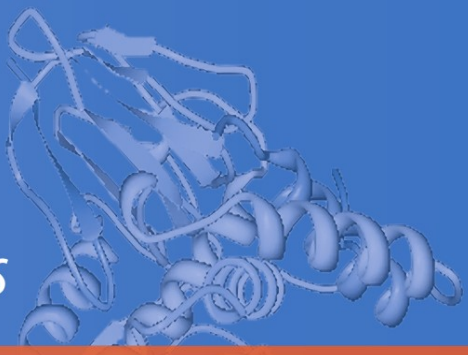
Subcellular Biochemistry 64

Xiaoyuan Wang
Jian Chen
Peter Quinn *Editors*



Reprogramming Microbial Metabolic Pathways

Springer

Reprogramming Microbial Metabolic Pathways

SUBCELLULAR BIOCHEMISTRY

Recent Volumes in this Series

Volume 48	**The Cronin Family of Proteins** Edited by Christoph S. Clemen and Ludwig Eichinger
Volume 49	**Lipids in Health and Disease** Edited by Peter J. Quinn and Xiaoyuan Wang
Volume 50	**Genome Stability and Human Diseases** Edited by Heinz-Peter Nasheuer
Volume 51	**Cholesterol Binding and Cholesterol Transport Proteins** Edited by Robin J. Harris
Volume 52	**A Handbook of Transcription Factors** Edited by Tim Hughes
Volume 53	**Endotoxins: Stricture, Function and Recognition** Edited by Xiaoyuan Wang and Peter J. Quinn
Volume 54	**Conjugation and Deconjugation of Ubiquitin Family Modifiers** Edited by Marcus Groettrup
Volume 55	**Purinergic Regulation of Respiratory Diseases** Edited by Maryse Picher and Richard C. Boucher
Volume 56	**Water Soluble Vitamins** Edited by Olaf Stanger
Volume 57	**Aging Research in Yeast** Edited by Michael Breitenbach, Michal S. Jazwinski and Peter Laun
Volume 58	**Phosphoinositides I: Enzymes of Synthesis and Degradation** Edited by Tamas Balla, Matthias Wymann and John D. York
Volume 59	**Phosphoinositides II: The Diverse Biological Functions** Edited by Tamas Balla, Matthias Wymann and John D. York
Volume 60	**Adherens Junctions: From Molecular Mechanisms to Tissue Development and Disease** Edited by Tony Harris

For further volumes:
http://www.springer.com/series/6515

Xiaoyuan Wang • Jian Chen • Peter J. Quinn
Editors

Reprogramming Microbial Metabolic Pathways

Springer

Editors
Xiaoyuan Wang
State Key Laboratory of Food Science
and Technology
Jiangnan University
Wuxi, China

Jian Chen
School of Biotechnology
Jiangnan University
Wuxi, Jiangsu, China

Peter J. Quinn
Division of Life Sciences
King's College London
London, UK

ISSN 0306-0225
ISBN 978-94-007-5054-8 ISBN 978-94-007-5055-5 (eBook)
DOI 10.1007/978-94-007-5055-5
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012951319

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Metabolism in microorganisms is at the core of cellular function, and involves hundreds of reactions that allow cells to convert raw materials into different metabolites. Some metabolites such as amino acids are required for microbial growth, while others such as antibiotics play no role in the growth, but usually have an important ecological function. Some microbes have a remarkable ability to utilize cheap sources of carbon and nitrogen to over-produce metabolites superfluous to their requirements. Such organisms can be exploited for large-scale production of valuable metabolites. Reprogramming microbial metabolic pathways could further increase the production of these metabolites.

In the past, microorganisms were genetically modified by chemically induced mutations to obtain strains capable of increased productivity of desired metabolites. However, mutations produced using such techniques were random and, as a result, the relevant enzymes modified were unknown leaving constraints on productivity unresolved. These hitherto hit or miss strategies have been replaced by rational approaches such as metabolic engineering in which metabolic pathways can be analyzed to determine the precise constraints and limitations on the production of desired metabolites. This knowledge can be applied to relieve these constraints, optimize the genetic and regulatory processes within cells resulting in increased production of a particular metabolite. Research on reprogramming microbial metabolic pathways to tune microorganisms exquisitely for particular applications serves to increase our understanding of the factors responsible for metabolic regulation and improve productivity of compounds of economic value.

Several approaches for metabolic manipulation can be used to reprogram microbial metabolic pathways to produce valuable metabolites. Firstly, the metabolic pathway for a specific metabolite can be set up by using reference books and online databases. Secondly, the theoretical yield of the metabolite or the reaction fluxes in the cell can be determined by analyzing and modeling the metabolic pathway. Finally, the yield of the desired metabolite can be maximized by altering the reaction fluxes using metabolic engineering.

This volume brings together contributions from researchers in the forefront of microbial metabolic engineering. The volume is divided into two parts. The first

part focuses on the emerging tools and methods for redesigning microorganisms and aims to explore the basic rules governing the operation of metabolic networks in microbes. Such rules are important for metabolic engineering applied to industrial microorganisms. Since different microorganisms may have different metabolic pathways and networks, the second part of the volume provides various examples for designing improved microbes to produce valuable chemicals and materials.

This volume would provide important information and underlying principles on how to reprogram metabolic networks in industrial microbes to create desired valuable products on an industrial scale in a cost effective manner. Since genome sequences of more and more microorganisms become available, reprogramming microbial metabolic pathways using methods described in this volume will be tasks that can be accomplished on a much wider scale.

Wuxi — Xiaoyuan Wang
Wuxi — Jian Chen
London — Peter J. Quinn

Contents

Contributors

Roee Amit Department of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Andreas Aurich Helmholtz-Center for Environmental Research Leipzig – UFZ, Leipzig, Germany

Gerold Barth Institute of Microbiology, Dresden University of Technology, Dresden, Germany

Christopher J. Brigham Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

Ross P. Carlson Chemical and Biological Engineering Department, Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA

Jian Chen School of Biotechnology, Jiangnan University, Wuxi, Jiangsu, China

Corinne P. Cluis Department of Biology, Concordia University, Montréal, Québec, Canada

Tore Dehli Center for Systems Microbiology, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

Xunyan Dong Key Laboratory of Industrial Biotechnology of Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China

Guocheng Du School of Biotechnology, Jiangnan University, Wuxi, China

Joshua T. Ellis Department of Biological Engineering, Utah State University, Logan, UT, USA

Adam B. Fisher Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA, USA

Stephen S. Fong Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA, USA

Chikara Furusawa Quantitative Biology Center, RIKEN, Osaka, Japan

Athanassios Giannis Institute of Organic Chemistry, University of Leipzig, Leipzig, Germany

Guoqiang Han Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China

Philipp Heretsch Institute of Organic Chemistry, University of Leipzig, Leipzig, Germany

Silvia Herold Research Division Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

Takashi Hirasawa Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

Jiří Holátko Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Martina Holz Institute of Microbiology, Dresden University of Technology, Dresden, Germany

Xiaoqing Hu State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, China

Peter Ruhdal Jensen Center for Systems Microbiology, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

Jongpill Kim Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan

Christian P. Kubicek Research Division Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

Yong Lai School of Biological Sciences, The University of Hong Kong, Hong Kong, China

Paloma Liras Department of Molecular Biology, University of León, León, Spain

Juan F. Martín Department of Molecular Biology, University of León, León, Spain

Vincent J. Martin Department of Biology, Concordia University, Montréal, Québec, Canada

Hailin Meng Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Charles D. Miller Department of Biological Engineering, Utah State University, Logan, UT, USA

Roland A. Müller Helmholtz-Center for Environmental Research Leipzig – UFZ, Leipzig, Germany

Jan Nešvera Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Olusegun J. Oshota Chemical and Biological Engineering Department, Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA

Christina Otto Institute of Microbiology, Dresden University of Technology, Dresden, Germany

Miroslav Pátek Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Dominic Pinel Department of Biology, Concordia University, Montréal, Québec, Canada

Peter J. Quinn Division of Life Sciences, King's College London, London, UK

Bernhard Seiboth Research Division Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

Yue Shan School of Biological Sciences, The University of Hong Kong, Hong Kong, China

Hiroshi Shimizu Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

Tomokazu Shirai RIKEN Biomass Engineering Program, Yokohama, Kanagawa, Japan

Ekaterina Shishatskaya Institute of Biophysics, Siberian Federal University, Krasnoyarsk, Russian Federation

Dieter Sicker Institute of Organic Chemistry, University of Leipzig, Leipzig, Germany

Ronald C. Sims Department of Biological Engineering, Utah State University, Logan, UT, USA

Anthony J. Sinskey Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

Engineering Systems Division, Massachusetts Institute of Technology, Cambridge , MA, USA

Division of Health Sciences Technology, Massachusetts Institute of Technology, Cambridge, MA, USA

Christian Solem Center for Systems Microbiology, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

Robert Specht Helmholtz-Center for Environmental Research Leipzig – UFZ, Leipzig, Germany

Ulrich Stottmeister Helmholtz-Center for Environmental Research Leipzig – UFZ, Leipzig, Germany

Adrie J.J. Straathof Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

Reed L. Taffs Chemical and Biological Engineering Department, Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA

Franziska A. Thomas Institute of Organic Chemistry, University of Leipzig, Leipzig, Germany

R. Adam Thompson Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, USA

Cong T. Trinh Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN, USA

Walter M.van Gulik Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

Niti Vanee Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA, USA

Tatiana G. Volova Institute of Biophysics, Siberian Federal University, Krasnoyarsk, Russian Federation

Jianfeng Wang Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Xiaoyuan Wang State Key Laboratory of Food Science and Technology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxiz, China

Yiguang Wang Key Laboratory of Biotechnology of Antibiotics, Ministry of Health, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Yong Wang Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Zhou Wang Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China

Zhiqiang Xiong Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Aixin Yan School of Biological Sciences, The University of Hong Kong, Hong Kong, China

Venelina Yovkova Institute of Microbiology, Dresden University of Technology, Dresden, Germany

Natalia Zhila Institute of Biophysics, Siberian Federal University, Krasnoyarsk, Russian Federation

Jingwen Zhou School of Biotechnology, Jiangnan University, Wuxi, China

Part I
Emerging Methods for Redesigning Microorganisms

Chapter 1
Towards Synthetic Gene Circuits with Enhancers: Biology's Multi-input Integrators

Roee Amit

Abstract One of the greatest challenges facing synthetic biology is to develop a technology that allows gene regulatory circuits in microbes to integrate multiple inputs or stimuli using a small DNA sequence "foot-print", and which will generate precise and reproducible outcomes. Achieving this goal is hindered by the routine utilization of the commonplace σ^{70} promoters in gene-regulatory circuits. These promoters typically are not capable of integrating binding of more than two or three transcription factors in natural examples, which has limited the field to developing integrated circuits made of two-input biological "logic" gates. In natural examples the regulatory elements, which integrate multiple inputs are called enhancers. These regulatory elements are ubiquitous in all organisms in the tree of life, and interestingly metazoan and bacterial enhancers are significantly more similar in terms of both Transcription Factor binding site arrangement and biological function than previously thought. These similarities imply that there may be underlying enhancer design principles or grammar rules by which one can engineer novel gene regulatory circuits. However, at present our current understanding of enhancer structure-function relationship in all organisms is limited, thus preventing us from using these objects routinely in synthetic biology application. In order to alleviate this problem, in this book chapter, I will review our current view of bacterial enhancers, allowing us to first highlight the potential of enhancers to be a game-changing tool in synthetic biology application, and subsequently to draw a road-map for developing the necessary quantitative understanding to reach this goal.

Keywords Bacterial enhancers • Biological computation • Digital computation • Gene regulatory networks • Synthetic enhancer

R. Amit (✉)
Department of Biotechnology and Food Engineering,
Technion – Israel Institute of Technology, Haifa 32000, Israel
e-mail: roeeamit@technion.ac.il

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_1,

Abbreviations

NRI	nitrogen regulation I
RNAP	RNA polymerase
TF	transcription factors
TSS	transcriptional start site

1.1 Introduction

The ability of living organisms from all branches of the tree of life to convert a complex palette of variable input signals into discrete output levels that in turn trigger cell differentiation, morphogenesis, stress responses, complex metabolic reactions, or a host of other cellular phenomenon is one of the great mysteries of modern biology. One way to achieve this involves an amalgam of gene network interactions with complex regulatory regions. While the gene regulatory subnetwork structure (Ackers et al. 1982; Bintu et al. 2005a, b; Bolouri and Davidson 2002; Buchler et al. 2003; Rosenfeld et al. 2005; Stathopoulos and Levine 2005) and the input/output relationship between different genes is becoming better defined for many systems, the structure, binding-site arrangement, and the underlying mechanisms responsible for the regulatory output remains for the most part undeciphered. Consequently, our ability to engineer novel gene-regulatory circuits is severely hindered by this knowledge gap, where the process of integration of multiple regulatory inputs remains poorly understood.

1.1.1 Gene Regulation in Bacteria

In bacteria, the prevailing view of transcriptional regulation is built around the idea of regulated recruitment of RNA polymerase and the dissociable sigma factor σ^{70}. In this picture, the presence or absence of RNA polymerase at a promoter of interest is dictated by the corresponding presence or absence of batteries of transcription factors that either increase (activators) or decrease (repressors) the probability of polymerase binding. An increasingly sophisticated understanding of this kind of regulatory response has resulted in an explosion of efforts in synthetic and systems biology research efforts built using a broad palette of different activators and repressors for a range of different promoters (Belyaeva et al. 1998; Bintu et al. 2005a; Elowitz and Leibler 2000; Gardner et al. 2000; Guido et al. 2006; Joung et al. 1993, 1994; Muller et al. 1996).

Another whole set of bacterial promoters utilize an alternative sigma factor (σ^{54}) which together with RNAP form a stable closed promoter complex that, unlike its σ^{70} counterpart, is unable to initiate transcription by itself (Amit et al. 2011; Buck

et al. 2000; Bulger and Groudine 2011; Ninfa et al. 1987; Rappas et al. 2007). This effectively causes the polymerase to be poised at the gene of interest awaiting the arrival of a transcription factor partner termed the "driver", which releases the polymerase. Consequently, these promoters are regulated in a different fashion than their recruitment counterparts. The activating or transcription driving complex is typically widely separated from the promoter (100–1,000 bp (Ninfa et al. 1987)), precluding it from forming direct contact with the poised polymerase. It has been shown that DNA looping (Amit et al. 2011; Huo et al. 2006; Schulz et al. 2000; Su et al. 1990) and ATP hydrolysis (Rappas et al. 2007) are required to induce open complex formation and transcription initiation. These regulatory regions belong to a different class of regulatory elements called bacterial enhancers, whose structure and function are similar to their eukaryotic counterparts.

1.1.2 Enhancers – General Structure and Mode of Action

Enhancer elements or cis regulatory modules are ubiquitous in all genomes (Buck et al. 2000; Bulger and Groudine 2011; Rappas et al. 2007). It is hypothesized that enhancers execute their regulatory program by making direct contact with the basal promoter via DNA or chromatin looping. In general, they are made up of contiguous genomic regions that stretch from tens to thousands of base-pairs, contain several binding sites for a variety of transcription factors (TF), and often their regulatory output is independent of their location or orientation relative to the basal promoter (Amit et al. 2011; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Huo et al. 2006; Ninfa et al. 1987). Furthermore, enhancers, like gene-regulatory networks themselves, can be viewed qualitatively (Amit et al. 2011) as modular entities, which in this case are made of three connected irreducible parts: the driver binding sites responsible for initiation of transcription, transcription factor binding sites responsible for the modulation of expression levels, and a basal promoter. In these systems, a basal promoter has the capability to generate little or no transcriptional output on its own (Aida et al. 2006; Boehm et al. 2003; Bulger and Groudine 2011; Gilmour 2009; Magasanik 1993; Muse et al. 2007; Ninfa et al. 1987; Rappas et al. 2007; Zeitlinger et al. 2007) but together with the rest of the enhancer it can express its full regulatory potential (Atkinson et al. 2002, 2003; Davidson 2001, 2006; Lee and Schleif 1989; Magasanik 1993; Small et al. 1992; Yuh et al. 2001). Even though many aspects associated with enhancer regulation are routinely studied in natural systems with state-of-the-art techniques in both bacteria (Amit et al. 2011; Atkinson et al. 2002, 2003; Huo et al. 2006; Ninfa et al. 1987) and higher eukaria (Bolouri and Davidson 2002; Davidson 2006; Stathopoulos and Levine 2005), the underlying mechanisms of regulatory "action-at-a-distance" responsible for integrating the various inputs in enhancers remain elusive. In order to fulfill the potential promised within synthetic gene regulatory circuits, we must rapidly close our knowledge gap between the relatively advanced understanding of dynamic phenomenon associated with gene subnetwork motifs with our meager

grasp of the underlying biophysical mechanisms that are responsible for producing the enhancer regulatory output. It is the purpose of this book chapter to outline the potential benefits of utilizing enhancers routinely in synthetic biology applications, and to draw a road-map that will guide the development of the necessary knowledge base to facilitate this capability.

1.2 Structure and Function of Bacterial Enhancers

1.2.1 Enhancer Architecture and Transcriptional Kinetics

Bacterial enhancers are highly modular objects, whose binding site architecture can be grossly divided into three distinct modules (Fig. 1.1a). The driver module is typically associated with either a tandem of or three specialized binding sites that are located between 50 and 500 bp upstream of the basal promoter. The driver binding sites facilitate the cooperative assembly of a hexameric ATPase (e.g. NRI/NtrC, PspF, etc.) belonging to the AAA+family. These ATPases exist in the cytoplasm as dimers, each capable of individually binding one of the binding sites in the driver module. The assembly of the hexameric complex, apparently occurs as a result of the binding of two dimers, to which a third cooperatively binds to complete the assembly. The cooperative nature of this binding ensures that the hexameric complex is highly stable, reminiscent of other AAA+DNA bound hexamers (e.g. RuvB Amit et al. 2004) that also have an increased binding affinity as an assembled complex vs. the cytoplasmic dimers.

The second module encompasses the region in between both the promoter and driver binding sites. This region typically contains a multitude of binding sites for several (1–5) transcription factors, and its main role is to modulate the expression level that would be generated if no proteins were bound. This modulation was demonstrated recently on one natural system (Atkinson et al. 2002), and with two synthetic enhancer systems (Amit 2012; Amit et al. 2011; Huo et al. 2006), showing that the expression level can either be inhibited (repressed) or amplified (activated) depending on the type of protein that binds, the number of binding sites, the location of the binding sites with respect to the promoter within this region, and the spacing between the binding sites.

Finally, the third module is the basal promoter, which in this case binds the σ^{54}-RNAP holoenzyme complex. This module is responsible for integrating all the inputs thereby generating a particular expression pattern at some integrated rate. The integration of the inputs takes place via a sequential kinetic mechanism, whereby the σ^{54}-RNAP holoenzyme complex binds the promoter, but is unable to initiate expression, and as a result remains paused at the Transcriptional Start Site (TSS). Simultaneously, the rest of the transcription apparatus assembles at the various binding sites on the enhancer. Transcription is facilitated when the upstream assembled driver complex (e.g. NtrC – the "driver" (Amit et al. 2011)) loops and makes directs contact with the poised σ^{54}-RNAP complex. The driver

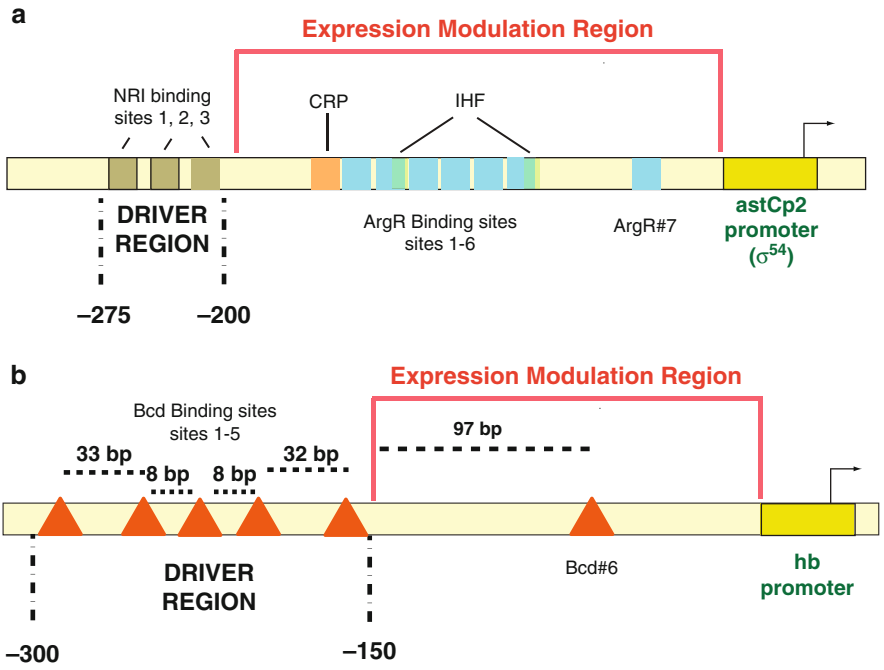


Fig. 1.1 Bacteria and Drosophila enhancers. The figure is a schematic designed to pictorially convey the similarities between a sample bacterial enhancer and a near-promoter Eukaryotic enhancer. (**a**) The astCp2 enhancer in *E. coli*, exhibiting a ~200 bp expression modulation region, at least ten binding sites for three different kinds of TFs, and three NRI~P driver binding sites (Kiupakis and Reitzer 2002). (**b**) The hb (hunchback) gene enhancer in D. melanogaster showing a very similar architecture (Driever and Nusslein-Volhard 1989; Driever et al. 1989) to bacterial enhancers in terms of binding sites, proximity to promoter, and binding site separation. Note, that in this case Bcd is also the driver in this system. This enhancer and others with similar binding-site architectures can serve as model systems for an initial Eukaryotic Rosetta stone algorithm

has a special amino acid loop termed GAFTGA (to signify the amino-acid content (Rappas et al. 2007; Zhang et al. 2002)), which effectively enters a specialized slot within the poised holoenzyme complex, similar to a "key in a hole" mechanism. Upon binding, subsequent ATP hydrolysis by the driver generates a conformational change within the holoenzyme complex, which in turn alleviates the poised state, allowing transcription to progress. The expression modulation region affects the looping rate by inducing certain structural effects that either increases the probability for a successful looping event or decreases it. Since the rates associated with ATP hydrolysis and subsequent conformational changes are fast, the rate of looping becomes the determining rate-limiting factor for transcription. Consequently, the bacterial enhancer architecture allows the promoter to modulate the expression rate based on the transcription factor content that is found upstream, thereby allowing it to function as a form of biological integrator (Amit 2012; Amit et al. 2011).

1.2.2 Biological Function

Most bacteria contain some version of the σ^{54} sigma factor. A few well-known examples include the nitrogen regulation protein C (NRI or NtrC), the nitrogen fixation protein A (NifA), the C4-dicarboxylic acidic transport protein D (DctD), the phage shock protein F (PspF), the xylene catabolism regulatory protein (XylR) and the 3,4-dimethylphenol catabolism regulatory protein (DmpR) (Xu and Hoover 2001 and references therein). A close examination of all of these examples indicates that the σ^{54} regulated genes are often activated in response to various stresses and growth inhibiting conditions (Buck et al. 2000). In such cases, bacterial cell responds to the stress by turning on a dormant metabolic pathway in order for it to cope successfully with the stress. Such a massive shifting of transcriptional resources is in many ways akin to a bacterial form of cell differentiation into a specialized cell-type designed to cope with the stressful environment.

