flow of water in a small stream that comes off a large river will have a negligible effect on the rate of flow in the large river.

Figure 11.2 summarizes the changes in sensitivities at a branch point.

11.2 Implicit Differentiation

As was previously done in Chapter 6 we can also compute the control coefficients for a branched system by implicit differentiation. We start by writing out the rate of change of S at steady state for a simple branch as follows:

$$\frac{dS}{dt} = v_1 - v_2 - v_3 = 0$$

Assuming we wish to compute the coefficients with respect to E_1, we ca write the equation as:

$$0 = v_1(S(E_1), E_1) - v_2(S(E_1)) - v_3 S(E_1))$$

Differentiating with respect to E_1 gives:

$$0 = \frac{\partial v_1}{\partial S}\frac{dS}{dE_1} + \frac{\partial v_1}{\partial E_1} - \frac{\partial v_2}{\partial S}\frac{dS}{dE_1} - \frac{\partial v_3}{\partial S}\frac{dS}{dE_1}$$

Scaling, setting $\varepsilon^1_{E_1} = 1$ and solving for $C^S_{E_1}$ yields:

$$C^S_{E_1} = \frac{1}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1}$$

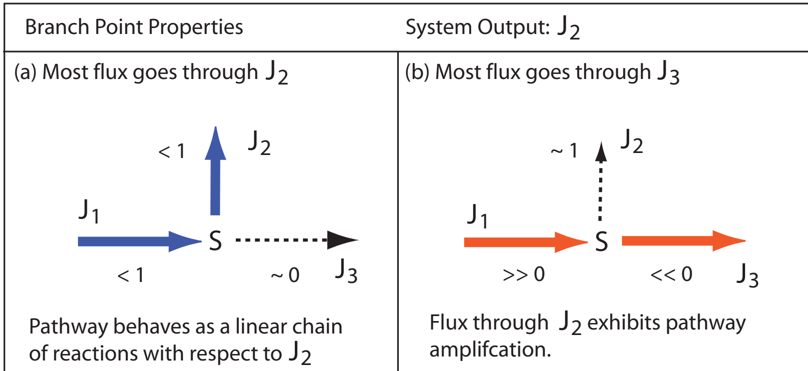


Figure 11.2 The figure shows two flux extremes relative to the flux through branch J_2. In case (a) where most of the flux goes through J_2, the branch reverts functionally to a simple linear sequence of reactions comprised of J1 and J_2. In case (b), where most of the flux goes through J_3, the flux through J_2 now becomes very sensitive to changes in activity at J_1 and J_3. Given the right kinetic settings, the flux control coefficients can become ŚultrasensitiveŠ with values greater than one (less than minus one for activity changes at J_3). The values next to each reaction indicates the flux control coefficient for the flux through J_2 with respect to activity at the reaction.

where as before $\alpha = J_2/J_1$. The control coefficients for E_2 and E_3 can be derived in a similar manner.

11.3 Futile or Substrate Cycles

Closely related to branched systems are cyclic pathways. A typical cyclic pathway in shown in Figure 11.3. For cycling to occur both forward and back reactions must operate. It is typical to find that the forward and reverse reactions are chemically distinct. Often one reaction will be driven by ATP while the other by the hydrolysis of phosphate groups. Typical examples in metabolism include the cycle between glucose and glucose-6-

phosphate and the cycling between fructose-6-phosphate and fructose 1.6-bisphosphate. Such cycles have often been called futile cycles (or better substrate cycles) because of the expenditure of free energy (as ATP) without any apparent benefit. A number of suggestions have been put forward to rationalize this apparent waste of energy. These include heat production, control of flux direction, metabolite buffering and more sensitive control of the net flux through the pathway. We will only consider the later here.

Sensitivity Control

Figure 11.3 shows a typical cyclic pathway embedded in a linear chain. Of interest is the sensitivity of the pathway flux, v_1 or v_4 to changes in

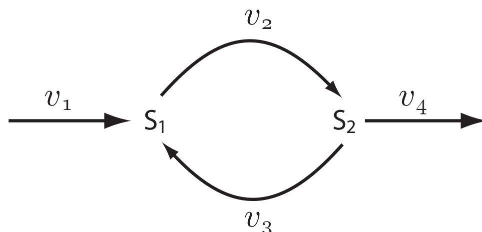


Figure 11.3 Cyclic Pathway.

v_2. The simplest assumption to make is that when we change v_2 there is no change in back flux, v_3. This could be for a number of reasons, for example v_3 is saturated by its substrate S_2.

Figure 11.4 illustrate two situations, a references state in panel a) and a perturbation of 5% to v_2 shown in panel b). Assuming that the entire flux changes appear in output flux v_4 and that v_3 is not changed, then the percentage change in v_4 (or v_1) is 100%, a twenty fold amplification.

This effect can be easily quantified as follows. First we note the flux constraint due to the cycle is:

$$v_1 = v_2 - v_3$$

We then assume that a perturbation in v_2 leads to the same change in v_1,

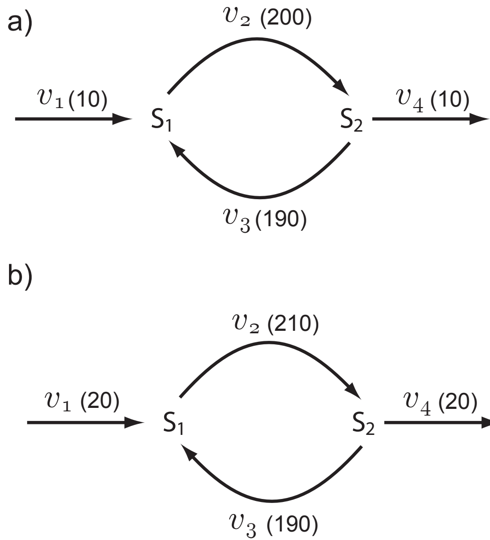


Figure 11.4 Amplification in a substrate cycle. Panel a) Reference state, values refer to fluxes at various points, note that $v_1 = v_2 - v_3$. Panel b) Activation of v_2 by 5% leads to a 100% change in v_1 and v_4. It assumes that v_3 is not activated by any changes in S_2.

that is:

$$\delta v_2 = \delta v_1$$

We can now compute the fractional changes in v_1 and v_2 as:

$$\frac{\delta v_1}{v_1} = \frac{\delta v_2}{v_2}\frac{v_2}{v_1}$$

The degree of amplification is then given by

$$\frac{\delta v_1/v_1}{\delta v_2/v_2} = \frac{v_2}{v_1}$$

Since $v_2 = v_1 + v_3$ then

$$\frac{\delta v_1/v_1}{\delta v_2/v_2} = \frac{v_1 + v_3}{v_1} = 1 + \frac{v_3}{v_1} \tag{11.1}$$

This result shows that the higher the cycling rate (v_3) compared to the through flux, the greater the amplification. This equation gives us the maximum degree of amplification possible. In practice, v_3 will not remain unchanged because S_2 rises. In addition S_1 will fall due to high consumption which will reduce v_2 but increase v_1 due to lower product inhibition. The resulting amplification is therefore a more complicated function than the one suggested by equation 11.1. However equation 11.1 gives the maximum possible ampilfication.

To carry out a more detailed analysis we must turn to metabolic control analysis. We can examine the flux control coefficient for $C_2^{J_1}$:

$$C_2^{J_1} = \frac{\varepsilon_1^1 \varepsilon_2^4 (1 + v_3/v_1)}{D}$$

$$D = \varepsilon_1^1 \varepsilon_2^4 - \left(1 + \frac{v_3}{v_1}\right)\left(\varepsilon_1^1 \varepsilon_2^2 + \varepsilon_2^4 \varepsilon_1^2\right) + \frac{v_3}{v_1}\left(\varepsilon_1^1 \varepsilon_2^3 + \varepsilon_2^4 \varepsilon_1^3\right)$$

Let us simplify this equation by assuming that there is little or no product inhibition from S_2 on to v_2 and S_1 on to v_3. This means that $\varepsilon_1^3 = 0$ and $\varepsilon_2^2 = 0$. If we also multiply top and bottom by v_1 and using the relation $v_1 + v_3 = v_2$, then we can simplify the control equation to:

$$C_2^{J_1} = \frac{\varepsilon_1^1 \varepsilon_2^4 v_2}{D}$$

$$D = \varepsilon_1^1 \varepsilon_2^4 v_1 - \varepsilon_2^4 \varepsilon_1^2 v_2 + \varepsilon_1^1 \varepsilon_2^3 v_3$$

Two things to note immediately from this equation. There must be product inhibition on the first step, ε_1^1, in order to get any sensitivity. If ε_1^1 is zero then so is $C_2^{J_1}$. This is because all control is now on the first step. This highlights again the danger of using rate laws in models that are product insensitive because the use of such rate laws often give misleading or trivial results of no real interest. The second relatively simple statement to make from the above equation is the importance of ε_2^3. This elasticity is the activation of the reverse arm with respect to S_2. The larger this elasticity the smaller the degree of amplification. This is expected because any flux that flows back along the reverse cycle instead of into v_4 reduces

the potential amplification factor. To analyze the equation further we can make additional simplifications.

We know that sensitivity increases when the cycling rate increases relative to the main flux, v_1 and v_4. If v_2 and v_3 are much greater than v_1 then we can simplify the equation further to:

$$C_2^{J_1} = \frac{v_2}{v_3 \varepsilon_2^3/\varepsilon_2^4 - v_2 \varepsilon_1^2/\varepsilon_1^1}$$

If the cycling rate is so high that v_2 and v_3 are almost indistinguishable then we can see that maximal sensitivity is achieved when:

$$\frac{\varepsilon_2^3}{\varepsilon_2^4} + \frac{\varepsilon_1^2}{\varepsilon_1^1} \ll 1$$

This tells us that substrate activation of v_4 by S_2 should be stronger than substrate activation of S_2 on v_3 and secondly that product inhibition of S_1 on v_1 must be stronger than substrate activation of S_1 on v_2. If we think about this in a thought experiment, these results are expected.

The requirements for amplification in substrate cycles is fairly complicated and questions remain whether real pathways use this mechanism in vivo.

At this point we leave the topic of branches and cycles. In a subsequent chapter we will consider the dynamic properties of conserved cycles.

Exercises

1. Given the simple branch in Figure 11.1 prove the following theorems:

$$C_{E_1}^{J_2}(1-\alpha) + C_{E_3}^{J_2} = 0$$

$$C_{E_1}^{J_3}\alpha + C_{E_2}^{J_3} = 0$$

2. Prove that the following two theorems are true for the branch point in Figure 11.1:

$$C^S_{E_1}(1-\alpha) + C^S_{E_3} = 0$$

$$C^S_{E_1}\alpha + C^S_{E_2} = 0$$

3. Derive the flux branch points for the following multibranched system:

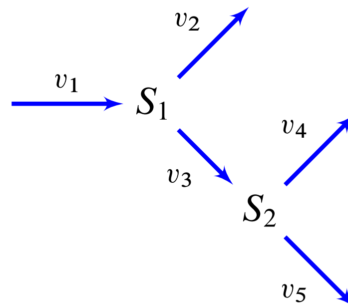


Figure 11.5 Multi-Branched Pathway.

Appendix

See Appendix **??** for more details of Jarnac.

```
p = defn cell
  var S;
  ext Xo, w;
  J1: $Xo -> S; Vm1/Km1*(Xo-S/Keq)/(1+Xo/Km1+S/Km2);
  J2: S -> $w; Vm2*S^4/(Km3+S^4);
  J3: S -> $w; Vm3*S/(Km4+S);
end;
p.Xo = 9;
p.S = 0.2;
p.Vm1 = 1.4;
```

```
p.Km1 = 0.4;
p.Keq = 4.5;

p.Km2 = 0.6;
p.Vm2 = 0.05;
p.Km3 = 0.8;
p.Vm3 = 2.3;
p.Km4 = 0.3;

// Due to the high sensitivity, change the evaluation
// method to a five-point difference method (default is three).
// Also decrease the step size to improve accuracy.
p.diffstepsize = 0.01; p.diffmethod = 1; p.ss.tol = 1E-9;
p.ss.eval;
println "Flux Control Coefficients:";
println p.cc (<p.J2>, p.Vm1);
println p.cc (<p.J2>, p.Vm2);
println p.cc (<p.J2>, p.Vm3);
println "Elasticities:";
e1 = p.ee (<p.J1>, p.S);
e2 = p.ee (<p.J2>, p.S);
e3 = p.ee (<p.J3>, p.S);
println e1, e2, e3;
println "Fluxes: ", p.J1, p.J2, p.J3;
```

Listing 11.1 Simple Branched Pathway showing Flux Amplification

12
Moiety Conservation Laws

12.1 Moiety Constraints

Many cell processes operate on different time scales. For example, metabolic processes tend to operate on a faster time scale than protein synthesis and degradation. Such time scale differences have a number of implications to model builders, software designers and model behavior. In this chapter we will examine these aspects in relation to species conservation laws. To introduce this topic consider a simple protein phosphorylation cycle such as the one shown in Figure 12.1. This shows a protein undergoing phosphorylation (upper limb) and dephosphorylation (lower limb) via a kinase and phosphatase respectively.

The depiction in Figure 12.1 is however a simplification. The ATP used during phosphorylation is not shown as well as the release of free phosphate during the dephosphorylation. In addition synthesis and degradation of protein is also absent. In many cases we can leave these aspects out of the picture. ATP for instance is held at a relatively constant level by strong

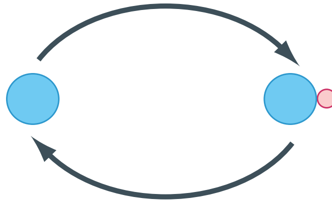


Figure 12.1 Phosphorylation and Dephosphorylation Cycle forming a Moiety Conservation Cycle between Unphosphorylated (left species) and Phosphorylated protein (right species).

homeostatic forces from metabolism so that within the context of the cycle, changes in ATP isn't something we need worry about. More interestingly is that within the time scale of phosphorylation and dephosphorylation we can assume that the rate of protein synthesis and degradation is negligible. This assumption leads to the emergence of a new property of the cycle called **moiety conservation** [67].

In chemistry a moiety is described as a subgroup of a larger molecule. In Figure 12.1 the moiety is a protein. During the interconversion between the phosphorylated and unphosphorylated protein, the amount of moiety (protein) remains constant. More abstractly we can draw a cycle in the following way (Figure 12.2), where S_1 and S_2 are the cycle species:

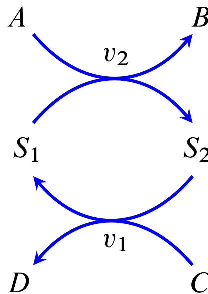


Figure 12.2 Simple Conserved cycle where $S_1 + S_2 =$ constant.

The two species, S_1 and S_2 are conserved because the total $S_1 + S_2$ re-

mains constant over time (at least over a time scale shorter than protein synthesis and degradation). Such cycles are collectively called **conserved cycles**.

Protein signalling pathways abound with conserved cycles such as these although many are more complex and may involve multiple phosphorylation reactions. In addition to protein networks other pathways also possess conservation cycles. One of the earliest conservation cycles to be recognized was the adenosine triphosphate (ATP) cycle. ATP is a chain of three phosphate residues linked to a nucleoside adenosine group, Figure 12.3.

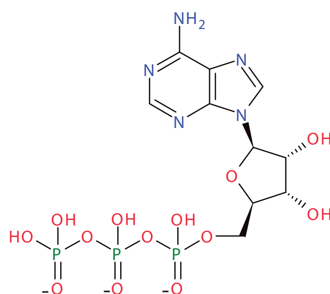
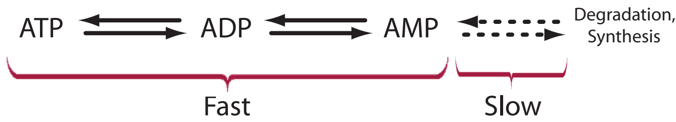


Figure 12.3 Adenosine Triphosphate: Three phosphate groups plus an adenosine subgroup.

The linkage between the phosphate groups involves phosphoric acid anhydride bonds that can be cleaved by hydrolysis one at a time leading in turn to the formation of adenosine diphosphate (ADP) and adenosine monophosphate (AMP) respectively. The hydrolysis provides much of the free energy to drive endergonic processes in the cell. Given the insatiable need for energy, there is a continual and rapid interconversion between ATP, ADP and AMP as energy is released or captured. One thing that is constant during these interconversions is the amount of adenosine group (Figure 12.4). That is adenosine is a conserved moiety. Over longer time scales there is also the slower process of AMP degradation and biosynthesis via the purine nucleotide pathway but for many models we assume that this process is negligible compared to ATP turn over by energy metabolism.



There are many other examples of conserved moieties such enzyme/enzyme-substrate complexes, NAD/NADH, phosphate and coenzyme A. In all these cases the basic assumption is that the interconversions of the subgroups is rapid compared to their net synthesis and degradation. We should emphasize that in reality conserved moieties do not exist since all molecular subgroups will at some point be subject to synthesis and degradation. However, over sufficiently short time scales, the sum total of these groups can be considered constant. In this chapter we will consider conserved moieties in detail. In particular we will look at how to detect them in our models and how they influence the design of simulation software. We will wait until Chapter 13 to discuss their effect on pathway dynamics.

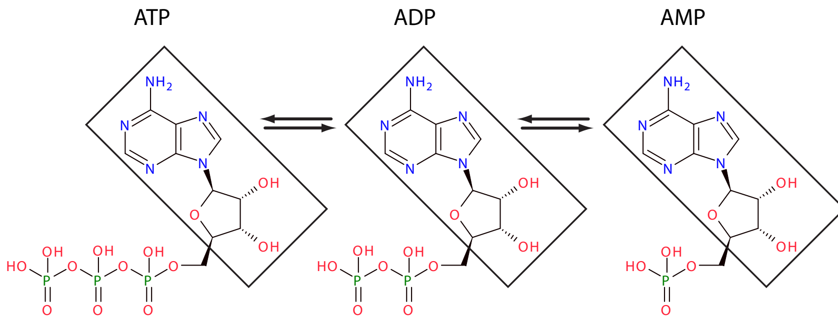


Figure 12.4 The adenosine moiety, indicated by the boxed molecular group, is conserved during the interconversion of ATP, ADP and AMP.

Moiety:	A subgroup of a larger molecule.
Conserved Moiety:	A subgroup whose interconversion through a sequence of reactions leaves it unchanged.

12.2 Moiety Conserved Cycles

Any chemical group that is preserved during a cyclic series of interconversions is called a **conserved moiety**. Examples of conserved moiety subgroups include species such as phosphate, acyl, nucleoside groups or covalently modifiable proteins, As a moiety gets redistributed through a network, the **total amount** of the moiety is constant and does not change during the time evolution of the system. For any particular subgroup, the total amount is determined solely by the initial conditions imposed on the model.

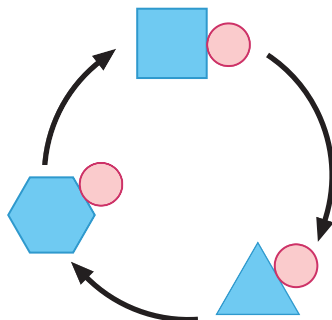


Figure 12.5 Conserved Moiety in a Cyclic Network. The blue species are modified as they traverse the reaction cycle, but the red subgroup (small circle) remains unchanged. This creates a conserved cycle, where the total number of moles of moiety (red subgroup) stays constant.

There are rare cases when a 'conservation' relationship arises out of a non-moiety cycle. This does not affect the mathematic analysis but only the physical interpretation of the relationship. For example, in Figure 12.6 the constraint $B - C = T$ applies even though there is no moiety involved.

