Pathway Analysis and Optimization in Metabolic Engineering

Rapid advances in functional genomics and proteomics have created a platform from which pressing problems in biotechnology can be addressed. Facility in the targeted manipulation of the genetic and metabolic composition of organisms, combined with unprecedented computational power, is forging a niche for a new subspecialty of biotechnology called metabolic engineering. This emerging field has metabolic pathways and gene networks as its targets, and optimization as its ultimate goal.

Pathway Analysis and Optimization in Metabolic Engineering introduces researchers and advanced students in biology and engineering to methods of optimizing biochemical systems of biotechnological relevance. It examines the development of strategies for manipulating metabolic pathways, demonstrates the need for effective systems models, and discusses their design and analysis, while placing special emphasis on optimization. The authors propose power-law models and methods of Biochemical Systems Theory toward these ends. All concepts are derived from first principles, and the text is richly illustrated with numerous graphs and examples throughout. Special features include:

- both nontechnical and technical introductions to models of biochemical systems
- a review of basic methods of model design and analysis
- concepts of optimization, both generic and specific to biotechnology
- detailed case studies

Applications of metabolic engineering range from the industrial production of organic acids and enzymes to antibiotics and agricultural products. It is a truly interdisciplinary field, and while biotechnologists will find this introductory book a highly valuable reference, it will also be of utility to other scientists and engineers interested in biological systems.

Néstor V. Torres is Associate Professor of Biochemistry and Molecular Biology at the Universidad de La Laguna, Canary Islands, Spain. His research group has pioneered research in metabolic control analysis and biochemical systems theory, and in the application of mathematical and computational modeling studies to the optimization of biochemical systems.

Eberhard O. Voit is Professor in the Department of Biometry and Epidemiology, and in the Department of Biochemistry and Molecular Biology, at the Medical University of South Carolina. He is the editor of *Canonical Modeling* (Van Nostrand/Reinholt, 1991) and the author of *Computational Analysis of Biochemical Systems* (Cambridge University Press, 2000).

Pathway Analysis and Optimization in Metabolic Engineering

NÉSTOR V. TORRES
Universidad de La Laguna, Canary Islands, Spain

EBERHARD O. VOIT
Medical University of South Carolina

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521177481

First published 2002
First paperback edition 2011

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Torres, Néstor V., 1958–
Pathway analysis and optimization in metabolic engineering / Néstor V. Torres,
Eberhard O. Voit.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-80038-2
1. Biochemical engineering – Mathematical models. 2. Mathematical optimization.
3. System theory. I. Voit, Eberhard O. II. Title.
TP248.3 .T67 2002
660.6′3 – dc21 2002071511

ISBN 978-0-521-80038-9 Hardback
ISBN 978-0-521-17748-1 Paperback

Additional resources for this publication at www.cambridge.org/9780521177481

To those who taught us and those who learn from us

Contents

Preface

If the 1990s were the decade of molecular biology, culminating in the completion of the Human Genome Project, the first decade of the twenty-first century can be envisaged as the time of applying that new knowledge to pressing problems in medicine, pharmaceutical sciences, agriculture, and, generally, the biotech industry. Advances in functional genomics and the targeted manipulation of the genetic composition of organisms, combined with unprecedented computational power and access, are in the process of forging a new subspecialty of biotechnology, called *Metabolic Engineering*. Metabolic engineering offers a different perspective, a fresh approach with a focus not only on the development of better tools for the biological manipulation of organisms, but also on a deeper understanding of metabolic pathways in a systemic, integrative way.

Metabolic engineering is an interdisciplinary science that has grown from numerous roots in molecular biology and biochemistry, genetics, chemical engineering, biotechnology, mathematical modeling, and systems analysis. Its targets are metabolic pathways and gene networks, often in microorganisms, and its goal is the optimization of the yield of some desired metabolite. Microorganisms are manipulated and grown at an industrial scale to produce energy (for instance, in the form of ethanol), organic solvents, gums, and pigments. The food industry relies heavily on preservatives such as citric acid, which is currently produced by microbes at a rate of several hundred thousand tons per year. The pharmaceutical industry uses microorganisms for the production of vitamins, amino acids, lipids, enzymes, and precursors for antibiotics.