In addition, σ^{54} promoters (Rappas et al. 2007; Xu and Hoover 2001; Zhang et al. 2002) are also over-represented in genes that play an important role in bacterial developmental-like processes. This includes the two-component nitrogen response pathway and related systems, which exhibits regulatory and signaling characteristics that are also reminiscent of a primitive developmental like process (Goldman et al. 2006; Magasanik 1993; Ninfa and Peng 2005). A more telling example is the involvement of σ^{54} promoters in the formation of *M. Xanthus* fruiting bodies (Goldman et al. 2006). In particular, a recent genomic analysis carried out on *M. Xanthus* genomes as compared with other bacterial genomes (Goldman et al. 2006; Jelsbak et al. 2005) revealed that the comparative number of σ^{54} promoters as a function of genome size is much larger as compared with other bacterial species, indicating that these promoters are likely associated with specialized biological functions in *M. Xanthus* fruiting body development.

Interestingly, in Eukaryotes, promoters that go through promoter proximal pausing gene activation were also found to be over-represented in developmentally-important or cell-differentiation type of processes. Recently, (Aida et al. 2006; Guenther et al. 2007; Muse et al. 2007; Rasmussen and Lis 1993; Zeitlinger et al. 2007; Zhang et al. 2007) have showed that in metazoan organisms ranging from humans to flies, "paused" genes, known to be off at particular developmental stages, tissues, or based on ambient environmental conditions were found to be occupied by an active PolII transcriptional complex (with nascent transcript) localized 20–60 nt from the Transcriptional Start Site (TSS). Release of a paused polymerase from its stalled state requires a secondary event of looping from an upstream region, which allows a specialized protein called pTEF-b (Cheng and Price 2007; Renner et al. 2001), to phosphorylate several sites on the paused PolII (Boehm et al. 2003), which in turn allows transcription to progress.

Consequently, a form of enhancer-regulated paused or poised transcription is ubiquitous to all biological kingdoms, and seems to be over-represented in genes that are known to play an important role in executing some sort of a "developmental-like" or "cell-differentiation" type program. Such programs also seem to be characterized by both precision and often synchronized initiation of transcription along a cluster of cells.

Therefore, it is tempting to speculate, that there may be some characteristic inherent to enhancer regulatory structure (Fig. 1.1b) as well as with the activation of sequential kinetics, which endows the enhancer regulatory response in all organisms with precision, discrete, and possible synchronized behavior.

1.3 Engineering Gene Circuits with Synthetic Enhancers

The main premise of synthetic biology is to use "biological parts" to construct novel biological composite objects for a variety of applications from basic research to personal medicine. In order to achieve this goal and to develop this technology, a more "engineering" friendly scheme had to be adopted to describe biological function. Given the rapid development and penetration of Information Technology via the "digital computer" in recent decades, it became quaint to compare a gene being turned on/off to a process of switching a bit from 0 to 1, which as a result led to the adoption of a computational language and Boolean algebra as a generalizing mechanism (Andrianantoandro et al. 2006). This scheme was particularly commensurate with the biological function of the more commonplace bacterial σ^{70} promoters, which generate transfer functions reminiscent of sigmoidal functions that characterize digital on/off switches. Consequently, this property has made them highly attractive for designing simple computational modules from elementary biological parts (i.e. genes, promoters, etc.), and amongst other reasons has led to almost exclusive utilization of these objects in the first generation of synthetic biology works.

Unlike σ^{70} bacterial promoters, the coupled enhancer-σ^{54} promoter systems have been completely ignored by the community in the early days of synthetic biology except for one notable exception (Atkinson et al. 2003). In addition to this work, bacterial enhancers so far have been used in three additional synthetic biology works (Amit et al. 2011; Huo et al. 2006) with no real applications as of yet. In this section, I will explain in computational terms why this underutilization of enhancers is expected to change as we move into developing the next phase of synthetic biology applications.

1.3.1 Biological Computation at the Gene Regulatory Level

One of the major challenges of synthetic biology is to engineer compact, yet complex gene regulatory networks capable of carrying out complex computational operations in a precise fashion. In this case "computation" means a type of calculation process that follows a well-defined model expressed as an algorithm, protocol, network topology, or any other set of predefined rules. From a biological perspective such processes may involve sensing and processing a whole palette of chemical input signals, deciding on where and when a particular gene should be expressed, dividing into particular cell types, regulatory responses, etc. Thus far, the major workhorse

used to demonstrate novel synthetic biological computational processes have been synthetic gene regulatory circuits implemented using standard bacterial σ^{70} promoters.

Promoters that belong to this family are often regulated (i.e. turned on and off) by transcription factors whose binding sites are either in the vicinity or over-lapping the RNAP binding region. Due to the transcription factors' binding sites proximity to their cognate promoters, these proteins regulate gene expression by either preventing RNAP from binding, or by recruiting RNAP and increasing the probability for transcription. As a result, the transfer functions that depict how these promoters activate gene expression as a function of intracellular transcription factor concentrations are highly reminiscent of sigmoidal switching behavior, which has been compared to a form of binary digital computation and attributed properties of buffer gates (Andrianantoandro et al. 2006; Gardner et al. 2000). This characteristic of gene expression has been one of the primary drivers for the engineering of gene-regulatory circuits that function as "noisy" biological binary-logic gates (Andrianantoandro et al. 2006 and references therein), and subsequent construction of simple circuits made of connected biological digital gates (Anderson et al. 2007). Since the gene products (proteins or RNA molecules) of these biological logic gates can be utilized with minimal effort to either feed-back on their own promoter, or participate in further down-stream regulation, such efforts have led to a plethora of implementations of composite biological circuits made of several interconnected biological gates that have been shown to be capable of executing simple computational operations akin to simple electronic circuits (Basu et al. 2005; Friedland et al. 2009; Tabor et al. 2009; Tasmir et al. 2011).

Despite the rapid progress achieved over the last 10 years with increasingly complex circuits capable of carrying out sophisticated computational algorithms, σ^{70} recruitment promoters are not capable of generating transfer functions that are sufficiently close to the digital ideal. First, the process of induction, which generates the transition between "gene-off" to "gene-on" states is typically spread over a wide-range of inducer or transcription factor concentrations. This, in turn, yields an extended range where a gradiated response is observed, which is characteristic of analog computational processes. Consequently, the sharp switching that characterizes electronic binary digital gates cannot be simply engineered with the biological versions.

Moreover, in order to execute complex computational algorithms using biological binary digital computation that often relies on "wiring" together a whole set of two-input digital gates (e.g. AND, OR, etc.) to carry out simple Boolean computations, many regulatory components are needed. Since σ^{70} promoters necessitate that the TF binding sites be present within a close proximity, individual promoters can integrate only one or two signals. This, in turn, means that in order to program cells to carry out complex computational processes, large gene regulatory network circuits with many nodes need to be designed, which translates to generating a need for engineering very large sequences, whose growth potential is limited by the biological vessel that will execute the computation. Since bacterial cells are capable of encasing ~1–10 Mbp of DNA, this suggests that very quickly a glass ceiling of computational complexity will be reached using the binary paradigm.

Finally, unlike electronic computers, which are not subject to thermal noise, biological computation is subject to large thermal noise effects. This in turns renders any biological computation operation a stochastic process, which is by definition subject to different modeling rules, than the deterministic processes that characterize conventional computational processes. In particular, the recruitment transcriptional process is particularly susceptible to molecular noise (Elowitz and Leibler 2000; Elowitz et al. 2002; Thattai and van Oudenaarden 2001), which makes this problem even more of an acute issue for these systems. Consequently, at present, any computational processes that are carried out by biological modules are not only limited in computational capability, but are also imprecise and noisy. Yet, despite the physical, energetic, and thermal limitations, natural biological computation is capable of executing tremendously complex and precise computational operations at the gene regulatory level. So the question remains how do we overcome these limitations and develop a technology that can carry out precise and reproducible molecular computation operations?

1.3.2 Biological Computation with Enhancers

Unlike σ^{70} promoters, σ^{54} are always coupled to bacterial enhancers, and in effect can be considered to be one large regulatory unit. This unit includes a multitude of binding sites for many transcription factors, which in turn can support the integration of many different input signals. Thus, enhancers provide a convenient platform for engineering Boolean digital gates with multiple inputs (n), which allows 2^2^n computational operations to be carried out at a single promoter as compared with approximately 16 for a σ^{70} promoter (e.g. an enhancer capable of integrating three or four input will support 256 and 65,536 different computational operations respectively.) As an example for the utility and compactness of enhancer-based computation as compared with σ^{70} recruitment promoter, consider constructing a three or four input AND gate. With the latter system this will require the utilization of at least two different chemically wired two-input gate promoters, while with enhancers these operations can be carried out at a single promoter.

Another advantage of enhancers is the capability to engineer interactions between transcription factors that are bound adjacently to one another. In the case of cooperative interactions between transcription factors, enhancer output will be characterized by transfer functions (Fig. 1.2a) whose transition region occupies smaller TF concentration ranges that are much closer to the digital ideal. Alternatively, anti-cooperative or mutually exclusive interactions between bound transcription factors on the enhancer can generate transfer functions with more than two "stable" states (Fig. 1.2b, c). Having more than two stable output states supports a non-Boolean digital computation model, where instead of a 0 or 1 output, the enhancer can generate a 0,1,2 or more output. Digital computation with more than two discrete input/output states offers a much larger computational flexibility (Table 1.1), as the number of possible algebraic operations with a 2 or 3-input gates increases exponentially

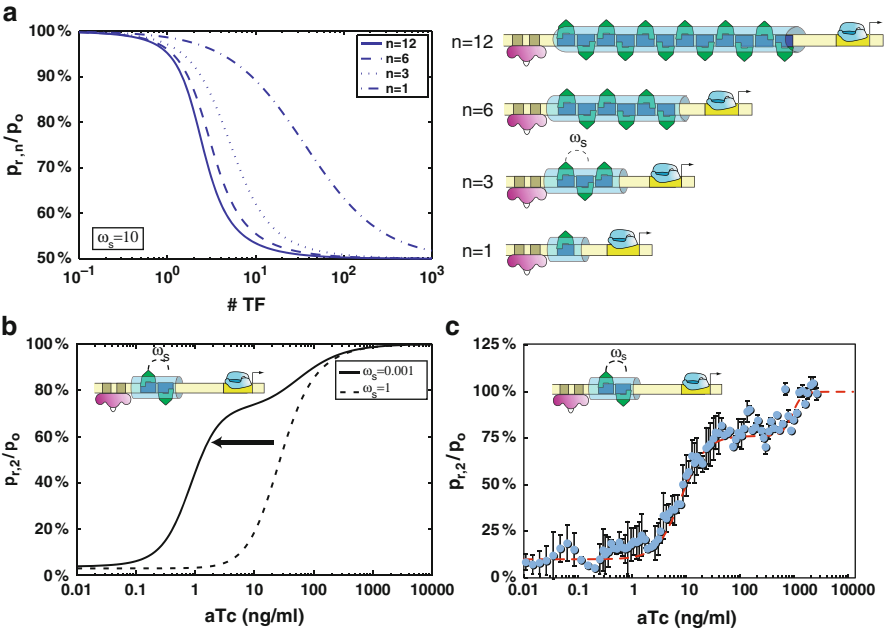


Fig. 1.2 Enhancer Transfer Function. (**a**) Transfer functions for bacterial enhancers characterized by two input signals: driver bound to a tandem of binding sites upstream of the poised RNAP, and a variable number of binding sites (1,3,6,12) in the expression modulation region for some Transcription Factor (TF). The TF is assumed to rigidify the DNA when bound to the DNA leading to repression (see SI of (Amit et al. 2011) for definition of repression and the values on the y-axis), and to bind the enhancer cooperatively (quantified by a protein-interaction parameter $\omega_s > 1$). The model shows that given these assumptions, it is possible to generate sharper transfer functions by simply increasing the number of binding sites. (**b**) Alternatively, one can generate a step-like response in the model using the exact same binding architecture by setting the protein-interaction parameter to some value $\omega_s \ll 1$. Therefore, a wide-array of possible transfer functions may be possible depending on a handful of characteristics such as type of protein bound, number of binding sites, spacing between binding sites, etc. (**c**) Data published previously (Amit et al. 2011) showing that by varying the number of active TetR proteins inside cells via the inducer anhydrous-tetracycline (aTc) and using an enhancer structure containing two TetR binding sites with 16 bp spacing between sites, a step-like response is generated. For further detail of model and experimental data see (Amit 2012; Amit et al. 2011)

with the number of possible output states. Consequently, the enhancer's capability to both integrate multiple inputs and to generate multiple stable state transfer functions endows them with a tremendous computational flexibility and complexity, which can only be produced by σ^{70} based gene circuits that are composed of multiple promoters and require a significantly larger sequence signature.

Finally, unlike electronic computers, where thermal noise plays a minor role, in biological systems thermal noise plays a critical role in regulation. The sources of noise of have been enumerated and quantified in several recent publications (Elowitz et al. 2002; Friedman et al. 2006; Golding et al. 2005; Ozbudak et al. 2002; Paulsson 2004;

Table 1.1 Computational complexity of different digital computational systems

# inputs (n)	Boolean-logic[a,b]	Combined-logic[c,d]	3-logic[e]
1	4	8	9
2	16	512	3^9
3	256	2^{27}	3^{27}
N	2^2^n	2^3^n	3^3^n

[a] The number of possible operations with conventional "Boolean-logic" defined as 2^2^n operations

[b] The left "2" in 2^2^n corresponds to the number of values each input channel can accept, the middle "2" to the number of output values that are possible, and n is the number of input channels

[c] The number of possible operations with "Combined-logic" defined as 2^3^n operations

[d] 2^3^n operations can be executed by a gate that can accept only two values (i.e. protein bound/unbound), but can output three discrete values (e.g. enhancer)

[e] The number of possible operations with "3-logic" defined as 3^3^n operations. "3-logic" is the "three" version of Boolean logic, which can also be implemented with enhancers

Pedraza and Paulsson 2008; Raser and O'Shea 2004; Sanchez et al. 2011; Thattai and van Oudenaarden 2001), and are mostly due to a small finite number of interacting objects, the kinetics of binding and unbinding, and variation of different molecular species from cell to cell (i.e. some cells may have more RNAP molecules available than others, etc.). A major challenge of synthetic biology is to not only construct synthetic circuits capable of carrying out complex computation, but to do so with minimal noise effects. Since noise is an additive quantity (Pedraza and Paulsson 2008; Sanchez et al. 2011), circuits with multiple promoters and components are inherently susceptible to thermal noise, and as such must constantly add elements that simultaneously mitigate the deleterious noise affects (Andrianantoandro et al. 2006). Enhancers, on the other hand, which presumably can carry out complex calculations at a single promoter, double almost by default as a noise-minimizing mechanism due to the compactness of the molecular design. Consequently, enhancers have the potential to not only allow us to code complex biological algorithm, but to do so in a noise-minimal fashion as well.

1.3.3 Putting It Altogether – Constructing Circuits

Despite their potential, to date only three synthetic biology works have utilized enhancers to generate novel regulatory effects. These works have either altered bacterial enhancers to generate novel regulatory effects (Amit et al. 2011; Huo et al. 2006) or wired two enhancers together to generate a damped oscillator characterized by a periodicity that was an order of magnitude or so larger than the standard time-scale associated with a bacterial cell-cycle (Atkinson et al. 2003).

In the former works, Huo et al. (2006)) showed that by careful positioning of a binding site for IHF at different locations along the enhancer, the regulatory effects

can either be sharply repressive or highly activating, with a periodicity that is commensurate with the DNA helical pitch. In a recent work, we carried out a systematic analysis of many synthetic enhancers, which showed that altering the enhancer's ability to loop using bound transcription factors affects regulation. We (Amit et al. 2011) were able to show transfer functions (Fig. 1.2c) that are characterized by multiple output levels with sharp transitions between states that pointed to a combined cooperative and anti-cooperative effect (Amit 2012) in the binding of TetR proteins. The next stage will be to utilize these libraries of characterized synthetic enhancers to engineer gene-circuits capable of carrying out complex computation in a noise minimizing and compact genomic architecture. My lab is advancing towards this goal with our current research.

Based on these early achievements, it seems that utilizing synthetic enhancers to construct synthetic gene circuits promises to generate some very interesting applications in the very near future. Complex circuits that can induce bacterial cell differentiation in response to stimuli, convert continuous input signals into some discrete output, and function as intra-cellular detectors are all possible. While it may be possible to develop such applications using the standard gene regulatory network coupled to the recruitment promoter tool kits, it will likely take up a larger space of sequence, and be composed of many more components. Finally, one can imagine constructing complex multi-enhancer synthetic circuits, adding another level of complexity, which can push us closer to the dream of building biological integrated circuits. Therefore, coupling a library of synthetic enhancers with characterized transfer functions to known circuit architectures can lead to a great advance in biological circuit capabilities.

1.4 Synthetic Enhancers as a Basic Research Tool for a Biological Rosetta Stone Algorithm

In order to reach this goal and to be able to engineer routinely gene circuits with synthetic or natural bacterial enhancers as regulatory code, we must first decipher the regulatory code encoded within natural enhancers so that they can provide a credible starting point. However, due to their modular architecture and large binding site content, enhancers are notoriously difficult to dissect, often requiring large and labor-intensive collaborative efforts. To understand the scope of the problem, consider the following example in eukaryotes: the regulatory region (Davidson 2006) of the gene *endo16* in the sea urchin *S. purpuratus.* This is a "run-of-the-mill" gene that participates in the endo-mesoderm formation in early sea-urchin development. It has a regulatory region that spans ~2.3 kbp, with purportedly seven cis-regulatory modules that play a role in defining the time and place of *endo16* expression. Of those modules only two modules have been quantitatively characterized using a "knock down and rescue" type of approach, which necessitated many years' worth of man-work. While the *endo16* analysis and similar works (Atkinson et al. 2002; Davidson 2006; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Small

et al. 1992) have led to provocative data that spawned a vibrant research field, the labor-intensive nature of the research has generated slow progress, which resulted in only a handful of enhancers (bacterial or Eukaryotic) that have been quantitatively characterized to this day. Consequently, one of the greatest challenges facing modern day biological research is to develop a high-throughput methodology for the decipherment of the regulatory programs encoded within enhancers.

Interestingly, as a result of the handful of examples dissected thus far (Atkinson et al. 2002; Davidson 2006; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Ninfa et al. 1987; Small et al. 1992; Yuh et al. 2001) an interesting pathway for a more rapid decipherment of the regulatory programs encoded within enhancers may have emerged. A close examination of the data indicates that there may be a regulatory code characterized by "grammar" (Datta and Small 2011) or design rules encoded into both metazoan and bacterial enhancers. These grammar rules, once deciphered, can in principle allow us to predict the regulatory output of an enhancer based on sequence information alone. If there is a regulatory code encoded into enhancer sequences, what is the best strategy to go about developing a decoding algorithm? One possible method is to work in "reverse": namely, try encoding "words" or "sentences" and testing the decoding algorithm to see if its output recovered the original information. For a biological application, this approach implies developing a synthetic biology strategy for the decipherment of the regulatory output encoded into enhancers. In effect, to engineer using synthetic biology a Biological Rosetta stone algorithm for the regulatory code.

Unlike the archaeological Rosetta Stone, which contained panels of identical messages written in three different scripts and two languages, the biological Rosetta Stone is still missing two of the three panels (Fig. 1.3). In order to develop a draft for this algorithm, we need to construct the two additional panels to complement the sequence panel that we want to decipher. One possible way to do this is by engineering a "collection" or library of simple synthetic enhancers from the ground up, which will be coupled to a high-throughput analysis platform. Results from this analysis can then be used as "training" tool for candidate Rosetta Stone algorithms. Unlike the traditional approach of "knock-out and rescue", building enhancers from the ground up allows one to systematically increase the complexity of the enhancer and enhancer circuit designs in a controlled fashion, which, in turn, provides the opportunity to reconstruct regulatory behavior revealed by quantitative analysis of natural enhancers in an insulated fashion. Therefore, the synthetic approach allows one to dissect quantitatively a multitude of enhancers in substantially less time and manpower.

While it may be difficult to develop such a strategy in metazoans, bacteria are perfectly suited for an initial development of this approach. Recently, we took the first step towards this goal (Amit 2012; Amit et al. 2011) by constructing a "rough sketch" of a bacterial Rosetta Stone algorithm (Fig. 1.3), which in turn had enabled us to formulate qualitative predictions for the expression level outputs of heretofore unexplored bacterial enhancers based purely on sequence analysis. If this strategy proves to be successful in bacteria, a similar strategy for Eukaryotic enhancers can be developed as the next step, while simultaneously

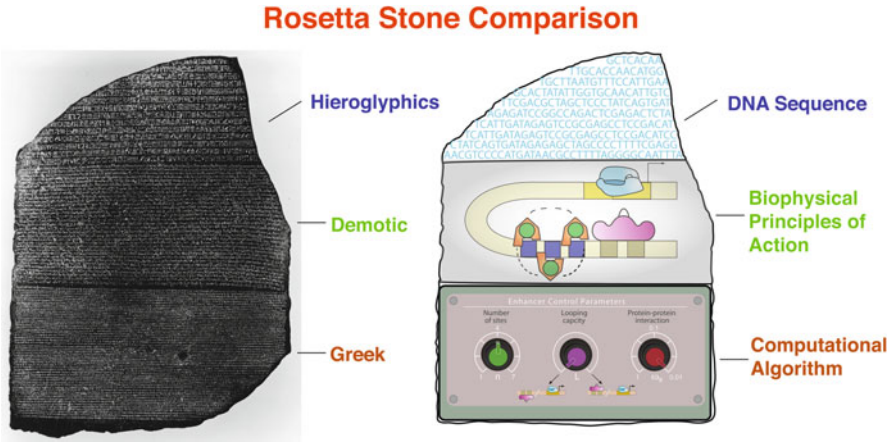


Fig. 1.3 The Biological Rosetta Stone. The Rosetta stone is an archeological artifact that contains three identically written segments in three different scripts and two languages. The Greek (*bottom*) and Demotic (*middle*) segments allowed researchers to interpret the Hieroglyphics script (*top*), which in turn provided archaeologists with a decoding "algorithm" that allowed them to read many previously undecipherable ancient Egyptian texts. The Biological Rosetta Stone strategy's is based on producing the regulatory code equivalent of the real Rosetta stone, where the *top panel* in this case is the DNA sequence. The *middle panel* is biophysical principles or "machine-code" deciphered via the synthetic enhancer experiments. The *bottom panel* is the computational algorithm executed by the enhancer, which is encoded within the sequence depicted by the *top panel*. Such a tool can then be used as a decoding algorithm to predict regulatory output from the sequence of naturally occurring enhancers

allowing us to progress in implementing this technology in bacterial applications. Since current methodologies for the decipherment of the regulatory code are dependent on the arduous and labor-intensive "knock-down and rescue" approaches, I am certain that complementing the standard dissection or reductionist approach with this synthetic methodology will substantially accelerate our ability to decipher the regulatory code, and as such will impact this field to a large extent. Consequently, the ability to construct synthetic enhancer gene regulated circuits in microbial organisms has the potential to not only spawn a quantum leap for a whole host of synthetic biology applications in therapeutics, environmental challenges, biofuel production, etc., but also to bring us a step closer to deciphering the significantly more complex Eukaryotic regulatory code. Successful implementation of such modalities will bring the field closer to fulfilling its great technological potential that had so far proven to be somewhat elusive.

1.5 Conclusions

Enhancers are a class of ubiquitous regulatory objects that potentially can alter the way by which we construct gene circuits. They are capable of executing complex molecular computational operations via a promiscuous architecture capable of integrating multiple

binding sites for several transcription factors. This compact architecture can potentially be used to engineer gene circuits capable of executing complex computational operations that are currently untenable with standard approaches. This ability to integrate multiple signals can be used a fine tuned spatio-temporal control of gene expression, a capability which may be crucial for most future synthetic biology applications from biofuels to smart drug designs.