The presence of conserved moieties is an approximation introduced into a model, however, over the time scale in which the conservations hold, their existence can have a profound effect on the dynamic behavior of the model. For example the hyperbolic response of a simple enzyme (in the

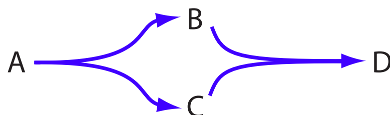


Figure 12.6 Conservation due to stoichiometric matching. In this system, $B - C =$ constant.

form of enzyme conservation between E and ES), or the sigmoid behavior observed in protein signalling networks is due in significant part to moiety conservation laws (see section 13).

Figure 12.7 illustrates the simplest possible network which displays a conserved moiety, the total mass, $S_1 + S_2$ is constant during the evolution of the network.

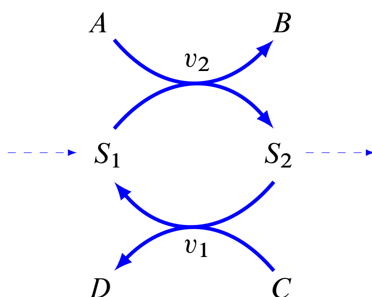


Figure 12.7 Simple Conserved cycle. The dotted lines signify negligible levels of synthesis and degradation, therefore over short time scales, $S_1 + S_2 =$ constant.

The system equations for the simple conserved cycle are easily written

down as:

$$\frac{dS_1}{dt} = v_1 - v_2$$

$$\frac{dS_2}{dt} = v_2 - v_1$$

From these equations it should be evident that the rate of appearance of S_1 must equal the rate of disappearance of S_2, that is $dS_1/dt = -dS_2/dt$. This means that when ever S_1 changes, S_2 must change in the opposite direction by **exactly** the same amount. During a simulation the sum of S_1 and S_2 will therefore remain unchanged.

Computationally we need only explicitly evaluate one of the differential equations because the other one can be computed from the conservation relation. Whichever differential equation is chosen however, the species left out must be computed algebraically using the conservation law. Therefore, the system can be reduced to one differential and one linear algebraic equation compared to the two differential equations in the original formulation.

$$S_2 = T - S_1$$

$$\frac{dS_1}{dt} = v_1 - v_2$$

The term T in the algebraic equation shown above refers to the total amount of S_1 and S_2. This value is computed from the initial amounts given to S_1 and S_2 at the start of a simulation.

12.3 Basic Theory

The question we want to address here is how to determine whether a given network contains conserved cycles and if so what are they. The key to this question is the stoichiometry matrix, $\mathbf{N}$. In the example shown in Figure 12.7 the stoichiometry matrix is given by:

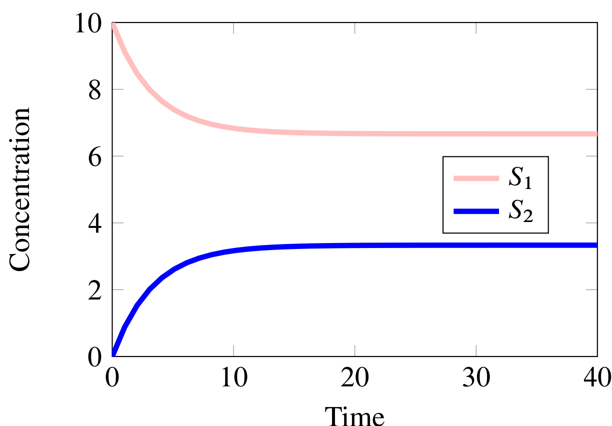


Figure 12.8 Simulation of the simple cycle shown in Figure 12.7. The total moiety remains constant at 10 concentration units. Model: `S1 -> S2; k1*S1; S2 -> S1; k2*S2; S1 = 10; k1=0.1; k2=0.2`

$$\mathbf{N} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The first thing to note is that since either row can be derived from the other by multiplication by -1, the rows are called **linearly dependent rows**, (See Box 7.0) and the **rank** of the matrix is therefore 1 (See Box 7.1). It is these dependencies that appear as linear relationships between the rates of change, dS/dt.

Whenever a network exhibits conserved moieties, there will be dependencies among the rows of $\mathbf{N}$, and the rank of $\mathbf{N}$ $rank(\mathbf{N})$, will be less than m, the number of rows of $\mathbf{N}$. The rows of $\mathbf{N}$ can be rearranged so that the first $rank(\mathbf{N})$ rows are linearly independent. The metabolites which correspond to these rows are called the **independent species** ($\boldsymbol{S_i}$). The remaining $m - rank(\mathbf{N})$ rows correspond to the **dependent species** ($\boldsymbol{S_d}$).

In the simple conserved cycle, Figure 12.7, there is one independent species, S_1 and one dependent species, S_2.

Box 7.0 Linear Dependence and Independent - Recap

One of the most important ideas in linear algebra is the concept of linear dependence and independence. Take three vectors, say $[1, -1, 2]$, $[3, 0, -1]$ and $[9, -3, 4]$. If we look at these vectors carefully it should be apparent that the third vector can be generated from a combination of the first two, that is $[9, -3, 4] = 3[1, -1, 2] + 2[3, 0, -1]$. Mathematically we say that these vectors are *linearly dependent*.

In contrast, the following vectors, $[1, -1, 0]$, $[0, 1, -1]$ and $[0, 0, 1]$, are independent because there is no combination of these vectors that can generate even one of them. Mathematically we say that these vectors are *linearly independent*.

Example 12.1

Figure 12.5 illustrates a three species cycle. What is the conservation law for this pathway? The stoichiometry matrix for this system is given by:

$$\mathbf{N} = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 \end{matrix} \\ \left[\begin{matrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{matrix} \right] \end{array} \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} \tag{12.1}$$

Inspection reveals that the sum of the three rows is zero meaning that

$$\frac{dS_1}{dt} + \frac{dS_1}{dt} + \frac{dS_1}{dt} = 0$$

or that the total $S_1 + S_2 + S_3$ is constant. There are no other relationships between the rows other than this one.

Example 12.2

A linear pathway has the following stoichiometry matrix:

$$\mathbf{N} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Does the pathway contain any conserved cycles? No, because neither row in the matrix can be derived from the other by a simple operation, the rows are linearly independent, therefore the pathway has no conserved cycles.

To illustrate this idea on a more complicated example, consider the pathway shown in Figure 12.9. This pathway includes four species, S_1, S_2, E and ES.

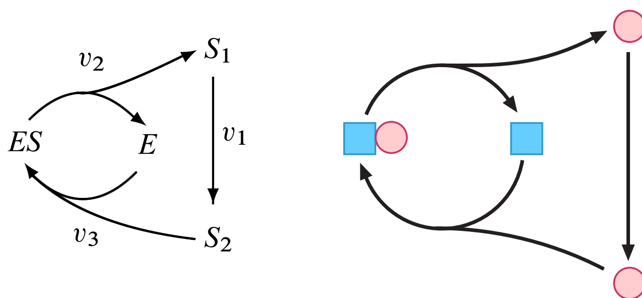


Figure 12.9 Linked Conserved Cycles. The network rendered on the right shows the moiety composition of the participating species.

The mass-balance equations of this model can be written down as:

$$\frac{dE}{dt} = v_2 - v_3 \qquad \frac{dES}{dt} = v_3 - v_2$$

$$\frac{dS_1}{dt} = v_2 - v_1 \qquad \frac{dS_2}{dt} = v_1 - v_3$$

A visual inspection of the mass-balance equations reveals the following two relationships:

$$\begin{aligned} \frac{dE}{dt} + \frac{dES}{dt} &= 0 \\ \frac{dES}{dt} + \frac{dS_1}{dt} + \frac{dS_2}{dt} &= 0 \end{aligned} \tag{12.2}$$

These relationships tell us that there are two conservation laws, $E + ES$ and $ES + S_1 + S_2$. This means that given the amount of ES, the amount

Box 7.1 The Rank of a Matrix - Recap

Closely related to linear independence (Box 7.0) is the concept of Rank. Consider the three vectors described in Box 7.0, $[1, -1, 2]$, $[3, 0, 1]$ and $[9, -3, 4]$ and stack them one atop each other to form a matrix:

$$\begin{bmatrix} 1 & -1 & 2 \\ 3 & 0 & 1 \\ 9 & -3 & 4 \end{bmatrix}$$

then the Rank is simply the number of linear independent vectors that make up the matrix. In this case the Rank is 2, because there are only two linear independent row vectors in the matrix.

of E can be computed. In addition, given the amount of ES and S_1, the amount of S_2 can be computed. Therefore ES and S_1 can be designed the independent species and E and S_2 the dependent species. What this means in practical terms is that in a modeling program only two differential equations need be solved instead of four. The reduced model equations will look like:

$$E = T_1 - ES$$

$$S_2 = T_2 - S_1 - ES$$

$$\frac{dES}{dt} = v_3 - v_2$$

$$\frac{dS_1}{dt} = v_2 - v_1$$

where T_1 is the total amount of E type moiety and T_2 is the total amount of S type moiety.

The stoichiometry matrix for the model in Figure 12.9 is given by:

$$\mathbf{N} = \begin{array}{c} \\ \end{array}\begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{array}{l} S_2 \\ ES \\ S_1 \\ E \end{array} \quad \text{(columns: } v_1, v_2, v_3\text{)} \tag{12.3}$$

Examining the stoichiometry matrix reveals conservation laws as relationships among the matrix rows. The 4th row (E) can be formed by multiplying the 2nd row (ES) by -1, and the 3rd row (S_1) can be formed by multiplying the first row by -1 and adding it to the 4th row (ES).

These simple examples show that it is possible to derive conservation laws by looking for dependencies among the rows of the stoichiometry matrix. For simple cases this can be done by inspection but for large pathways this approach is not practical. Instead a more systematic theory for deriving the conservation laws must be developed.

12.4 Computational Approaches

There are a number of related methods for computing the conservation laws of a given pathway, some are simple such as the one shortly to be described, while others are more sophisticated and are used to determine the conservation laws in very large stoichiometry matrices.

The easiest method to derive conservation laws is to use row reduction [63, 15, 16]. This is based on forward elimination which is the first part of Gaussian Elimination. Gaussian Elimination is a traditional way to solve simultaneous linear equations by eliminating one unknown at a time and is a technique often taught in high school. Elimination is carried out by applying a series of simple manipulations called elementary operations. These operations include interchanging two equations (exchange), multiplying an equation through by a nonzero number (scaling) and adding an equation one or more times to another equation (replacement). In practice the equations are recast into a matrix form so that the elementary operations are applied to the values in the matrix where each row of the matrix

represents an equation. Thus interchanging two equations is equivalent to swapping two rows in the matrix. The elementary operations are carried out on the matrix until a particular arrangement, called the echelon form, is established (See Box 7.3).

Elementary operations are often represented in matrix form and are then called elementary matrices (See Box 7.2). Applying a particular elementary operation then becomes equivalent to multiplying by an elementary matrix.

The technique for finding conservation laws works as follows. Consider the network in Figure 12.9. The system equation for this network is:

$$\begin{matrix} S_2 \\ ES \\ S_1 \\ E \end{matrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} dS_2/dt \\ dES/dt \\ dS_1/dt \\ dE/dt \end{bmatrix}$$

We will recast the equation in the following form where an identity matrix has been added to the right-hand side.

$$\mathbf{N}\boldsymbol{v} = \boldsymbol{I}\frac{d\mathbf{s}}{dt}$$

Written out fully the system equation will look like:

$$\begin{matrix} S_2 \\ ES \\ S_1 \\ E \end{matrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dS_2/dt \\ dES/dt \\ dS_1/dt \\ dE/dt \end{bmatrix}$$

Let us now apply forward elimination to the stoichiometry matrix. To do this we apply a series of elementary operations to the left-hand side such that the stoichiometry matrix is reduced to echelon form. For consistency we apply the same set of elementary operations to the right-hand side so that the identity matrix records whatever operations we carried out. This amounts to multiplying both sides by a set of elementary matrices. We

Box 7.2 Elementary Matrices - Recap

Elementary matrix operations such as row exchange, row scaling or row replacement can be represented by simple matrices called elementary matrices, called Type I, II and III respectively. Elementary matrices can be constructed from the identity matrix. For example a scaling operation can be represented out by replacing one of the elements of the main diagonal of an identity matrix by the scaling factor. The following matrix represents a type II matrix which will scale the second row of a given matrix by the factor k:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Type I elementary matrices will exchange two given rows in a given matrix and are constructed from an identity matrix where rows in an identity matrix are exchanged that correspond to the rows exchanged in the target matrix. The following type I matrix will exchange rows 2 and 3 in a target matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Type III elementary matrices will add/subtract a given row in a target matrix to another row in the same matrix. Type III matrices are constructed from an identity matrix where a single off diagonal element is set to the multiplication factor and the specific location represents the two rows to combine. If an elementary matrix adds a row i to a row j multiplied by a factor α, then the identity matrix with entry i, j is set to α. In the following example, the type III elementary matrix will subtract five times the 2nd row from the 3rd row.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -5 & 0 \end{bmatrix}$$

A particularly important property of elementary matrices is that they can all be inverted. In addition, pre-multiplying by an elementary matrix will modify the rows of a target matrix while post-multiplying will operate on the columns.

only need to reduce the matrix to its row echelon form not to its reduced echelon form.

Reducing a matrix to echelon form raises the possibility of generating zero rows in the matrix if there are dependencies in the rows (See Box 7.3).

This being the case the system equation after forward elimination can be expressed in the following way:

$$\begin{bmatrix} \boldsymbol{M} \\ \mathbf{0} \end{bmatrix} \boldsymbol{v} = \boldsymbol{E} \frac{d\mathbf{s}}{dt} \tag{12.4}$$

where the identity matrix has been shown transformed into the matrix $\boldsymbol{E}$ which represents the product of all elementary operations that were applied to the left-hand side. The left-hand side has itself been transformed into an echelon form which is represented as a partitioned matrix. The $\boldsymbol{E}$ matrix can also be partitioned row-wise to match the partitioning in the echelon matrix, that is:

$$\begin{bmatrix} \boldsymbol{M} \\ \mathbf{0} \end{bmatrix} \boldsymbol{v} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \frac{d\mathbf{s}}{dt} \tag{12.5}$$

Multiplying out the lower partition one obtains:

$$\boldsymbol{Y} \frac{d\mathbf{s}}{dt} = \mathbf{0} \tag{12.6}$$

This general result is equivalent to the equations shown in 12.2, that is 12.6 represents the set of conservation laws. Determining the conservation laws therefore involves reducing the stoichiometry matrix and extracting the lower portion of the modified identity matrix.

Let us now proceed with an example to illustrate this method. We will use the stoichiometry matrix from equation 12.3. For convenience the stoichiometry and identity matrix are placed next to each other in the following sequence of elementary operations. An elementary operation carried out on the stoichiometry matrix is simultaneously applied to the identity matrix.

Box 7.3 Echelon Forms - Recap

There are two kinds of matrices that one frequently encounters in the study of linear equations. These are the **row echelon** and **reduced echelon forms**. Both matrices are generated when solving sets of linear equations. The row echelon form is derived using forward elimination and the reduced echelon form by Gauss-Jordan Elimination.

A **row echelon matrix** is defined as follows:

1. All rows that consist entirely of zeros are at the bottom of the matrix.
2. In each non-zero row, the first non-zero entry is a 1, called the leading one.
3. The leading 1 in each row is to the right of all leading 1's above it. This means there will be zeros below each leading 1.

The following three matrices are examples of row echelon forms:

$$\begin{bmatrix} 1 & 4 & 3 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 5 & 3 & 0 \\ 0 & 1 & 7 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The **reduced echelon form** has one additional characteristic:

4. Each column that contains a leading one has zeros above and below it. The following three matrices are examples of reduced echelon forms:

$$\begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Sometimes the columns of a reduced echelon can be ordered such that each leading one is immediately to the right of the leading one above it. This will ensure that the leading 1's form an identity matrix at the front of the matrix. The reduced echelon form will therefore have the following general block structure:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{A} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

It is always possible to reduce any matrix to its echelon or reduced echelon form by an appropriate choice of elementary operations. The function `rref()` implemented in many math applications will generate a reduced row echelon.

1. Stoichiometry matrix on the left and identity matrix on the right.

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2. Add the 1st row to the third row to yield:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. Add the 2nd row to the third and forth rows to yield:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

4. Multiply the second row by -1 to yield the final echelon form:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

The final operation achieves the goal of reducing the stoichiometry matrix to an echelon form (in this case it happens to be a reduced echelon form). Note that the operation has resulted in two zero rows appearing in the reduced stoichiometry matrix. These two rows correspond to the $\boldsymbol{Y}$ partition in equation 12.5. The lower two rows can be extracted from the right-hand matrix (what was once the identity matrix) to construct equation 12.6, thus

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} dS_2/dt \\ dES/dt \\ dS_1/dt \\ dE/dt \end{bmatrix} = 0$$

Or:

$$\frac{dS_2}{dt} + \frac{dES}{dt} + \frac{dS_1}{dt} = 0$$

$$\frac{dES}{dt} + \frac{dE}{dt} = 0$$

From the above equations the following conservation laws should be evident:

$$\begin{aligned} S_2 + ES + S_1 &= T_1 \\ ES + E &= T_2 \end{aligned} \tag{12.7}$$

In summary the algorithm for deriving the conservation laws is as follows:

1. Apply elementary operations to the stoichiometry matrix until the matrix is reduced to its row echelon form. Simultaneously apply the elementary operations to an identity matrix. The size of the identity matrix should be equal to the number of rows in the stoichiometry matrix.

2. If there are zero rows at the bottom of the reduced stoichiometry matrix then there are conservation laws in the network otherwise there are not. The number of conservation laws will be equal to the number of zero rows.

3. Extract the rows in the transformed identity matrix that correspond to the position of the zero rows in the reduced stoichiometry matrix. The extracted rows represent the conservation laws.

There are two points worth making when applying this algorithm. The first is that any row swaps made using the row reduction in the stoichiometry matrix will not translate to swaps in the names of the species on the right-hand side of the equation. This means that when reading the conservation rows, the names on the columns are not changed by any row exchanges in the stoichiometry matrix. The second point to make is that when carrying out the elementary row operations, it is recommended to eliminate, whenever possible, terms below a leading entry by adding rather than subtracting. This will ensure that entries in the transforming identity matrix remain positive and that the resulting conservation laws will be made up of positive terms. Sometimes the ability to add will not be possible and

subtractions will be necessary. This will result in negative terms appearing in the conservation laws which may make them more difficult to interpret physically.

A useful strategy that can be used to avoid negative terms in the conservation equations is to order the rows of the stoichiometry matrix such that any species that is likely to appear in more than one conservation relationship should be placed at the bottom of the stoichiometry matrix. In the case of the previous example we would make sure that ES is located to the bottom row of the stoichiometry matrix. This ordering ensures that the independent species (top rows) are represented by the free variables and the dependent species (bottom rows) by the shared variables. This means that the shared or dependent variables (i.e. complexes) will then be a function of the free variables which is more likely to result in positive terms [73]. A more brute force method is to try all permutations of the matrix rows until a positive set of conservation laws is found. For small models (< 10 species) this approach is a viable option.

Although it is possible to manually reduce a stoichiometry matrix, it is far easier to use specialized math software such Scilab, Octave, Matlab and Mathematica or even advanced modern desktop calculators. All these tools offer a `rref()` command for generating a reduced row echelon. The following examples will illustrate the use of the freely available Scilab application (www.scilab.org) to compute the conservation laws.