Metabolic engineering has two major components. One is the development of strategies for manipulating pathways in organisms, and the other is the actual biotechnological implementation of such strategies. This book exclusively treats the strategic component and does not discuss the astonishing advances in molecular biology and biotechnology that make it worthwhile even developing strategies for pathway manipulation. The numerous novel options for specifically altering the genetic or genomic make-up of organisms, the seemingly unlimited possibilities of modern biology,

the methods of fermentation technology, and chemical process control reach far beyond the scope of what we want to present in this text.

Much of this book will directly or indirectly discuss optimization applied to metabolic networks. Of course, neither the development of optimization methods nor the study of metabolic networks is new, but the specific and customized application of one to the other was not much considered until about a decade ago. In the olden days of fermentation technology, yield in microorganisms was improved successively through chemically induced mutagenesis and subsequent selection of the best-suited strains. Combined with simultaneous innovations in experimental conditions and growth media, the strategy of mutation and selection was very fruitful, often increasing the yield of the original parent strain more than a hundredfold. In recent years, however, the performance of well-studied systems like ethanol or citric acid production has begun to stagnate, and the mutation-selection approach in general has apparently reached an impasse in many cases. The former strategies of removing, one by one, the most restrictive bottlenecks in a metabolic pathway have become insufficient.

Faced with this situation, some forward-looking metabolic engineers have shifted their attention toward a deeper mechanistic understanding of the biochemical processes that are involved in the synthesis of a desired product. Initially targeting specific biochemical or transport steps that seemed to limit the microbial production process, they soon realized that a more global view of metabolism is required. This view must consider integrated pathways or, going one step further, networks of pathways, because alterations in one pathway very often have direct or indirect consequences that affect other pathways. If a microbe uses larger than normal amounts of NAD^+ to generate ethanol, the demand for NAD^+ is increased and will most likely be satisfied by other pathways. The metabolic engineering question thus becomes how one should use available kinetic information to redirect material flux toward a desired end product, without compromising other pathways and, thereby, the viability of the organism.

The realization of the intricate connectedness of enzyme-catalyzed reactions, individual pathways, and integrated networks of pathways makes it clear that mathematical and computational methods must be employed. Mathematics facilitates the bookkeeping issues arising from systems with many constituents and allows the quantitative specification of processes with a power that is simply unachievable by the unaided human mind. Efficacious computational tools must accompany the theoretical advances in mathematical representation and analysis. Only a few decades ago, the mathematical analysis of a realistically sized metabolic system would have been infeasible or even impossible, because the underlying equations were unsolvable. And although computer power has doubled about every eighteen months, we are still not at a point where every mathematical problem can be solved numerically in a reasonable period of time. This implies that the choice of "good" mathematical representations and effective methods of analysis is an important issue. As an example, nonlinear regression and the search for maxima in nonlinear systems are theoretically well understood in principle. However, even large computers and the

best available software do not always solve such problems in an expeditious and satisfactory manner.

These computational challenges are especially relevant for the analysis and optimization of metabolic pathways, as they contain many substrates, intermediate metabolites, products, enzymes, cofactors, and modulators. Additionally, enzyme-catalyzed steps have nonlinear characteristics that preclude simple scaling. For instance, most proceed quickly for low substrate concentrations and tend to saturate for high concentrations. Activators and inhibitors affect reactions in a nonlinear fashion and may lead to damped or increasing oscillations, stable limit cycles, or chaotic time courses. All these phenomena are found in living systems, so a valid mathematical description of a metabolic system must be able to account for them. Thus, if our goal is the optimization of some flux or metabolite under practically relevant conditions, it is clear that we are faced with a nonlinear search problem.