Even though enhancers are more commonly associated with Eukaryotic regulation, their prevalence and utilization in similar biological function in bacteria points to a largely untapped potential in utilization for synthetic biology applications. However, at present the difficulty in using these components is not rooted in our ability to produce large sequences of DNA, but rather in our ignorance as to the basic operating principles that underlie many of the regulatory effects that are generated by these modules. Hence, before progressing to actively constructing circuits from these objects, we must first develop a better understanding of the basic design rules that guide enhancer regulatory function. One such approach is to develop a biological Rosetta stone via synthetic enhancers to try to distill the rules, which can then be applied on natural bacterial enhancers to test the level of our new understanding. Once we have this tool in place, the engineering of gene circuits with enhancers as basic regulatory modules can finally tap the nearly unlimited computational potential provided by these modules.

References

Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A 79:1129–1133

Aida M, Chen Y, Nakajima K, Yamaguchi Y, Wada T, Handa H (2006) Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. Mol Cell Biol 24:2444–2454

Amit R (2012) TetR anti-cooperative and cooperative interactions in synthetic enhancers. J Comp Biol 19:115–125

Amit R, Gileadi O, Stavans J (2004) Direct observation of RuvAB-catalyzed branch migration of single Holliday junctions. Proc Natl Acad Sci U S A 101:11605–11610

Amit R, Garcia HG, Phillips R, Fraser SE (2011) Building enhancers from the ground up: a synthetic biology approach. Cell 146:105–118

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3:133

Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2:1–14

Atkinson MR, Pattaramanon N, Ninfa AJ (2002) Governor of the glnAp2 promoter of *Escherichia coli*. Mol Microbiol 46:1247–1257

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. Cell 113:597–607

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434:1130–1134

Belyaeva TA, Rhodius VA, Webster CL, Busby SJ (1998) Transcription activation at promoters carrying tandem DNA sites for the *Escherichia coli* cyclic AMP receptor protein: organisation of the RNA polymerase alpha subunits. J Mol Biol 277:789–804

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R (2005a) Transcriptional regulation by the numbers: applications. Curr Opin Genet Dev 15:125–135
Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R (2005b) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15:116–124
Boehm AK, Saunders A, Werner J, Lis JT (2003) Transcription factor and polymerase recruitment, modification, movement on dhsp70 in vivo in the minutes following heat shock. Mol Cell Biol 23:7628–7637
Bolouri H, Davidson EH (2002) Modeling DNA sequence-based cis-regulatory gene networks. Dev Biol 246:2–13
Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100:5136–5141
Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD (2000) The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J Bacteriol 182:4129–4136
Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327–339
Cheng B, Price DH (2007) Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem 282:21901–21912
Datta RR, Small S (2011) Gene regulation: piecing together the puzzle of enhancer evolution. Curr Biol 21:R542–R543
Davidson EH (2001) Genomic regulatory systems: development and evolution. Academic, New York
Davidson EH (2006) The regulatory genome. Elsevier, Burlington
Driever W, Nusslein-Volhard C (1989) The bicoid protein is a positive regulator of hunchback transcription in the early drosophila embryo. Nature 337:138–143
Driever W, Thoma G, Nusslein-Volhard C (1989) Determination of spatial domains of zygotic gene expression in the drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature 340:363–367
Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338
Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297:1183–1186
Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ (2009) Synthetic gene networks that count. Science 324:1199–1202
Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett 97:168302
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342
Gilmour DS (2009) Promoter proximal pausing on genes in metazoans. Chromosoma 118:1–10
Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123:1025–1036
Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB (2006) Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci USA 103:15200–15205
Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130:77–88
Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) A bottom-up approach to gene regulation. Nature 439:856–860
Huo YX, Tian ZX, Rappas M, Wen J, Chen YC, You CH, Zhang X, Buck M, Wang YP, Kolb A (2006) Protein-induced DNA bending clarifies the architectural organization of the sigma54-dependent glnAp2 promoter. Mol Microbiol 59:168–180
Jelsbak L, Givskov M, Kaiser D (2005) Enhancer-binding proteins with a forkhead- associated domain and the σ54 regulon in myxococcus Xanthus fruiting body development. Proc Natl Acad Sci U S A 102:3010–3015

Joung JK, Le LU, Hochschild A (1993) Synergistic activation of transcription by *Escherichia coli* cAMP receptor protein. Proc Natl Acad Sci U S A 90:3083–3087
Joung JK, Koepp DM, Hochschild A (1994) Synergistic activation of transcription by bacteriophage lambda cI protein and *E. coli* cAMP receptor protein. Science 265:1863–1866
Kiupakis AK, Reitzer LJ (2002) ArgR-independent induction and ArgR-dependent superinduction of the astCADBE operon in *Escherichia coli*. J Bacteriol 184:2940–2950
Lee DH, Schleif RF (1989) In vivo DNA loops in araCBAD: size limits and helical repeat. Proc Natl Acad Sci U S A 86:476–480
Magasanik B (1993) The regulation of nitrogen utilization in enteric bacteria. J Cell Biochem 51:34–40
Muller J, Oehler S, Muller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. J Mol Biol 257:21–29
Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K (2007) RNA polymerase is poised for activation across the genome. Nat Genet 39: 1507–1511
Ninfa AJ, Peng J (2005) PII signal transduction proteins:sensors of α-ketoglutarate that regulate nitrogen metabolism. Curr Opin Microbiol 8:168–173
Ninfa AJ, Reitzer LJ, Magasanik B (1987) Initiation of transcription at the bacterial glnAp2 promoter by purified *E. coli* components is facilitated by enhancers. Cell 50:1039–1046
Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nat Genet 31:69–73
Paulsson J (2004) Summing up the noise in gene networks. Nature 427:415–418
Pedraza JM, Paulsson J (2008) Effects of moelcular memory and bursting on fluctuations in gene expression. Science 319:339–343
Rappas M, Bose D, Zhang X (2007) Bacterial enhancer-binding proteins: unlocking sigma-54 dependent gene transcription. Curr Opin Struct Biol 17:110–116
Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304:1811–1814
Rasmussen EB, Lis JT (1993) In vivo transcriptional pausing and cap formation on three drosophila heat shock genes. Proc Natl Acad Sci U S A 90:7923–7927
Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH (2001) A highly purified RNA polymerase II elongation control system. J Biol Chem 276:42601–42609
Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. Science 307:1962–1965
Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J (2011) Effect of promoter architecture on the cell-to-cell variability in gene expression. PLOS Comput Biol 7:e1001100
Schulz A, Langowski J, Rippe K (2000) The effect of the DNA conformation on the rate of NtrC activated transcription of *Escherichia coli* RNA polymerase sigma(54) holoenzyme. J Mol Biol 300:709–725
Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the drosophilla embryo. EMBO J 11:4047–4057
Stathopoulos A, Levine M (2005) Genomic regulatory networks and animal development. Dev Cell 9:449–462
Su W, Porter S, Kustu S, Echols H (1990) DNA-looping and enhacer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. Proc Natl Acad Sci U S A 87:5504–5508
Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD (2009) A synthetic genetic edge detection program. Cell 137:1272–1281
Tasmir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 469:212–215
Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98:8614–8619
Xu R, Hoover TR (2001) Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol 4:138–144

Yuh CH, Bolouri H, Davidson EH (2001) Cis-regulatory logic in the endo16 gene: switching from specification to a differentiation mode of control. Development 128:617–629

Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA (2007) RNA polymerase stalling at developmental control genes in the drosophila melanogaster embryo. Nat Genet 39:1512–1516

Zhang X, Chaney M, Wigneshweraraj SR, Schumacher J, Bordes P, Cannon W, Buck M (2002) Mechanochemical ATPases and transcriptional activation. Mol Microbiol 45:895–893

Zhang Z, Klatt A, Gilmour DS, Henderson AJ (2007) Negative elongation factor NELF represses human immunodeficiency virus transcription by pausing the RNA polymerase II complex. J Biol Chem 282:16981–16988

Chapter 2
Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Reprograming Microbial Metabolic Pathways

Cong T. Trinh and R. Adam Thompson

Abstract Elementary mode analysis is a useful metabolic pathway analysis tool to characterize cellular metabolism. It can identify all feasible metabolic pathways known as elementary modes that are inherent to a metabolic network. Each elementary mode contains a minimal and unique set of enzymatic reactions that can support cellular functions at steady state. Knowledge of all these pathway options enables systematic characterization of cellular phenotypes, analysis of metabolic network properties (e.g. structure, regulation, robustness, and fragility), phenotypic behavior discovery, and rational strain design for metabolic engineering application. This chapter focuses on the application of elementary mode analysis to reprogram microbial metabolic pathways for rational strain design and the metabolic pathway evolution of designed strains.

Keywords Elementary mode analysis • Engineering • Genetic knockout analysis • Metabolic minimal metabolic functionality • Rational strain design

Abbreviations

CASOP	computational approach for strain optimization aiming at high productivity
cMCS	constrained minimal cut set

C.T. Trinh (✉)
Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN, USA
e-mail: ctrinh@utk.edu.

R.A. Thompson
Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, USA

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_2,

EM	elementary mode
EMA	elementary mode analysis
ExPa	extreme pathway
FBA	flux balance analysis
MFA	metabolic flux analysis
MPA	metabolic pathway analysis

2.1 Introduction

Advances in sequencing technology have generated a large database of genome sequences of different organisms across the three kingdoms of bacteria, eukaryote, and archaea (Pagani et al. 2012). Based on these sequences, comparative genomics enable the automated reconstruction of drafted metabolic networks of these sequenced organisms in a high throughput manner (Henry et al. 2010). A standard protocol has been recently developed to curate these draft metabolic networks to improve the model accuracy (Thiele and Palsson 2010). A reconstructed metabolic network provides a useful link between cellular genotype and phenotype that enables a systematic analysis of cellular metabolism (Durot et al. 2009). Based on the first principle of mass conservation, different metabolic network analysis tools have been developed to predict the fates of metabolites and metabolic flux distributions that define cellular phenotypes under environmental and genetic perturbations (Lewis et al. 2012). A list of curated metabolic network models for some organisms suitable for metabolic network analysis can be found in the BiGG database (Schellenberger et al. 2010).

Metabolic network analysis can be classified into dynamic and structural metabolic network modeling. An accurate dynamic metabolic network model can help elucidate complex dynamic cellular phenotypes under different environmental and genetic perturbations. A general framework has been recently proposed to develop a dynamic model for a genome-scale metabolic network (Jamshidi and Palsson 2008). Currently its development is hampered by the unavailability of kinetic parameters and is limited to small-size metabolic networks (typically <100 reactions). Several approaches have been developed to generate kinetic parameters under uncertainty to enable dynamic metabolic network modeling, for instance, cybernetic kinetic modeling and ensemble metabolic network modeling through various assumptions (Machado et al. 2012; Song and Ramkrishna 2012; Tran et al. 2008; Wang et al. 2004; Young et al. 2008). However, this topic is not the focus of this chapter and will not be discussed further.

Due to lack of kinetic parameters, structural metabolic network modeling has been widely applied for analyzing cellular metabolism under steady-state. Depending on what assumptions are made and whether experimental data are required, different techniques have been developed to analyze the invariant of metabolic networks such as metabolic flux analysis (MFA), flux balance analysis (FBA), and metabolic pathway analysis (MPA) including elementary mode and extreme pathway analyses (Lewis et al. 2012; Stephanopoulos et al. 1998; Trinh et al. 2009).

This chapter will first discuss the formulation of common structural metabolic network modeling techniques and their differences in addressing biological questions through a demonstration of a simple metabolic network. Next it will present the application of elementary mode analysis (EMA), one of the useful MPA tools for characterizing cellular metabolism with an emphasis on reprogramming microbial metabolic pathways for rational strain design and the metabolic pathway evolution of designed strains.

2.2 Different Techniques to Analyze a Metabolic Network

The theory behind the analysis of metabolic networks is built upon the first principle of mass conservation of internal metabolites within a biological system (Reder 1988; Schuster and Schuster 1993; Stephanopoulos et al. 1998). A system can be defined by a single cell or a cell compartment that contains internal metabolites. These metabolites are transformed to others through an intricate web of enzyme-catalyzed reactions. The reactions that transform metabolites within the system are defined as internal reactions while reactions involving the transport of metabolites in and out of the system are considered exchange reactions (Schilling et al. (2000a, b); Schuster and Hilgetag 1994). The general equation to describe the mass conservation of metabolites within a system of defined volume is

$$\frac{d}{dt}\underline{C} = \underline{\underline{S}} \cdot \underline{r} - \mu \cdot \underline{C} \tag{2.1}$$

where $\underline{C}$ (mol/L) is the column vector ($m \times 1$) of internal metabolite concentrations; $\underline{\underline{S}}$ is the stoichiometric matrix ($m \times n$) of the metabolic network whose rows and columns represent metabolites and reactions, respectively; $\underline{r}$ (mol/L/h) is the flux (reaction rate) vector ($n \times 1$); and μ (1/h) is the dilution effect, e.g., volume expansion during cell growth. Within the dimensions of a single cell, since the dilution rate is much slower than reaction rates of enzymes (Stephanopoulos et al. 1998), the contribution of volume change to the change in metabolite concentrations can be considered negligible. In addition, due to the fast turnover rate of internal metabolites, they can be assumed not to accumulate in the system (Fell 1992; Stephanopoulos et al. 1998) and one can simplify Eq. (2.1) to

$$\underline{\underline{S}} \cdot \underline{r} = 0 \tag{2.2}$$

The reversibility of a reaction is defined by the thermodynamic constraint. If a reaction is irreversible, it is constrained to carry out positive flux, i.e.

$$r_i \geq 0 \tag{2.3}$$

Figure 2.1a shows how the problem is formulated for a simple metabolic network. The network consists of 11 reactions, 2 of which are reversible, and contains 11 metabolites, 6 of which are internal. Typically, Eq. (2.2) is an underdetermined system

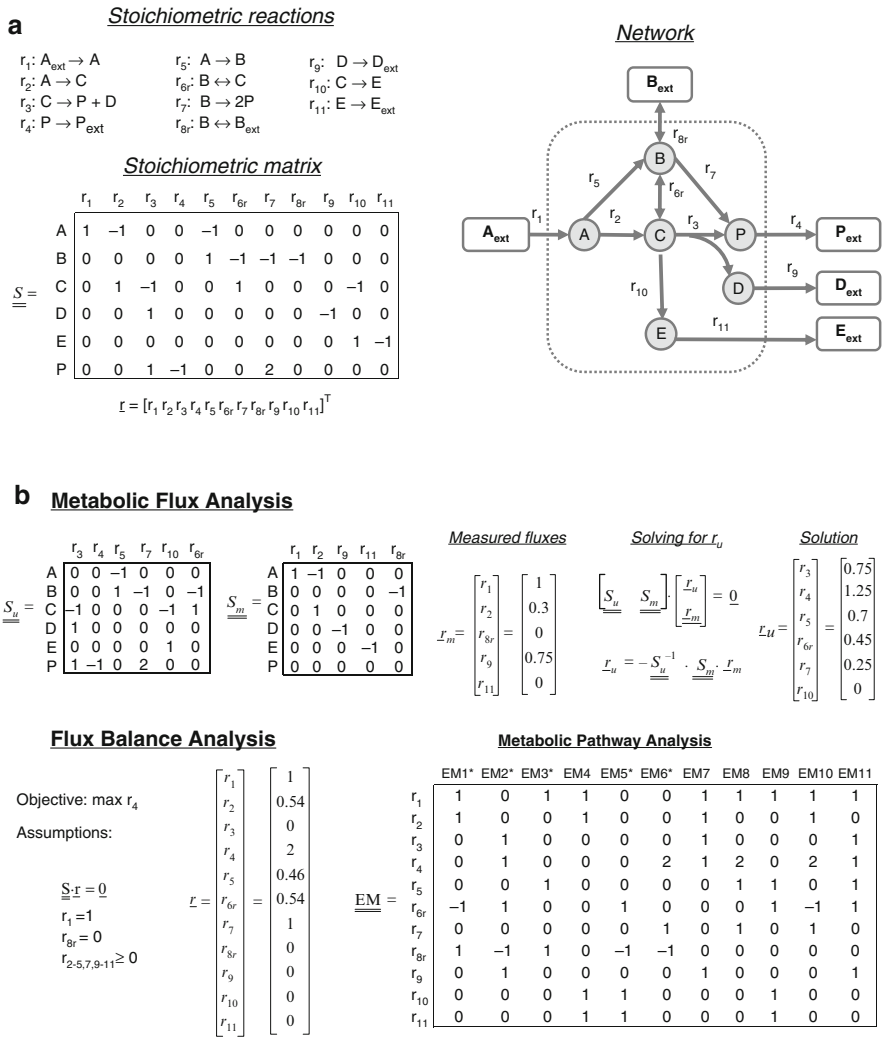


Fig. 2.1 Tools for metabolic network analysis. (**a**) Problem Statement: The simple depicted network is comprised of internal metabolites A, B, C, D, E, and P linked through the internal reactions (e.g. r_2, r_3, r_5, r_{6r}, r_7, and r_{10}) and external metabolites (e.g. A_{ext}, B_{ext}, E_{ext}, and P_{ext}) which are transported into or out of the system through the exchange reactions (e.g. r_1, r_4, r_{8r}, r_9, and r_{11}). The reactions r_{6r} and r_{8r} are reversible which can proceed in a forward or reverse direction. The stoichiometric matrix $\underline{\underline{S}}$ is constructed from the stoichiometric reactions where the rows denote internal metabolites, columns correspond to reactions, and each element s_{ij} within the matrix represents the stoichiometric coefficient of metabolite *i* participating in reaction *j* (e.g. $s_{ij}>0$ for products and $s_{ij}<0$ for reactants). (**b**) Metabolic Flux Analysis: The stoichiometric matrix S is partitioned into $[S_u\ S_m]$ and the metabolic flux vector $\underline{r}$ is partitioned into $[\underline{r}_u\ \underline{r}_m]$ where u and m denote unmeasured and measured, respectively. Calculating $\underline{r}_u$ is feasible if $\underline{r}_m$ is known. Flux Balance Analysis: The objective function is chosen to maximize the production of metabolite P while the constraints include metabolite A used as the only substrate, thermodynamic restriction for some reactions (e.g. $r_{2\text{-}5,7,9\text{–}11} \geq 0$), and

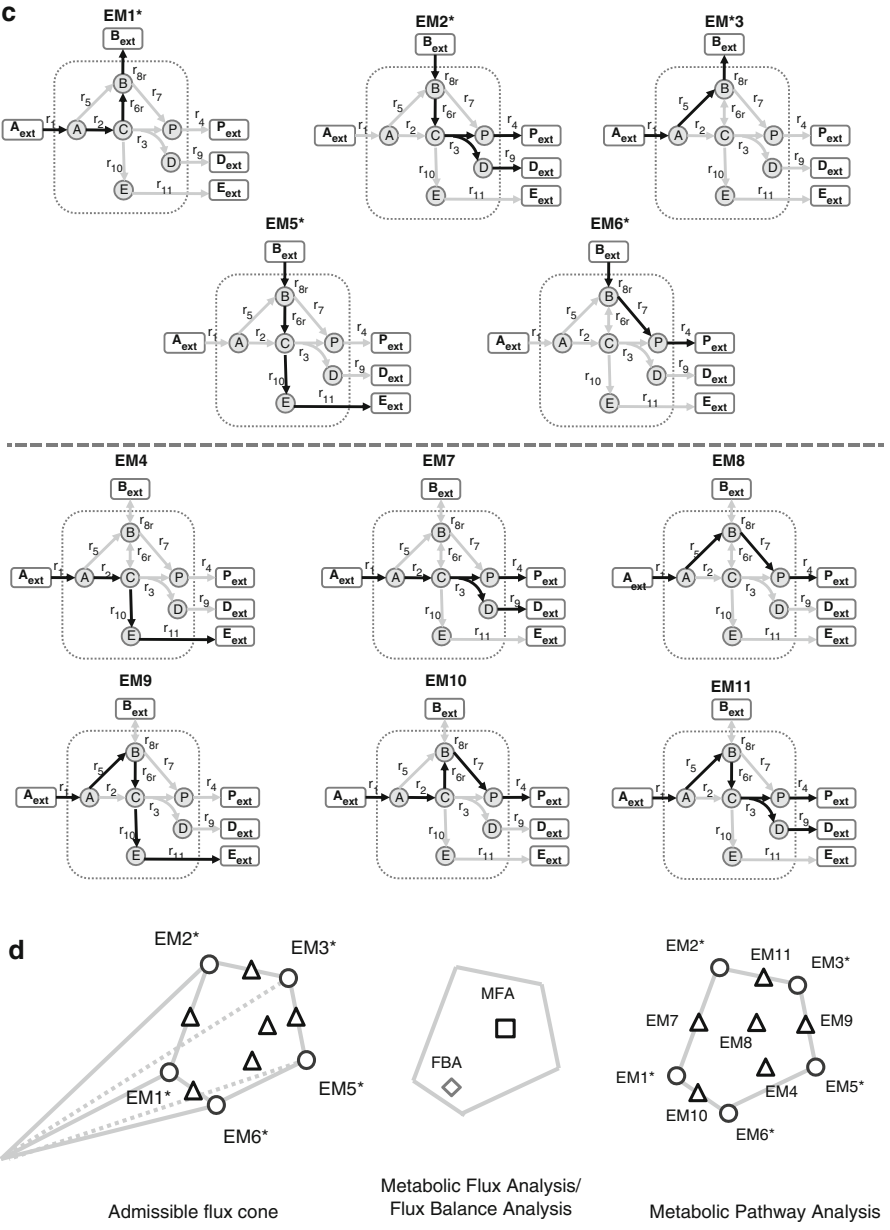


Fig. 2.1 (continued) fixed flux $r_1 = 1$. Metabolic Pathway Analysis: The elementary mode matrix EM of this network calculated by METATOOL contains 11 unique EMs, 5 of which are ExPas (shown in asterisks). (**c**) Pictorial Representation: ExPas (*above*) and EMs (*below*) calculated for this metabolic network. (**d**) Geometric Interpretation of Analysis: The admissible polyhedral flux cone represents all possible pathways within the network spanned by 5 ExPas. MFA identifies a pathway that lies within the cone and is in accordance with the measured fluxes, while FBA also finds a single pathway that solves the objective function. MPA finds all genetically independent pathways within the admissible flux cone. ExPas (*circles*) and EMs (*triangles*)

where the number of metabolites is much less than the number of reactions. Depending on the invariant structure of S and whether some experimentally measured fluxes are required for model input, three main techniques have been developed to solve the system of linear equations (2.2) together with the inequality constraint (2.3) for the metabolic flux vector $\underline{r}$. These methods are metabolic flux analysis (MFA), flux balance analysis (FBA), and metabolic pathway analysis (MPA). Their differences are briefly discussed below and illustrated in Fig. 2.1.