Example 12.3

Row reduction using Scilab/Matlab. Given the following stoichiometry matrix, use Scilab functions to row reduce and extract the conservation laws.

$$\mathbf{N} = \begin{matrix} S_2 \\ ES \\ S_1 \\ E \end{matrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Enter the stoichiometry matrix into the software:

```
-->n = [1 0 -1; 0 -1 1; -1 1 0; 0 1 -1];
```

Augment the matrix with the identity matrix, this will allow us to record row reduction operations in the identity matrix part of the augmented matrix.

```
-->ni = [n, eye(4,4)]
ni  =
    1.    0.  - 1.    1.    0.    0.    0.
    0.  - 1.    1.    0.    1.    0.    0.
  - 1.    1.    0.    0.    0.    1.    0.
    0.    1.  - 1.    0.    0.    0.    1.
-->
```

Row reduce the augmented matrix:

```
-->rni = rref (ni)
 rni  =
    1.    0.  - 1.    0.    0.  - 1.    1.
    0.    1.  - 1.    0.    0.    0.    1.
    0.    0.    0.    1.    0.    1.  - 1.
    0.    0.    0.    0.    1.    0.    1.
```

The left partition of the reduced matrix contains two zero rows, therefore there are two conservation laws. These laws correspond to the two bottom rows in the right partition. We extract the rows in the right partition to yield:

```
-->c = rni(3:4,4:7)
 c  =
    1.    0.    1.  - 1.
    0.    1.    0.    1.
```

The species column order is the same as the species row order in the original matrix, that is S_2, ES, S_1 and E, therefore:

$$
\begin{aligned}
S_2 + S_1 - E &= T_1 \\
ES + E &= T_2
\end{aligned}
$$

Note the negative E term in the first conservation law. At first glance this does not appear to be the same set of conservation laws that were derived earlier. However, if we substitute E from the second equation into the first we will get the same set of conservation laws: $S_1 + S_2 + ES = T$, showing us that the two sets are identical. To avoid negative terms appearing in the conservation laws, we can use the rule that all complex species (that is shared species), such as ES be moved to the bottom of the matrix (See next example).

Example 12.4

Row reduction using Scilab/Matlab. Given the following stoichiometry matrix, use Scilab functions to row reduce and extract the conservation laws. In this example, the shared species ES has been moved to the bottom of the matrix.

$$\mathbf{N} = \begin{array}{l} S_2 \\ S_1 \\ E \\ ES \end{array} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

The reduced augmented matrix is now:

```
-->rni = rref (ni)
 rni  =
    1.    0.  - 1.    0.  - 1.    0.  - 1.
    0.    1.  - 1.    0.    0.    0.  - 1.
    0.    0.    0.    1.    1.    0.    1.
    0.    0.    0.    0.    0.    1.    1.
```

Once again there are two zero rows but this time the corresponding conservation laws all have positive entries, yielding the following equations:

$$\begin{aligned} S_2 + S_1 + ES &= T_1 \\ ES + E &= T_2 \end{aligned}$$

The following Scilab/Matlab code will find the conservation laws for any stoichiometry matrix.

```
// Compute Conservation Laws
// -------------------------

// Enter the stoichiometry matrix first

n = [1 0 -1; 0 -1 1; -1 1 0; 0 1 -1];
nRows = size(n, 1);
// Create the augmented matrix
ni = [n, eye(nRows,nRows)];
// Carry out row reduction
rni = rref (ni);
r = rank (n);
// Extract the conservation rows
c = rni(r+1:nRows,size(n,2)+1:size(ni,2));
// Display result
c
```

Figure 12.10 General purpose Scilab/Matlab code to determine conservation laws using row reduction.

Row reduction of the augmented stoichiometry is probably the easiest way to derive the conservation laws. The main advantage of this method includes simplicity and significantly the ability to direct the calculation by setting the order of rows in the initial stoichiometry. However it has one disadvantage which is potential numerical instability for large systems. In particular for large genomic style stoichiometry models [62] that involve many hundreds or even thousands of reactions and species, the method can suffer dramatic failures due to rounding errors during row reduction. In a subsequent section more robust methods will be described that rely on QR factorization [84] and Singular Value Decomposition (SVD). The main disadvantage of these other methods is that sometimes, depending on the particular algorithm, the row order can not be easily prescribed. In any event there are some simple tests one can do to check that the computed conservation laws are correct, one such test will be described next.

Null Space of $\mathbf{N}^{\mathrm{T}}$

To complete this section let us consider in more detail the algebraic nature of the $\boldsymbol{Y}$ partition in equation 12.6.

The elementary matrix, $\boldsymbol{E}$, reduced the stoichiometry matrix to a row echelon form, that is to:

$$\boldsymbol{E}\mathbf{N} = \begin{bmatrix} \boldsymbol{M} \\ \mathbf{0} \end{bmatrix} \tag{12.8}$$

The $\boldsymbol{E}$ matrix corresponds to the same $\boldsymbol{E}$ matrix in equation 12.5, so that we can partition the elementary matrix, $\boldsymbol{E}$ row-wise into $\boldsymbol{X}$ and $\boldsymbol{Y}$ partitions (equation 12.5).

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \mathbf{N} = \begin{bmatrix} \boldsymbol{M} \\ \mathbf{0} \end{bmatrix}$$

From which we can immediately see that:

$$\boldsymbol{Y}\mathbf{N} = \mathbf{0}$$

Taking the transpose we obtain

$$\mathbf{N}^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}} = \mathbf{0}$$

The $\boldsymbol{Y}$ partition is therefore the null space of the transpose of the stoichiometry matrix [1]. This is a significant result for a number of reasons. It gives a very concise definition of the conservation matrix but more importantly it opens up the possibility of using other computational approaches.

The other point of interest is that this result can be used to test whether a set of conservation laws were correctly derived or not. To do this we simply multiply the transpose of $\mathbf{N}$ by the transpose of the conservation matrix $\boldsymbol{Y}$ and make sure the product equals zero.

[1] *cf.* Chapter 4, Section Computing the Null Space in Introduction to Linear Algebra for Systems Biology, Sauro)

Many software packages such as Matlab, Scilab or Mathematica supply commands to compute the null space. This makes is easy to compute the conservation laws by simply computing the null space of the transpose of the stoichiometry matrix. For example the following session shows how we can use Scilab to compute the conservation laws for the example matrix we used in previous examples.

```
-->N = [1 0 -1; -1 1 0; 0 1 -1; 0 -1 1]
 N  =
    1.    0.  - 1.
    0.  - 1.    1.
  - 1.    1.    0.
    0.    1.  - 1.
--> ns = kernel (N')
ans  =
    0.          0.6324555
    0.          0.6324555
    0.7071068  - 0.3162278
    0.7071068    0.3162278
--> // Convert the orthonormal set
--> // into a rational basis using rref
-->rref (ns')'
ans  =
   1.    0.
   1.    0.
   0.    1.
   1.    1.
```

The null space command in Scilab is `kernel`, in Matlab it is `null` and in Mathematical it is `NullSpace`. Like many null space commands implemented in mathematical software, the kernel command in Scilab has the drawback of generating an orthonormal set. In order to generate a rational basis we must row reduce the kernel, this results in a more interpretable set of conservation laws. In Matlab it is possible to use the modified null space command, `null (N, 'r')` which will automatically generate a rational basis (Neither Octave or Scilab support this format). Interestingly, Mathe-

matica's (7.0) null space function does generate a rational basis, however, the algorithm that Mathematica uses is unknown which raises its own issues.

Given that we can now compute the conservation laws for arbitrary networks, one question to consider is whether conservation laws have any behavioral consequences. The answer to this question will be considered in Chapter 13.

12.5 Summary

Of particular interest is to compare these results with equation 4.8. Whereas the flux balance relationships are derived from the stoichiometry matrix, the moiety conservation laws are derived from the transpose of the stoichiometry matrix. Thus to summarize:

Moiety Conservation Laws:

$$\begin{bmatrix} -\boldsymbol{L_0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{N_R} \\ \boldsymbol{N_0} \end{bmatrix} = \mathbf{0} \qquad \mathbf{N}^\top \boldsymbol{\Gamma}^\top = \mathbf{0}$$

Flux Balance Laws:

$$\begin{bmatrix} \boldsymbol{N_{DC}} & \boldsymbol{N_{IC}} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{K_0} \end{bmatrix} = \mathbf{0} \qquad \boldsymbol{N_R}\,\boldsymbol{K} = \mathbf{0}$$

Further Reading

1. Hofmeyr JH, Kacser H, van der Merwe KJ. (1986) Metabolic control analysis of moiety-conserved cycles. Eur J Biochem. 155(3):631-41.

2. Sauro HM and Ingalls B (2004) Conservation analysis in biochemical networks: computational issues for software writers, Biophys Chem, 109, 1–15

3. Sauro HM Introduction to Linear Algebra for Systems Biology

4. Vallabhajosyula RR, Chickarmane V and Sauro HM (2006) Conservation analysis of large biochemical networks, Bioinformatics, 22(3), 346-353

5. Cornish-Bowden and Hofmeyr J-HS (2002) The Role of Stoichiometric Analysis in Studies of Metabolism: An Example J. theor. Biol 216, 179-191

13

Moiety Conserved Cycles

13.1 Moiety Conserved Cycles

In this chapter the topic of moiety conserved cycles and their impact on behavior will be examined. There are some more advanced concepts in this chapter which require some prerequisite knowledge of bistability. The reader is refereed to the Appendix for a brief review of bistability. Alternatively the section that requires knowledge of bistability can be omitted.

In Chapter **??** we introduced the idea of moiety conceived cycles. To recap, Figure 13.1 shows a typical moiety-conserved cycle. In this cycle there are two species, one might be a protein and the other the phosphorylated protein. If we assume that the synthesis and degradation rates for the protein is low in comparison to the dynamics of phosphorylation and dephosphorylation, then we can assume that the total amount of protein, phosphorylated plus unphosphorylated is constant. Figure 13.2 shows a more abstract form of a moiety-conserved cycle.

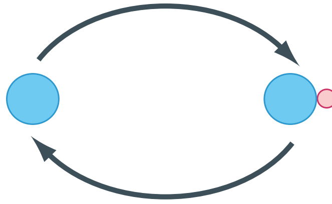


Figure 13.1 Phosphorylation and Dephosphorylation Cycle forming a Moiety Conservation Cycle between Unphosphorylated (left species) and Phosphorylated protein (right species).

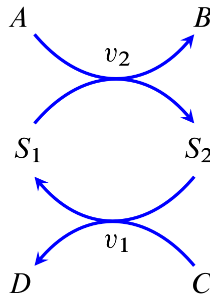


Figure 13.2 Simple Conserved cycle where $S_1 + S_2 =$ constant.

13.2 Species Limits

One of the simplest effects of a moiety-conserved cycle is that it puts upper limits on the concentrations of the participants. In the simple cycle shown in Figure 13.2 where the total amount of mass in the cycle is fixed at $S_1 + S_2$, the upper limit that either S_1 of S_2 can reach is $S_1 + S_2$. This effect was made very clear in a study of glycolysis in Trypanosoma brucei (Figure 13.3). What is unusual about the pathway is that much of glycolysis resides in a single membrane organelle called the glycosome. Many of the metabolites in the glycosome are phosphorylated, for example, glucose-6-phosphate, glyceraldehyde-3-phosphate and it is this that creates a constraint on the levels of phosphate. In addition to the glyco-

some, the mitochondrion of Trypanosoma brucei appears to do very little other than oxidizing the glycerol 3-phosphate via oxygen utilization.

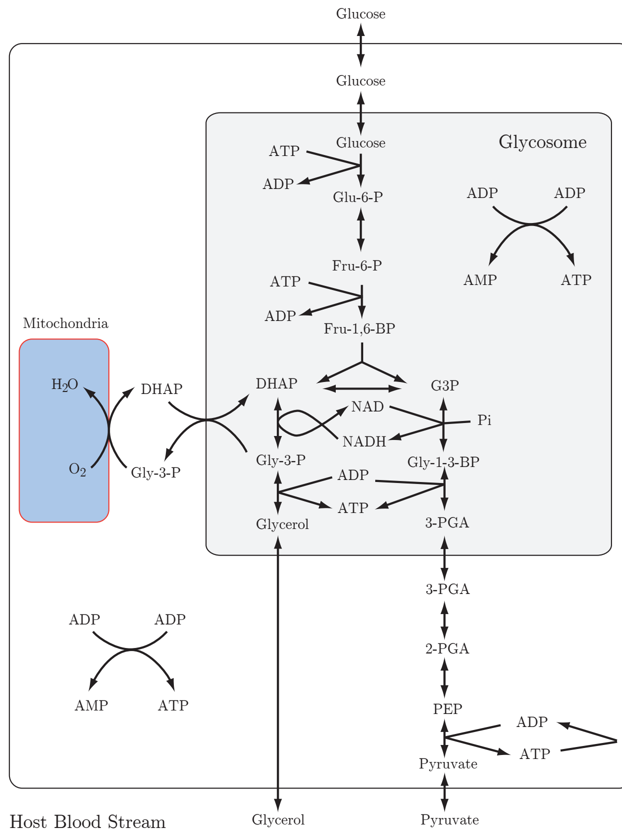


Figure 13.3 Energy Metabolism of Trypanasoma brucei

What is of interest is that an analysis of the stoichiometry using the techniques described in Chapter **??** indicates the presence of four conservation

laws, these include:

1: $\text{ATP}_c + \text{ADP}_c + \text{AMP}_c$

2: $\text{ATP}_g + \text{ADP}_g + \text{AMP}_g$

3: $\text{NAD}_g + \text{NADH}_g$

4: glycerol 3-phosphate$_c$ + dihydroxyacetone phosphate$_c$ +
glycerol 3-phosphate$_g$ + dihydroxyacetone phosphate$_g$ +
glucose 6-phosphate$_g$ + fructose 6-phosphate$_g$ +
fructose 1,6-bisphosphate$_g$ + glyceraldehyde 3-phosphate+
1,3-bisphosphoglycerate + ATP_g + ADP_g

where the subscript $_c$ means cytoplasm and $_g$ means glycosome. Figure 13.4 shows the same metabolic map as Figure 13.3 but with the conserved moieties highlighted. The fact that phosphate is a conserved moiety means that any species that includes the moiety will be constrained by the total amount of phosphate. As pointed out by Eisenthal and Cornish-Bowden [19] in relation to the work of Bakker et al [3], there are two possible ways to disrupt an organism metabolically. One can either reduce a flux to a very low level or increase one or more metabolite levels to such high levels that they become toxic. Bakker's analysis [3] of Trypanosoma metabolism showed that much of the flux control was on glucose transport. This limits the number of potential sites for flux disruption. As for disrupting concentrations only one step had a significant concentration control coefficient, Pyruvate transport [19] again limited the choice for drug targets. The reason why pyruvate transport is a susceptible target is because it is one of the few steps not involved in the conservation laws.

13.3 Ultrasensitivity

Another area where moiety-conserved cycles can have a marked behavioral impact is in generating ultrasensitivity, a term used to describe the amplification of a signal. For example, if a change in an input of 1% results in an output change of 2% then the input is amplified two fold. In engineering this effect is often referred to as the **gain** of the system.

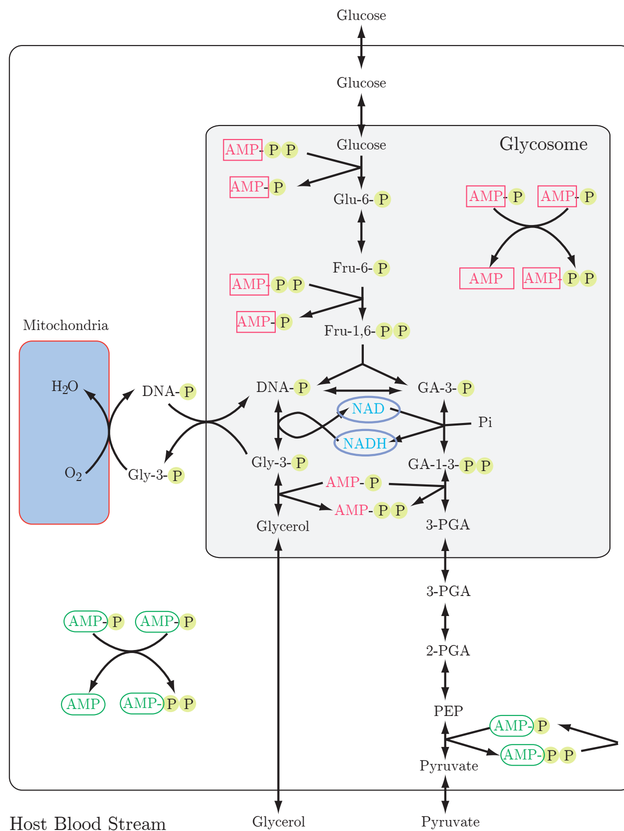


Figure 13.4 Energy Metabolism of Trypanasoma brucei

Amplification of signals is a fundamental operation in both biology and engineering. Many inputs in biology are small magnitude that they must be amplified before they can be acted upon. Examples include changes in hormonal levels and amplification of sensory inputs from external stimuli such as nutrient gradients or light levels. There are also other reasons to amplify signals, for example to generate sustained oscillations, implement effective responses in feed-forward networks.

The classic mechanism for generating gains greater than one is via multimeric cooperativity. We will not discuss this here but refer the reader to

other texts [71] where the topic is discussed in much more detail. In this chapter we will discuss a number of other mechanisms for generating high gains such as zero-order sensitivity in covalent cycles and titration based ultrasensitivity.

Quantifying Gain

A common way to measure the gain of a system is measure the ratio of fractional change between the output, Y, and input, X, of the system:

$$R = \frac{d \ln Y}{d \ln X}$$

This is clearly related to the control coefficients discussed previously. If $R > 1$ then the system is ultrasensitive.

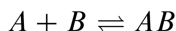
13.4 Saturation

Perhaps the simplest way to amplify a signal is to use two enzyme catalyzed steps where the second step is saturable and the first step irreversible. If the first step is increased in activity (by adding more active enzyme) there will come a point where the intermediate concentration, S, becomes very sensitive to the first step, Figure 13.5.

$$X_o \xrightarrow{v_1} S \xrightarrow{v_2} X_1$$

13.5 Sequestration

There is a very simple mechanism involving sequestration that can generate an ultrasensitivity response. The effect depends again on a conservation law between the active participant, A and a sequester molecule B. The reaction is just:



$$A + B \rightleftharpoons AB$$

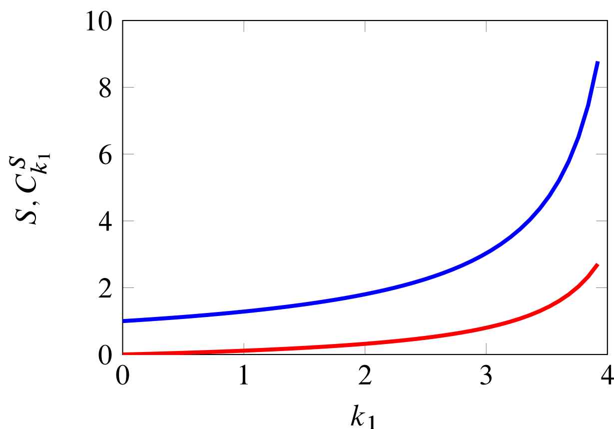


Figure 13.5 Sensitivity in a two step pathway. First step irreversible, second step saturable. When the second step is near saturation, the intermediate concentration is very sensitive to change in the activity of the first step. `Xo -> S; k1*Xo; S -> ; Vm*S/(Km + S); Xo = 1; k1 = 0.1; Km = 0.41 Vm = 4.5`. Upper curve is $C^S_{k_1}$, lower curve is concentration of S.