The optimization of nonlinear systems is notoriously difficult, but metabolic engineers have developed strategies to minimize or even avoid nonlinearities. After the brief discussion in the previous paragraph, one may think that such attempts would be futile or head in the wrong direction. However, this conclusion is faulty. By focusing on specific properties of metabolic networks and by using prudent approximations, different groups of researchers have proposed two types of linear approaches, each with a number of extensions and generalizations. One approach is based on the stoichiometry of the metabolic network. It studies fluxes characterizing the production and degradation of individual metabolites. At steady state, where the concentrations of all metabolites are constant and all incoming fluxes exactly balance all outgoing fluxes, the stoichiometry leads to a linear system that is readily analyzed with methods of matrix algebra. Extensions of this stoichiometric strategy include the consideration of physiological, physicochemical, and thermodynamic constraints that the organism must satisfy in addition to the purely stoichiometric balance. Other extensions study features of hypothetical paths through the metabolic network and the feasibility of optimally rerouting and redistributing flux. Very importantly, the linearity of the stoichiometric system allows this approach to be scaled up to essentially any size.

The stoichiometric approach more or less ignores the nonlinearities that govern enzymatic steps. On one hand, one may be willing to pay this price for the advantages of linearity and unlimited size. On the other hand, one must wonder to what degree it is valid to ignore the regulatory features of pathways, which obviously exist and often lead to other systems responses than in an unregulated system. One approach that fully acknowledges the nonlinearities, yet ultimately leads to linear optimization tasks, is based on a modeling framework called *Biochemical Systems Theory* (BST). The hallmark of this theory is approximation of rate laws and other processes with products of power-law functions. Mathematically, this type of representation is very convenient and solidly grounded in Taylor's theory of approximating functions with polynomials. Further theoretical support is found in Bode analysis of control engineering and in the laws of allometry, which pervade biology from biochemical and physiological levels to phenomena in organismic growth and population biology.

Power laws also appear in elemental chemical kinetics and have time and again been shown to provide valid representations of cellular processes in intact organisms.

In this book, we propose power-law approximations and methods of BST as tools for pathway analysis and optimization in metabolic engineering. Pathway analysis with methods of BST began in the late 1960s and has been successfully applied to a wide variety of phenomena within and outside biology and biotechnology. The optimization aspect is more recent, with the first article on the subject published in 1992. As of yet, most optimization work has been of a theoretical or predictive nature, but experimental confirmations of predictions based on optimization with methods of BST have lately begun to appear in the literature.

As a highly interdisciplinary venture, metabolic engineering requires some basic knowledge in several areas of biology, chemistry, mathematics, computing, and engineering. A text on a subject this widespread creates the challenge of suitable presentation. Indeed, we expect that there will be few readers who will not find some parts of this book too easy or too difficult. Our personal involvement with biologists influenced our decision to write the book with a quantitatively interested biologist in mind who has had some basic calculus or at least is not afraid of unfamiliar symbols and terminologies. Of course, we also hope to attract other scientists and engineers interested in metabolic pathways. Quantitatively trained readers may find some material on optimization methods rather trivial, but then again, they may enjoy a quick refresher or the option of simply skipping the material.

We have organized the book in the following fashion. Chapter 1 discusses the search for useful models. What are we looking for when we want to analyze and optimize pathways? What are our options? How can we distinguish between a useful model and a purely academic exercise in applied mathematics? The chapter describes some established and some more recent modeling strategies for biochemical and metabolic networks and explores to what degree these models are sufficient for the tasks of metabolic engineering.

Chapter 2 reviews key properties and methods of BST. It outlines the process of model design and discusses which components should be included and to what degree simplifying assumptions are valid and necessary. The first step of model design leads to a symbolic model, which is subsequently parameterized with information from the lab or the literature. Once a numerical model is established, it is subjected to an array of analyses. Some address the features of the model at steady state, some assess the robustness and stability of the model, and some diagnose dynamic responses to perturbations or structural changes.