2.2.1 Metabolic Flux Analysis

A metabolic network typically contains more unknown fluxes than mass balance equations for metabolites ($m<n$), so there are n-m degrees of freedom assuming that $m=\text{rank}\,(\underline{\underline{S}})$. MFA can calculate the metabolic flux vector $\underline{r}$ by measuring n-m fluxes (Stephanopoulos et al. 1998). By partitioning $\underline{r}$ into measured (e.g. $\underline{r}_m$) and unmeasured (e.g., $\underline{r}_u$) vectors and the stoichiometric matrix $\underline{\underline{S}}$ into the corresponding $\underline{\underline{S}}_m$ and $\underline{\underline{S}}_u$, one can easily calculate $\underline{r}_u$ as follows:

$$\underline{r}_u = -\underline{\underline{S}}_u^{-1} \cdot \underline{\underline{S}}_m \cdot \underline{r}_m \tag{2.4}$$

MFA is set up to calculate the metabolic flux vector $\underline{r}$ corresponding to a physiological state of the cell under a given condition. It requires a large set of experimental data to solve for $\underline{r}_u$ especially for a large metabolic network, which can become experimentally infeasible. MFA usually handles small and simplified metabolic networks (Martínez et al. 2010), which may be inaccurate. Figure 2.1b illustrates the concept of MFA.

The ^{13}C-based metabolic flux analysis is a more advanced technique that calculates the metabolic flux vector r by additionally using the ^{13}C-labelled pattern of the stable and abundant protein-bound amino acids determined by either gas chromatography coupled with mass spectroscopy (GC/MS) and/or nuclear magnetic resonance (NMR) (Christensen et al. 2002; Sauer 2006; Wiechert et al. 2001; Zamboni et al. 2005, 2009). The most sophisticated technique known as kinetic flux profiling has recently been developed to calculate the metabolic flux vector $\underline{r}$ by measuring the dynamic incorporation of labeled substrates (e.g. ^{13}C, ^{15}N) into downstream intermediate metabolites. This measurement can be subsequently used to calculate metabolic fluxes (rates) directly without relying on the simplified metabolic network like the traditional MFA approach (Yuan et al. 2006, 2008, 2010).

2.2.2 Flux Balance Analysis

Similar to MFA, flux balance analysis (FBA) determines a metabolic flux vector $\underline{r}$ corresponding to a physiological state for the cell under a given condition but can

be applied for a genome-scale metabolic network ($m << n$). The method is based on the convex analysis theory and imposes objective function(s) (e.g. maximizing specific growth rate, minimizing ATP utilization, maximizing product formation, etc.) (Feist and Palsson 2010; Schuetz et al. 2007) that are constrained by several reaction fluxes in $\underline{r}$ (e.g. substrate uptake rates, product secretion rates, or thermodynamic constraints) to calculate the metabolic flux vector $\underline{r}$ (Feist and Palsson 2008; Kauffman et al. 2003; Lewis et al. 2012; Price et al. 2004). Therefore, some experimentally measured fluxes (typically fewer than n-m) should be provided to perform FBA. The more experimentally measured fluxes are available, the better the model prediction will perform to determine the metabolic flux vector $\underline{r}$. The OMICS data can also be used to develop gene regulatory constraints on $\underline{r}$ to improve the model prediction (Chandrasekaran and Price 2010; Jensen et al. 2011; Park et al. 2007). Figure 2.1b demonstrates the concept of FBA, which determines the flux vector $\underline{r}$ in the example network with the objective function of maximizing the formation of the product P from only the substrate A.

Some challenging issues in applying FBA are to justify the assumption of objective function(s) to predict the physiological state of the cell under a given condition (Schuster et al. 2008) and account for all alternative solutions that may represent actual physiological states of the cell (Lee et al. 2000; Mahadevan and Schilling 2003; Reed and Palsson 2004). Although facing potential drawbacks, one of the useful applications of FBA is to implement the bilevel programming for reengineering microbial metabolic pathways for rational strain design, e.g. OPTKNOCK, OPTGENE, OPTSTRAIN, OPTREG, OPTORF, BiMOMA, SimOptStrain, and EMiLiO (Burgard et al. 2003; Kim and Reed 2010; Kim et al. 2011; Patil et al. 2005; Pharkya et al. 2004; Pharkya and Maranas 2006; Yang et al. 2011).

2.2.3 Metabolic Pathway Analysis

Different from both MFA and FBA, metabolic pathway analysis (MPA) identifies all admissible metabolic flux vectors existing in a metabolic network by solving the Eq. (2.2) coupled with the inequality (2.3) without requiring any input of measured metabolic fluxes or objective functions for the analysis. From the convex analysis theory, solutions of the Eq. (2.2) constrained by the inequality (2.3) belong to the convex polyhedral flux cone and are infinite (Rockafellar 1970). The inequality (2.3) specifies directions of reactions in the network to be either reversible or irreversible and is governed by the reaction thermodynamics (Schuster et al. 1994, 2002). To obtain a finite number of solutions that span the convex polyhedral cone, additional constraints such as non-decomposability (Schuster et al. 1994) and systematic independence (Schilling et al. 2000a, 2000b) are imposed on the Eq. (2.2) together with the inequality (2.3). Depending on what constraints are used, two techniques have been developed: elementary mode analysis (EMA) and extreme pathway analysis (EPA).

Table 2.1 A list of software packages to calculate elementary modes

Softwares	URL	Capabilities	Refs
CellNetAnalyzer	http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html/	EMs, ExPas	Klamt et al. (2007)
COPASI	http://www.copasi.org/tiki-view_articles.php	EMs	Hoops et al. (2006)
efmtool	http://www.csb.ethz.ch/tools/efmtool	EMs	Terzer and Stelling (2008)
ElMoComp	http://elmocomp.sourceforge.net/elmocomp.php	EMs	
ExPa	http://systemsbiology.ucsd.edu	ExPas	Bell and Palsson (2005)
METATOOL	http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool/metatool5.0/metatool5.0.html	EMs, ExPas	Pfeiffer et al. (1999) and von Kamp and Schuster (2006)
ScrumPY	http://mudshark.brookes.ac.uk/ScrumPy	EMs	Poolman (2006)
YANAsquare	http://www.biozentrum.uni-wuerzburg.de/yana.html	EMs, ExPas	Schwarz et al. (2007)

EMA uses the non-decomposability constraint that reduces the infinite number of solutions of $\underline{r}$ to a finite set. The constraint states that let $C(\underline{r})$ be a set of indices of non-zero fluxes of the admissible flux vector $\underline{r}$ ($\in R^n$); $\underline{r}$ is an EM iff there is no other EMs $\underline{r}^*$ ($\in R^n$) such that $C(\underline{r}) \subset C(\underline{r}^*)$ (Schuster et al. 1994). This constraint makes $\underline{r}$ contain a minimal set of enzymatic reactions satisfying the Eq. (2.2) and inequality (2.3) and be unique up to a scalar multiple. Thus, deletion of any reaction belonging to an EM will disrupt its function, so that each EM can be regarded as a "genetically independent" pathway, a unique property for rational strain design.

EPA uses the same constraints as EMA in addition to the systematic independence specifying that any ExPa cannot be expressed as a non-linear combination of two other different ExPas (Schilling et al. 2000a, b). All ExPas constitute the generating vectors that span the entire polyhedral convex cone. For a given metabolic network, all ExPas form a subset of EMs (Klamt and Stelling 2003; Papin et al. 2004). It is straight forward to determine which EMs are ExPas by using the nullity test (Jevremovic et al. 2010). The nullity of the stoichiometric matrix $\underline{\underline{S}}$ defines the dimension of the null space of $\underline{\underline{S}}$. This test can determine whether any feasible metabolic flux vector that satisfies the Eq. (2.2) and inequality (2.3) is an EM or ExPa even before the computation of all EMs. Let $\underline{\underline{S}}$ and $\underline{r}$ in the Eq. (2.1) be partitioned into $[S_a\ S_b\ S_c\ S_d]$ and $[\underline{r}_a\ \underline{r}_b = 0\ \underline{r}_c\ \underline{r}_d = 0]^T$ where the columns of $[S_a\ S_b]$ and $[S_c\ S_d]$ correspond to irreversible and reversible reactions, respectively. A feasible flux vector $\underline{r}$ is defined as an ExPa iff nullity $([S_a\ S_c\ S_d]) = 1$ or an EM iff nullity $([S_a\ S_c]) = 1$.

Figure 2.1b illustrates the application of MPA for analyzing a simple metabolic network. EMA shows that there are a total of 11 EMs inherent to the metabolic network and 5 of them are ExPas (Fig. 2.1c). These EMs can be computed by any software package shown in Table 2.1 and ExPas can be simply evaluated by applying the nullity test for all computed EMs. Figure 2.1d shows the differences among

different structural metabolic network modeling techniques, e.g., MFA, FBA, and MPA. Either MFA or FBA computes a metabolic flux vector $\underline{r}$ corresponding to a physiological state of the cell under a given condition and can be expressed as a linear combination of ExPas or EMs. In comparison, ExPas represent the edge of the convex polyhedral cone while some EMs can lie within the cone.

As stated above, Table 2.1 shows a list of different software packages that have been developed to compute EMs and ExPas. The fastest software to compute elementary modes to date is efmtool that employs the bit pattern trees algorithm (Terzer and Stelling 2008). Different techniques have also been developed to compute EMs for a large-scale metabolic network based on determination of the K-shortest elementary flux mode(s) (de Figueiredo et al. 2009), elementary flux patterns (Kaleta et al. 2009), and parallel computation (Centler et al. 2010; Jevremovic et al. 2011).

2.3 Applications of EMA for Rational Strain Design

The ability to identify all feasible EMs inherent to a metabolic network with unique properties makes EMA a useful metabolic pathway analysis tool for a wide range of applications including systematic characterization of cellular phenotypes (Beurton-Aimar et al. 2011; Carlson and Srienc 2004a, b; Kurata et al. 2007; Larhlimi and Bockmayr 2009; Nookaew et al. 2007; Poolman et al. 2003, 2004; Rajvanshi and Venkatesh 2011; Schwartz and Kanehisa 2005, 2006; Wiback et al. 2003, 2004; Wlaschin et al. 2006; Zhao and Kurata 2010), analysis of metabolic network properties (Ay and Kahveci 2010; Flynn et al. 2012; Papin et al. 2002a, b; Peres et al. 2011; Poolman et al. 2007; Price et al. 2003), regulation (Cakir et al. 2004, 2007; Carlson 2007; Stelling et al. 2002; Wessely et al. 2011), robustness (Larhlimi et al. 2011; Min et al. 2011; Stelling et al. 2002, 2004), fragility (Behre et al. 2008; Klamt 2006; Klamt and Gilles 2004; Tepper and Shlomi 2010; Wilhelm et al. 2004), phenotypic behavior discovery (Kaleta et al. 2011; Kenanov et al. 2010; Rügen et al. 2012; Schauble et al. 2011), identification of pharmaceutical or therapeutic targets (Beuster et al. 2011), and rational strain design for metabolic engineering applications (Table 2.2). Several comprehensive reviews have discussed some of these applications in detail (Medema et al. 2012; Schaeuble et al. 2011; Schuster et al. 1999, 2006; Trinh et al. 2009) so this chapter will focus on the reprogramming of microbial metabolic pathways for rational strain design and the metabolic pathway evolution of designed strains.

From the entire set of EMs computed for a given metabolic network, it is straight forward to select a subset of EMs that can provide the highest product yields. To enforce the operation of only these EMs, a naïve approach is to delete all reactions whose fluxes are null in the selected subsets. The number of knockout candidates can be significantly large due to the fact that (1) some reaction fluxes are inactive under a given growth condition considered in the analysis, e.g. growth on glucose does not utilize reactions involved in other sugar uptakes and (2) reactions belonging to a linear pathway will have null fluxes if only one of them is deleted. For instance, assume that EM8 is the most efficient and feasible pathway to produce P from S

Table 2.2 A list of EMA applications for rational strain design

Species	Products	Description	Exp verified?	References
Asperillus niger	Fructofuranosidase	Identifying genetic modifications for enhancing recombinant fructofuranosidase production based on EMA.	Yes	Driouch et al. (2012)
Corynebacterium glutamicum	L-valine	Improving valine production by increasing NADPH regeneration guided by EMA.	Yes	Bartek et al. (2010)
	L-glutamate	Using EMA to identify key metabolic nodes that can be manipulated by rational dissolved-oxygen control strategy to improve L-glutamate production.	Yes	Chen et al. (2009)
	1,5-Diaminopentane	Using EMA to predict yield improvement of 1,5-diaminopentane synthesis in *C. glutamicum* by using *E. coli* transhydrogenase enzyme to replenish NADPH pool.	No	Kind and Wittmann (2011)
	L-methionine	Using EMA to guide substrate optimization for improving L –methionine production.	No	Kromer et al. (2006)
	L-lysine	Using EMA to identify genetic modifications to enhance cell growth and L-lysine production by co-utilizing lactate-glucose mixture.	Yes	Neuner and Heinzle (2011)
Escherichia coli	Succinate	Using EMA to identify genetic modifications and guide dissolved-oxygen supply to improve succinate production from glycerol.	No	Chen et al. (2010)
	Ethanol	Using EMA to design, construct, characterize, and validate engineered *E. coli* strains for the efficient conversion of sugars and glycerol into ethanol.	Yes	Trinh and Srienc (2009) and Trinh et al. (2008)
	Isobutanol	Using EMA to design, construct, characterize and validate an engineered *E. coli* strain for anaerobic isobutanol production.	Yes	Trinh et al. (2011)
	DHAP	Using EMA to identify the optimal pathways for efficient DHAP production.	Yes	Liao et al. (1996)
	Diapolycopendioic acid	Using EMA to design, construct, and characterize an engineered *E. coli* strain for enhanced diapolypendioic acid production.	Yes	Unrean et al. (2010)
	Recombinant proteins	Using EMA to design an engineered *E. coli* strain for improved recombinant protein production.	No	Vijayasankaran et al. (2005)

Klebsiellia pneumoniae	1,3-Propanediol	Using EMA to investigate the metabolic capabilities of *K. pneumonia* for producing 1,3-propanediol from glycerol under aerobic and anaerobic conditions.	No	Zhang and Xiu (2009)
	1,3-Propanediol	Using EMA to guide the increased supply of reducing equivalent NADPH for enhanced 1,3-propandiol production.	Yes	Chen et al. (2011)
Psedomonas putida/ Ralstonia eutropha	Cyanophycin	Applying EMA to simulate growth conditions to optimize cyanophycin production.	Yes	Diniz et al. (2006)
Saccharomyces cerevisiae	Poly-β-hydroxybutyrate	Using EMA to analyze a recombinant yeast capability for poly-β-hydroxybutyrate production.	Yes	Carlson et al. (2002)
	Higher alcohols	Using EMA to guide strain design for improved higher alcohols production.	No	Matsuda et al. (2011)
Thermoanaerobacter saccharolyticum	Ethanol	Using EMA to design, construct, characterize, and validate an engineered *T. saccharolyticum* for efficient ethanol production and predict the metabolic pathway evolution of the engineered and evolved mutant.	Yes	Unrean and Srienc, (2011) and Unrean and Srienc (2012)

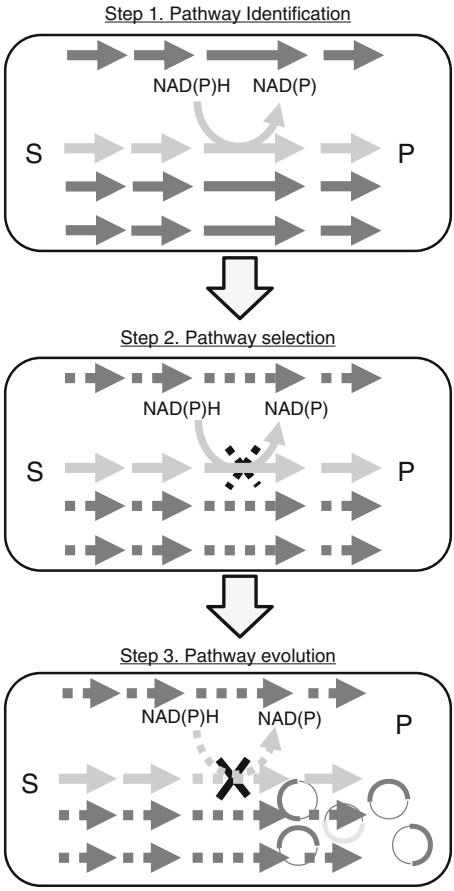


Fig. 2.2 EMA-based rational strain design strategy using the MMF technique (Adapted from Trinh et al. 2009)

with a maximum yield of 2 (mol P/mol S) (Figs. 2.1b and 2.3e). The naïve approach to constrain the operation of only this pathway is to block a large set of reactions {r2, r3, r6r, r8r, r9, r10, r11} but the alternative approach is to select only a subset of these reactions {r2, r6r, r8r} to delete and constrain the operation of this EM8. Therefore, it is desirable to find a minimum set of deleted reactions to reduce the entire set of EMs to a subset that yields the most desirable phenotypes. Different algorithms have been developed to implement the rational strain design.

2.3.1 Minimal Metabolic Functionality Technique

Minimal **M**etabolic **F**unctionality (MMF) technique focuses on finding a minimum set of multiple deleted reactions to reprogram metabolic pathways of a rationally designed strain to overproduce target product(s). Figure 2.2 shows the three steps

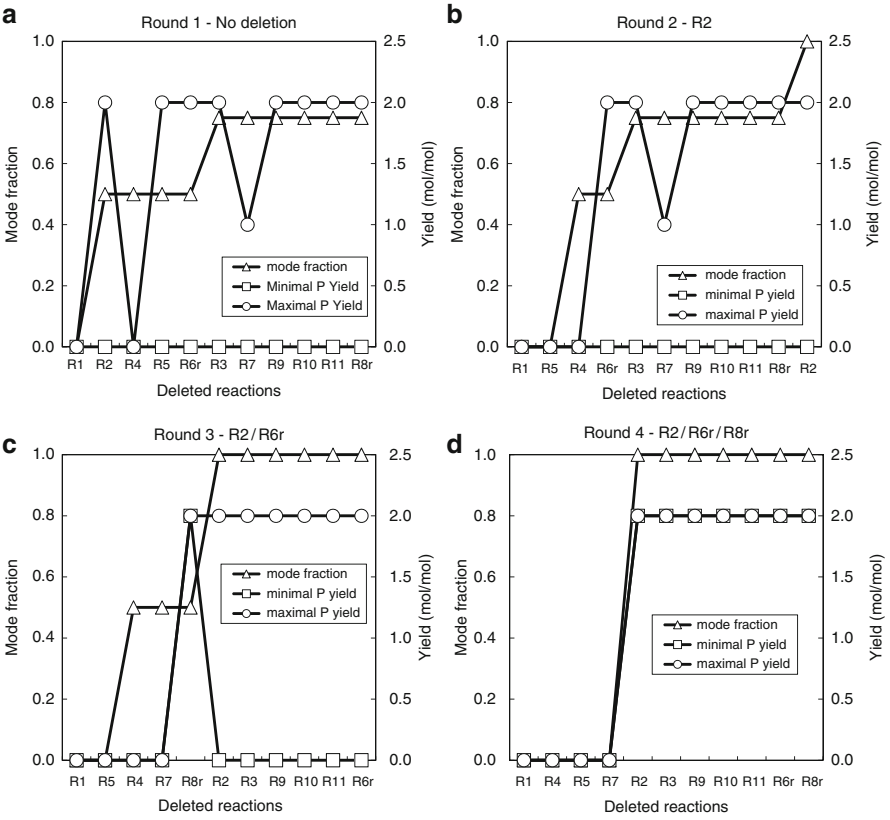
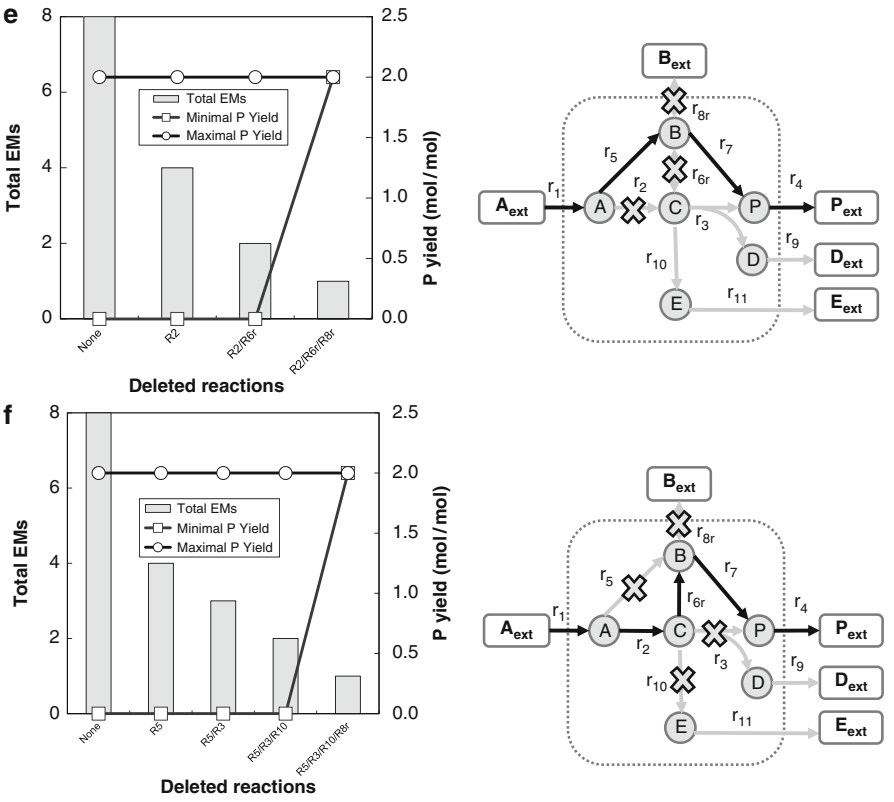


Fig. 2.3 Rational strain design through the execution of multiple reaction deletions, using substrate A to produce metabolite P. (**a**) Original network analysis of single deletions with the remaining mode fraction sorted in the increasing order. (**b**) Single deletions following the initial deletion of r_2. (**c**) Single deletions following deletion of r_2 and r_{6r}. (**d**) Final single deletion analysis following sequential deletion of r_2, r_{6r}, and r_{8r}. (**e**) Summary of multiple reaction deletions and the effect on number of EMs and the yield range, and the pathway illustrated (*right*). (**f**) Alternative solution based on EMA that achieve the same optimal $Y_{P/A}$

involved in the rational strain design by using MMF (Trinh et al. 2009) and Fig. 2.3 demonstrates the implementation of these steps through a simple metabolic network.

Step 1: Metabolic Pathway Identification (Fig. 2.2). This step aims to identify all possible feasible pathways inherent to a metabolic network that a cell can utilize to function. In principle, the wild type strain can utilize any of these pathways to function under a given condition. It is recommended to use EMA to carry out this step instead of EPA. By closely examining 5 ExPas among the complete set of 11 EMs in the example network, it is observed that none of ExPas could convert P into A at the theoretical yield. This observation suggests that the set of ExPas may not be sufficient to be used for rational strain design by applying MMF.



Step 2: Metabolic Pathway Selection. This step aims to select and strictly enforce the operation of the most efficient pathway(s) of interest. It identifies minimum multiple genetic modifications (e.g. gene knock-out/knock-in, gene over-/down-expression) to enable the engineered strain to function only according to the most desirable pathways based on three criteria. The first criterion is to eliminate as many undesirable pathways as possible to reduce the entire set of EMs to a small desirable subset. In the second criterion, EMs belonging to the desirable subset can accommodate cell growth and target product synthesis at the highest possible yield during the cell growth phase. For the third criterion, the selected EMs can achieve the maximum theoretical product yield during the no-growth phase (Trinh et al. 2008).