The response begins with a low level of active molecule A. As A is increased, most free A is sequestered by B. However eventually A reaches a point where the amount of free B is so small that free A can not longer be removed. At this point the concentration of A rises rapidly. This is the point where ultrasensitivity is seen. Figure 13.6 illustrates a number of plots that show ultrasensitivity between $A_T = 8$ and $A_T = 10$. In this model, the active participant A is also used to activate a saturable promoter, which is why the curve eventually plateaus at high total A. Buchler and Cross

13.6 Zero-Order Ultrasensitivity

The MAPK (mitogen-activated protein kinase) pathways, are highly conserved and common components in signal transduction pathways (Chang

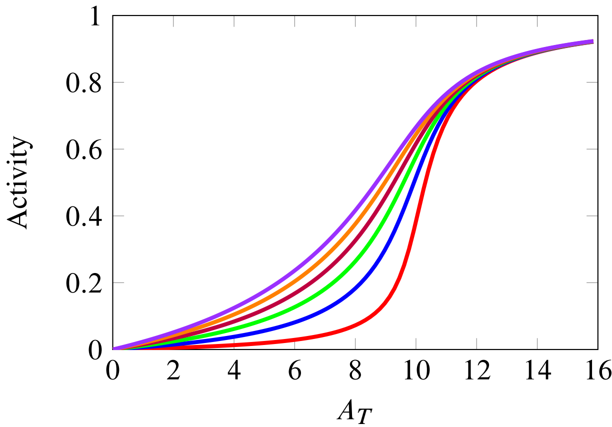


Figure 13.6 Ultrasensitivity by Simple Sequestration: `A + B <-> AB; k1*A*B - k2*AB`.

& Karin, 2001). Virtually all eukaryotic cells that have been examined (ranging from yeast to man) possess multiple MAPK pathways each of which responds to multiple inputs. In mammalian systems MAPK pathways are activated by a wide range of input signals including a variety of growth factors and environmental stresses such as osmotic shock and ischemic injury (Kyriakis & Avruch, 2002; Gomperts et al., 2002). Once the MAPK pathways have integrated these signals, they coordinately activate gene transcription with resulting changes in protein expression leading to cell cycling, cell death and cell differentiation.

Consider the simple conserved cycle shown in Figure 13.8. As discussed in Chapter **??**, the two species, S_1 and S_2 are conserved because the total $S_1 + S_2$ remains constant over time (at least over a time scale shorter than protein synthesis and degradation). Let us assume that the kinetics governing each cycle arm is simple first order mass-action kinetics.

If we plot the steady state concentration of S_1 and S_2 versus the kinetic constant k_1 we get the response curves shown in Figure 13.9. The response curves are in fact hyperbolic. For example, S_2 rises linearly then levels off to 10 concentration units in the limit. What is happening is that as k_1

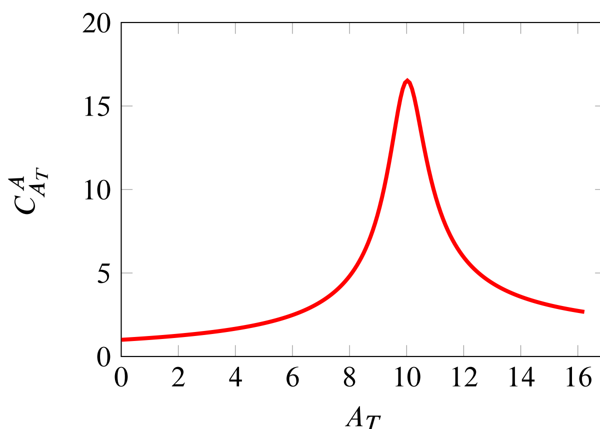


Figure 13.7 Sensitivity, $C^A_{A_T}$ as a function of A_T

increases more and more S_1 is converted to S_2 leading to a rise in S_2 and a fall in S_1. The limit is reached because there is only a limited amount of mass in the cycle.

Simple Cycle with Non-Linear Kinetics

If we now take the simple cycle model from the last section and instead of linear kinetics we now use non-linear kinetics, for example Michaelis-Menten kinetics on the forward and reverse arms then additional changes in behavior will be observed.

The response is now sigmoidal rather than hyperbolic. The reason for this is explained in Figure 13.11. The intersection points marked by a grey marker represents the corresponding steady state point ($v_1 = v_2$). A perpendicular dropped from these indicates the corresponding steady state concentration of S_1. If the activity of v_1 is increased by increasing k_1 by 20% then the v_1 curve moves up. The left intersection point indicates how much the steady state concentration moves as a result, shown by ΔS. The closer the steady state point is to the saturated point of the curve, the more the steady state will move. This shows that the response in S_1 can

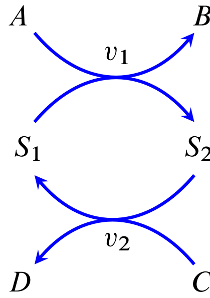


Figure 13.8 Simple Conserved cycle where $S_1 + S_2 = \text{constant} = M_t$.

be very sensitive in changes in k_1. Because k_1 is a linear term in the rate law we could replace it with the concentration of the enzyme implied in the Michaelis-Menten law. In practice such a cycle could represent a phosphorylation/dephophsorylation cycle where the implied enzyme is now a kinase. The kinase in turn could be controlled by other processes so that changes in the kinase activity results in sigmoid (or switch like) behavior in the cycle dynamics. In the literature such behavior was studied by [27, 28] and has been observed experimentally [40].

Analytical Approach

It is possible to use the machinery of metabolic control analysis derive the conditions for ultrasensitivity in a covalent modification cycle without recourse to specific kinetic laws [27, 28]. This was work was originally carried out by Small and Fell [78]. To investigate ultrasensitivity we need to evaluate the concentration control coefficient, $C_{v_1}^{S_2}$. To do this we would normally write down the connectivity theorem (6.4) and summation theorem (6.6), and solve for the appropriate coefficient. In this example we cannot easily do this because S_1 and S_2 are not independent and the normal connectivity theorems do not apply. As a result we must first derive a modified connectivity theorem that is specific to covalent modification cycles [20, 39].

Assume that a small change, δS_1 is made in S_1. In order to maintain

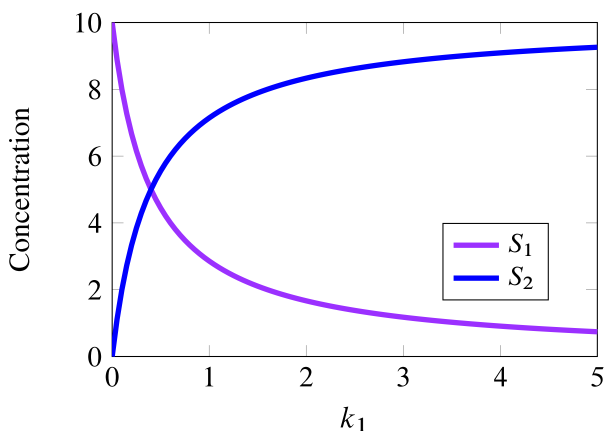


Figure 13.9 Simulation of the simple cycle with linear kinetics. Plot shows the steady state concentration of each species as a function of k_1. Model: `S1 -> S2; k1*S1; S2 -> S1; k2*S2; S1=10; k1=0.1; k2=0.4`

a constant amount of $S_1 + S_2$, we also make a compensating change in δS_2 equal to $-\delta S_1$. These changes will cause both v_1 and v_2 to change, however we can make changes to E_1 and E_2 such that the reaction rates are unchanged. We can express this thought experiment using the following local equations:

$$
\begin{aligned}
\frac{\delta v_1}{v_1} &= \frac{\delta E_1}{E_1} + \varepsilon^1_1 \frac{\delta S_1}{S_1} + \varepsilon^1_2 \frac{\delta S_2}{S_2} = 0 \\
\frac{\delta v_2}{v_2} &= \frac{\delta E_2}{E_2} + \varepsilon^2_2 \frac{\delta S_2}{S_2} + \varepsilon^2_1 \frac{\delta S_1}{S_1} = 0
\end{aligned}
$$

To make things simpler let us assume that each forward reaction is product insensitive, that is $\varepsilon^2_1 = 0$ and $\varepsilon^1_2 = 0$ so that the local equations are now

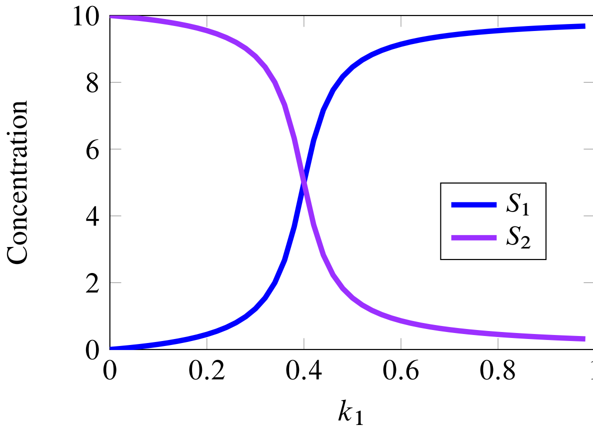


Figure 13.10 Simulation of the simple cycle with non-linear kinetics illustrating sigmoid or ultrasensitive behavior. Model: `S1 -> S2; k1*S1/(Km1+S1); S2 -> S1; k2*S2/(Km2+S2); S1=10; k1=0.1; Km1=0.5; k2=0.4; Km2=0.5`

reduced to:

$$\frac{\delta E_1}{E_1} = -\varepsilon^1_1 \frac{\delta S_1}{S_1}$$

$$\frac{\delta E_2}{E_2} = -\varepsilon^2_2 \frac{\delta S_2}{S_2}$$

We can also express the thought experiment in terms of the systems equation:

$$\frac{\delta S_2}{S_2} = C^{S_2}_{E_1} \frac{\delta E_1}{E_1} + C^{S_2}_{E_2} \frac{\delta E_2}{E_2}$$

We can now substitute the local equations into the systems equation:

$$-\frac{\delta S_2}{S_2} = C^{S_2}_{E_1} \varepsilon^1_1 \frac{\delta S_1}{S_1} + C^{S_2}_{E_2} \varepsilon^2_2 \frac{\delta S_2}{S_2} \tag{13.1}$$

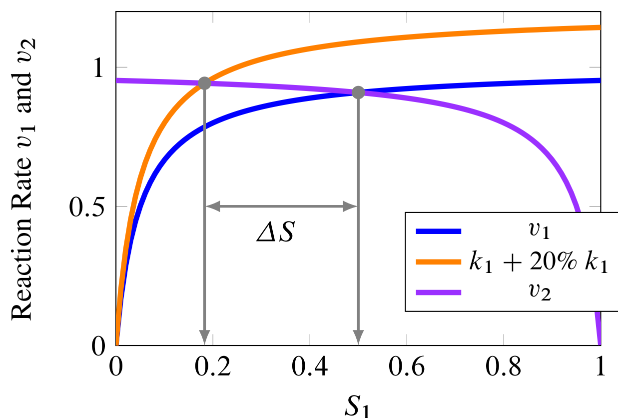


Figure 13.11 Plots the two cycle rates, v_1 and v_2 for the simple cycle with non-linear kinetics. Model: `S1 -> S2; k1*S1/(Km1+S1); S2 -> S1; k2*S2/(Km2+S2); S1=1; k1=1; Km1=0.05; k2=1; Km2=0.05.` The intersection points marked by a grey marker represents the steady state point ($v_1 = v_2$). See main text for explanation.

The thought experiment added the constraints that $\delta S_1 = -\delta S_2$, so that:

$$-\frac{\delta S_2}{S_2} = -C^{S_2}_{E_1} \varepsilon^1_1 \frac{\delta S_2}{S_2} \frac{S_2}{S_1} + C^{S_2}_{E_2} \varepsilon^2_2 \frac{\delta S_2}{S_2}$$

Canceling the $\delta S_2 / S_2$ terms yields the covalent modification connectivity theorem:

$$-1 = -C^{S_2}_{E_1} \varepsilon^1_1 \frac{S_2}{S_1} + C^{S_2}_{E_2} \varepsilon^2_2$$

One way to reexpress this is to divide both sides by S_2 and rearrange so that:

$$\frac{1}{S_2} = C^{S_2}_{E_1} \varepsilon^1_1 \frac{1}{S_1} - C^{S_2}_{E_2} \varepsilon^2_2 \frac{1}{S_2} \tag{13.2}$$

We now combine this new connectivity theorem with the normal summation theorem:

$$C^{S_2}_{E_1} + C^{S_2}_{E_2} = 0$$

Solving for $C^{S_2}_{E_1}$ we obtain:

$$C^{S_2}_{E_1} = \frac{S_1}{\varepsilon^1_1 S_2 + \varepsilon^2_2 S_1} \tag{13.3}$$

If we assume that v_1 and v_2 operate far below saturation than both elasticities, ε^1_1 and ε^2_2 are approximately one, that:

$$C^{S_2}_{E_1} = \frac{S_1}{S_2 + S_1}$$

This value will always be equal to or less than one. Therefore there is no possibility of ultrasensitivity when both enzymes are operating first-order. However if we assume that the enzymes are operating near saturation, then $\varepsilon^1_1 < 1$ and $\varepsilon^2_2 < 1$. For example if we set $S_1 = 9$ and $S_2 = 1$ and set both elasticities to 0.5 then the control coefficient, $C^{S_2}_{E_1} = 1.8$. If we reduce the elasticities further $\varepsilon^1_1 = 0.2$ and $\varepsilon^2_2 = 0.2$, the response of the system rises to 4.5. That is a 1% increase in E_1 will result in a 4.5% increase in S_2. In most cases like this, S_2 is itself a protein, often a kinase. So that a small change in one protein, E_1 can have a large effect on another, S_2.

Cascades

It is common in eukaryotic organisms to find cascades of covalent modification cycles. Figure 13.12 shows one such cascade. The overall gain of the system, between the input S and output, S_4 is given by the product of the individual sensitivities.

$$R^{S_4}_{S} = C^{S_2}_{E_1} C^{S_4}_{S_2}$$

Dual Cycle

We can also consider double cycles such as the one shown in Figure 13.13. We can write out the stoichiometry matrix for the double cycle as:

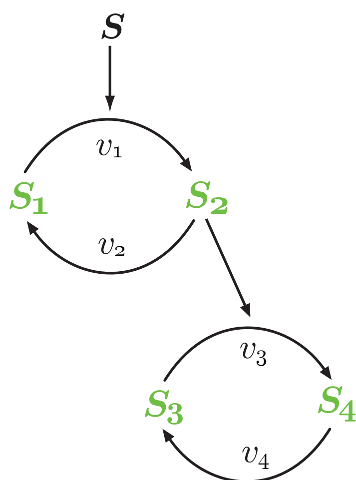


Figure 13.12 Two cycles in a cascade.

$$\mathbf{N} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

From this it is possible to show that there is one conservation law given by the relation:

$$S_1 + S_2 + S_3 = T$$

If we assume simple linear mass-action kinetics for each of the reactions, simulation will reveal that the concentration of S_3 shows sigmoid behavior with respect to the stimulus signal S. We can assume that the stimulus signal, S, operates on the rate constants, k_1 and k_3 by the same factor, that is an increase in S by $x\%$ results in a change in k_1 and k_3 by $x\%$. What is of interest is that we no longer need non-linear kinetics to generate sigmoidal behavior but can instead rely on only a small increase in the complexity of the conservation laws.

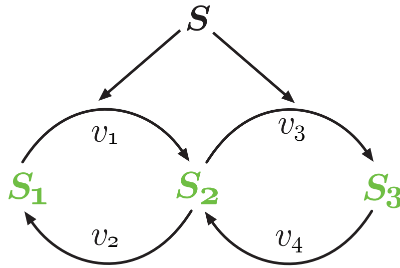


Figure 13.13 Two cycles connected by a common intermediate, S_2. The rate laws for each step is given by $v_1 = k_1 S_1$, $v_2 = k_2 S_2$, $v_3 = k_3 S_2$, $v_4 = k_4 S_3$. S is the stimulus signal which acts by increasing k_1 and k_3 by the same factor.

$$C_S^{S_3} = \frac{S_1(\varepsilon_2^3 + \varepsilon_2^2) + S_2 \varepsilon_1^1}{S_1 \varepsilon_2^2 \varepsilon_3^4 + S_2 \varepsilon_1^1 \varepsilon_2^4 + S_3 \varepsilon_1^1 \varepsilon_2^3} \tag{13.4}$$

If we assume all reactions are first-order then all the elasticities equal one. Under these conditions the above equation reduces to:

$$C_S^{S_3} = \frac{2S_1 + S_2}{S_1 + S_2 + S_3}$$

This shows that given the right ratios for S_1, S_2 and S_3, it is possible for $C_S^{S_3} > 1$. Therefore unlike the case of a single cycle where near saturation is required to achieve ultrasensitivity, multiple cycles can achieve ultrasensitivity with simple linear kinetics

The cyclic models considered here assume negligible sequestration of the cycle species by the catalyzing kinase and phosphatase. In reality this is not likely to be the case because experimental evidence indicates that the concentrations of the catalyzing enzymes and cycle species are comparable (See [7] for a range of illustrative data). In such situations additional effects are manifest [21, 70], of particular interest is the emergence of new regulatory feedback loops which can alter the behavior quite markedly (See [57] and [61]).

Appendices

A

Kinetics in a Nutshell

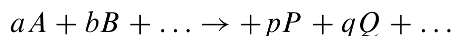
Definition

Reaction Kinetics is the study of how fast chemical reactions take place, what factors influence the rate of reaction and what mechanisms are responsible.

Stoichiometric Amount

This is defined as the number of molecules of ap articular reactant or product taking part in a reaction.

Depicting Reactions



$$aA + bB + \ldots \rightarrow +pP + qQ + \ldots$$

where $a, b, \ldots$ are stoichiometric amounts.

Rates of Change

The rate of change is defined as the rate of change in concentration or amount of a designated molecular species.

$$\text{Rate Of Change} = \frac{dS}{dt}$$

Stoichiometric coefficients

The **stoichiometric coefficient**, c_i, for a molecular species A_j, is the difference between the molar amount of the species on the product side and the molar amount of the species on the reactant side.

$$c_i = \text{Molar Amount of Product} - \text{Molar Amount of Reactant}$$

In the reaction, $2A \longrightarrow B$, the molar amount of A on the product side is zero while on the reactant size it is two. Therefore the stoichiometric coefficient of A is given by $0-2=-2$. In many cases a particular species will only occur on the reactant or product side and it is not to common to find situations where a species occurs simultaneously as a product and a reactant. As a result, reactant stoichiometric coefficients tend to be **negative** while product stoichiometric coefficients tend to be **positive**.

Reaction Rates

The reaction rate, often denoted by the symbol v, is measured with respect to a given molecular species normalized by the species' stoichiometric coefficient. This definition ensures that no matter which molecular species in a reaction is measures, the reaction rate is uniquely defined for that reaction. More formally the reaction rate for the given reaction:

$$aA + bB + \ldots \rightarrow pP + qQ + \ldots$$

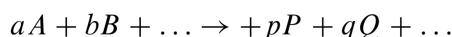
$$v = \frac{1}{c_a}\frac{dA}{dt} = -\frac{1}{c_b}\frac{dB}{dt}\ldots = \frac{1}{c_p}\frac{dP}{dt} = \frac{1}{c_q}\frac{dQ}{dt}\ldots$$

where c_x are the stoichiometric coefficients. Alternative we can write the rate of change in terms of the reaction rate as follows:

$$\frac{dA}{dt} = c_a v \tag{A.1}$$

Elementary mass-action kinetics

An elementary reaction is one that cannot be broken down into simpler reactions. Such reaction will often have simple kinetics called mass-action kinetics. For a reaction of the form



$$aA + bB + \ldots \rightarrow + pP + qQ + \ldots$$

the mass-action kinetic rate law is given by:

$$v = k_1 A^a B^b \ldots - k_2 P^p Q^q \ldots$$

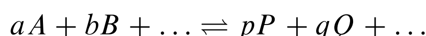
where k_1 and k_2 are the forward and reverse rate constants respectively.

Chemical Equilibrium

In principle all reactions are reversible, meaning transformations can occur from reactant to product or product to reactant. The net rate of a reversible reaction is the difference between the forward and reverse rates. At chemical equilibrium the forward and reverse rates are equal. Chemical equilibrium is then given by:

$$\frac{B}{A} = K_{eq} \tag{A.2}$$

This ratio has special significance and is called the **equilibrium constant**, denoted by K_{eq}. The equilibrium constant is also related to the ratio of the rate constants, k_1/k_2. For a general reversible reaction such as:



$$aA + bB + \ldots \rightleftharpoons pP + qQ + \ldots$$

and using arguments similar to those described above, the ratio of the rate constants can be easily shown to be:

$$K_{eq} = \frac{P^p Q^q \ldots}{A^a B^b \ldots} = \frac{k_1}{k_2} \tag{A.3}$$

where the exponents are the stoichiometric **amounts** for each species.