Chapter 3 applies the methods of model design and analysis to a specific pathway of relevance, namely citric acid production in the mold *Aspergillus niger*. The chapter begins with the construction of a model that ultimately turns out to be unsatisfactory. We consider the inclusion of this "failed attempt" beneficial, because it reflects the reality of modeling, yet is seldom documented in the literature. Furthermore, even though the initial model is not suited for practical application, it contains a number of good and bad results that provide clues for modifying the model. The good results indicate which information and model structures should be retained. The bad results

are just as important. They pinpoint problem areas in the initial model, the diagnosis of which provides guidance for the revision process toward an improved model. In practice, the modeling process cycles through many rounds of design, analysis, interpretation, and the inevitable realization that some aspects of the investigated phenomenon are still not adequately addressed. Chapter 3 shows only two models, but these should suffice to illustrate the procedure of amending and fine-tuning models.

Chapter 4 prepares us for the optimization of metabolic networks that are discussed in later chapters. This chapter begins with the basics of optimization, discussing the determination of maxima and minima in simple univariate functions. Most readers will remember this topic from a college calculus class or even from high school. It is included here because the determination of maxima and minima in multivariate functions is quite similar in principle. Thus, the optimization of big systems should be an extension of straightforward calculus. However, the theoretical development reveals how quickly the barriers to an actual optimization of a large network become overwhelming. As an alternative to the theoretical approach, Chapter 4 introduces the concept of numerical optimization of large nonlinear systems. The algorithms that implement these concepts should hypothetically solve problems of any size, but practice again demonstrates the challenges of nonlinear systems.

Happily, Chapter 5 provides relief. After all the complications uncovered in Chapter 4, this chapter explains how metabolic systems can be optimized with linear methods. In the end, the stoichiometric and the BST approach both lead to a constrained optimization task known in the field as a *linear program*. Linear programs can be solved with numerous software packages, and it is not the intent of the chapter to discuss computational details. Instead, the philosophy is that the reader should develop a feel for the concepts of linear programs and either consult the rich literature for mathematical and computational details or trust the available methods and software packages. The chapter presents the components of linear programs with special emphasis on objective functions and constraints that are relevant in metabolic engineering.

Equipped with knowledge about biochemical systems theory from Chapter 2 and at least superficial expertise in linear programming from Chapter 5, Chapter 6 describes a full-blown optimization of the citric acid pathway discussed in Chapter 3. It derives suitable objective functions, discusses in detail constraints on metabolites and fluxes, as well as issues of metabolic burden, and analyzes a variety of results that are obtained under different conditions. It also touches on issues of experimental implementation.

Chapter 7 presents a second detailed case study, the glycolytic pathway, and poses different tasks for optimization. First, ethanol yield is the target of optimization, but subsequently, the same system is used for the optimization of glycerol and carbohydrate production. The chapter furthermore discusses methods for simultaneously satisfying multiple and often contradictory tasks. Initially, two tasks are reformulated in a manner that still allows optimization of a single objective that combines the two

tasks. This strategy and its results are then compared with a true multiobjective optimization.

The concluding chapter is an attempt to predict the immediate future of pathway analysis and optimization in metabolic engineering. It seems rather evident that modern techniques of genome analysis will strongly influence metabolic engineering. The chapter indicates where we see the role of these techniques in terms of pathway analysis and optimization.

Many individuals have helped us create this book, and we would be remiss if we neglected to mention at least those who directly and significantly assisted us. The following have given us particularly valuable feedback and comments, and we would like to express to them our sincere gratitude: Fernando Alvarez-Vasquez, Julio Avila, Julio Banga, Agnieszka Bronowska, Carlos M. Gónzález-Alcón, W. Chip Hood, Alberto Marín-Sanguino, Kellie Sims, and Julio Vera.

Uncounted others have influenced our thinking, sparked our interest, and supported us throughout our careers, allowing us to develop and share the ideas proposed in this book. Thank you all!

Néstor V. Torres
La Laguna, Tenerife

Eberhard O. Voit
Charleston, SC