For example, assume that the most desirable pathway in the simple network is the one that can achieve maximum yield of P on A (2 mol P/mol A) and there is no biomass term involved in the network for simple demonstration purpose. Figure 2.3 shows the sequential steps of identifying multiple deleted reactions to achieve the maximum $Y_{P/A}$ based on the above three criteria. Figure 2.3a shows the first round of identifying a target deleted reaction and the effect of single deleted reactions on the remaining mode fraction (arranged in increasing order) and the minimum and

maximum $Y_{P/A}$. In this first round, the choice of knocking out r_2, r_5, or r_{6r} is preferable because it can reduce a large fraction of undesirable pathways and does not affect maximum $Y_{P/A}$. The process of identifying the multiple target deleted reactions is repeated as shown in Fig. 2.3b–d until the strain design criteria is met. By deleting r_2, r_{6r}, or r_{8r}, there is only one EM left that maximize $Y_{P/A}$. For practical metabolic network that also contains the biomass synthesis reaction, it is very often not feasible to reduce the entire set of EMs to a single EM but only a subset of desirable EMs that are constrained to couple growth and product formation, and achieve a narrow range of high product yields.

MMF was used to rationally design E. coli strains for the most efficient biomass production from glucose (Trinh et al. 2006), simultaneous utilization of mixed sugars and glycerol to ethanol (Trinh and Srienc 2009; Trinh et al. 2008), anaerobic production of butanol and isobutanol from glucose (Trinh 2012; Trinh et al. 2011), and enhanced production of carotenoids (Unrean et al. 2010). Recently, MMF was applied to design a thermophilic bacterium *Thermoanaerobacter saccharolyticum* for efficient conversion of glucose to ethanol at 60 °C (Unrean and Srienc 2011). All of the designed strains were constructed, characterized, and validated with model prediction. It should be pointed out that in most of the designed strains reported, the tight coupling between growth and product formation was proven by removing the function of desirable pathways, e.g. ethanol and isobutanol production and observing the subsequent inability of cell growth.

Step 3: Metabolic Pathway Evolution (Fig. 2.2). This step is an optional step but can be employed as a powerful tool to overcome metabolic flux imbalance in the reprogrammed metabolic pathways. Due to genetic modifications by design, the engineered strain can be constrained to operate in new reprogrammed pathways whose enzyme capacities may not yet been evolved to accommodate flux changes in new pathways. As a result, some metabolite intermediates in the reprogrammed metabolic pathways may be accumulated inside the cell, which causes cytotoxicity, severely affects cell growth, and reduces the target product synthesis. However, since the engineered strain is designed to tightly couple growth and desirable pathways tailored to efficient target product synthesis, metabolic pathway evolution can be carried to overcome the rate liming steps to balance metabolic fluxes and increase product formation just simply based on simple growth selection. This approach has been successfully demonstrated for the efficient conversion of glycerol to ethanol in *E. coli* at 37 °C (Trinh and Srienc 2009) and the efficient conversion of glucose to ethanol in *T. saccharolyticum* at 60 °C (Unrean and Srienc 2011). Through the metabolic pathway evolution, the engineered and evolved *E. coli* grew faster with higher ethanol yields, titers and productivities (Trinh and Srienc 2009). A similar trend was observed for the thermophilic strain *T. saccharolyticum* at 60 °C (Unrean and Srienc 2011). By applying the metabolic flux control analysis, it was determined that the rate limiting step responsible for improved performance of the engineered and evolved strain *T. saccharolyticum* was due to the mutated enzyme PGI (phosphoglucose isomerase) (Unrean and Srienc 2012). Interestingly, it is suggested that the engineered strain *T. saccharolyticum* evolved toward the state of maximum total entropy production (Unrean and Srienc 2011).

It should be pointed out that a similar approach to the MMF technique, FBA-based technique such as OPTKNOCK also seeks to design engineered strains to couple growth and product formation and employ the metabolic pathway evolution strategy to improve the strain performance, e.g. development of an engineered and evolved *E. coli* strains for improved production of lactate (Fong et al. 2005), 1,4-butanediol (Yim et al. 2011), and various flavonoids (Fowler et al. 2009).

2.3.2 Other EMA-Based Techniques for Rational Strain Design

CASOP, the **C**omputational **A**pproach for **S**train **O**ptimization Aiming at high **P**roductivity was developed to design optimal strains for optimizing target product synthesis (Hädicke and Klamt 2010). This method devised a systematic approach to identify the rating values for each reaction in a metabolic network that can be chosen for deletion, down-regulation, or over-expression. This method finds a trade-off between high yield EMs and high flexibility (e.g. capacity or pathway redundancy) to determine the rating values for each reaction by using two input parameters. The first parameter $\gamma \in [0, 1]$ specifies the desirable fraction of target product formation in competition with biomass synthesis. The second parameter $k \geq 0$ adjusts the magnitude of weighting factors for each EM that can lead to high target product formation. The CASOP is similar to MMF when a large k is chosen. This technique was demonstrated by developing engineered *E. coli* strains for efficient succinate and lactate production. The CASOP can be performed by using the CellNetAnalyzer software package (Table 2.1).

The constrained Minimal Cut Set (cMCS) is an useful technique for rational strain design. The original formulation of MCS focuses on finding minimal cut sets of reactions that can block a target metabolite or reaction flux. This formulation is later extended to identify what MCSs are required to eliminate undesirable pathways but optimize a target metabolite or reaction flux. Each cut set contains a unique and minimum set of deleted reactions that either disrupt or retain a target (e.g. a metabolite or reaction). The cMCS technique is similar to MMF but seeks to find all other alternative solutions. The cMCS for a metabolic network can be computed by using the CellNetAnalyzer software package (Table 2.1).

Similar to CASOP, the Flux Design technique identifies reactions in a metabolic network that need to be amplified or down-regulated to enhance the target product formation. By defining a target objective (e.g. a reaction flux that maximizes a target metabolite production), Flux Design finds the slope of linear correlation between the flux through the target reaction and each of other reactions in all EMs of a metabolic network. If the slope is positive or negative with the cutoff of $r^2 > 0.7$, the reaction can be targeted for amplification or down-regulation, respectively. This technique was applied to identify candidate reaction steps for improving lysine production in *C. glutamicum* and protein production (e.g. fructofuranosidase, glucoamylase, and epoxide hydrolyase) in *A. niger* (Driouch et al. 2012; Melzer et al. 2009).

2.4 Conclusion

EMA is a useful metabolic pathway analysis tool for rational strain design and metabolic pathway evolution of designed strains. EMA is entirely based on the structural analysis of invariant metabolic networks without requiring kinetics parameters or experimental flux data, such as substrate uptake or secretion rates, which enables the prediction of genetic modifications (e.g. gene deletion and over- or down-expression) to reprogram microbial metabolic pathways and achieve desirable phenotypes.

Acknowledgements This work is supported in parts by the laboratory start-up, SEERC seed, and JDRD funds for CT from the University of Tennessee, Knoxville.

References

Ay F, Kahveci T (2010) Functional similarities of reaction sets in metabolic pathways. In: Proceedings of the first ACM international conference on bioinformatics and computational biology, ACM, Niagara Falls, New York, pp 102–111

Bartek T, Blombach B, Zonnchen E, Makus P, Lang S, Eikmanns BJ, Oldiges M (2010) Importance of NADPH supply for improved L-valine formation in corynebacterium glutamicum. Biotechnol Prog 26:361–371

Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S (2008) Structural robustness of metabolic networks with respect to multiple knockouts. J Theor Biol 252:433–441

Bell SL, Palsson BO (2005) Expa: a program for calculating extreme pathways in biochemical reaction networks. Bioinformatics 21:1739–1740

Beurton-Aimar M, Beauvoit B, Monier A, Vallee F, Dieuaide-Noubhani M, Colombie S (2011) Comparison between elementary flux modes analysis and 13c-metabolic fluxes measured in bacterial and plant cells. BMC Syst Biol 5:95

Beuster G, Zarse K, Kaleta C, Thierbach R, Kiehntopf M, Steinberg P, Schuster S, Ristow M (2011) Inhibition of alanine aminotransferase in silico and in vivo promotes mitochondrial metabolism to impair malignant growth. J Biol Chem 286:22323–22330

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657

Cakir T, Kirdar B, Ulgen KO (2004) Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. Biotechnol Bioeng 86:251–260

Cakir T, Kirdar B, Onsan ZI, Ulgen KO, Nielsen J (2007) Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in saccharomyces cerevisiae. BMC Syst Biol 1:18

Carlson RP (2007) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics 23:1258–1264

Carlson R, Srienc F (2004a) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. Biotech Bioeng 85:1–18

Carlson R, Srienc F (2004b) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. Biotech Bioeng 86:149–162

Carlson R, Fell D, Srienc F (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotech Bioeng 79:121–134

Centler F, Kaleta C, di Speroni Fenizio P, Dittrich P (2010) A parallel algorithm to compute chemical organizations in biological networks. Bioinformatics 26:1788–1789

Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and mycobacterium tuberculosis. Proc Natl Acad Sci U S A 107:17845–17850

Chen N, Du J, Liu H, Xu QY (2009) Elementary mode analysis and metabolic flux analysis of L-glutamate biosynthesis by corynebacterium glutamicum. Ann Microbiol 59:317–322
Chen Z, Zhang JA, Liu DH (2010) Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by *Escherichia coli*. J Biomed Biotechnol 518743
Chen Z, Liu H, Liu D (2011) Metabolic pathway analysis of 1,3-propanediol production with a genetically modified klebsiella pneumoniae by overexpressing an endogenous NADPH-dependent alcohol dehydrogenase. Biochem Eng J 54:151–157
Christensen B, Karoly Gombert A, Nielsen J (2002) Analysis of flux estimates based on 13c-labelling experiments. Eur J Biochem 269:2795–2800
de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25:3158–3165
Diniz SC, Voss I, Steinbuchel A (2006) Optimization of cyanophycin production in recombinant strains of pseudomonas putida and ralstonia eutropha employing elementary mode analysis and statistical experimental design. Biotechnol Bioeng 93:698–717
Driouch H, Melzer G, Wittmann C (2012) Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. Metab Eng 14:47–58
Durot M, Bourguignon P-Y, Schachter V (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev 33:164–190
Feist A, Palsson B (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotech 26:659–667
Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol 1(3):344–349
Fell DA (1992) Metabolic control analysis: a survey of its theoretical and experimental development. Biochem J 286(Pt 2):313–330
Flynn CM, Hunt KA, Gralnick JA, Srienc F (2012) Construction and elementary mode analysis of a metabolic model for shewanella oneidensis Mr-1. Biosystems 107:120–128
Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 91:643–648
Fowler ZL, Gikandi WW, Koffas MAG (2009) Increased malonyl coenzyme a biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. Appl Environ Microbiol 75:5831–5839
Hädicke O, Klamt S (2010) Casop: a computational approach for strain optimization aiming at high productivity. J Biotechnol 147:88–101
Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotech 28:977–982
Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) Copasi – a complex pathway simulator. Bioinformatics 22:3067–3074
Jamshidi N, Palsson BO (2008) Formulating genome-scale kinetic models in the post-genome era. Mol Syst Biol 4:171
Jensen P, Lutz K, Papin J (2011) Tiger: toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst Biol 5:147
Jevremovic D, Trinh CT, Srienc F, Boley D (2010) On algebraic properties of extreme pathways in metabolic networks. J Comput Biol 17:107–119
Jevremovic D, Trinh CT, Srienc F, Sosa CP, Boley D (2011) Parallelization of nullspace algorithm for the computation of metabolic pathways. Parallel Comput 37:261–278
Kaleta C, de Figueiredo LF, Schuster S (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res 19:1872–1883
Kaleta C, de Figueiredo LF, Werner S, Guthke R, Ristow M, Schuster S (2011) In Silico Evidence for gluconeogenesis from fatty acids in humans. PLoS Comput Biol 7:e1002116
Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol 14:491–496

Kenanov D, Kaleta C, Petzold A, Hoischen C, Diekmann S, Siddiqui RA, Schuster S (2010) Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary flux mode analysis. FEBS J 277:1023–1034

Kim J, Reed J (2010) Optorf: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 4:53

Kim J, Reed JL, Maravelias CT (2011) Large-scale Bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One 6:e24162

Kind S, Wittmann C (2011) Bio-based production of the platform chemical 1,5-diaminopentane. Appl Microbiol Biotechnol 91:1287–1296

Klamt S (2006) Generalized concept of minimal Cut sets in biochemical networks. Bio Syst 83:233–247

Klamt S, Gilles ED (2004) Minimal cut sets in biochemical reaction networks. Bioinformatics 20:226–234

Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis? Trends Biotechnol 21:64–69

Klamt S, Saez-Rodriguez J, Gilles ED (2007) Structural and functional analysis of cellular networks with cellnetanalyzer. BMC Syst Biol 1:2

Kromer JO, Wittmann C, Schroder H, Heinzle E (2006) Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and corynebacterium glutamicum. Metab Eng 8:353–369

Kurata H, Zhao Q, Okuda R, Shimizu K (2007) Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. BMC Syst Biol 1:31

Larhlimi A, Bockmayr A (2009) A New constraint-based description of the steady-state flux cone of metabolic networks. Discret Appl Math 157:2257–2266

Larhlimi A, Blachon S, Selbig J, Nikoloski Z (2011) Robustness of metabolic networks: a review of existing definitions. Biosystems 106:1–8

Lee S, Phalakornkule C, Domach MM, Grossmann IE (2000) Recursive milp model for finding all the alternate optima in Lp models for metabolic networks. Comput Chem Eng 24:711–716

Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nat Rev Micro 10:291–305

Liao JC, Hou SY, Chao YP (1996) Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. Biotechnol Bioeng 52:129–140

Machado D, Costa RS, Ferreira EC, Rocha I, Tidor B (2012) Exploring the gap between dynamic and constraint-based models of metabolism. Metab Eng 14:112–119

Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276

Martínez I, Bennett GN, San K-Y (2010) Metabolic impact of the level of aeration during cell growth on anaerobic succinate production by an engineered *Escherichia coli* strain. Metab Eng 12:499–509

Matsuda F, Furusawa C, Kondo T, Ishii J, Shimizu H, Kondo A (2011) Engineering strategy of yeast metabolism for higher alcohol production. Microb Cell Fact 10:70

Medema MH, van Raaphorst R, Takano E, Breitling R (2012) Computational tools for the synthetic design of biochemical pathways. Nat Rev Micro 10:191–202

Melzer G, Esfandabadi M, Franco-Lara E, Wittmann C (2009) Flux design: in silico design of cell factories based on correlation of pathway fluxes to desired properties. BMC Syst Biol 3:120

Min Y, Jin X, Chen M, Pan Z, Ge Y, Chang J (2011) Pathway knockout and redundancy in metabolic networks. J Theor Biol 270:63–69

Neuner A, Heinzle E (2011) Mixed glucose and lactate uptake by corynebacterium glutamicum through metabolic engineering. Biotechnol J 6:318–329

Nookaew I, Meechai A, Thammarongtham C, Laoteng K, Ruanglek V, Cheevadhanarak S, Nielsen J, Bhumiratana S (2007) Identification of flux regulation coefficients from elementary flux modes: a systems biology tool for analysis of metabolic networks. Biotechnol Bioeng 97:1535–1549

Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC (2012) The genomes online database (gold) V.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40:D571–D579

Papin JA, Price ND, Edwards JS, Palsson BBO (2002a) The genome-scale metabolic extreme pathway structure in haemophilus influenzae shows significant network redundancy. J Theor Biol 215:67–82

Papin JA, Price ND, Palsson BO (2002b) Extreme pathway lengths and reaction participation in genome-scale metabolic networks. Genome Res 12:1889–1900

Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405

Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci U S A 104:7797–7802

Patil K, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinform 6:308

Peres S, Vallee F, Beurton-Aimar M, Mazat JP (2011) Acom: a classification method for elementary flux modes based on motif finding. Biosystems 103:410–419

Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S (1999) Metatool: for studying metabolic networks. Bioinformatics 15:251–257

Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8:1–13

Pharkya P, Burgard A, Maranas C (2004) Optstrain: a computational framework for redesign of microbial production systems. Genome Res 14:2367–2376

Poolman MG (2006) Scrumpy: metabolic modelling with python. IEE Proc Syst Biol 153:375–378

Poolman MG, Fell DA, Raines CA (2003) Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. Eur J Biochem 270:430–439

Poolman MG, Venkatesh KV, Pidcock MK, Fell DA (2004) A method for the determination of flux in elementary modes, and its application to lactobacillus rhamnosus. Biotechnol Bioeng 88:601–612

Poolman MG, Sebu C, Pidcock MK, Fell DA (2007) Modular decomposition of metabolic systems via null-space analysis. J Theor Biol 249:691–705

Price ND, Reed JL, Papin JA, Famili I, Palsson BO (2003) Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. Biophys J 84:794–804

Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2:886–897

Rajvanshi M, Venkatesh K (2011) Phenotypic characterization of corynebacterium glutamicum under osmotic stress conditions using elementary mode analysis. J Ind J Ind Microbiol Biotechnol 38:1345–1357

Reder C (1988) Metabolic control theory: a structural approach. J Theor Biol 135:175–201

Reed JL, Palsson BØ (2004) Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14:1797–1805

Rockafellar RT (1970) Convex analysis, 28th edn. Princeton University Press, Princeton

Rügen M, Bockmayr A, Legrand J, Cogne G (2012) Network reduction in metabolic pathway analysis: elucidation of the key pathways involved in the photoautotrophic growth of the green alga *Chlamydomonas reinhardtii*. Metab Eng 14(4):458–467

Sauer U (2006) Metabolic networks in motion: ^{13}C-based flux analysis. Mol Syst Biol 2:62

Schaeuble S, Schuster S, Kaleta C (2011) Hands-on metabolism: analysis of complex biochemical networks using elementary flux modes. In: Jameson D, Verma, M, Westerhoff, HV (eds) Methods in enzymology, vol 500. Elsevier/Academic, San Diego, pp 437–456

Schauble S, Heiland I, Voytsekh O, Mittag M, Schuster S (2011) Predicting the physiological role of circadian metabolic regulation in the green alga *Chlamydomonas reinhardtii*. PLoS One 6:e23026

Schellenberger J, Park J, Conrad T, Palsson B (2010) Bigg: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinform 11:213

Schilling CH, Edwards JS, Letscher D, Palsson BO (2000a) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol Bioeng 71:286–306

Schilling CH, Letscher D, Palsson BO (2000b) Theory for the systemic definition of metabolic pathways and their Use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 203:229–248

Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119

Schuster S, Hilgetag S (1994) On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2:165–182

Schuster R, Schuster S (1993) Refined algorithm and computer program calculating All Non-negative fluxes admissible in steady states of biochemical reaction systems with and without some fluxes rates fixed. CABIOS 9:79–85

Schuster S, Hilgetag C, Woods JH, Fell DA (1994) Elementary modes of functioning in biochemical networks. In: Cuthbertson R, Holcombe M, Paton R (eds) Computation in cellular and molecular biological systems. World Scientific, Singapore, pp 151–165

Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60

Schuster S, Hilgetag C, Woods JH, Fell DA (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J Math Biol 45:153–181

Schuster S, Kamp A, Pachkov M (2006) Understanding the roadmap of metabolism by pathway analysis. Method Mol Biol 358:199–226

Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252:497–504

Schwartz J-M, Kanehisa M (2005) A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. Bioinformatics 21:ii204–205

Schwartz J-M, Kanehisa M (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinform 7:186

Schwarz R, Liang C, Kaleta C, Kuehnel M, Hoffmann E, Kuznetsov S, Hecker M, Griffiths G, Schuster S, Dandekar T (2007) Integrated network reconstruction, visualization and analysis using Yanasquare. BMC Bioinform 8:313

Song H-S, Ramkrishna D (2012) Prediction of dynamic behavior of mutant strains from limited wild-type data. Metab Eng 14:69–80

Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193

Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J (2004) Robustness of cellular functions. Cell 118:675–685

Stephanopoulos G, Aristidou AA, Nielsen JH (1998) Metabolic engineering: principles and methodologies. Academic, San Diego

Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26:536–543

Terzer M, Stelling J (2008) Large scale computation of elementary flux modes with Bit pattern trees. Bioinformatics 24(19):2229–2235

Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5:93–121

Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. Biophys J 95:5606–5617

Trinh C (2012) Elucidating and optimizing *E. coli* metabolisms for obligate anaerobic butanol and isobutanol production. Appl Microbiol Biotechnol 95(4):1083–1094

Trinh CT, Srienc F (2009) Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. Appl Environ Microbiol 75:6696–6705

Trinh CT, Carlson R, Wlaschin A, Srienc F (2006) Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab Eng 8:628–638

Trinh CT, Unrean P, Srienc F (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74:3634–3643

Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826
Trinh CT, Li J, Blanch HW, Clark DS (2011) Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. Appl Environ Microbiol 77:4894–4904
Unrean P, Srienc F (2011) Metabolic networks evolve towards states of maximum entropy production. Metab Eng 13:666–673
Unrean P, Srienc F (2012) Predicting the adaptive evolution of thermoanaerobacterium saccharolyticum. J Biotechnol 158:259–266
Unrean P, Trinh CT, Srienc F (2010) Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. Metab Eng 12:112–122
Vijayasankaran N, Carlson R, Srienc F (2005) Metabolic pathway structures for recombinant protein synthesis in *Escherichia coli*. Appl Microbiol Biotechnol 68:737–746
von Kamp A, Schuster S (2006) Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22:1930–1931
Wang L, Birol I, Hatzimanikatis V (2004) Metabolic control analysis under uncertainty: framework development and case studies. Biophys J 87:3750–3763
Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Mol Syst Biol 7:515
Wiback SJ, Mahadevan R, Palsson BO (2003) Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J Theor Biol 224:313–324
Wiback SJ, Mahadevan R, Palsson BO (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. Biotechnol Bioeng 86:317–331
Wiechert W, Möllney M, Petersen S, de Graaf AA (2001) A universal framework for 13c metabolic flux analysis. Metab Eng 3:265–283
Wilhelm T, Behre J, Schuster S (2004) Analysis of structural robustness of metabolic networks. Syst Biol 1:114–120
Wlaschin AP, Trinh CT, Carlson R, Srienc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. Metab Eng 8:338–352
Yang L, Cluett WR, Mahadevan R (2011) Emilio: a fast algorithm for genome-scale strain design. Metab Eng 13:272–281
Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7:445–452
Young JD, Henne KL, Morgan JA, Konopka AE, Ramkrishna D (2008) Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. Biotechnol Bioeng 100:542–559
Yuan J, Fowler WU, Kimball E, Lu W, Rabinowitz JD (2006) Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. Nat Chem Biol 2:529–530
Yuan J, Bennett BD, Rabinowitz JD (2008) Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc 3:1328–1340
Yuan Y, Hoon Yang T, Heinzle E (2010) ^{13}C metabolic flux analysis for larger scale cultivation using Gas chromatography-combustion-isotope ratio mass spectrometry. Metab Eng 12:392–400
Zamboni N, Fischer E, Sauer U (2005) Fiatflux – a software for metabolic flux analysis from ^{13}C-glucose experiments. BMC Bioinform 6:209
Zamboni N, Fendt SM, Ruhl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protoc 4:878–892
Zhang Q, Xiu Z (2009) Metabolic pathway analysis of glycerol metabolism in Klebsiella pneumoniae incorporating oxygen regulatory system. Biotechnol Prog 25:103–115
Zhao Q, Kurata H (2010) Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis. J Biosci Bioeng 110:254–261

Chapter 3
Evolutionary Engineering for Industrial Microbiology

Niti Vanee*, Adam B. Fisher*, and Stephen S. Fong

Abstract Superficially, evolutionary engineering is a paradoxical field that balances competing interests. In natural settings, evolution iteratively selects and enriches subpopulations that are best adapted to a particular ecological niche using random processes such as genetic mutation. In engineering desired approaches utilize rational prospective design to address targeted problems. When considering details of evolutionary and engineering processes, more commonality can be found. Engineering relies on detailed knowledge of the problem parameters and design properties in order to predict design outcomes that would be an optimized solution. When detailed knowledge of a system is lacking, engineers often employ algorithmic search strategies to identify empirical solutions. Evolution epitomizes this iterative optimization by continuously diversifying design options from a parental design, and then selecting the progeny designs that represent satisfactory solutions. In this chapter, the technique of applying the natural principles of evolution to engineer microbes for industrial applications is discussed to highlight the challenges and principles of evolutionary engineering.