Mass-action and Disequilibrium Ratio

Although in closed systems, reactions will tend to equilibrium, reactions occurring in living cells are generally out of equilibrium and the ratio of the products to the reactants *in vivo* is then called by the **mass-action ratio**, Γ. The ratio of the mass-action ratio to the equilibrium constant is often called the **disequilibrium ratio**:

$$\rho = \frac{\Gamma}{K_{eq}} \tag{A.4}$$

At equilibrium, the mass-action ratio will be equal to the equilibrium constant and $\rho = 1$. If the reaction is away from equilibrium ($B/A < K_{eq}$) then $\rho < 1$.

For a simple unimolecular reaction it was shown previously that the equilibrium ratio of product to reactant, B/A, is equal to the ratio of the forward and reverse rate constants. Substituting this into the disequilibrium ratio gives:

$$\rho = \Gamma \frac{k_2}{k_1} = \frac{B}{A}\frac{k_2}{k_1}$$

Therefore

$$\rho = \frac{v_r}{v_f} \tag{A.5}$$

That is the disequilibrium ratio is the ratio of the reverse and forward rates. If $\rho < 1$, then the net reaction must in the direction of product formation. If ρ is zero then the reaction is as out of equilibrium as possible with no product present.

Modified Mass-Action Rate Laws

A typical reversible mass-action rate law will require both the forward and the reverse rate constants to be fully defined. Often however, only one rate constant may be known. In these circumstances it is possible to express the reverse rate constant in terms of the equilibrium constant.

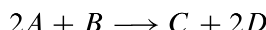
For example, given the simple unimolecular reaction, $A \rightleftharpoons B$. it is possible to derive the following:

$$
\begin{aligned}
v &= k_1 A - k_2 B \\
v &= k_1 A \left(1 - \frac{k_2 B}{k_1 A}\right) \\
\text{Since } K_{eq} &= \frac{k_1}{k_2} \\
v &= k_1 A \left(1 - \frac{\Gamma}{K_{eq}}\right)
\end{aligned} \tag{A.6}
$$

where Γ is the mass-action ratio. This can be generalized to an arbitrary mass-action reaction to give:

$$
v = k_1 A^a B^b \dots \left(1 - \frac{\Gamma}{K_{eq}}\right) = k_1 A^a B^b \dots (1 - \rho)
$$

where $A^a B^b \dots$ represents the product of all reactant species, a and b are the **corresponding** stoichiometric amounts, and ρ is the disequilibrium ratio. For example, for the reaction:



$$
2A + B \longrightarrow C + 2D
$$

where k_1 is the forward rate constant, the modified reversible rate law is:

$$
v = k_1 A^2 B \, (1 - \rho)
$$

The modified formulation demonstrates how a rate expression can be divided up into functional parts that include both kinetic and thermodynamic components [37]. The kinetic component is represented by the term

$k_1 A^a B^b \ldots$ while the thermodynamic component is represented by the expression $1 - \rho$.

We can also derive the modified rate law in the following way. Given the net rate of reaction $v = v_f - v_r$, we can write this expression in the following way:

$$v = v_f \left(1 - \frac{v_r}{v_f}\right)$$

That is:

$$v = v_f (1 - \rho)$$

B

Enzyme Kinetics in a Nutshell

Enzymes

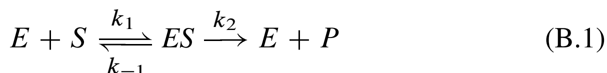
Enzymes are protein molecules that can accelerate a chemical reaction with changing the equilibrium constant of the reaction of themselves.

Enzyne Kinetics

Enzyme kinetics is a branch of science that deals with the many factors that can affect the rate of an enzyme-catalysed reaction. The most important factors include the concentration of enzyme, reactants, products, and the concentration of any modifiers such as specific activators, inhibitors, pH, ionic strength, and temperature. When the action of these factors is studied, we can deduce the kinetic mechanism of the reaction. That is, the order in which substrates and products bind and unbind and the mechanism by which modifiers alter the reaction rate.

Michaelis-Menten Kinetics

The standard model for enzyme action, describes the binding of free enzyme to the reactant forming an **enzyme-reactant complex**. This complex undergoes a transformation, releasing product and free enzyme. The free enzyme is then available for another round of binding to new reactant.



$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\longrightarrow} E + P \tag{B.1}$$

where k_1, k_{-1} and k_2 are rate constants, S is substrate, P is product, E is the free enzyme, and ES the enzyme-substrate complex.

By either assuming rapid equilibrium between enzyme, substrate and the substrate complex, or assuming a steady state condition on the enzyme substrate complex, an aggregate rate law, often called the Michaelis-Menten equation in the case the rapid equilibrium assumption or the Briggs-Haldane equation when using the steady state assumption is given by:

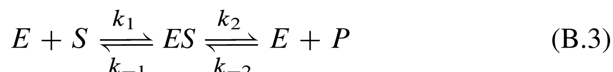
$$v = \frac{Vm\ S}{K_m + S} \tag{B.2}$$

where V_m is the maximal velocity and KM_m the substrate concentration that yield have the maximum velocity.

Product Inhibition

Reversible Rate laws

An alternative and more realistic model is the reversible form:



$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \tag{B.3}$$

The aggregate rate law for the reversible form of the mechanism can also be derived and is given by:

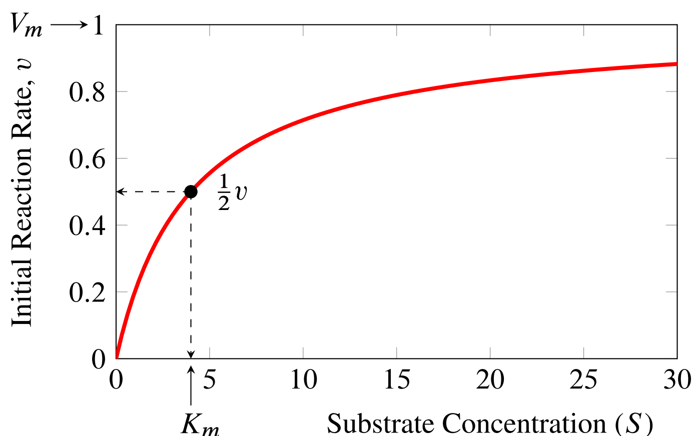


Figure B.1 Relationship between the initial rate of reaction and substrate concentration for a simple Michaelis-Menten rate law. The reaction rate reaches a limiting value called the V_m. K_m is set to 4.0 and V_m to 1.0. The K_m value is the substrate concentration that gives half the maximal rate.

$$v = \frac{V_f \; S/K_S - V_r \; P/K_P}{1 + S/K_S + P/K_P} \tag{B.4}$$

Haldane Relationship

For the reversible enzyme kinetic law there is an important relationship:

$$K_{eq} = \frac{P_{eq}}{S_{eq}} = \frac{V_f \; K_P}{V_r \; K_S} \tag{B.5}$$

and shows that the four kinetic constants, V_f, V_r, K_P and K_S are not independent. Haldane relationships can be used to eliminate one of the kinetic constants by substituting the equilibrium constant in its place. This is useful because equilibrium constants tend to be known compared to kinetic constants. By incorporating the Haldane relationship we can eliminate the

reverse maximal velocity (V_r) to yield the equation:

$$v = \frac{V_f/K_S(S - P/K_{eq})}{1 + S/K_S + P/K_P} \tag{B.6}$$

Separating out the terms makes it easier to see that the above equation can be partitioned into a number of distinct terms:

$$v = V_f \cdot (1 - \Gamma/K_{eq}) \cdot \frac{S/K_s}{1 + S/K_S + P/K_P} \tag{B.7}$$

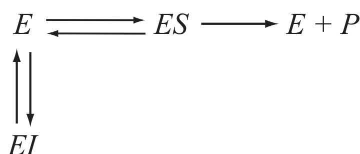
where $\Gamma = P/S$. The first term, V_f is the maximal velocity; the second term, $(1 - \Gamma/K_{eq})$ indicates the direction of the reaction according to thermodynamic considerations and the last terms refers to the fractional saturation with respect to substrate. We thus have a maximal velocity, a thermodynamic term and a saturation term. We will see this breakdown into distinct terms repeatedly as we consider other enzyme kinetic rate laws.

Competitive Inhibition

There are many molecules capable of slowing down or speeding up the rate of enzyme catalyzed reactions. Such molecules are called enzyme inhibitors and activators. One common type of inhibition, called **competitive inhibition**, occurs when the inhibitor is structurally similar to the substrate so that it competes for the active site by forming a dead-end complex.

The kinetic mechanism for a pure competitive inhibitor is shown in Figure B.2(a), where I is the inhibitor and EI the enzyme inhibitor complex. If the substrate concentration is increased, it is possible for the substrate to eventually out compete the inhibitor. For this reason the inhibitor alters the enzyme's apparent K_m but not the V_m.

$$\begin{aligned} v &= \frac{V_m\, S}{S + K_m\left(1 + \dfrac{I}{K_i}\right)} \\ &= \frac{V_m\, S/K_m}{1 + S/K_m + I/K_i} \end{aligned} \tag{B.8}$$



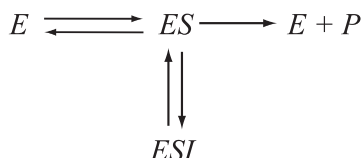


Figure B.2 Competitive and uncompetitive inhibition. P is the concentration of product, E is the free enzyme, ES the enzyme-substrate complex, and ESI the enzyme-substrate-inhibitor complex.

At $I = 0$, the competitive inhibition equation reduces to the normal irreversible Michaelis-Menten equation. Note that the term $K_m(1 + I/K_i)$ in the first equation more clearly shows the impact of the inhibitor, I, on the K_m. The inhibitor has no effect on the V_m.

The reversible form of the competitive rate law can be derived from equation (**??**) by setting $a \gg 1$ and $b = 0$ and is shown below:

$$v = \frac{\dfrac{V_m}{K_s}\left(S - \dfrac{P}{K_{\text{eq}}}\right)}{1 + \dfrac{S}{K_s} + \dfrac{P}{K_p} + \dfrac{I}{K_i}} \tag{B.9}$$

where V_m is the forward maximal velocity, and K_s and K_p are the substrate and product half saturation constants.

Sometimes reactions appear irreversible, that is no discernable reverse rate is detected, and yet the forward reaction is influenced by the accumulation of product. This effect is caused by the product competing with substrate for binding to the active site and is often called **product inhibition**. Given that product inhibition is a type of competitive inhibition we will briefly discuss it here. An important industrial example of this is the conversion of lactose to galactose by the enzyme $\beta-$galactosidase where galactose competes with lactose, slowing the forward rate [26].

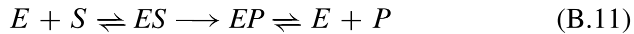
To describe simple product inhibition with rate irreversibility, we can set

the P/K_{eq} term in the reversible Michaelis-Menten rate law (B.4) to zero. This yields:

$$v = \frac{V_m S}{S + K_m \left(1 + \dfrac{P}{K_p}\right)} \tag{B.10}$$

It is not surprising to discover that equation (B.10) has exactly the same form as the equation for competitive inhibition (B.8). Figure **??** shows how the reaction rate responds to increasing product concentration at a fixed substrate concentration. As the product increases, it out competes the substrate and therefore slows down the reaction rate.

We can also derive the equation by using the following mechanism and the rapid equilibrium assumption:



$$E + S \rightleftharpoons ES \longrightarrow EP \rightleftharpoons E + P \tag{B.11}$$

where the reaction rate, $v \propto ES$.

Cooperativity

Many proteins are known to be oligomeric, that is they are composed of more than one identical protein subunit where each subunit has one or more binding sites. Often the individual subunits are identical.

If the binding of a ligand (a small molecule that binds to a larger macromolecule) to one site alters the affinity at other sites on the same oligomer then this is called **cooperativity**. If ligand binding increases the affinity of subsequent binding events, it is termed **positive cooperativity** whereas if the affinity decreases then it is termed **negative cooperativity**. One characteristic of positive cooperativity is that it results in a sigmoidal response instead of the usual hyperbolic response.

The simplest equation that displays sigmoid like behavior is the Hill equation:

$$v = \frac{Vm\ S^n}{K_d + S^n} \tag{B.12}$$

One striking feature of many oligomeric proteins is the way individual monomers are physically arranged. Often one will find at least one axis of symmetry. The individual protein monomers are not arranged in a haphazard fashion. This level of symmetry may imply that the gradual change in the binding constants as ligands bind, as suggested by the Adair model, might be physically implausible. Instead one might envisage transitions to an alternative binding state that occurs within the entire oligomer complex. The original authors laid out the following criteria for the MWC model:

1. The protein is an oligomer.
2. Oligomers can exist in two states: R (relaxed) and T (tense). In each state, symmetry is preserved and all subunits must be in the same state for a given R or T state.
3. The R state has a higher ligand affinity than the T state.
4. The T state predominates in the absence of ligand S.
5. The ligand binding microscopic association constants are all identical, this is in complete contrast to the Adair model

Given these criteria, the MWC model assumes that an oligomeric enzyme may exist in two conformations, designated T (tensed, square) and R (relaxed, circle) with an equilibrium between the two states with equilibrium constant, $L = T/R$, also called the allosteric constant. If the binding constants of ligand to the two states are different, then the distribution of the R and T forms can be displaced either towards one form or the other. By this mechanism, the enzyme displays sigmoid behavior. A minimal example of this model is shown in Figure B.3.

In the **exclusive model** (Figure B.3), the ligand can only bind to the relaxed form (circle). The mechanism that generates sigmoidicity in this model works as follows. When ligand binds to the relaxed form it displaces the equilibrium from the tense form to the relaxed form. In doing so, additional ligand binding sites are made available. Thus one ligand

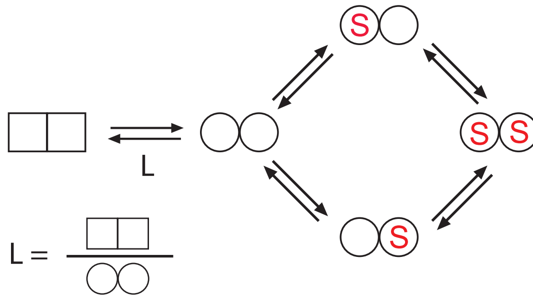


Figure B.3 A minimal MWC model, also known as the exclusive model, showing alternative microscopic states in the circle (relaxed) form. L is called the allosteric constant. The square form is called the tense state.

binding may generate four or more new binding sites. Eventually there are no more tense states remaining at which point the system is saturated with ligand. The overall binding curve will therefore be sigmoidal and will show positive cooperativity. Given the nature of this model, it is not possible to generate negative cooperativity. By assuming equilibrium between the various states it is possible to derive an aggregate equation for the dimer case of the exclusive MWC model:

$$v = V_m \frac{\dfrac{S}{k_R}\left(1 + \dfrac{S}{k_R}\right)}{\left(1 + \dfrac{S}{k_R}\right)^2 + L}$$

This also generalizes to n subunits as follows:

$$Y = \frac{\dfrac{S}{k_R}\left(1 + \dfrac{S}{k_R}\right)^{n-1}}{\left(1 + \dfrac{S}{k_R}\right)^n + L} \tag{B.13}$$

For more generalized a reversible rate laws the exhibit sigmoid behavior the reversible Hill equation is a good option to use.

Invoking the rapid-equilibrium assumption we can write the various complexes in terms of equilibrium constants to give:

$$v = \frac{V_f \alpha \, (1 - \rho) \, (\alpha + \pi)}{1 + (\alpha + \pi)^2}$$

where $\rho = \Gamma / K_{eq}$. For an enzyme with h (using the authors original notation) binding sites, the general form of the reversible Hill equation is given by:

$$v = \frac{V_f \alpha \;\; (1 - \rho) \;\; (\alpha + \pi)^{h-1}}{1 + (\alpha + \pi)^h} \tag{B.14}$$

Allostery

An allosteric effect is where the activity of an enzyme or other protein is affected by the binding of an effector molecule at a site on the protein's surface other than the active site. The MWC model described previously can be easily modified to accomodate allosteric action.

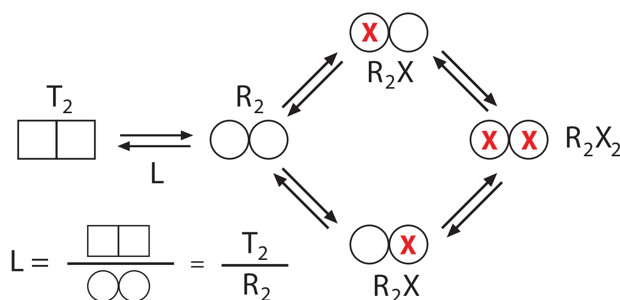


Figure B.4 Exclusive MWC model based on a dimer showing alternative microscopic states in the form of T and R states. The model is exclusive because the ligand, X, only binds to the R form.

The key to including allosteric effectors is the equilibrium between the tense (T) and relaxed (R) states (See Figure B.4). To influence the sig-

moid curve, an allosteric effector need only displace the equilibrium between the tense and relaxed forms. For example, to behave as an activator, an allosteric effector needs to preferentially bind to the R form and shift the equilibrium away from the less active T form. An allosteric inhibitor would do the opposite, that is bind preferentially to the T form so that the equilibrium shifts towards the less active T form. In both cases the V_m of the enzyme is unaffected.

The net result of this is to modify the normal MWC aggregate rate law to the following if the effector is an inhibitor:

$$v = V_m \frac{\alpha\,(1+\alpha)^{n-1}}{(1+\alpha)^n + L(1+\beta)^n} \tag{B.15}$$

where $\alpha = S/K_s$ and $\beta = I/K_I$. K_s and K_I are kinetic constants related to each ligand. A MWC model that is regulated by other an inhibitor and an activator is represented by:

$$v = V_m \frac{\alpha\,(1+\alpha)^{n-1}}{(1+\alpha)^n + L\dfrac{(1+\beta)^n}{(1+\gamma)^n}}$$

There are also reversible forms of the allosteric MWC model but is fairly complex. Instead it is possible to modify the reversible Hill rate law to include allosteric ligands.

$$v = \frac{V_f \alpha \left(1 - \dfrac{\Gamma}{K_{eq}}\right) (\alpha+\pi)^{h-1}}{\dfrac{1+\mu^h}{1+\sigma\mu^h} + (\alpha+\pi)^h} \tag{B.16}$$

where

$$\begin{array}{ll} \sigma < 1 & \text{inhibitor} \\ \sigma > 1 & \text{activator} \end{array}$$

C
Math Fundamentals

C.1 Notation

Sum and Product:

$$a_1 + a_2 + a_3 + \ldots + a_n = \sum_{i=1}^{n} a_i$$

$$a_1 \times a_2 \times a_3 \times \ldots \times a_n = \prod_{i=1}^{n} a_i$$

Vectors and Matrices:

Bold lower case letters indicate vectors, for example: $\boldsymbol{v}, \mathbf{s}$

Bold upper case letters indicate matrices, for example: $\mathbf{N}, \boldsymbol{X}$.