Keywords Cellular objectives • Directed evolution • Diversity • Engineering objectives • Evolutionary engineering • Fitness landscapes • Industrial microbiology • Screening and selection

*The authors Niti Vanee and Adam B. Fisher are contributed equally.

N. Vanee • A.B. Fisher • S.S. Fong (✉)
Department of Chemical and Life Science Engineering, Virginia Commonwealth University, 601 W. Main Street, Room 422, 843028, Richmond, VA 23284-3028, USA
e-mail: ssfong@vcu.edu

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_3,

Abbreviations

ADAM	array-based discovery of adaptive mutations
CO	cellular objectives
EMS	ethyl methane sulfonate
EO	engineering objectives
EvoEng	evolutionary engineering
MAGE	multiplex automated genome engineering
NTG	nitroso-methyl guanidine
Oligo(s)	oligonucleotide(s)
RNAseq	RNA sequencing
SELEX	selectable evolution of ligands by exponential enrichment
SS	solution space
StEP	staggered extension process
TRMR	trackable multiplex recombineering

3.1 Introduction

The interaction of humans and microorganisms has a long history including the domestication of microbes as early as the fifth century BC for a variety of fermentation processes such as baking and viticulture. With the progression of time and scientific knowledge, microbiology has been applied to many industrial sectors including food, waste treatment, health and medicine, and more recently, energy. The microbes used in these processes have all been advantageous over other production methodologies due to the unique properties of life: a self-replicating system, capable of organizing highly complex, chaotic chemistry in response to constraints imposed by the surrounding state or environment. In this light, industrial microbiology utilizes the beneficial properties of microorganisms by programming cellular biochemistry to maximize production of a target chemical.

An additional interesting and potentially complicating aspect of microorganisms is that they evolve. Evolution, under natural conditions, is the process of change throughout a population over time: cycling between generation of diversity, and subsequent selection for the most 'fit' subpopulations. Classically, biology generates diversity through genetic mutation while natural selection acts as an evaluator of competitive fitness. In its most basic form, evolution can be considered to be an optimization function, finding maxima through iterative trial-and-error. Nature changes the genotypic composition of the cell through various non-directed phenomena such as point mutations, genome rearrangements and recombination (as well as proposed non-genetic mechanisms, i.e. epigenetics). These inherited changes are key determinants to the ultimately displayed phenotypic variations. However, the staggering complexity of even the smallest genomes makes it difficult to draw conclusive cause-and-effect relationships between the genotype and phenotype. Such unpredictability occludes ambitions of rational design and engineering of complex phenotypes without significant a priori knowledge; an important goal for industrial microbiology.

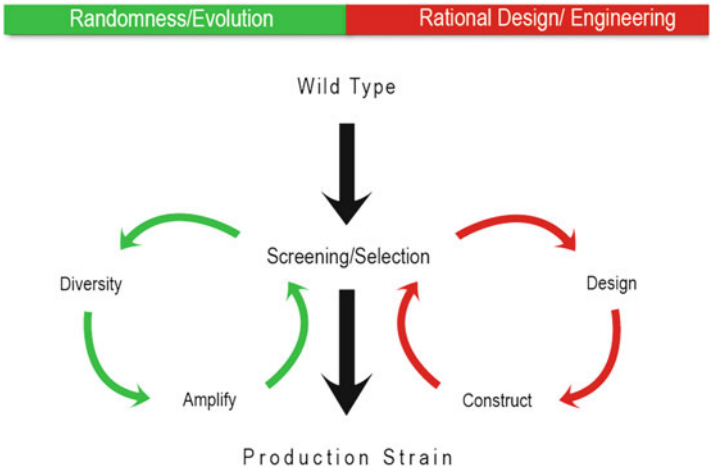


Fig. 3.1 The work flow for Evolutionary Engineering draws on two different, but related cycles – the evolutionary optimization cycle and the engineering cycle. First a wild-type strain or seed sequence is selected as a platform to begin the Evolutionary engineering workflow. The seed sequence proceeds one of two ways: it can proceed to the engineering cycle, where alterations will be made rationally, or the seed sequence can be optimized through the evolutionary cycle. Both cycles return to the screening/selection node at which the resulting characterization of fitness confers either a satisfactory production strain or a seed sequence for further iteration

Historically, one of the most powerful methods to produce a desired phenotype has involved leveraging the innate optimization properties of evolution. Often industrial microbiology goals are at odds with normal cellular function, so aligning evolutionary and industrial goals can lead to productive results. By utilizing a diverse population and selecting subpopulations for incorporation in subsequent rounds of evolution, it is possible to employ evolution as a search function for optimal genotypes with respect to the desired phenotype. This practice is known as evolutionary engineering, and will be the focus of the following chapter. As alluded by the name, evolutionary engineering is conceptually characterized by parameters of both the evolutionary process and the engineering principles (Fig. 3.1). As such several specific evolutionary engineering (EvoEng) methodologies and examples will be discussed with particular emphasis placed upon theory and applications to the design of industrially-relevant microbes. Additionally, EvoEng will be evaluated for its implications in systems biology and synthetic biology, as well as the implications of advancements in DNA sequencing and synthesis technologies on the future of EvoEng.

3.2 Optimization

In connecting evolutionary concepts to industrial microbiology design, an objective (e.g. production titer of biofuel) can be represented by a relative measure of fitness. This would allow for a visual and mathematical means of monitoring increases in a

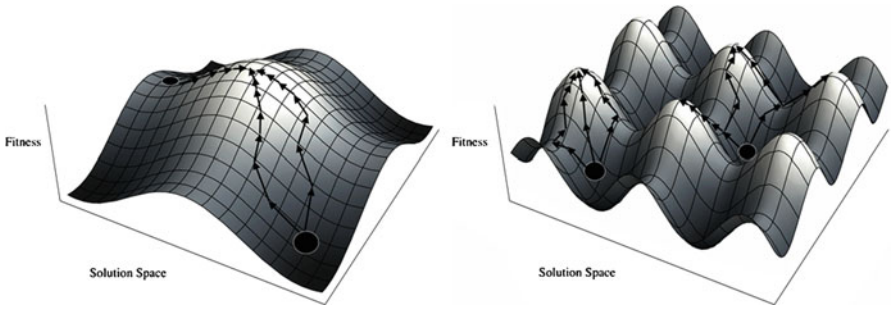


Fig. 3.2 Fitness landscapes. The three-dimensional fitness landscapes projected here represent the relationship between genotype and phenotype, with the two-dimensional Solution Space (SS) representing the genotype and all of its possible variations. During the evolutionary optimization process, genotypes diversify from a starting or seed sequence (*Black dots*) to related genotypes surrounding the seed sequence in the SS, while selection pressures to improve fitness provide a directionality to the evolutionary paths (*Black arrows*). As the topology is traversed, the global landscape and seed sequence will directly impact the final convergence. In (**a**) the fitness landscape is very homogenous, with one central peak, so all starting points will converge upon this global maxima, however, in (**b**) the multitude of local maximas allows many different paths of convergence even for the same seed sequence, although not every peak is accessible to the same seed sequence. In (**b**) this diversity of possible evolutionary paths is caused by the homogeneity of the rugged landscape; the slopes and peak heights are all identical, so trajectories have no bias in directionality and the different optima have no selectable fitness bias

desired objective. The preferred method for visualizing adaptation toward fitness is a *fitness landscape* (Fig. 3.2), a three-dimensional fitness map where the X- and Y-axes are chosen to represent underlying contributing factors to fitness. The possible cellular functions that are depicted on a fitness landscape represent the *solution space* (SS) or the design space. With accurate fitness landscapes models can describe the topological landscape of the correlation between the X and Y projection to fitness (Kauffman and Levin 1987). These models are then compiled and used to make predictions about the behavior of future designs. Thereby, it becomes impossible to achieve accuracy or precision in the *de novo* design of optimized systems without, first, accurate data regarding the behavior of parts within the system.

If a framework can be developed to describe the fitness landscape/SS, then it may also be possible to utilize mathematical or algorithmic approaches for interrogating the space for design purposes. There are a variety of different approaches that could be employed including expectation-maximization algorithms, simulated annealing algorithms, or cost-benefit analysis modeling approaches (Dekel and Alon 2005; Gillespie 1984). These types of approaches could not only help identify a global maximum in a fitness landscape, but also help direct the best methods to evolutionarily reach that maximum.

3.2.1 The Fitness Landscape

The fitness landscape is classically a three-dimensional plot, which has a "fitness" score displayed on the Z-axis as a dependence of the X and Y variables. First introduced

by Wright (1988, 1931), the fitness landscape in evolution uses the genome sequence as the X- and Y- axes, where the area of the two-dimensional plane formed by these two parameters represents the total SS for that particular genome. Within the total SS for an organism, each point represents one particular variant of the genome sequence. This projection of the fitness landscape can manifest as a monotonic (Fig. 3.2a) or a rugged landscape (Fig. 3.2b), which will greatly affect the movements around the SS. Each immediately neighboring point then represents one individual change in the genomic sequence. Given an *Escherichia coli* genome consists of ~4.6 million base pairs of DNA, which have four possible outcomes (or informational units), the possible SS would be $4^{4,600,000}$, as calculated by:

$$N = \lambda^{L}$$

Where N is the number of possible sequences, λ is the number of informational units, and L is the length of the genome. Obviously, this is a fantastically large number that is impossible to explore by empirical experimental characterization of each possible variants.

3.2.1.1 Parameters of Fitness Landscapes

Fitness landscapes are an extremely useful tool in evolutionary analysis when they are created using tightly controlled constants. These fitness landscapes are generated under the assumption that a phenotype is correlated only to changes in genotype, however this is clearly untrue. Abiotic environmental factors can drastically change cellular dynamics and thereby, phenotype. From an only slight increase in incubation temperature, the global phenotype of a microbe could radically switch to a heat-shock response – which will create a very different projection of cellular fitness. Indeed, many of these global changes of phenotypes will also change the evolutionary parameters of the microbe; in times of stress such as heat-shock or starvation microbial populations will increase the rate of mutation (through error-prone DNA repair systems). These changes to the evolutionary parameters alters the topology of the fitness landscape, as well as the movement of subpopulations around the SS. Thus, the fitness landscape is fundamentally linked to genotype, but the detailed fitness landscape project can change based upon any number of factors that can influence gene expression.

Similarly, fitness landscapes are susceptible to the influences of coevolution. It is always important to remember that the process of evolution happens on a population-scale: any measure of an individual's fitness is a relative quantification based upon the competitiveness of that particular variant against the other populations present. However, relationships may emerge between two variants that cause the exhibited phenotype to depend upon the presence of both strains. To observe such an evolutionary event it becomes important to maintain a level of heterogeneity of the culture, but have distinct separation of variants during the screening process.

In short, the parameters of the fitness landscape encompass a variety of factors that can alter the phenotype (fitness). These parameters that uncouple phenotype from genotype, have become the subject Epigenetics, coined by Conrad Waddington

in the 1940s and conceptually illustrated using fitness landscapes (Siegal and Bergman 2002; Waddington 1942, 1959, 1960) Focusing on instances of non-coupled phenotypic and genotypic variation, epigenetics can alter an overall cellular fitness landscape, acting as a sort of underlying epigenetic landscape. After all, each cell in the human body contains identical genetic information, yet liver tissue is very different from gut mucosal tissue. For phenotypic optimization it is important to remain cognizant of these epigenetic parameters, which may be underlying the perceived genetic changes.

3.2.1.2 Traversing the Fitness Landscape Through Evolution

Optimization in a fitness landscape becomes the process of climbing peaks to find fitness maxima. This process is possible through the capability of evolution to (i) move about the SS, (ii) sense improvements in fitness and (iii) select for subpopulations with an improved fitness. The movement around the SS is dependent upon the rate of diversification, the topology of the landscape, and the *seed sequences* – the starting points in the SS. For a moment, imagine being a blind mountain climber with the goal of climbing to the highest peak in an uncharted mountain range. Without direction, the climber would attempt to achieve their goal by continually climbing upward. In this way we could climb upward in any direction until we reach a peak – however – much to our chagrin, we reach the top of what we thought was the highest peak, only to realize there was a taller peak hidden behind the peak we just climbed! Now we have no way of reaching the higher peak without descending into a valley. Without changing the process (always moving upward) one way to reach a different peak would be to have a different starting point.

This is a simple analogy, but it is easy to see what we have encountered is the problem of local maxima – we will not be able to reach the global maximum from our initial conditions. If we had, perhaps, started in a different position or not had the first mountain blocking the larger peak behind it, we could've climbed the tallest peak the first time. Another possible solution is to be able to move greater distances with each successive step – possibly allowing us to "jump" across valleys. These changes in simulation conditions reflect favorable selection of seed sequences or a tuning of the mutational rate. Either *a priori* knowledge of the system or utilizing many starting points can drastically improve identification of seed sequences; for many phenotypes this can be as general as selecting the proper organism (or *chassis*) for evolution. Changing the search movements around the SS by tuning mutational rates can be done through many *in vivo* (i.e. sexuality) and *in vitro* (i.e. *in vitro* recombination) methodologies, which will be covered in Sect. 3.4.

3.2.2 Length Scales

The shape and step-size for traversing a fitness landscape can change in relation to different types of mechanisms of generating genetic diversity. While the foundation

for building a fitness landscape is DNA, the conceptual framework discussed to this point has focused on variations that occur within the context of an organism's wild-type genome. All organisms have a basal genetic mutation rate for point mutations. Point mutations constitute the smallest length-scale of genetic change. Point mutations will generally make small, if any, change to the shape of the fitness landscape and represents the smallest step size for traversing the fitness landscape. Different length-scales of genetic modifications can be used to make larger changes to the shape of the fitness landscape and to more quickly traverse through the landscape. Whole gene deletions or additions represent a next level of genetic change followed by introduction/deletion of pathways. Recent approaches to industrial microbiology have explored the design and use of microbial consortia for directed production. This represents a broad scale approach where the fitness landscape would be defined by the capabilities of two independent genomes. Looking ahead, the broadest scope would be to consider design and the fitness landscape from a metagenomic perspective where any genetic information is possible and synthetic biology can be used to experimentally implement completely novel gene combinations.

3.3 Design Parameters

3.3.1 Rationality vs. Randomness

Adaptation over a period of time to confer a favorable and stable functional state is at the core of biological evolution. As mentioned previously, natural biological evolution occurs by the generation of genetic diversity (via random mutations) and selection according to improvements in fitness associated with survival and replication. The genetic information of the organism not only codes for the information related to how the organism should function in the current environment but also the potential for evolving with changing environments. Genotypic alterations that are positively selected by the environmental factors/selection are at the core of Adaptive Evolution (AE) (Atwood et al. 1951). Besides environment, engineering designs can be developed to implement evolution for real-world applications to attain desired characteristic traits or engineering objectives in parallel to maintaining the basic cellular objectives. This approach is one form of EvoEng and is defined as a rational approach toward the design and fabrication of cells to obtain a stable phenotype. These phenotypic objectives play a primary role in quantification of evolution for respective genotypic changes. In other words, there exists a one-way relationship between the genotypic and phenotypic components (i.e. the changes at the genetic level are translated into phenotype). EvoEng, attempts to simultaneously maximize engineering objectives and cellular objectives by incorporating rational design to select for the engineering objective and randomness of evolution to search across the SS.

Ultimately, the source of genetic variation is mutation. Mutations may occur spontaneously or be induced by external mutagens to achieve diversity to address desired cellular and or metabolic objective. Spontaneous mutations that occur naturally in the form of point mutations, genome rearrangements or horizontal gene

transfers are considered to be relatively stable and occur at a low basal rate (Drake 1999). Natural environmental adversities such as nutrient deficiency or metabolic stress modulate the rate of such mutations. On the other hand there are external mutagens that can cause considerable changes in the environment of the organism leading to the phenomenon of frame shifts, deletion or insertions of the nucleotides.

Past researchers have proven through environment-dependent mutagenesis that engineering with evolution is sensitive to two major drawbacks. First, the rate of adaptation may not directly correlate to the rate of mutation. While comparing sexually reproducing populations and asexually reproducing populations it is known that the rate of mutation is elevated in asexual populations, yet does not alone accelerate the speed of evolutionary adaptation, for then sexuality would be outcompeted as a phenotype. Second, isolation of a mutant strain with desirable traits is highly dependent upon the selection and screening for the desired trait, which requires traits that are differential and quantifiable. To deal with these concerns, it is necessary to incorporate rationality in the design of an efficient and successful EvoEng investigation. When introducing rationality to evolution, there is a trade-off as rationality can constrain the possible evolutionary trajectories.

Metabolic engineering and Synthetic biology approaches utilize available gene-function information to take a fully rational approach to the design and construction of desired strains (Yokobayashi et al. 2002). In particular, Synthetic biology strives to establish bioengineering as a classical engineering discipline by developing methods for standardization, modularity and abstraction of biological parts (Valente and Fong 2011). This will enable biological design to occur in a high-throughput, rational fashion with all the benefits of a forward-engineering discipline.

Ideally, a completely rational approach would be taken in improving and altering production strains. In a complete rational design approach every cellular function could be designed to attain the optimal balance of cellular objectives with engineering objectives. However, this level of complete rational design requires absolute knowledge of the biological system – given the limitations of biological knowledge, this is not currently possible. At its core, EvoEng blends rationale with randomness by attempting to direct function towards a goal (rationale) by using the native cellular processes involved with evolution (randomness).

3.3.2 Establishing the Solution Space

The SS represents possible phenotypes based upon biological parameters. In Sect. 3.2, fitness landscapes were described as a method for visualizing and modeling local and global maxima of cell physiology, with the genotype considered as the underlying parameter that dictates the shape of the landscape. Establishing the shape of the fitness landscape is contingent upon a satisfactory knowledge of the underlying parameters. In this case, comprehensive evaluations of fitness landscapes were difficult prior to whole genome sequencing. However, a breadth of experiments

probing AE without full genome sequencing has contributed rough fitness projections. In the study by Lenski et al., where *Escherichia coli* was evolved in a laboratory setting for more than 10,000 generations by simple serial dilutions of batch culture. In these studies, the average fitness (growth) of the derived genotype was increased by approximately 50% relative to the ancestor (Lenski et al. 1998). The shape of the fitness landscape is determined by the genomic sequence of *E. coli,* movement across the landscape is related to mutations arising in the population, which are selected and characterized to explain the nature of the improved phenotype (Fong et al. 2005a, b). This process of AE and strain improvement is predicated by the starting genotype – what we have considered the seed sequence.

The best seed sequence is the one that produces a phenotype closest to the desired phenotypic maxima. This way the prospects for achieving an improved outcome are greatly increased as the required time to achieve the desired outcome is decreased. One proven method to improve seed sequence is to choose a fitting *chassis*. Chassis selection plays an important role in defining the SS as well as the culture conditions, which will be developed to underlying the industrial conditions. This way a design is optimized for certain environments, for example: enzymes have been evolved for thermostability by heterologous expression and selection in a thermostable chassis (Steipe 1999). As more AE experiments are conducted, results can be collectively analyzed to extrapolate genotype-phenotype relationships. In addition, collection of high-throughput "-omics" data provides detailed molecular-level gene expression data related to the global landscape. Improved knowledge of mechanisms leading to a desired trait can then be used to facilitate bottom-up specification of seed sequences for subsequent EvoEng or rational design approaches.

Besides favoring the improved starting point there is also a need to monitor the process of evolution with time to keep a track of each stage and collect data from each evolved generation. Keeping a record of the process allows us to monitor progress towards objectives as well as possibly collect information about major changes in trajectories. One particular 10,000 generation adaptation experiment illustrates a plot of competitive fitness generated through exemplary step measurement and monitoring (Lenski et al. 1998). With the rapid development of DNA technologies, sequencing has become much more feasible even allowing full genome re-sequencing projects to identify genetic mutations that occur across the genome as a result of AE (Atsumi et al. 2010; Barrick et al. 2009; Conrad et al. 2009; Gresham et al. 2008; Gabriel et al. 2006; Herring et al. 2006).

3.3.3 Competing Objectives

When considering the fitness landscape, the highest fitness is related to the phenotypic function with respect to a specific objective. In terms of EvoEng, there are normally two different objectives that often are at odds with each other, a cellular objective (e.g. growth) and an engineering objective (e.g. production of a target chemical).

3.3.3.1 Cellular Objective

In natural evolution organisms strive to maximize their competitiveness for an ecological niche – for microbes, this is usually through growth. We will refer to growth here as the *cellular objective* (CO), which is to maximize representation through increased reproduction or efficiency of biomass utilization. A cell must first incorporate nutrients to fuel cellular metabolic pathways, to finally carry out replication functions. For an entire cell, using the CO acts a measure of global fitness, which as we will see, may not always be the most optimal production strain.

3.3.3.2 Engineering Objective

Designing for industry requires the definition of an *engineering objective* (EO), which is the fitness score for industrial potential. Using either selection or screening (Sect. 3.4) the EO must be increased in each round. This objective can abstractly represent any phenotype, from resistance to a toxin to production of a useful intermediate, but the EO must overlap the CO if evolutionary selection toward the EO is to occur (Valente and Fong 2011). Further, maximizing both the CO and EO maximizes the fitness of a strain, but can be difficult to accomplish. When the target molecule for production is not a critical biomass component, the EO will not sufficiently overlap the CO and the cell will attempt to utilize its cellular resources to maximize its CO. Many of the target products of industry are secondary metabolites, which, by definition, fall into this latter category of targets. Hereby, in EvoEng the challenge is to design a selection system which can tie the EO to the CO. This may seem a difficult task, but consider the potential advantages of approaching engineering in a whole-cell optimization using the cells natural ability to grow and evolve.

3.3.4 Evolutionary Selection

Natural evolution processes achieve the diversity and complexity in the biosphere however the stability in this diversity is a function of natural selection. By having a continuous selection pressure, there is ongoing selection of beneficial characteristics, so a genotype with a fitness advantage will be maintained. Explaining in terms of the most simple AE, an evolved species acclimates to an environment and thrives, eventually replacing less adapted subpopulations (Hardin 1960). A major challenge faced by EvoEng is overlapping cellular COs that occur during evolution with EOs. By manipulating the growth environment and cellular interactions, natural selection is coupled to an engineering objective to perform selection in either in step-wise batch culture or continuous culture.

3.3.4.1 Selection Considerations

Using step-wise batch culturing, the culture is susceptible to several phenomena such as "Müller's ratchet". "Müller's ratchet" appears in asexual populations through the accumulation of deleterious mutations as a side effect of nonspecific mutagenic targeting. This creates a strain overfit to its selection; the strain may satisfy the EO and be selected for, but becomes severely crippled from accumulation of deleterious mutations towards the cellular phenotype permitted to "hitchhike" along (Müller 1964). Competition-based selection may also create mutational dynamics in which several subpopulations rise to existence, each possessing different beneficial alleles. In a process known as "clonal interference" (Müller 1964) one subpopulation may gain a slight advantage and overrun the other populations – preventing incorporation of possible other beneficial mutations. By dominating these other mutations, there may be a perceived reduction in the mutational rates, as noted by Luria and Delbruck (1943). Step-wise batch culture is susceptible to the effects of genetic drift – as cultures are restarted in fresh media from a small portion of the original culture subpopulations may be lost, including adapted variants. Clonal interference and genetic drift can be minimized in chemostat-driven experimentation when a higher number of mutants remain in the population (Muller 1932). In contrast, serial dilution of batch cultures causes homogeneity as selective pressures favor sweeps from clonal interference and drift, purging the diversity from culture (Conrad et al. 2011).