Derivatives:

One the left is Leibniz's notation and on the right Lagrange's notation:

$$\frac{df}{dx} \equiv f'(x)$$

$$\frac{d^2 f}{dx^2} \equiv f''(x)$$

$$\frac{d^n f}{dx^n} \equiv f^{(n)}(x)$$

C.2 Short Table of Derivatives

$$\frac{d}{dx}[c] = 0 \qquad \frac{d}{dx}[x] = 1$$

$$\frac{d}{dx}[cu] = c\ \frac{du}{dx} \qquad \frac{d}{dx}[u+v] = \frac{du}{dx} + \frac{dv}{dx}$$

$$\frac{d}{dx}[uv] = u\ \frac{dv}{dx} + v\ \frac{du}{dx} \qquad \frac{d}{dx}[u/v)] = \frac{v\ \frac{du}{dx} - u\ \frac{dv}{dx}}{v^2}$$

$$\frac{d}{dx}[u^n] = nu^{n-1}\frac{du}{dx} \qquad \frac{d}{dx}[f(u)] = \frac{df}{du}f(u)\ \frac{du}{dx}$$

$$\frac{d}{dx}[\ln u] = \frac{1}{u}\ \frac{du}{dx} \qquad \frac{de^u}{dx} = e^u\frac{du}{dx}$$

$$\frac{d}{dx}[\sin(u)] = \cos(u)\ \frac{du}{dx} \qquad \frac{d}{dx}[\cos(u)] = -\sin(u)\ \frac{du}{dx}$$

C.3 Logarithms

$$\log(AB) = \log(A) + \log(B)$$
$$\log(A/B) = \log(A) - \log(B)$$
$$\log(A^n) = n \ \log(A)$$
$$x^n \times x^m = x^{n+m}$$
$$\frac{x^n}{n^m} = x^{n-m}$$
$$(x^n)^m = x^{n \times m}$$

C.4 Partial Derivatives

If the value of a given function depends on two variables then we write this function in the form:

$$u = f(x, y)$$

If it is possible to change x without affecting y then x and y are called independent variables. The rate of change of u with respect to x when x varies but y remains constant is called the **partial derivative** of u with respect to x. Partial derivatives are denoted using the partial symbol, ∂. That is the partial derivative of u with respect to x is:

$$\frac{\partial u}{\partial x}$$

Likewise the partial derivative of u with respect to y is given by:

$$\frac{\partial u}{\partial y}$$

To find a partial derivative, we simply differentiate with respect to the variable of interest while treating the remaining variables as constants. For example, the reaction rate for a given reaction is $v = k_1 S - k_2 P$, where

S is the reactant, P the product and k_1 and k_2 the rate constants. In a controlled environment we should in principle be able to change S and P independently. Therefore we can write down the partial derivatives of the reaction rate with respect to S and P as follows:

$$\frac{\partial v}{\partial S} = k_1$$
$$\frac{\partial v}{\partial P} = -k_2$$

In order to indicate what variables are kept constant in the partial derivative the following notation is sometimes used, particulary in thermodynamics:

$$\left(\frac{\partial v}{\partial S}\right)_P = k_1$$
$$\left(\frac{\partial v}{\partial P}\right)_S = -k_2$$

Or for function with many variables, $x, y, z, \ldots$, the notation would extend to:

$$\left(\frac{\partial u}{\partial x}\right)_{y,z,\ldots}$$

Like derivatives, partial derivatives are defined in terms of a limit. For example the partial derivatives for the function, $f(x, y)$ are defined as:

$$\frac{\partial f(x, y)}{\partial x} = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$
$$\frac{\partial f(x, y)}{\partial y} = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

The graphical interpretation of a partial derivative, $\partial f(x, y)/\partial x$ is that it represents the slope of the function, $f(x, y)$ in the x direction.

C.5 Differential Equations

Differential equations are equations that contain derivatives. For example, the following is a differential equation:

$$\frac{dy}{dx} + y^2 = 0$$

An **ordinary differential equation** is where the derivatives are functions of the same variable. For example, the following equations are ordinary differential equations:

$$\frac{dy}{dx} = ay$$
$$\frac{dy}{dx} = 2x + 3y - 8$$
$$\frac{d^2y}{dx^2} - x\frac{du}{dx} = 0$$

A differential equation expressed in terms of the first derivative (dy/dx) is called a first-order differential equation. A differential equation that is expression in terms of second order derivatives (d^2y/dx^2) is called a second-order differential equation. When solving differential equations the objective is to find the function $y(x)$ such that when differentiated gives the original differential equation. For example the solution to:

$$\frac{dy}{dx} = ay$$

is

$$y = y_o e^{ax} \quad \text{(C.1)}$$

If we differentiate the solution (C.1) we get back the original differential equation.

Differential equations are used very often to model physical systems where they describe the rate of change of some variable with respect to time, t. The reason why they are used is because we may not explicitly know the

solution $y(t)$ but we will often know the rate of change the variable has at any given moment in time, dy/dt. This means we can at least obtain a numerical solution to $y(t)$ even if the analytical solution is unobtainable.

Differential equations can be further classified as autonomous or non-autonomous. Autonomous differential equations are the most common in biochemical models. These equations do not depend on time, that is the right-hand side of the differential equation has no terms relating explicitly to time. For example equation C.2 is autonomous equation C.3 is non-autonomous:

$$\frac{dx}{dt} = x^2 + 10 \tag{C.2}$$

$$\frac{dx}{dt} = x^2 + t - 5 \tag{C.3}$$

A **partial differential equation** is one where the derivatives are functions of more than one derivative. For example the equation is a partial differential equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{\partial p}{\partial x}$$

Note the use of the partial d (∂) in the partial differential equation to indicate that the function u is differentiated with respect to more than one variable. The partial derivative also indicates than when a derivative is made, other other constants are assumed to be held constant.

C.6 Taylor Series

Expressions like $1+2x+6x^2$ and $2+4x+x^2-3x^3$ that consist of the sum of a number of terms raised to a positive power are called polynomials. The only operations allowed in a polynomial are addition, subtraction, multiplication and non-negative integer powers. One of the simplest polynomials is the straight line, $y = a + bx$, termed a polynomial of first degree. The coefficients, a and b can be chosen so that the line will pass through any two points. That is we can express any straight line using $y = a + bx$. Similarly for a polynomial of second degree, $y = a+bx+cx^2$, a parabola,

we can choose the constants, a, b and c so that the curve can pass through any three points.

It follows that we can find a polynomial equation of nth degree that will pass through any $n + 1$ points. If the polynomial has an infinite number of terms, then we can imagine that the polynomial can be made to follow any function, $f(x)$ by suitable adjustment of the polynomial coefficients. Although this statement may not always be true, in many cases it is which makes the polynomial series very useful.

A polynomial of infinite degree is called a polynomial series:

$$f(x) = c_o + c_1 x + c_2 x^2 + c_3 x^3 + \dots$$

The question is, how can we find the polynomial series that will represent a given function, for example $\sin(x)$? To answer this we have to determine the constants, c_o, c_1 etc. in the polynomial equation. Let us assume that we wish to know the value of $\sin(x)$ at $x = 0$ using a polynomial series. At $x = 0$, all terms vanish except for c_o, therefore at $x = 0$:

$$f(0) = c_o$$

We can therefore interpret the first constant, c_o as the value of the function at $x = 0$. What about c_1? Let us take the derivative of the series, that is:

$$f'(x) = c_1 + 2c_2 x + 3c_3 x^2 + \dots$$

If we set $x = 0$, we find that:

$$f'(0) = c_1$$

That is the second constant, c_1, in the polynomial series is the first derivative of the function. If we take the second derivative we can also show that at $x = 0$, $f''(x0) = 2c_2$, that is $c_2 = f''(0)/2$. For the third derivative we can show that $f'''(0) = 3(2)c_3$, that is $c_3 = f'''(0)/(3!)$. This pattern continues for the remaining terms in the polynomial so that we can now write:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$

Function	Second order approximation
$\frac{1}{1+x}$	$1+x+x^2$
$\sqrt{1+x}$	$1+\frac{x}{2}+\frac{x^2}{8}$
$\sin(x)$	x

Table C.1 Examples of common approximations

This series is called the **Maclaurin series** for the function, $f(x)$. It approximates the function around the specific value of $x = 0$. To illustrate the use of the Maclaurin series consider expanding $\sin(x)$ around $x = 0$. $f(0)$ will equal $\sin(0) = 0$. $f'(0) = cos(0) = 1$ and so on. We can therefore write the series as:

$$\sin(x) = 0 + 1x + 0 - \frac{1}{3!}x^3 + 0 + \frac{1}{5!}x^5 - \dots$$
$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

What if we wanted to approximate a function about an arbitrary value of x? To do this we would use the Taylor series which is a generalization of the Maclaurin series. The **Taylor series** is defined by:

$$f(x) = f(x_o) + \frac{\partial f}{\partial x_o}(x - x_o) + \frac{1}{2!}\frac{\partial^2 f}{\partial x_o^2}(x - x_o)^2 + \dots + \frac{1}{n!}\frac{\partial^n f}{\partial x_o^n}(x - x_o)^n + \dots \quad \text{(C.4)}$$

where the approximation is now centered on x_o. If we set x_o equal to zero we will obtain the Maclaurin series.

C.7 Total Derivative

Consider the function:

$$f(t) = f(x(t), y(t))$$

The derivative of $f(t)$ with respect to t, is given by the chain rule:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Note the use of partial derivatives. This equation is often abbreviated to:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

where it is called the **total derivative**. Often the variable, t is not specified in the total derivative. Operationally the total derivative computes the change in f, given small changes in x and y.

C.8 Eigenvalues and Eigenvectors

A square matrix such as $\boldsymbol{A}$ can be use to transform a given vector, $\boldsymbol{v}$ in specific ways. For example, if the matrix $\boldsymbol{A}$ is:

$$\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

then the result of multiplying $\boldsymbol{A}$ into $\boldsymbol{v}$ will yield a vector that is similar to $\boldsymbol{v}$ but where the first element is scaled by 2 and the second element by 4.

For an arbitrary square matrix, if it is possible to find a vector $\boldsymbol{v}$ such than when we multiple the vector by $\boldsymbol{A}$ we get a scaled version of $\boldsymbol{v}$, then we call the vector $\boldsymbol{v}$ the **eigenvector** of $\boldsymbol{A}$ and the scaling value, the **eigenvalue** of $\boldsymbol{A}$. For a matrix of dimension n, there will be at most n eigenvalues and n eigenvectors. In the case of the simple example above the eigenvalues must be 2 and 4 respectively while the two eigenvector will be:

$$\begin{bmatrix} \alpha \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$$

The definition of an eigenvector and eigenvalue is often given in the form:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$$

We can rearrange this equation as follows:

$$\begin{aligned} \boldsymbol{A}\boldsymbol{v} &= \lambda \boldsymbol{I}\boldsymbol{v} \\ \boldsymbol{A}\boldsymbol{v} - \lambda \boldsymbol{I}\boldsymbol{v} &= 0 \\ (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} &= 0 \end{aligned}$$

From linear algebra we know that there will be non-zero solutions to $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = 0$ if $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$. We can use this observation to compute the eigenvalues and eigenvectors of a matrix. For example consider the matrix:

$$\begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}$$

Computing $\boldsymbol{A} - \lambda \boldsymbol{I}$ yields:

$$\begin{aligned} \boldsymbol{A} - \lambda \boldsymbol{I} &= \begin{bmatrix} 3-\lambda & 6 \\ 1 & 4-\lambda \end{bmatrix} \\ \det(\boldsymbol{A} - \lambda \boldsymbol{I}) &= (3-\lambda)(4-\lambda) - 6 \\ &= \lambda^2 - 7\lambda + 6 \\ &= (\lambda - 6)(\lambda - 1) \end{aligned}$$

The eigenvalues are therefore 6 and 1. With two eigenvalues there will be two eigenvectors. First we consider $\lambda = 6$.

$$\begin{aligned} (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} &= 0 \\ \left(\begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix} - \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} \right) \boldsymbol{v} &= 0 \\ \begin{bmatrix} -3 & 6 \\ 1 & -2 \end{bmatrix} \boldsymbol{v} &= 0 \end{aligned}$$

By inspection we can see that the eigenvector is:

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

satisfied this equation. Likewise we can do the same for the other eigenvalue, $\lambda = 1$ where the corresponding eigenvector is found to be:

$$\begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Further Reading

1. Smail LL (1953) Analytical Geometry and Calculus. Appleton-Century-Crofts ISBN: 978-0982477311

D

Control Equations

D.1 Linear Pathways

Two Step pathway

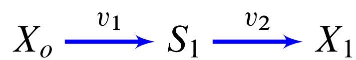


Figure D.1 Two Step Pathway.

$$C^J_{E_1} = \frac{\varepsilon^2_S}{\varepsilon^2_S - \varepsilon^1_S}$$

$$C^J_{E_2} = -\frac{\varepsilon^1_S}{\varepsilon^2_S - \varepsilon^1_S}$$

$$C^{S_1}_{E_1} = \frac{1}{\varepsilon^2_S - \varepsilon^1_S}$$

$$C^{S_1}_{E_2} = -\frac{1}{\varepsilon^2_S - \varepsilon^1_S}$$

Three Step pathway

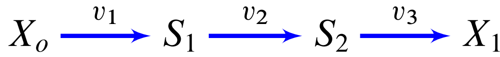


Figure D.2 Three Step Pathway.

The denominator, D is given by:

$$D = \varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^1_1\varepsilon^2_2$$

$$C^J_{E_1} = \frac{\varepsilon^2_1\varepsilon^3_2}{D} \qquad C^J_{E_2} = -\frac{\varepsilon^1_1\varepsilon^3_2}{D} \qquad C^J_{E_3} = \frac{\varepsilon^1_1\varepsilon^2_2}{D}$$

$$C^{S_1}_{E_1} = \frac{\varepsilon^3_2 - \varepsilon^2_2}{D} \qquad C^{S_2}_{E_1} = \frac{\varepsilon^2_1}{D}$$

$$C^{S_1}_{E_2} = -\frac{\varepsilon^3_2}{D} \qquad C^{S_2}_{E_2} = -\frac{\varepsilon^1_1}{D}$$

$$C^{S_1}_{E_3} = \frac{\varepsilon^2_2}{D} \qquad C^{S_2}_{E_3} = \frac{\varepsilon^1_1 - \varepsilon^2_1}{D}$$

Three Step pathway with Negative Feedback

$$D = \varepsilon^1_1\varepsilon^2_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^2_1\varepsilon^3_2 - \varepsilon^1_2\varepsilon^2_1$$

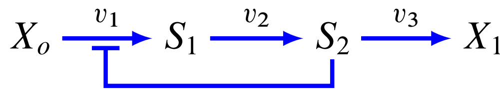


Figure D.3 Three Step Pathway.

$$C^J_{E_1} = \frac{\varepsilon^2_1\varepsilon^3_2}{D} \qquad C^J_{E_2} = -\frac{\varepsilon^1_1\varepsilon^3_2}{D} \qquad C^J_{E_3} = \frac{\varepsilon^1_1\varepsilon^2_2 - \varepsilon^1_2\varepsilon^2_1}{D}$$

$$C^{S_1}_{E_1} = \frac{\varepsilon^3_2 - \varepsilon^2_2}{D} \qquad C^{S_1}_{E_2} = -\frac{\varepsilon^1_2 - \varepsilon^3_2}{D} \qquad C^{S_1}_{E_3} = \frac{\varepsilon^2_2 - \varepsilon^1_2}{D}$$

$$C^{S_2}_{E_1} = \frac{\varepsilon^2_1}{D} \qquad C^{S_2}_{E_2} = -\frac{\varepsilon^1_1}{D} \qquad C^{S_2}_{E_3} = \frac{\varepsilon^1_1 - \varepsilon^2_1}{D}$$

Three Step pathway with Feedforward Loop

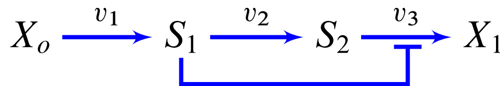


Figure D.4 Three Step Pathway.

$$D = \varepsilon^1_1\varepsilon^2_2 + \varepsilon^2_1\varepsilon^2_2 - \varepsilon^1_1\varepsilon^3_2 + \varepsilon^2_1\varepsilon^3_2$$

$$C^J_{E_1} = \frac{\varepsilon^2_2\varepsilon^3_1 + \varepsilon^2_1\varepsilon^3_2}{D} \qquad C^J_{E_2} = -\frac{\varepsilon^1_1\varepsilon^3_2}{D} \qquad C^J_{E_3} = \frac{\varepsilon^1_1\varepsilon^2_2}{D}$$

$$C^{S_1}_{E_1} = \frac{\varepsilon^3_2 - \varepsilon^2_2}{D} \qquad C^{S_1}_{E_2} = -\frac{\varepsilon^3_2}{D} \qquad C^{S_1}_{E_3} = \frac{\varepsilon^2_2}{D}$$

$$C^{S_2}_{E_1} = \frac{\varepsilon^2_1 + \varepsilon^3_1}{D} \qquad C^{S_2}_{E_2} = \frac{\varepsilon^1_1 + \varepsilon^3_1}{D} \qquad C^{S_2}_{E_3} = \frac{\varepsilon^1_1 - \varepsilon^2_1}{D}$$

D.2 Cycles

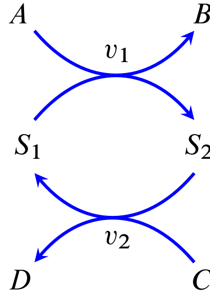


Figure D.5 Simple Conserved cycle where $S_1 + S_2 = \text{constant} = M_t$.

$$C_{E_1}^{S_2} = \frac{S_1}{\varepsilon_1^1 S_2 + \varepsilon_2^2 S_1}$$

D.3 Branches

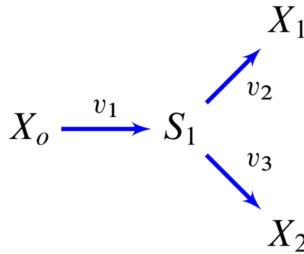


Figure D.6 Three Step Pathway.

Let $\alpha = \frac{v_2}{v_1}$.

$$C^{J_2}_{E_1} = \frac{\varepsilon_2}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} > 0$$

$$C^{J_2}_{E_2} = \frac{\varepsilon_3(1-\alpha) - \varepsilon_1}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} > 0$$

$$C^{J_2}_{E_3} = \frac{-\varepsilon_2(1-\alpha)}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1} < 0$$

And for the concentration control coefficients:

$$C^{S}_{E_1} = \frac{1}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1}$$

$$C^{S}_{E_2} = \frac{-\alpha}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1}$$

$$C^{S}_{E_3} = \frac{-(1-\alpha)}{\varepsilon_2\alpha + \varepsilon_3(1-\alpha) - \varepsilon_1}$$

E
Modeling with Python

In this appendix a brief description of the Python programming language will be given plus a brief introduction to the Antimony reaction network format and libRoadRunner.

Python Python is an easy to learn general purpose interactive programming language. It has similar usability characteristics to Matlab or Basic. As such it is a good language to use for doing pathway simulations and is easily learned by new users. In recent years Python has also become more widely used as a general purpose scientific programming language and now supports many useful libraries and tools for modelers. All the scripts we provide in this book are written in Python.

Antimony SBML has become a de facto standard for exchanging models of biological pathways. Any tool we use should therefore be able to support SBML. However SBML is a computer readable language and it is not easy for humans to read or write SBML. Instead more human readable formats have been developed. In this text book we

will be using the Antimony pathway description language [?]. Models can be described in Antimony then converted to SBML or vice versa.

libRoadRunner To support SBML from within Python we developed a C/C++ simulation library called libRoadRunner [?] that can read and run models based on SBML. In order to use libRoadRunner within Python, we also provide a Python interface that makes it easy to carry out simulations with Python.

Spyder Integration of the various tools including Python is achieved by using spyder2 (`https://code.google.com/p/spyderlib/`). Spyder2 offers a Matlab like experience in a friendly, cross-platform environment.

E.1 Introduction to Python

One great advantage of the Python language is that is runs on many computer platforms, most notably Windows, Mac and Linux and is freely downloadable from the Python web site. To execute Python code we will need access to what is often referred to as a Python IDE (Integrated Development Environment). In the Python world there are many IDEs to choose from, ranging from very simple consoles to sophisticated development systems that includes documentation, debuggers and other visual aids. In this book we use the cross-platform IDE called spyder2 (`https://code.google.com/p/spyderlib/`).

The best way to learn Python is to download a copy and start using it. We have prepared installers that install all the relevant components you need, these can be found at `tellurium.analogmachine.org`. The Tellurium distribution includes some additional helper routines which can make life easier for new users. The Tellurium version can be downloaded for Mac and Windows computers. We will use the Windows version here. To download the installer go to the web site `tellurium.analogmachine.org`, and click on the first link you see called *Download Windows version here*. Run the installer and follow the instructions.

Once Tellurium is installed go to the start menu, find Tellurium and select the application call Tellurium spyder. If successful you should see something like the screen shot in Figure E.1 but without the plotting window. The screen-shot shows three important elements, on the left we see an editor, this is where models can be edited. On the lower right is the Python console where Python commands can be entered. At the top right we show plotting window that illustrates some output from a simulation. For those familiar with IPython, the latest version of spyder2 supports the IPython console directly.

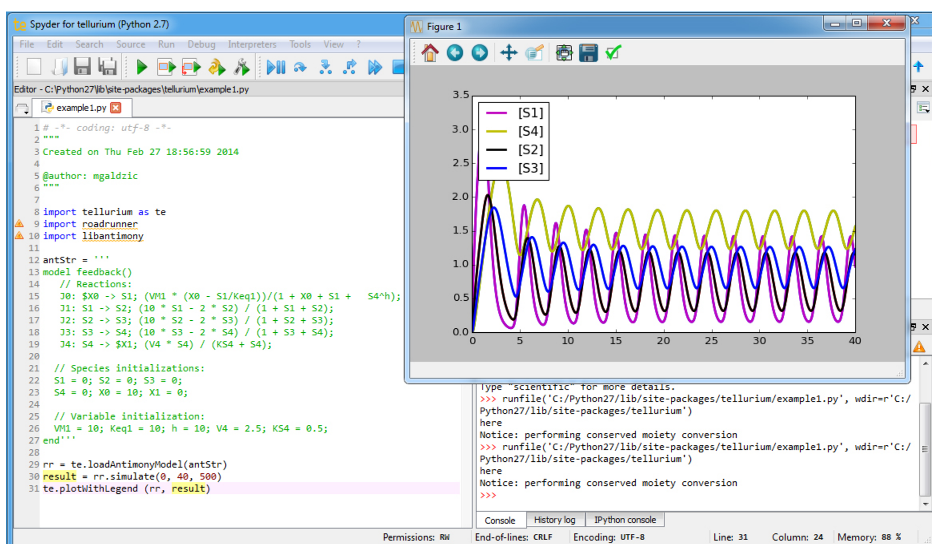


Figure E.1 Screen-shot of Tellurium, showing editor on the left, Python console bottom right and plotting window top-right.

Once you have started the Tellurium IDE, let us focus on the Python console at the bottom right of the application. A screen-shot of the console is shown in Figure E.2.

The >>> symbol marks the place where you can type commands. The following examples are based on Python 2.7. To add two numbers, say 2 + 5, we would type the following:

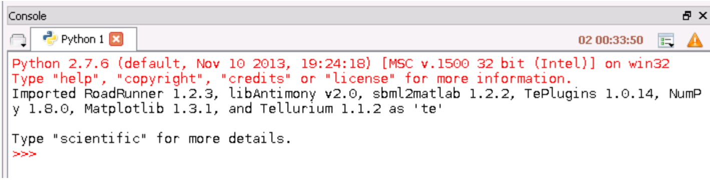


Figure E.2 Screen-shot of Tellurium, focusing on the Python console.

```
>>> print 2 + 5
7
>>>
```

Listing E.1 Simple Arithmetic

Just like Matlab or Basic we can assign values to variables and use those variables in other calculations:

```
>>> a = 2
>>> b = 5
>>> c = a + b
>>> print c
7
>>>
```

Listing E.2 Assigning values to variables

The types of values we can assign to variables include values such as integers, floating point numbers, Booleans (True or False), strings and complex numbers.

```
>>> a = 2
>>> b = 3.1415
>>> c = False
>>> d = "Hello Python"
>>> e = 3 + 6j
>>>
```

Listing E.3 Different kinds of values

Many functions in Python are accessible via modules. For example to compute the sin of a number we can't simply type `sin (30)`. Instead we must first load the math module. We can then call the sin function:

```
>>> import math
>>> print sin (3.1415)
9.265358966049026e-05
>>>
```

Listing E.4 Importing modules (libraries) into Python

In Tellurium we preload some libraries including the math library.

Repeating Calculations

One of the commonest operations we do in computer programming is iteration. We can illustrate this with a simple example that loops ten times, each time printing out the loop index. This example will allow us to introduce the IDE editor. The editor is the panel on the left side of the IDE. In the editor we can type Python code, for example we could type:

```
a = 4.0
b = 8.0
c = a/b
print "The answer is:", c
```

Listing E.5 Writing a simple program in the IDE editor

When we've finished typing this in the editor window, we can save our little program to a file (Select Menu: File/Save As...) and run the program by clicking on the green arrow in the tool bar of the IDE (Figure E.3). If we run this program we will see:

```
The answer is: 0.5
>>>
```

Listing E.6 Writing a simple program in the IDE editor

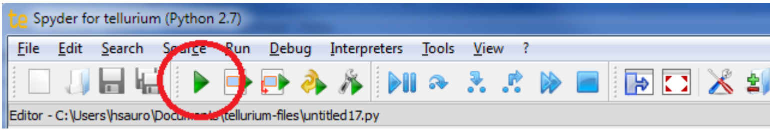


Figure E.3 Screen-shot of Tellurium, focusing on the Toolbar with the run button circled.

The IDE allows a user to have as many program files open at once, each program file is given its own tab so that it is easy to move from one to the other. This is useful if one is working on multiple models at the same time.

We will now use the editor to write the simple program that loops ten times, this is shown below:

```
for i in range (10):
    print i,
```

Listing E.7 A simple loop in python

This will generate the sequence:

```
0 1 2 3 4 5 6 7 8 9
```

Listing E.8 Result from simple loop program

There are a number of new concepts introduced in this small looping program. The first line contains the `for` keyword can be translated into literal English as "for all elements in a list, do this". The list is generated from the `range()` function and in this case generates a list of 10 numbers starting at 0. `i` is the loop index and within the loop, `i` can be used in other calculations. In this case we will just print the value of `i` to the console. Each time the program loops it extracts the next value from the list and assigns it to `i`.

Two things are important to note in the print line. The first and most important is that the line has been indented four spaces. This isn't just for aesthetic reasons but is actually functional. It tells Python what code should

be executed *within* the loop. To elaborate we could add more lines to the loop, such as:

```
for i in range (10):
    a = i
    b = a*2
    print b,
print "Finished Loop"
```

Listing E.9 A simple loop illustrating multiple statements

In this example there are three indented lines, this means that these three lines will be executed within the loop. The last line which prints a message, is not indented and therefore will not be executed within the loop. This means we only see the message appear once right at the end. The output for this little program is shown below.

```
0 2 4 6 8 10 12 14 16 18 Finished Loop
```

Another important point worth noting is the use of the `,` after the loop print statement. The comma is used to suppress a newline. This is why the output appears on one line only. If we had left out the comma each print statement would be on its own line.

A final word about `range()`. Range takes up to three arguments. In the example we only gave one argument, 10. A single argument means create a list starting at zero, incrementing one for each item until the incremented value reaches 10. A second argument such as `range (5, 10)` means start the list at 5 rather than zero. Finally, a third argument can be used to specify the increment size. For example the command `range (1, 10, 2)` will yield the list:

```
[1, 3, 5, 7, 9]
```

The easiest way to try out the various options in range is to type them at the console to get immediate feedback.

The use of variables, printing results, importing libraries and looping are

probably the minimum concepts one needs to start using Python. However there are a huge range of resources online to help learn Python. Of particular interest is the codecademy web site (http://www.codecademy.com/). This site offers an interactive means to learn Python (including other programming languages).

E.2 Describing Reaction Networks using Antimony

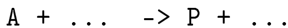
The code shown in the panel below illustrates the description of a very simple model using the Antimony syntax [?] followed by two lines of Python that uses libRoadRunner to run a simulation of the model. In this section we will briefly describe the Antimony syntax. A more detailed description of Antimony can be found at http://antimony.sourceforge.net/index.html.

```
import tellurium as te

rr = te.loada ('''
  S1 -> S2; k1*S1;
  S1 = 10; k1 = 0.1
''')

rr.simulate (0, 50, 100)
rr.plot()
```

Listing E.10 Simple model Antimony and simulated using libRoadRunner

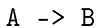
The main purpose of Antimony is to make it straight forward to specify complex reaction networks using a familiar chemical reaction notation.

A chemical reaction can be an enzyme catalyzed reaction, a binding reaction, a phosphorylation, a gene expressing a protein or any chemical process that results in the conversion of one of more species (reactants) to a set of one or more other species (products). In Antimony, reactions are described using the notation:



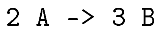
```
A + ...  -> P + ...
```

where the reactants are on the left side and products on the right side. The left and right are separated by the -> symbol. For example:



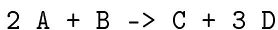
```
A -> B
```

describes the conversion of reactant A into product B. In this case one molecule of A is converted to one molecule of B. The following example shows non-unity stoichiometry:



```
2 A -> 3 B
```

which means that two molecules of A react to form three molecules of B. Bimolecular and other combinations can be specified using the + symbol, that is:



```
2 A + B -> C + 3 D
```

tells us that two molecules of A combine with one molecule of B to form one molecule of C and three molecules of D.

To specify species that do not change in time (boundary species), add a dollar character in front of the name, for example:

```
$A + B -> C
```

means that during a simulation A is fixed.

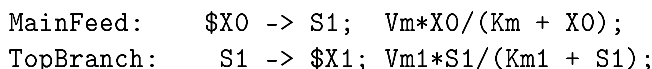
Reactions can be named using the syntax J1:, for example:

```
J1:  A + B -> C
```

means the reaction has a name, J1. Named reaction are useful if you want to refer to the flux of the reaction; kinetic rate laws come immediately after the reaction specification. If only the stoichiometry matrix is required, it is not necessary to enter a full kinetic law, a simple `...  -> S1; v;` is sufficient. Here is an example of a reaction that is governed by a Michaelis-Menten rate law:

```
A -> B; Vm*A/(Km + A);
```

Note the semicolons. Here is a more complex example involving multiple reactions:



```
MainFeed:    $X0 -> S1;  Vm*X0/(Km + X0);
TopBranch:    S1 -> $X1; Vm1*S1/(Km1 + S1);
```

```
    BottomBranch: S1 -> $X2; Vm2*S1/(Km2 + S1);
```

There is no need to pre-declare the species names shown in the reactions or the parameters in the kinetic rate laws. Strictly speaking, declaring the names of the floating species is optional, however this feature is for more advanced users who wish to define the order of rows that will appear in the stoichiometry matrix. For normal use there is no need to pre-declare the species names. To pre-declare parameters and variables see the example below:

```
    const Xo, X1, X2; // Boundary species
    var S1;           // Floating species

    MainFeed:     $X0 -> S1;  Vm*X0/(Km + X0);
    TopBranch:     S1 -> $X1; Vm1*S1/(Km1 + S1);
    BottomBranch: S1 -> $X2; Vm2*S1/(Km2 + S1);
```

We can load an Antimony model into libRoadRunner using the short-cut command `loada`. For example:

```
rr = te.loada ('''
    const Xo, X1, X2; // Boundary species
    var S1;           // Floating species

    MainFeed:     $X0 -> S1;  Vm*X0/(Km + X0);
    TopBranch:     S1 -> $X1; Vm1*S1/(Km1 + S1);
    BottomBranch: S1 -> $X2; Vm2*S1/(Km2 + S1);
''')
```

To reference model properties and methods, the property or method must be proceeded with the roadrunner variable. e.g. `rr.S1 = 2.3;`

When loaded into libRoadRunner the model will be converted into a set of differential equations. For example, consider the following model:

```
    $Xo -> S1;  v1;
     S1 -> S2;  v2;
     S2 -> $X1; v3;
```

will be converted into:

$$\frac{dS_1}{dt} = v_1 - v_2$$

$$\frac{dS_2}{dt} = v_2 - v_1$$

Note that there are no differential equations for X_o and X_1. This is because they are fixed and do not change in time. If the reactions have non-unity stoichiometry, this is taken into account when the differential equations are derived.

E.2.1 Initialization of Model Values

To initialize the concentrations and parameters in a model we can add assignments after the network is declared, for example:

```
MainFeed:    $X0 -> S1;  Vm*X0/(Km + X0);
TopBranch:    S1 -> $X1; Vm1*S1/(Km1 + S1);
BottomBranch: S1 -> $X2; Vm2*S1/(Km2 + S1);

X0 = 3.4;  X1 = 0.0;
S1 = 0.1;
Vm = 12; p.Km = 0.1;
Vm1 = 14; p.Km1 = 0.4;
Vm2 = 16; p.Km2 = 3.4;
```

E.3 Using libRoadRunner in Python

libRoadRunner is a high performance simulator [?] that can simulate models described using SBML. In order to use Antimony with libRoadRunner it is necessary to first convert an Antimony description into SBML and then load the SBML into libRoadRunner. Telluirum provides a handy routine called `loadAntimonyModel` to help with this task (The short-cut name is

loada). To load an Antimony model we first assign an Antimony description to a string variable, for example:

```
model = '''
   S1 -> S2; k1*S1;

   S1 = 10; k1 = 0.1;
'''
```

We now use the `loadAntimonyModel (model)` or `loada` to load the model into libRoadRunner.

```
>>> rr = te.loadAntimonyModel (model)
```

Listing E.11 Loading an Antimony model

In this book we generally use the short-cut command as follows:

```
rr = te.loada ('''
    S1 -> S2; k1*S1;

    S1 = 10; k1 = 0.1;
 ''')
>>>
```

Listing E.12 Loading an Antimony model using the short-cut command

Note that `loadAntimonyModel` and `loada` are part of the Tellurium Python package supplied with the Tellurium installer. If the Tellurium packages hasn't been loaded, use the following command to load the Tellurium package:

```
>>> import tellurium as te
```

Listing E.13 Importing the Tellurium Package

E.3.1 Time Course Simulation

Once a model has been loaded into libRoadRunner, performing a simulation is very straight forward. To simulate a model we use the libRoadRunner simulate method. This method has many options but for everyday use four options will suffice. The following panel illustrates a number examples of how to use simulate.

```
>>> result = rr.simulate ()
>>> result = rr.simulate (0, 10)
>>> result = rr.simulate (0, 10, 100)
>>> result = rr.simulate (0, 10, 100, ['time', 'S1'])
```

Listing E.14 Calling the simulate method

Argument	Description
1st	Start Time
2nd	End Time
3rd	Number of Points
4th	Selection List

Let us focus on the forth version of the simulate method that takes four arguments. This call will run a time course simulation starting at time zero, ending at time 10 units, and generating 100 points. The results of the run are deposited in the matrix variable, `result`. At the end of the run, the `result` matrix will contain columns corresponding to the time column and all the species concentrations as specified by the forth argument. The forth argument can be used to change the columns that are returned from the simulate method. For example:

```
>>> result = rr.simulate (0, 10, 1000, ['S1'])}
```

will return a matrix 1,000 rows deep and one column wide that corresponds to the level of species `S1`.

Note that the special variable `Time` is available and represents the independent time variable in the model.

To visualize the output in the form of a graph, one can pass the matrix of results to the plot command. In the following example we return one species level, S1 and three fluxes. Finally we plot the results.

```
result = rr.simulate (0, 10, 1000, ['Time', 'S1', 'J1', 'J2', 'J3']);
te.plotWithLegend (rr, result)
```

or if we are not interested in the result data itself we can use the libRoad-Runner plot:

```
rr.simulate (0, 10, 1000, ['Time', 'S1', 'J1', 'J2', 'J3']);
rr.plot()
```

It is possible to set the output column selections separately using the command:

```
rr.selections = ['time', 'S1']
```

This can save some typing each time a simulation needs to be carried out. By default the selection is set to time as the first column followed by all molecular species concentrations. As such it is more common to simply enter the command:

```
>>> result = rr.simulate (0, 10, 50)
```

In fact even the start time and end time and number of points are optional and if missing, simulate will revert to its defaults.

```
>>> result = rr.simulate()
```

E.3.2 Plotting Simulation Results

Tellurium comes with Matplotlib which is a common plotting package used by many Python users. To simplify its use we provide two simple plotting calls:

```
te.plot (array)
te.plotWithLegend (rr, array)
```

The first takes the resulting array generated by a call to `simulate` and uses the first column as the x axis and all subsequent columns as y axis data. The second call takes the roadrunner variable as well as the array and does the same kind of plot but this time adds a legend to the plot. We will use the first plotting command in the next section where we merge together multiple simulations.

E.3.3 Applying Perturbations to a Simulation

Often in a simulation we may wish to perturb a species or parameter at some point during the simulation and observe what happens. One way to do this in Tellurium is to carry out two separate simulations where a perturbation is made in between the two simulations. For example, let's say we wish to perturb the species concentration for a simple two step pathway and watch the perturbation decay. First, we simulate the model for 10 time units; this gives us a transient and then a steady state.

```
import numpy # Required for vstack
import tellurium as te

rr = te.loada ('''
     $Xo -> S1;  k1*Xo;
      S1 -> $X1; k2*S1;

     Xo = 10; k1 = 0.3; k2 = 0.15;
''')

m1 = rr.simulate (0, 40, 50)
```

We then make a perturbation in `S1` as follows:

```
rr.S1 = rr.S1 * 1.6
```

which increases `S1` by 60%. We next carry out a second simulation:

```
m2 = rr.simulate (40, 80, 50)
```

Note that we set the time start of the second simulation to the end time of the first simulation. Once we have the two simulations we can combine the matrices from both simulations using the Python command `vstack`

```
% Merge the two result array together
m = numpy.vstack ((m1, m2))
```

Finally, we plot the results, screen-shot shown in Figure E.4.

```
te.plotArray (m)
```

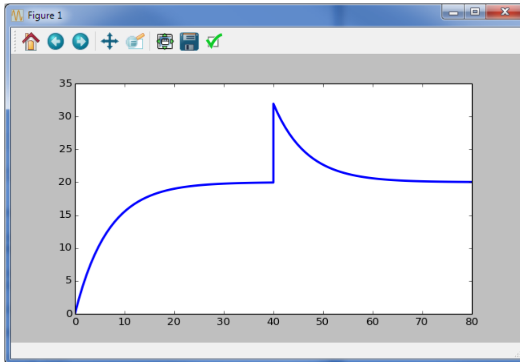


Figure E.4 Screen-shot from Matplotlib showing effect of perturbation in S1.

E.3.4 Steady State and Metabolic Control

To evaluate the steady-state first make sure the model values have been previously initialized, then enter the following statement at the console.

```
>>> rr.getSteadyState()
```

This statement will attempt to compute the steady state and return a value

indicating how effective the computation was. It returns the norm of the rate of change vector (i.e. sqrt (Sum of dydt)). The closer this is to zero, the better the approximation to the steady state. Anything less that 10^{-4} usually indicates that a steady state has been found.

Once a steady state has been evaluated, the values of the metabolites will be at their steady state values, thus `S1` will equal the steady state concentration of `S1`.