While some phenomena (clonal interference and drift) can promote population homogeny, other phenomena such as cross-feeding, can lead to stable subpopulation diversity. Cross-feeding features an evolved, stable commensalism between the metabolism of two or more subpopulations. For instance, it has been shown that *E. coli* mutants grown in glucose-limited media over 773 generations will yield a cross-feeding adaptation where a majority of the population represents a glucose-feeder/acetate-excreting phenotype, while a smaller slower-growing part of the population uptakes and metabolizes acetate (Rosenzweig et al. 1994; Helling et al. 1987). This particular cross-feeding example has been replicated, and shown to sometimes evolve as a diauxic switch in the acetate-feeding strain. Here the acetate-feeder metabolizes glucose at a slower rate than the native glucose-scavengers, but has an advantage of switching to an acetate-based metabolism more quickly than the wild-type glucose-respiring strain (Friesen et al. 2004). In industry, this form of co-metabolism is usually unfavorable as the culture phenotype will be a synthesis of two separate genomic sequences – meaning there is no individual variant with the desired phenotype. However there does exist circumstances for the design of microbial consortia. Cross-feeding adaptations that arise from random mutations are often undesired, but rationally-designed cross-feeding between populations can be advantageous as part of system design. Examples include a study where a hydrogen-consuming *Methanococcus maripaludis* was evolved along the hydrogen-excreting, lactate-fermenter, *Desulfovibrio vulgaris*. The consumption

of hydrogen by *M. maripaludis* fueled the thermodynamic driving force for the growth of *D. vulgaris* through lactate-fermentation (Hillesland and Stahl 2010).

3.3.4.2 Continuous-Culture Selection

Extended culture growth in chemostats has resulted in a variety of phenotypic adaptations (Helling et al. 1987; Sorgeloos et al. 1976; Atwood et al. 1951; Novick and Szilard 1950). In *E. coli* and *S. cerevisiae* carbon-limited chemostats have been used to increase biomass yield, growth rate and resistance to adverse conditions (Parekh et al. 2000; Vinci and Byng 1999; Rowlands 1984). The choice of limiting nutrient will dictate the evolutionary paths, and are therefore, central to selection. The results of limitation by nitrogen, phosphate, potassium, sulfur and other non-carbon source nutrients have been shown to be linked to the overproduction of various metabolic by-products (Dawson 1985). A very good review of adaptive laboratory evolution and continuous culture by U. Sauer, details the various formats of continuous culture including chemostats and their variants, batch culture, and microcolonization (Sauer 2001).

Since Environmental conditions can alter the evolutionary landscape so radically, the repeatability of EvoEng projects rely considerably on an experimenter's ability to control and report culture conditions (selection pressure). Furthermore, in industrial-scale fermentations repeatability can be the most essential quality of a robust production strain. This starts by cultivating investigations as close to industrial conditions as possible. This includes, but is not limited to; aeration, carbon sources, nutrient sources, pH, osmolarity, temperature, light-exposure, or cell density.

Eventually, advances in synthetic biology may address some of the current challenges with screening and selection. There continues to be a challenge in EvoEng of having two different objectives (cellular and engineering) to consider. It may be possible to utilize different synthetic constructs (RNA apatmers, ribozymes, etc.), designed biosensors, or genetic circuits to detect multiple desired inputs and convey selective fitness advantages.

3.4 Methodology

Traditionally, EvoEng has been strongly associated with "Classical Strain Improvement", through continuous culturing of a production-associated organism under selective conditions – usually paralleling industrial production conditions (Santos and Stephanopoulos 2008). However, like any science or engineering field, EvoEng has enjoyed significant advancements correlated with increases in technology. Recent improvements in sequencing, cloning, and high-throughput technologies have opened doors for researchers to investigate fitness landscapes and cellular optimization in unprecedented ways. This section will cover specific methodologies that have "evolved" to expand the EvoEng toolbox.

3.4.1 Generating Diversity

3.4.1.1 Point Mutations

Genetic diversity can occur by different mutagenic processes (SNPs, Insertions, and Deletions) that result in movement through the fitness landscape. If relying upon spontaneous mutagenesis as the main mechanism for phenotypic improvement, it is possible to conduct an evolutionary experiment with little foreknowledge of the system. A seminal piece of work, using only spontaneous improved penicillin titers 4,000-fold through selection on solid-media plates (Rowlands 1984). Naturally-occurring mutations actually occur at a nominally low and stable rate in normal cellular physiology, with DNA replicating faithfully (Drake 1999). This rate can be increased through the use of environmental influences or genetically-modified 'mutator' strains. Environmental factors can range from conditions which induce a stress-state in the cell (stationary-phase, glucose-limitation) to chemical mutagens: ethyl methane sulfonate (EMS) and nitroso-methyl guanidine (NTG) to ultraviolet irradiation (UV); all of which enhance various specific mutations (Rowlands 1982). For instance, it is known that NTG typically mutates close to the replication fork while UV irradiation is known to cause pyrimidine-dimers (Witkin 1976) – therefore mutagens are often varied in long-term experimentation. Utilizing these classic methods of spontaneous mutation has great utility as a fine-adjustment to the genotypic sequences. A mutation can occur at any part of the genome, the changes are completely independent of each other, and the rates of mutation can be controlled.

The original method of directing evolution has been most successful through the application of polymerase chain reaction (PCR). Using either 'leaky' DNA polymerase or a substitution of catalyzing ions (manganese as opposed to magnesium) to cause error-prone DNA polymerization. By designing primers to target your gene of interest, a large diversity of the enzyme is created which may be screened for desired phenotypes. More than likely, a wild-type enzyme will not be optimized for commercial purposes, but by utilizing PCR mutagenesis properties have been engineered into enzymes, such as thermostability, tolerance, novel catabolic activites, enantioselectivity, and substrate/product inhibition (Luetz et al. 2008).

3.4.1.2 Gene Modifications

Spontaneous mutagenesis can result in the silencing or overexpression of genes, but it is often difficult or costly to identify these mutations after they have occurred. For the purpose of quickly and inexpensively tracking genetic changes, transposons are widely available for use (de Lorenzo et al. 1998). These DNA elements are capable of self-catalyzing their insertion and movement across the genome or extra-chromosomal elements. Very often this will lead to inactivation of a gene on the chromosome, but some of these transposons feature highly expressive promoters, so it is also possible for transposon movement to result in gene overexpression if a transposon insertion is properly aligned next to a coding sequence (Schneider and Lenski 2004). However,

their real faculty lies in the fact that these transposons represent unique sequence, which can be used to trace observed phenotype back to the cognate genotypic change.

Another way the genome can be randomly overexpressed is through a collection of an overexpression library. First, the genome is sheared into smaller pieces that are inserted into plasmid vectors. These vectors can then be reinserted into strains to select for variants advantaged by overexpression of the inserted genomic sequence. This is conceptually similar to the common genetics technique, complementation. Methods of utilizing knockouts or overexpressions are commonly used to establish the seed sequence for an evolutionary or metabolic engineering investigation. Some of the recent work that has demonstrated the utility of augmenting gene expression was conducted in *E. coli* for the production of lycopene (Jin and Stephanopoulos 2007). In fact, it was shown that screening a library of plasmid-encoded genomic segments inside of a previously evolved knockout strain of *E. coli* increased production of lycopene further over application of either method in isolation (Jin and Stephanopoulos 2007).

Recombination is a powerful genetic tool for generating general genetic diversity or for specifically targeting desirable traits. In EvoEng, the potential for recombination during sexual reproduction can be a useful tool to produce recombinant phenotypes between parental strains bearing beneficial phenotypes. Eukaryotic organisms such as *S. cerevisiae*, which can exists in diploid or haploid states, are capable of *in vivo* recombination by the fusion of two haploids to create a chimeric diploid. For example, attributes of a highly-specialized production strain might be combined with a fast-growing industrial strain. Indeed, industrial fungal production has utilized the mating of yeast to reintroduce accelerated growth rates and more efficient biomass conversion in previously crippled strains (Rowlands 1982). More importantly, recombination can allow two mutually beneficial mutations to recombine – possibly resulting in a strain with compounded effects to the beneficial traits.

We have already seen that *in vivo* recombination can occur in clonal populations from the self-excising movements of transposons. However, the normal methods by which new DNA sequences arrive inside a prokaryote involves conjugation, transformation, or transduction. Most *in vitro* expansion of diversity will utilize transformation to introduce desired genetic information on an overexpression plasmid, but other constructs can be used. Conjugative plasmids have been used in the dairy industry (Vinci and Byng 1999) and phage-based transduction has been utilized for allelic replacement (Esvelt et al. 2011).

In vitro recombination allows for generation of PCR products that are recombinant versions of the parental template strains. In one of the more simple and widely used processes, staggered extension process (StEP), the templates are amplified by outer primers, but is staggered by abbreviated elongation cycles (Zhao et al. 1998). These abbreviated cycles yield only partially extended templates that can bind to heterologous template strands with limited homology yielding final products that are chimeric versions of the original templates. In the original study using this method by Zhao et al. the authors were able to isolate a version of Subtilin E that showed thermostability 25–50 times that of the wild-type enzyme (Zhao et al. 1998).

One benefit of these methodologies lies in their ability to address the phenomena of epistasis – for instance, a scenario were two mildly deleterious mutations on their own can combine to render a beneficial phenotype (Conrad et al. 2011, 2009; Dykhuizen and Hartl 1980). Also, allelic replacement can be used as a means of influencing the outcome of movement through a fitness landscape by altering the starting point of a strain. While it would be difficult to try and isolate these mutations together in a spontaneous manner, in allelic replacement these options are fully investigable. Epistatic interactions manifest as complexities in an organism's network topology and constrain the viable paths of evolution (Poelwijk et al. 2007).

3.4.1.3 Large-Scale Modifications

Genome shuffling can be viewed as a form of forced, accelerated recombination, as the process is leveraged for its ability to generate chimeras from a set of parental strains. This process, known as recursive protoplast fusion, relies upon the removal of the cell wall by polyethylene glycol (PEG) – a process used to prepare Gram-positive organisms for transformation of DNA. Protoplasts can also be fused, resulting in a diploid state in which the genomes exchange loci through homologous-crossing over. Since both genomes come from the same species the high level of homology will permit significant recombination, but by recursively fusing progeny of previous selections the SS searched grows considerably. When recursive protoplast fusion experiments were conducted in *Streptomyces fradiae* to investigate improved yield of the macrolide, tylosin, the authors we able to generate the same titers in two rounds of genome shuffling as accomplished by 20 rounds of classical mutagenesis (Zhang et al. 2002). In this way genome shuffling offers a strategy to do *in vivo* recombination in a variety of clonal populations to yield "progeny" with possible additive adaptive mutations (Petri and Schmidt-Dannert 2004).

Global Transcriptional Machinery Engineering

In adaptive laboratory evolution experimentation, mutations in the global transcriptional machinery can appear spontaneously (Conrad et al. 2010, 2009). Cultures of MG1655 *E. coli* grown in minimal glycerol M9 media showed mutations to *rpoC* of the β' subunit of the RNA polymerase holoenzyme in ~80 % sequenced mutants (Conrad et al. 2010).

While the core holoenzyme of RNA polymerase is the most conserved transcriptional machinery, targeting the sigma factors or factors giving the RNA polymerase its DNA specificity provide a greater modularity of control over distinct phenotypes. Using PCR mutagenesis, minor alterations in the DNA recognition motifs of these sigma factors (or their homologs) can greatly affect the RNA polymerase transcriptional kinetics toward the genes under control of the sigma factor. Using this approach the sigma 70 subunit in *E. coli* (normally controls housekeeping genes) was mutated to generate strains with increased ethanol tolerance, simultaneous

sodium dodecyl sulfate and ethanol tolerance, and lycopene overproduction (Alper and Stephanopoulos 2007). Similarly, in yeast the genes encoding the TATA-binding protein, *SPT15*, and its associated factors, *TAF25*, were mutated to yield a strain of yeast with the industrially-relevant profile of high-glucose/ethanol tolerance, with an 70 % improvement in volumetric ethanol production (Alper et al. 2006). To emphasize the validity of targeting transcriptional hubs further, this process was repeated in a strain of *Lactobacillus plantarum*. Using growth/colony size as the phenotypic measurement, the approach of targeting transcriptional machinery was compared directly to mutagenesis by nitrosoguanidine chemical mutagenesis. In this study the investigators showed that targeting transcriptional machinery was able to create a greater diversity of phenotypic profiles than PCR mutagenesis alone. The greater degree of diversity generated by modifying aspects of transcriptional machinery made it possible to more rapidly isolate a strain of *L. plantarum* with an increased tolerance to malic acid.

Ribosomal Engineering

Just as the transcriptional machinery could be mutated to generate altered cellular phenotypic landscapes, the ribosomal machinery can be mutated to alter the global translational profile. A long history of studying the ribosome utilizing inhibitory antibiotics targeting the ribosome has continued as the preferred method for generating ribosomal variants. Termed “ribosome engineering”, variants are isolated by plating resistant cultures on differing concentrations of ribosomal targeting antibiotics. This application has yielded strains of *Streptomyces* capable of increased antibiotics production, increased α-amylase and protease production in *Bacillus subtilis*, and increased tolerance to aromatic compounds in *Pseudomonas putida* (Ochi 2007; Ochi et al. 2004). The varying phenotypes due to mutated ribosomes are postulated to correlate with an activation of the ‘stringent response’ when stress signals of stationary phase lead to increased protein production from select stationary-phase loci such as sporulation, alternative carbon source utilization, and production of secondary metabolites.

MAGE

Multiplex Automated Genome Engineering, or MAGE, has received considerable attention recently for the speed and scope of genetic changes that can be achieved. MAGE represents an extremely high-throughput methodology to perform major alterations across the entire genome. Mediated by homologous recombination, the process utilizes a fairly complex cultivation system featuring a series of growth chambers, an electrophoresis machine, a computer controller and a library of synthetic oligos (Wang et al. 2009). The synthetic oligo library contains the target mutations that are moved into the cell by electrotransformation, and incorporated through homologous recombination. This continuous culture system can run

indefinitely to eventually generate all possible variants represented in the oligo library. In its initial application, MAGE was used to create *E. coli* pools containing over 15 billion genetic variants, targeting 24 separate genes to increase lycopene production 500% over wild-type lycopene producing strains (Wang et al. 2009). Ultimately, MAGE may be a powerful tool to modulate multiple genome targets simultaneously and with complete control. One of the current limiting factors of MAGE is the price of the technology, which currently still proves restrictive to academia, but may represent a viable opportunity for large-scale industrial projects.

TRMR

A variant of allelic replacement, TRMR (pronounced tremor) or trackable multiplex recombineering, has been used to construct libraries of up- and down-regulated variants of ~96 % of the entire *E. coli* genome in less than a single week and for less than $1 per target (Warner et al. 2010). Yet this staggering diversity only represents half of the full potential of this approach. A specific strength of this approach is that all of the generated mutations were completely trackable by the incorporation of DNA barcodes. By hybridizing the each of the generated mutant strains to a DNA microarray, the relative levels of each mutation can be quantitatively measured. TRMR is then capable of rapidly identifying gene interactions that could then be used to structure further optimization projects. It has been pointed out that combining TRMR as an initial coarse-grained investigation of optimization, with the MAGE approach as a fine-grained adjustment could generate genome-wide, highly precise optimization (Tipton and Dueber 2010).

3.4.2 Functional Characterization

3.4.2.1 Whole-Genome [Re] Sequencing

Ultimately, the assumptions of EvoEng boil down to the relationship between sequence information and the resultant phenotype. Accordingly, the most fundamental method of screening relies on investigating the sequence information of the resulting mutants. Many clever selection schemes exist to screen populations by automatically removing undesired populations, but eventually an isolated population must be sequenced to identify genetic changes that may have arisen by natural random mutation. Again, directed evolution can help constrain the final sequencing requirements, but this means possible beneficial mutations inherited in the cellular background may be ignored.

As next-generation sequencing technologies have expanded the feasible limits on data collection and screening have expanded accordingly. Many recent EvoEng explorations have utilized next-gen sequencing to find changes on a genomic scale by comparing whole-genome sequencing of evolved strains to their ancestral strain.

Evolutionary paths have been tracked by uncovering changes through single-nucleotide substitutions, insertions, deletions, and genomic rearrangements (Araya et al. 2010; Atsumi et al. 2010; Charusanti et al. 2010; Kishimoto et al. 2010; Lee and Palsson 2010; Lee et al. 2010; Barrick et al. 2009; Conrad et al. 2009; Gresham et al. 2008; Friedman et al. 2006; Herring et al. 2006; Velicer et al. 2006; Albert et al. 2005). Information gained by whole genome re-sequencing can be a useful source of information to feed into methods like MAGE or TRMR for a rational investigation into recombination adaptation. Further, whole genome sequencing preserves information and data that may be useful to the elucidation of cellular physiology to provide a holisitic systems biology view of cellular function (Conrad et al. 2011; Fong 2009).

Another approach to identifying genetic changes is a reapplication of DNA microarrays for array-based discovery of adaptive mutations (ADAM). In an EvoEng context, Goodzari et al. utilized a selectable marker, linked to a functional mutation (such as an insertion inactivation). Minimally, the ADAM approach requires a library of selectable markers transposed throughout the parental strain's DNA, a mechanism for transferring markers from the parental strain to the evolved strain in such a way that the sequence surrounding the marker replaces the corresponding DNA in the evolved strain, and a method for measuring the frequency of markers throughout the evolved population (Goodarzi et al. 2009). Basically, if the newly evolved strain replaced a beneficial mutation with the parental sequence – a reversion in effect – then this evolved/parental chimeric would show a decrease in fitness. By then hybridizing to separate microarrays the fitness of the evolved strains and the "revertants" could be compared, pinpointing advantageous mutations by a decrease in signal from the "revertant" populations.

3.4.2.2 Additional High-Throughput Data

As related to the EvoEng approach, high-throughput data can be useful as a means of characterizing the phenotypic state of a cell in detail. This provides a means of gaining information to connect genotypic changes to whole-cell phenotypic changes.

Transcriptomics ushered in the advent of system-level quantitation of fundamental cellular components. While gene expression microarrays are the more established method of measuring transcriptomics data, recent studies have shifted more towards using RNA sequencing (RNAseq), especially since RNASeq data has been shown to correlate with microarray hybridization techniques in reproducibility and relative quantification (Alexeyev and Shokolenko 1995). Due to their lower cost, gene expression arrays are still advantageous for revealing the enrichment or depletion of clones as a consequence of selection (Kao 1999); accessing genes which confer a selective advantage or disadvantage.

In addition to transcriptomic data, proteomic and metabolomics data would prove valuable for investigating and understanding cellular function. Both of these data types have improved as technological advances have increased the reliability and scope of measurable data. For instance, a proteome analysis of *Sacchromyces cerevisiae* response to carbon and nitrogen limitation was done using multidimensional

protein identification technology (MudPIT), combined with the labeling of proteins showed an up-regulation of protein in response to glucose limitation that was transcriptionally controlled, while the up-regulation in the presence of nitrogen occurred from regulation of a post-transcriptional nature (Kolkman et al. 2006). An example of a metabolomics study using a coupled detection of electrospray ionization in tandem with mass spectrometry (ESI-MS) was used to identify and measure of up to 84 % the metabolome of *S. cerevisiae* (Højer-Pedersen et al. 2008). In fact, recently a group published a high-throughput metabolomics workflow for investigating yeast in a multi-well format (Ewald et al. 2009).

3.4.3 Synthetic Biology

As a field, synthetic biology allows for better-controlled modification of a biological system. This started with the demonstration of synthetically generated genetic circuits (toggle switch and repressilator) that behaved in a designed fashion (Elowitz and Leibler 2000; Gardner et al. 2000). One application of synthetic biology can be to utilize designed synthetic constructs to sense and modify cellular function by developing novel circuits. For EvoEng, it may become possible to more directly couple COs and EOs using synthetic circuits.

3.4.3.1 SELEX

RNA is an extremely unique molecule in that it has been found in nature to be functional as an informational molecule, but also as a catalyst. Importantly, RNA polymers have the ability to form complex tertiary structures capable of specifically binding to small molecules and performing catalytic functions, such as self-cleavage. In 1990, the laboratories of G.E. Joyce (1989), J.W. Szostak (Ellington and Szostak 1990), and L. Gold (Tuerk and Gold 1990) independently developed a technique which allows the simultaneous screening of more than 10^{15} individual nucleic acid molecules for different functionalities (1–3 below). The selectable evolution of ligands by exponential enrichment (SELEX) is an EvoEng process in itself. It works by expanding a DNA library of oligos by PCR mutagenesis then selecting for variants capable of binding a target molecule. These isolates can be amplified through PCR methods, and the process can be iteratively repeated. For more details the reader is directed to several good reviews on the process (Sinha et al. 2011; Klug and Famulok 1994), or Aptamer base (Cruz-Toledo et al. 2012) (www.aptamer.freebase.com), a database of apatmer experimentation performed to date.

3.4.3.2 Designing Riboregulators

SELEX is an excellent tool to create RNA that binds a target molecule, altering a relatively small amount of sequence to search through a complexity of tertiary structures.

However, as an *in vivo* biosensor the ability to bind target molecules alone does not convey a measurable signal. The *trans*-acting aptamer, or riboregulator, still requires a mechanism to alter its expressional state (Roth and Breaker 2009; Henkin 2008). In short, the RNA needs to have *trans*-binding activity for the target molecule and *cis*-regulatory properties. Much work has been done in the field to determine methods by which to integrate RNA aptamers with a regulatory motif in order to create programmable, *in vivo* biosensors (Lucks et al. 2011; Win and Smolke 2007; Isaacs et al. 2004). Recently, Qi et al., reported a method by which to program non-coding RNAs to suppress the expression of a target mRNA when an upstream aptamer domain was bound to a target molecule – building a post-translational NOR gate (Qi et al. 2012). All these design mechanisms will allow the construction of target specific, tunable, and orthogonal *in vivo* biosensors or regulators.

3.5 Interpreting the Results

Having discussed the application and utility of the natural evolution process and EvoEng it is important to consider the process for utilizing results for two major reasons: (1) To monitor whether or not we are moving in the right direction and (2) to record the evolutionary paths through genotypic variations and their effect at each level of cellular organization.

Tracking results and traits during evolution is important as explained in Sect. 3.4 to ensure that the system and selection pressure is leading to an outcome consistent with desired cellular and engineering objectives. The second consideration provides details of genetic changes and allows them to be correlated to phenotype i.e. collecting the data for each stage of step-wise continuous process of evolution. (Lenski et al. 1998) Monitoring the history of adaptation and using this information to model a particular cellular system enables researchers to make hypotheses for subsequent experiments.

3.5.1 Systems Biology

Systems-level modeling, especially the constraint-based approach, has proven to be a robust method to account for biological complexity while integrating high-throughput data. Using this tool in the cyclic process of evolutionary adaptation is discussed by many authors in past (Valente and Fong 2011; Fong 2009; Sauer 2001). Metabolic modeling, in conjunction with experimental high-throughput data discussed in Sect. 3.2, can be used to analyze potential targets for strain improvements to yield optimized productivity. Modeling simulations can reduce the cost and time needed for the EvoEng process(Fong et al. 2005a, b) by employing algorithmic analyses including: OptKnock (Burgard et al. 2003), objective tilting (Feist et al. 2010), RobustKnock (Tepper and Shlomi 2010) and OptGene (Patil et al. 2005). For example, metabolic flux analysis was applied to help understand the carbon

metabolism of several *E.coli* strains under different growth conditions(Sauer et al. 1999). As for industrial applications, Lewis et al. have discussed methods that have been developed to predict, *in silico*, the growth-coupled desired phenotypes by performing deletions or insertions in the metabolic networks (Lewis et al. 2010). By encompassing data from previous iterations, it is possible to infer evolutionary trajectories as pointers for subsequent targets of adaptation.