The fluxes through the individual reactions can be obtained by either referencing the name of the reaction (e.g. `J1`), or via the short-cut command`rv`. The advantage to looking at the reaction rate vector is that the individual reaction fluxes can be accessed by indexing the vector (see example below). **Note that indexing is from zero.**

```
>>> print rr.J1, rr.J2, rr.J3
3.4, ...
>>> for i in range (0, 2):
...     print rr.rv()[i]
3.4
etc
->
```

To compute control coefficients use the statement:

`getCC (Dependent Measure, Independent parameter)`

The dependent measure is an expression usually containing flux and metabolite references, for example, `S1, J1`. The independent parameter must be a simple parameter such as a Vmax, Km, ki, boundary metabolite (X0), or a conservation total such as `cm_xxxx`. Examples include:

```
  rr.getCC ('J1', 'Vmax1')
  rr.getCC ('J1', 'Vm1') + rr.getCC ('J1', 'Vm2')
  rr.getCC ('J1', 'X0')
  rr.getCC ('J1', 'cm_xxxx')
```

To compute elasticity coefficients use the statement:

`getEE (Reaction Name, Parameter Name)`

For example:

```
rr.getEE ('J1', 'X0')
rr.getEE ('J1', 'S1')
```

Since `getCC` and `getEE` are built-in functions, they can be used alone or as part of larger expressions. Thus, it is easy to show that the response coefficient is the product of a control coefficient and the adjacent elasticity by using:

```
R = rr.getCC ('J1', 'X0')
print R - rr.getCC ('J1', 'Vm') * rr.getEE ('J1', 'X0')
```

To obtain the conservation matrix for a model use the model method, `getConservationMatrix`. Note that in the Antimony text we use the `var` word to predeclare the species so that we can set up the rows of the stoichiometry matrix in a certain order if we wish. This allows us to obtain conservation matrices with only positive terms.

```
import tellurium as te

rr = te.loada ('''
   var ES, S1, S2, E;

  J1: E + S1 -> ES; v;
  J2: ES -> E + S2; v;
  J3: S2 -> S1; v;
''')

print rr.getConservationMatrix()
print rr.fs()

# Output
[[ 1.  1.  1.  0.]
 [ 1.  0.  0.  1.]]
['ES', 'S1', 'S2', 'E']
```

The result given above indicates that the conservation relations, `ES + S1 + E` and `E + ES` exist in the model. As a result, Tellurium would generate two internal parameters of the form `cm` corresponding to the two relations.

E.3.5 Other Model Properties of Interest

There are a number of predefined objects associated with a reaction network model which might also be of interest. For example, the stoichiometry matrix, `sm`, the rate vector `rv`, the species levels vector and `dv` which returns the rates of change.

```
print rr.sm()
print rr.rv()
print rr.sv()
print rr.dv()
```

The names for the parameters and variables in a model can be obtained the short-cuts:

```
print rr.fs()  # List of floating species names
print rr.bv()  # List of boundary species names
print rr.ps()  # List of parameter names
print rr.rs()  # List of reaction names
print rr.vs()  $ List of compartment names
```

The jacobian matrix can be returned using the command: `rr.getFullJacobian()`).

E.4 Generating SBML and Matlab Files

Tellurium can import and export standard SBML [41] as well as export Matlab scripts for the current model. To load a model in SBML, load it directly into libRoadRunner. For example:

```
>>> rr = roadrunner.RoadRunner ('mymodel.xml')
>>> result = rr.simulate (0, 10, 100)
```

There are two ways to retrieve the SBML, one can either retrieve the original SBML loaded using `rr.getSBML()` or retrieve the *current* SBML using `rr.getCurrentSBML()`. Retrieving the current SBML can be useful

if the model has been changed. To save the SBML to a file we can use the Tellurium helper function `saveToFile ()`, for example:

```
>>> te.saveToFile ('mySBMLModel.xml', rr.getCurrentSBML())
```

To convert an SBML file into Matlab, use the `getMatlab` method:

```
import tellurium as te

rr = te.loada ('''
    S1 -> S2; k1*S1;
    S2 -> S3; k2*S2;
    S1 = 10; k1 = 0.1; k2 = 0.2;
''')

# Save the SBML
te.saveToFile ('model.xml', rr.getSBML())

# Save the Matlab
te.saveToFile ('model.mat', rr.getMatlab())
```

E.5 Exercise

Figure E.5 shows a two gene circuit with a feedforward loop. Assume the following rate laws for the four reactions:

$$
\begin{aligned}
v_1 &= k_1 X_o \\
v_2 &= k_2 x_1 \\
v_3 &= k_3 X_o \\
v_4 &= k_4 x_1 x_2
\end{aligned}
$$

Assume that all rate constants are equal to one and that $X_o = 1$. Assume X_o is a fixed species.

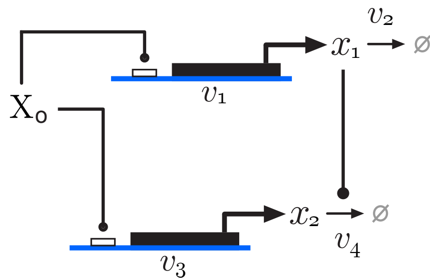


Figure E.5 Two gene circuit with feedfoward loop.

1. Use Tellurium to model this system.
2. Run a simulation of the system from 0 to 10 time units.
3. Next, change the value of X_o to 2 (double it) and rerun the simulation for another 10 time units from where you left off in the last simulation. Combine both simulations and plot the result, that is time on the x-axis, and X_o and x_2 on the y-axis.
4. What do you see?
5. Write out the differential equations for x_1 and x_2.
6. Show algebraically that the steady state level of x_2 is independent of X_o.

F

Answers to Questions

References

[1] Acerenza, L., and A. Cornish-Bowden. 1997. Biochem J **327 (Pt 1)**:217–224.

[2] Aparicio, O., Joseph V Geisberg, and Kevin Struhl. 2004. Curr Protoc Cell Biol **Chapter 17**:Unit 17.7.

[3] Bakker, B. M., P. A. Michels, F. R. Opperdoes, and H. V. Westerhoff. 1997. J Biol Chem **272** (6):3207–3215.

[4] Barabási, A. L., and Z N Oltvai. 2004. Nat Rev Genet **5** (2):101–113.

[5] Biondi, E. G., Sarah J Reisinger, Jeffrey M Skerker, Muhammad Arif, Barrett S Perchuk, Kathleen R Ryan, and Michael T Laub. 2006. Nature **444** (7121):899–904.

[6] Blackman, F. F. 1905. Ann. Botany **19**:281–295.

[7] Blüthgen, N., F J Bruggeman, S Legewie, H Herzel, H V Westerhoff, and B N Kholodenko. 2006. FEBS J **273** (5):895–906.

[8] Bode, A. M., and Zigang Dong. 2004. Nat Rev Cancer **4** (10):793–805.

[9] BRETT, D., H. POSPISIL, J. VALCARCEL, J. REICH, and P. BORK. 2002. Nature genetics **30** (1):29–30.

[10] Brilli, M., Marco Fondi, Renato Fani, Alessio Mengoni, Lorenzo Ferri, Marco Bazzicalupo, and Emanuele Biondi. 2010. BMC Systems Biology **4** (1):52.

[11] Burrell, M. M., P. J. Mooney, M. Blundy, D. Carter, F. Wilson, J. Green, K. S. Blundy, and T. ap Rees. 1994. Planta **194**:95——101.

[12] Burton, A. C. 1936. Journal Cellular and Comparative Physiology **9(1)**:1–14.

[13] Chance, B., and G R Williams. 1955. J Biol Chem **217** (1):429–438.

[14] Cohen, P. 2000. Trends in Biochemical Sciences **25** (12):596–601.

[15] Cornish-Bowden, A., and J. H. Hofmeyr. 2002. J. theor. Biol **216**:179–191.

[16] Cornish-Bowden, A., J. H. Hofmeyr, and M. Cardenas. 2002. Biochemical Society Transactions **30**:43–47.

[17] de Graauw, M. , (ed.), 2009. *Phospho-Proteomics*, vol. 527 of *Methods in Molecular Biology*. Humana Press.

[18] Dickson, R. C., and M. D. Mendenhall. , (ed.), 2004. *Signal Transduction Protocols*, vol. 284 of *Methods in Molecular Biology*. Humana Press, 2nd Edition.

[19] Eisenthal, R., and A. Cornish-Bowden. 1998. Journal of Biological Chemistry **273** (10):5500–5505.

[20] Fell, D. A., and H. M. Sauro. 1985. Eur. J. Biochem. **148**:555–561.

[21] Fell, D. A., and H. M. Sauro. 1990. Eur. J. Biochem. **192**:183–187.

[22] Fields, S., and O Song. 1989. Nature **340** (6230):245–246.

[23] Flint, H. J., R. W. Tateson, I. B. Bartelmess, D. J. Porteous, W. D. Donochie, and H. Kacser. 1981. Biochem. J. **200**:231–246.

[24] Gama-Castro, S., Verónica Jiménez-Jacinto, Martín Peralta-Gil, Alberto Santos-Zavaleta, Mónica I Peñaloza-Spinola, Bruno Contreras-Moreira, Juan Segura-Salazar, Luis Muñiz-Rascado, Irma Martínez-Flores, Heladia Salgado, César Bonavides-Martínez, Cei Abreu-Goodger, Carlos Rodríguez-Penagos, Juan Miranda-Ríos, Enrique Morett, Enrique Merino, Araceli M Huerta, Luis Treviño-Quintanilla,

and Julio Collado-Vides. 2008. Nucleic Acids Res **36** (Database issue):D120–D124.

[25] Gavin, A.-C., Patrick Aloy, Paola Grandi, Roland Krause, Markus Boesche, Martina Marzioch, Christina Rau, Lars Juhl Jensen, Sonja Bastuck, Birgit Dümpelfeld, Angela Edelmann, Marie-Anne Heurtier, Verena Hoffman, Christian Hoefert, Karin Klein, Manuela Hudak, Anne-Marie Michon, Malgorzata Schelder, Markus Schirle, Marita Remor, Tatjana Rudi, Sean Hooper, Andreas Bauer, Tewis Bouwmeester, Georg Casari, Gerard Drewes, Gitte Neubauer, Jens M Rick, Bernhard Kuster, Peer Bork, Robert B Russell, and Giulio Superti-Furga. 2006. Nature **440** (7084):631–636.

[26] Gekas, V., and M. Lopez-Leiva. 1985. Process biochemistry **20** (1):2–12.

[27] Goldbeter, A., and D. E. Koshland. 1981. Proc. Natl. Acad. Sci **78**:6840–6844.

[28] Goldbeter, A., and D. E. Koshland. 1984. J. Biol. Chem. **259**:14441–7.

[29] Goodyear, C., and G.J. Silverman. 2008. Cold Spring Harbor Protocols **2008** (9).

[30] Groen, A. K., R. J. A. Wanders, H. V. Westerhoff, R. van der Meer, and J. M. Tager. 1982. J. Biol. Chem. **257**:2754–2757.

[31] Hearon, J. Z. 1952. Physiol. Rev. **32**:499–523.

[32] Heinisch, J. 1986. Mol. Gen Genet. **202**:75——82.

[33] Heinrich, R., and T. A. Rapoport. 1974. Eur. J. Biochem. **42**:89–95.

[34] Heinrich, R., and S Schuster. 1996. *The Regulation of Cellular Systems*. Chapman and Hall.

[35] Higgins, J. 1963. Ann. N. Y. Acad. Sci. **108**:305–321.

[36] Hoefnagel, M., A Van Der Burgt, DE Martens, J Hugenholtz, and JL Snoep. 2002. Molecular biology reports **29** (1-2):157–161.

[37] Hofmeyr, J.-H. 1995. J Bioenerg Biomembr **27** (5):479–490.

[38] Hofmeyr, J.-H., and A Cornish-Bowden. 1991. Eur J Biochem **200** (1):223–236.

[39] Hofmeyr, J.-H., H. Kacser, and K. J. van der Merwe. 1986. Eur. J. Biochem. **155**:631–641.

[40] Huang, C. F., and J. E. Ferrell. 1996. Proc. Natl. Acad. Sci **93**:10078–10083.

[41] Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le NovÃÆŠÂÂÍre, L. M. Loew, D. Lucio, P. Mendes, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. 2003. Bioinformatics **19**:524–531.

[42] Huerta, A. M., H. Salgado, D. Thieffry, and J. Collado-Vides. 1998. Nucleic Acids Res **26** (1):55–59.

[43] Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. 2001. Proc Natl Acad Sci U S A **98** (8):4569–4574.

[44] Jeong, H., S. P. Mason, A. L. Barabási, and Z. N. Oltvai. 2001. Nature **411** (6833):41–42.

[45] Kacser, H. 1983. Biochem. Soc. Trans. **11**:35–40.

[46] Kacser, H., and J. A. Burns. 1973. *In:* Davies, D. D. , (ed.), Rate Control of Biological Processes, vol. 27 of *Symp. Soc. Exp. Biol.* p. 65–104. Cambridge University Press.

[47] Karp, P. D., I. M. Keseler, A. Shearer, M. Latendresse, M. Krummenacker, S. M. Paley, I. Paulsen, J. Collado-Vides, S. Gama-Castro, M. Peralta-Gil, A. Santos-Zavaleta, M. I. Peñaloza-Spínola,

C. Bonavides-Martinez, and J. Ingraham. 2007. Nucleic Acids Res **35** (22):7577–7590.

[48] Keseler, I. M., Julio Collado-Vides, Alberto Santos-Zavaleta, Martin Peralta-Gil, Socorro Gama-Castro, Luis Muñiz-Rascado, César Bonavides-Martinez, Suzanne Paley, Markus Krummenacker, Tomer Altman, Pallavi Kaipa, Aaron Spaulding, John Pacheco, Mario Latendresse, Carol Fulcher, Malabika Sarker, Alexander G Shearer, Amanda Mackie, Ian Paulsen, Robert P Gunsalus, and Peter D Karp. 2011. Nucleic Acids Res **39** (Database issue):D583–D590.

[49] Kroeze, W. K., Douglas J Sheffler, and Bryan L Roth. 2003. Journal of Cell Science **116** (24):4867–4869.

[50] Krogan, N. J., Gerard Cagney, Haiyuan Yu, Gouqing Zhong, Xinghua Guo, Alexandr Ignatchenko, Joyce Li, Shuye Pu, Nira Datta, Aaron P Tikuisis, Thanuja Punna, José M Peregrín-Alvarez, Michael Shales, Xin Zhang, Michael Davey, Mark D Robinson, Alberto Paccanaro, James E Bray, Anthony Sheung, Bryan Beattie, Dawn P Richards, Veronica Canadien, Atanas Lalev, Frank Mena, Peter Wong, Andrei Starostine, Myra M Canete, James Vlasblom, Samuel Wu, Chris Orsi, Sean R Collins, Shamanta Chandran, Robin Haw, Jennifer J Rilstone, Kiran Gandi, Natalie J Thompson, Gabe Musso, Peter St Onge, Shaun Ghanny, Mandy H Y Lam, Gareth Butland, Amin M Altaf-Ul, Shigehiko Kanaya, Ali Shilatifard, Erin O'Shea, Jonathan S Weissman, C. James Ingles, Timothy R Hughes, John Parkinson, Mark Gerstein, Shoshana J Wodak, Andrew Emili, and Jack F Greenblatt. 2006. Nature **440** (7084):637–643.

[51] LaPorte, D. C., K Walsh, and D E Koshland. 1984. J Biol Chem **259** (22):14068–14075.

[52] Le Novère, N., M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M.I. Aladjem, S.M. Wimalaratne, *et al.*. 2009. Nature biotechnology **27** (8):735–741.

[53] Lee, T. I., Nicola J Rinaldi, François Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, Julia Zeitlinger, Ezra G

Jennings, Heather L Murray, D. Benjamin Gordon, Bing Ren, John J Wyrick, Jean-Bosco Tagne, Thomas L Volkert, Ernest Fraenkel, David K Gifford, and Richard A Young. 2002. Science **298** (5594):799–804.

[54] Macek, B., F. Gnad, B. Soufi, C. Kumar, J.V. Olsen, I. Mijakovic, and M. Mann. 2008. Molecular & Cellular Proteomics **7** (2):299.

[55] Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. 2000. Science **298**:1912–1934.

[56] Mardis, E. R. 2007. Nat Methods **4** (8):613–614.

[57] Markevich, N. I., J B Hoek, and B. N. Kholodenko. 2004. J. Cell Biol. **164**:353–9.

[58] Mattick, J. 2004. Pharmacogenomics J **4**:9–16.

[59] Monod, J., J Wyman, and J. P. Changeux. 1965. J Mol Biol **12**:88–118.

[60] Morales, M. F. 1921. Journal Cellular and Comparative Physiology **30**:303–313.

[61] Ortega, F., J L Garcés, F Mas, B N Kholodenko, and M Cascante. 2006. FEBS J **273** (17):3915–3926.

[62] Palsson, B. O. 2007. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press.

[63] PARK, D. 1988. Computers & chemistry **12** (2):175–188.

[64] Phizicky, E., P.I.H. Bastiaens, H. Zhu, M. Snyder, and S. Fields. 2003. Nature **422** (6928):208–215.

[65] Ptacek, J., G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, *et al.*. 2005. Nature **438** (7068):679–684.

[66] Ptacek, J., and M. Snyder. 2006. Trends in Genetics **22** (10):545–554.

[67] Reich, J. G., and E. E. Selkov. 1981. *Energy metabolism of the cell.* Academic Press, London.

[68] Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young. 2000. Science **290** (5500):2306–2309.

[69] Sauro, H. M. 1990. *In:* Cornish-Bowden, A., and M. L. Cárdenas. , (ed.), Control of Metabolic Processes, NATO ASI Series chapter 17, p. 225–230. Plenum Press, New York.

[70] Sauro, H. M. 1994. BioSystems **33**:15–28.

[71] Sauro, H. M. 2011. *Enzyme Kinetics for Systems Biology*. Ambrosius Publishing. First Edition.

[72] Sauro, H. M. 2012. *Enzyme Kinetics for Systems Biology*. Ambrosius Publishing. 2nd Edition.

[73] Sauro, H. M., and D. A. Fell. 1991. Mathl. Comput. Modelling **15**:15–28.

[74] Sauro, H. M., and B. N. Kholodenko. 2004. Prog Biophys Mol Biol. **86**:5–43.

[75] Savageau, M. A. 1972. Curr. Topics Cell. Reg. **6**:63–130.

[76] Seshasayee, A. S. N., P. Bertone, G. M. Fraser, and N.M. Luscombe. 2006. Current Opinion in Microbiology **9** (5):511–519.

[77] Shen-Orr, S. S., R. Milo, S. Mangan, and U. Alon. 2002. Nature Genetics **31**:64–68.

[78] Small, J. R., and D. A. Fell. 1990. Eur. J. Biochem. **191**:405–411.

[79] Smith, G. P. 1985. Science **228** (4705):1315–1317.

[80] Stitt, M., W. P. Quick, U. Schurr, E.-D. Schulze, S. R. Rodermel, and L. Bogorad. 1991. Planta **183**:555–566. 10.1007/BF00194277.

[81] Taft, R., and JS Mattick. 2004. Arxiv preprint q-bio.GN/0401020 .

[82] Toledo, F., and Geoffrey M Wahl. 2006. Nat Rev Cancer **6** (12):909–923.

[83] Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. 2000. Nature **403** (6770):623–627.

[84] Vallabhajosyula, R. R., V Chickarmane, and H M Sauro. 2006. Bioinformatics **22** (3):346–353.

[85] Waley, S. G. 1964. Biochem. J. **91** (3):514–0.

[86] Wang, E., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Nature **456**:470–476.

[87] Webb, J. 1963. *Enzyme and Metabolic Inhibitors: General principles of inhibition. Volume 1*. Number v. 1. Academic Press.

[88] Wilkinson, D. J. 2012. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC Press, Boca Raton, Florida, 2nd edition.

[89] Woods, J. H., and H. M. Sauro. 1997. Comput Appl Biosci **13** (2):123–130.

[90] Wright, S. 1934. The American Naturalist **68**:24–53.

History

1. VERSION: 0.1

   Date: 2011-04-2

   Author(s): Herbert M. Sauro

   Title: Introduction to Metabolic Control Analysis

   Modification(s): First Edition Release

2. VERSION: 0.4

   Date: 2012-02-12

   Author(s): Herbert M. Sauro

   Title: Introduction to Metabolic Control Analysis

   Modification(s): Added new appendices and chapters