Currently, constraint-based modeling approaches have been limited to metabolism and the biomass-assimilating pathways, ignoring portions of cellular physiology. In order to add detail to these models, efforts have been put forth to incorporate transcriptional regulation. (Gianchandani et al. 2006; Chandrasekaran and Price 2010; Thiele et al. 2010) and other cellular processes (Molenaar et al. 2009; Covert et al. 2008). This however, faces a major challenge of extrapolating the altered binding kinetics and motifs of the proteins because of the mutations performed in the regulatory proteins and ultimately affecting the topology of the regulatory network.

High-Throughput Biologically Optimized Search Engineering (HT-BOSE) introduced by Valente and Fong (Valente and Fong 2011), describes a method of leveraging EvoEng in conjunction with Systems biology and Synthetic biology. Systems-level information can provide a foundation to plan and conduct three-part cyclic process of evolutionary engineering. This biologically optimized search engine starts from a "Registry of Seed Designs", similar to the "Registry of Biological Parts" (partsregistry.org). This Seed Design Registry would include phylogenetic information on how designs were evolved, as well as on what subsequent designs originated from them. Fitness scores are assigned to these designing depending upon how far they are from the desired objectives, yet the selection of a seed sequence from the registry should retain evolutionary flexibility. High fitness assays quantifying seed sequences satisfying the EOs, may reflect local maximums. It is therefore, important to switch between the fitness assays and a "supporting assay", where the later is inclined toward the identification of evolutionary flexibility. Finally, the search strategy should maximum biological information by exploring SS at multiple length-scales.

3.6 Patenting Evolutionary Engineering

Throughout collaboration between academia, research, and large-scale industrial sectors there exists a movement away from publishing results in journals to prioritizing the invention protection by filing a patent, particularly in engineering and "applied" sciences (Leimkühler and Meyers 2004, 2005). While the timeline of the patent application process (Fig. 3.3) is extensive, patentability requirements and regulation under European and U.S. patent laws prescribe the exclusivity of invention to provide protection for a span of 20 years (an additional 5 years for pharmaceutical drug molecules). The validity of patent claims are predicated upon three features: (1) Novelty as defined in European Patent Convention (EPC) Article 54 and US regulation 35 U.S.C. § 102; (2) Inventiveness as defined in EPC Article 56 and 35 U.S.C. § 103 (a); (3) Industrial application as defined in EPC Article 57 and 35 U.S.C. § 101.

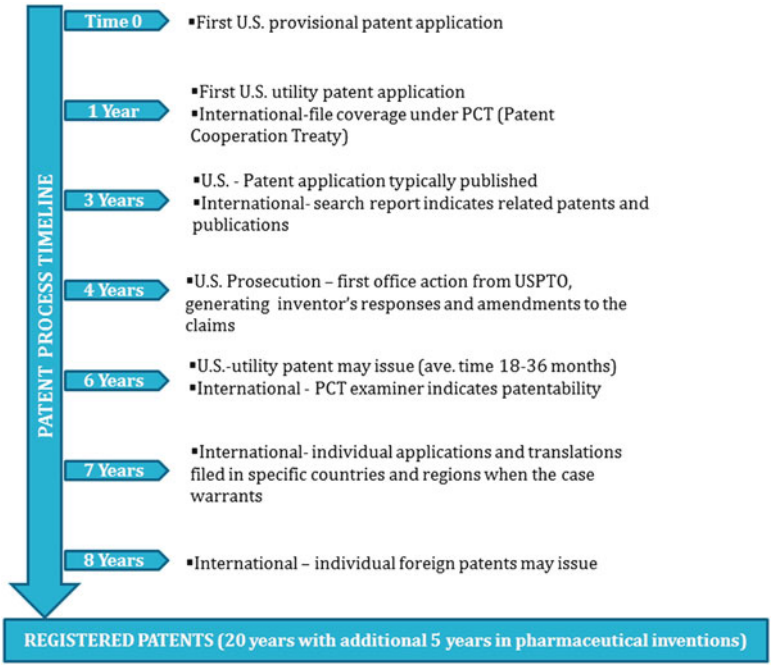


Fig. 3.3 Patent process timeline

The typical patent application consists of the claim itself, a precise description of specifications, and the rationale for protection. In the specific field of EvoEng patents filed to date have claimed produced molecules with known function and utility, state-of-the-art production processes of molecules with known utility, or a selection process for useful traits in an already characterized strain. As explained in Fig. 3.1, the process of EvoEng consists of 3 steps: amplification, diversification and selection. Protection claims may be directed towards the steps in the cycle, specific conditions of the cycle, or improvements to previous patents (e.g. the specific and defined mutation conditions used to obtain variation and/or definite selection with respect to fitness toward an EO). The penultimate example of EvoEng intellectual property is the SELEX technique (WO 91/19813), discussed in Sect. 3.3.1 and its follow-up patents regarding variations and improvements (Leimkühler and Meyers 2004, 2005).

3.7 Conclusion

Recent supportive technological advancements have played an integral role in aiding the application of EvoEng to industrial targets. Rapid and inexpensive whole genome sequencing and other omics analyses have made it feasible to characterize the genotype of the strain with high fidelity and to associate the genotype to the phenotype (Lee et al. 2011). In the effort to reach a $1,000-genome (DeFraccesco

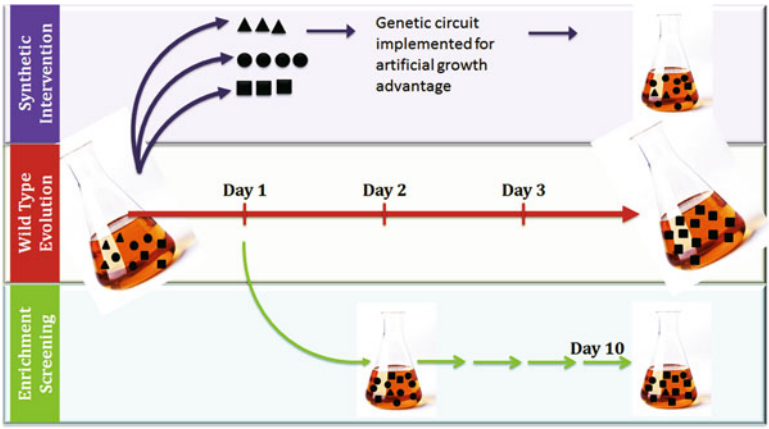


Fig. 3.4 Natural evolutionary process support the growth of square colonies over the period of evolution. Performing enrichment screening to obtain circled colonies might temporarily endows the culture with larger amount of those colonies however, because of squares' evolutionary advantage they will continually outcompete the circled colonies. In contrast to the enrichment screening approach, synthetic intervention or selection can be useful in isolating clones with optimized for both cellular and engineering objectives

2012), there has been exponential progress from Sanger sequencing to fluorophores (Braslavsky et al. 2003), to Nanopore (Church et al. 1998; Kasianowicz et al. 1996) and other next-generation sequencing technologies. This progress has been compared to "Moore's Law" and is predicted to continue in this manner into the immediate future with promises of technologies allowing over five kilobase contiguous reads (Hayden 2012; Clarke et al. 2009). Similarly, the technology for synthesizing DNA improves every day, with synthesis companies currently offering synthesis and delivery of 500 base pair double-stranded DNA oligos in 3–4 business days and for less than 100 US dollars (www.idtdna.com). Concurrent advances in cloning have enable even the smallest laboratories to assemble synthesized oligos up to several kilobases in length in under an hour (Gibson et al. 2009). Laboratory automation, microfluidics and other emergent screening technologies and have enabled the extreme increases in throughput and precision of rate-limiting steps in EvoEng such as serial dilution enrichment and chemostat selection (Grabar et al. 2006; Tyo et al. 2006; Zhou et al. 2006; Sonderegger and Sauer 2003). These impressive technologies open possibilities for whole genome sequencing and editing, in conjugation with massive "-omics" data collection. However, the success of implementing these technologies will remain dependent on our ability to interpret our findings – it is nice to own a fast car, but it does you no good when there is a low speed limit.

Metabolic engineering and forward-engineering cannot be reliably pursued without concrete genotype-phenotype correlations which equivocate with a priori knowledge of the system. Currently our level of knowledge of even the most elucidated model systems proves limiting. Selection for robust phenotypes by evolutionary engineering helps minimize instabilities such as those posed by genetic drift and clonal interference pictured in Fig. 3.4. For example, in order to improve the production of

aromatic compounds, the phosphotransferase system (PTS) of *E. coli* was deleted and spontaneous glucose-utilizing revertants with increased aromatic titers were selected (Flores et al. 1996). Evolving the non-PTS system in E coli presumably preserved more phosphoenolpyruvate, a precursor of the aromatic targets. However, when a non-PTS heterologous system was rationally engineered into the PTS knockout, the improvement in target titers were not observed (Chen et al. 1997). Even with the phenotype successfully capitulated by rational engineering, there is been no optimization for evolutionary stability – a trait absolutely necessary for industrial continuous-culture. To address this pitfall, a hybrid-approach known as "inverse metabolic engineering" has been proposed (Bailey et al. 2002). Employing a rational-approach, targeted phenotypes can be isolated rapidly as seed sequences for a subsequent Evolutionary engineering approach for optimization. This rational "constraint" on SS to display the phenotype significantly reduces the time invested in optimization and identification of a seed sequence, particularly when transferring heterologous traits to a new chassis. Evolutionary engineering, inverse metabolic engineering nor any other bioengineering approach represents a full-proof approach to the design of industrial microbes. Many of the evolutionary engineering techniques presented in this chapter could be conjugated into more powerful hybrid systems capable of improved search functionality. Ultimately however, the future lies in innovating upon current selection procedures, improving the knowledge base of genotype-phenotype correlations, and harnessing advances in technologies.

References

Albert TJ, Dailidiene D, Dailide G, Norton JE, Kalia A, Richmond TA, Molla M, Singh J, Green RD, Berg DE (2005) Mutation discovery in bacterial genomes: metronidazole resistance in helicobacter pylori. Nat Method 2:951–953

Alexeyev MF, Shokolenko IN (1995) Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria. Gene 160:59–62

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9:258–267

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565–1568

Araya CL, Payen C, Dunham MJ, Fields S (2010) Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomic 11:88

Atsumi S, Wu TY, Machado IM, Huang WC, Chen PY, Pellegrini M, Liao JC (2010) Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. Mol Syst Biol 6:449

Atwood KC, Schneider LK, Ryan FJ (1951) Periodic selection in *Escherichia coli*. Proc Natl Acad Sci U S A 37:146–155

Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS (2002) Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnol Bioeng 79:568–579

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247

Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A 100:3960–3964

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657

Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 107:17845–17850

Charusanti P, Conrad TM, Knight EM, Venkataraman K, Fong NL, Xie B, Gao Y, Palsson BO (2010) Genetic basis of growth adaptation of *Escherichia coli* after deletion of pgi, a major metabolic gene. PLoS Genet 6:e1001186

Chen R, Hatzimanikatis V, Yap WM, Postma PW, Bailey JE (1997) Metabolic consequences of phosphotransferase (PTS) mutation in a phenylalanine-producing recombinant *Escherichia coli*. Biotechnol Prog 13:768–775

Church GM, Deamer DW, Branton D, Baldarelli R, Kasianowicz J (1998) Characterization of individual polymer molecules based on monomer-interface interactions. Harvard College EP0815438

Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270

Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO (2009) Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol 10:R118

Conrad TM, Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R, Palsson BO (2010) RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proc Natl Acad Sci U S A 107:20500–20505

Conrad TM, Lewis NE, Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol 7:509

Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. Bioinformatics 24:2044–2050

Cruz-Toledo J, McKeague M, Zhang X, Giamberardino A, McConnell E, Francis T, Derosa MC, Dumontier M (2012) Aptamer base: a collaborative knowledge base to describe aptamers and SELEX experiments. Database J Biol Databases Curation 2012: bas006

Dawson PSS (1985) Continuous cultivation of microorganisms. CRC Crit Rev 2:315–372

de Lorenzo V, Herrero M, Sánchez JM, Timmis KN (1998) Mini-transposons in microbial ecology and environmental biotechnology. FEMS Microbiol Ecol 27:211–224

DeFraccesco L (2012) Life technologies promises $1,000 genome. Nat Biotechnol 14:126

Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592

Drake JW (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci 870:100–107

Dykhuizen D, Hartl DL (1980) Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics 96:801–817

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338

Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. Nature 472:499–503

Ewald JC, Heux S, Zamboni N (2009) High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. Anal Chem 81:3623–3629

Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO (2010) Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metab Eng 12:173–186

Flores N, Xiao J, Berry A, Bolivar F, Valle F (1996) Pathway engineering for the production of aromatic compounds in *Escherichia coli*. Nat Biotechnol 14:620–623

Fong SS (2009) Evolutionary engineering of industrially important microbial phenotypes. Metab Pathw Eng Handb 1:1–15

Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO (2005a) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 91:643–648

Fong SS, Joyce AR, Palsson BO (2005b) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Res 15:1365–1372

Friedman L, Alder JD, Silverman JA (2006) Genetic changes that correlate with reduced susceptibility to daptomycin in staphylococcus aureus. Antimicrob Agents Chemother 50:2137–2145

Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. Evolution 58:245–260

Gabriel A, Dapprich J, Kunkel M, Gresham D, Pratt SC, Dunham MJ (2006) Global mapping of transposon location. PLoS Genet 2:e212

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342

Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. PLoS Comput Biol 2:e101

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Method 6:343–345

Gillespie JH (1984) Molecular evolution over the mutational landscape. Evolution 38:1116–1129

Goodarzi H, Hottes AK, Tavazoie S (2009) Global discovery of adaptive mutations. Nat Method 6:581–583

Grabar TB, Zhou S, Shanmugam KT, Yomano LP, Ingram LO (2006) Methylglyoxal bypass identified as source of chiral contamination in l(+) and d(−)-lactate fermentations by recombinant *Escherichia coli*. Biotechnol Lett 28:1527–1535

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet 4:e1000303

Hardin G (1960) The competitive exclusion principle. Science 131:1292–1297

Hayden EC (2012) Sequencing set to alter clinical landscape. Nature 482(7385):288

Helling RB, Vargas CN, Adams J (1987) Evolution of *Escherichia coli* during growth in a constant environment. Genetics 116:349–358

Henkin TM (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. Genes Dev 22:3383–3390

Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38:1406–1412

Hillesland KL, Stahl DA (2010) Rapid evolution of stability and productivity at the origin of a microbial mutualism. Proc Natl Acad Sci U S A 107:2124–2129

Højer-Pedersen J, Smedsgaard J, Nielsen J (2008) The yeast metabolome addressed by electrospray ionization mass spectrometry: initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry. Metabolomics 4(4):393–405

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22:841–847

Jin YS, Stephanopoulos G (2007) Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. Metab Eng 9:337–347

Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. Gene 82:83–87

Kao CM (1999) Functional Genomic technologies: creating new paradigms for fundamental and applied biology. Biotechnol Prog 15:304–311

Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci U S A 93:13770–13773

Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 128:11–45

Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M, Okubo M, Suzuki S, Mori K, Kashiwagi A, Furusawa C, Ying BW, Yomo T (2010) Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. PLoS Genet 6:e1001164

Klug SJ, Famulok M (1994) All you wanted to know about SELEX. Mol Biol Rep 20:97–107

Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM, Pronk JT, Slijper M, Heck AJ (2006) Proteome analysis of yeast response to various nutrient limitations. Mol Syst Biol 2:2006.0026

Lee DH, Palsson BO (2010) Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. Appl Environ Microbiol 76:4158–4168

Lee HH, Molla MN, Cantor CR, Collins JJ (2010) Bacterial charity work leads to population-wide resistance. Nature 467:82–85

Lee JW, Kim TY, Jang YS, Choi S, Lee SY (2011) Systems metabolic engineering for chemicals and materials. Trends Biotechnol 29:370–378

Leimkühler M, Meyers H (2005) Patenting in evolutionary biotechnology evolutionary methods in Biotechnology. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp 191–209

Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, Gerrish PJ, Schmidt TM (1998) Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? Antonie van Leeuwenhoek 73:35–47

Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, Konig R, Smith RD, Palsson BO (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6:390

Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci U S A 108:8617–8622

Luetz S, Giver L, Lalonde J (2008) Engineered enzymes for chemical production. Biotechnol Bioeng 101:647–653

Luria SE, Delbruck M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511

Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol 5:323

Muller HJ (1932) Some genetic aspects of Sex. Am Nat 66:118–138

Müller HJ (1964) The relation of recombination to mutational advance. Mutat Res 106:2–9

Novick A, Szilard L (1950) Description of the chemostat. Science 112:715–716

Ochi K (2007) From microbial differentiation to ribosome engineering. Biosci Biotechnol Biochem 71:1373–1386

Ochi K, Okamoto S, Tozawa Y, Inaoka T, Hosaka T, Xu J, Kurosawa K (2004) Ribosome engineering and secondary metabolite production. Adv Appl Microbiol 56:155–184

Parekh S, Vinci VA, Strobel RJ (2000) Improvement of microbial strains and fermentation processes. Appl Microbiol Biotechnol 54:287–301

Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinform 6:308

Petri R, Schmidt-Dannert C (2004) Dealing with complexity: evolutionary engineering and genome shuffling. Curr Opin Biotechnol 15:298–304

Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445:383–386

Qi L, Lucks JB, Liu CC, Mutalik VK, Arkin AP (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res 40:5775–5786

Rosenzweig RF, Sharp RR, Treves DS, Adams J (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137:903–917

Roth A, Breaker RR (2009) The structural and functional diversity of metabolite-binding riboswitches. Annu Rev Biochem 78:305–334

Rowlands RT (1982) Industrial fungal genetics and strain selection. In: Smith JE, Berry DR, Kristiansen B (eds) The filamentous fungi", volume 4 of "fungal technology". Edward Arnold (Publishers) Ltd, London, pp 346–372

Rowlands RT (1984) Industrial strain improvement: mutagenesis and random screening procedures. Enzyme Microb Technol 6:3–10

Santos CNS, Stephanopoulos G (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. Curr Opin Chem Biol 12:168–176

Sauer U (2001) Evolutionary engineering of industrially important microbial phenotypes. Adv Biochem Eng Biotechnol 73:129–169

Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, Wuthrich K, Bailey JE (1999) Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. J Bacteriol 181:6679–6688

Schneider D, Lenski RE (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. Res Microbiol 155:319–327

Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528–10532

Sinha J, Topp S, Gallivan JP (2011) From SELEX to cell dual selections for synthetic riboswitches. Methods Enzymol 497:207–220

Sonderegger M, Sauer U (2003) Evolutionary engineering of saccharomyces cerevisiae for anaerobic growth on xylose. Appl Environ Microbiol 69:1990–1998

Sorgeloos P, Van Outryve E, Persoone G, Cattoir-Reynaerts A (1976) New type of turbidostat with intermittent determination of cell density outside the culture vessel. Appl Environ Microbiol 31:327–331

Steipe B (1999) Evolutionary approaches to protein engineering. Curr Top Microbiol Immunol 243:55–86

Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26:536–543

Thiele I, Fleming RM, Bordbar A, Schellenberger J, Palsson BO (2010) Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery. Biophys J 98:2072–2081

Tipton KA, Dueber J (2010) Shaking up genome engineering. Nat Biotechnol 28:812–813

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510

Tyo KE, Zhou H, Stephanopoulos GN (2006) High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. Strain PCC6803. Appl Environ Microbiol 72:3412–3417

Valente AXCN, Fong SS (2011) High-throughput biologically optimized search engineering approach to synthetic biology. eprint arXiv:1103.5490

Velicer GJ, Raddatz G, Keller H, Deiss S, Lanz C, Dinkelacker I, Schuster SC (2006) Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. Proc Natl Acad Sci U S A 103:8107–8112

Vinci VA, Byng G (1999) Strain improvement by non recombinant methods. In: Demain AL, Davies JE (eds) Manual of industrial microbiology and biotechnology, second edition. ASM Press, Washington, DC, pp 103–113

Waddington CH (1942) Canalization of development and the inheritance of acquired characters. Nature 150:563–565

Waddington CH (1959) Evolutionary systems; animal and human. Nature 183:1634–1638

Waddington CH (1960) Experiments in canalising selection. Genet Res 1:140–150

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898

Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol 28:856–862

Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104:14283–14288

Witkin EM (1976) Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*. Bacteriol Rev 40:869–907
Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159
Wright S (1988) Surfaces of selective value revisited. Am Nat 131:115–123
Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. Proc Natl Acad Sci U S A 99:16587–16591
Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP, del Cardayre SB (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415:644–646
Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16:258–261
Zhou S, Shanmugam KT, Yomano LP, Grabar TB, Ingram LO (2006) Fermentation of 12 % (w/v) glucose to 1.2 M lactate by *Escherichia coli* strain SZ194 using mineral salts medium. Biotechnol Lett 28:663–670

Chapter 4
Monitoring Microbial Diversity of Bioreactors Using Metagenomic Approaches

Joshua T. Ellis, Ronald C. Sims, and Charles D. Miller

Abstract With the rapid development of molecular techniques, particularly 'omics' technologies, the field of microbial ecology is growing rapidly. The applications of next generation sequencing have allowed researchers to produce massive amounts of genetic data on individual microbes, providing information about microbial communities and their interactions through *in situ* and *in vitro* measurements. The ability to identify novel microbes, functions, and enzymes, along with developing an understanding of microbial interactions and functions, is necessary for efficient production of useful and high value products in bioreactors. The ability to optimize bioreactors fully and understand microbial interactions and functions within these systems will establish highly efficient industrial processes for the production of bioproducts. This chapter will provide an overview of bioreactors and metagenomic technologies to help the reader understand microbial communities, interactions, and functions in bioreactors.

Keywords Bioproducts • Bioreactors • Metagenomics • Microbial diversity

Abbreviations

BLAST	basic local alignment search tool
COG	clusters of orthologous groups
IMG	integrated microbial genomes
JGI	joint genome institute
mcrA	methyl coenzyme-M reductase
NCBI	national center for biotechnology information

J.T. Ellis • R.C. Sims • C.D. Miller (✉)
Department of Biological Engineering, Utah State University,
4105 Old Main Hill, Logan, UT 84322-4105, USA
e-mail: charles.miller@usu.edu

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_4,

RDP10	ribosomal database project
SAMS	sequence analysis and management system
Y	yield
Y p/N	mass of bioproduct per mass of nitrogen utilized
Y p/O_2	mass of bioproduct per mass of oxygen utilized
Y p/s	mass of bioproduct produced per mass of carbon substrate utilized

4.1 Introduction

Bioreactors are typically considered vessels that produce products in a controlled manner via biological conversion. These systems convert materials (substrates) by influencing metabolic pathways to transform materials into products of interest (Williams 2002). Microbial bioreactors have been used for many years to produce products such as cheese, wine, beer, and bread through traditional fermentation, a process that was studied in depth by the famous microbiologist Louis Pasteur. Current technologies not only produce these products, but also a variety of other products such as industrial solvents (biofuels), biogas, acids, sugars, vitamins, antibiotics, and enzymes for bioconversion processes, as well as other primary and secondary metabolites (Williams 2002; Ullmann 2007).

Several bioreactor designs are used to produce bioproducts, and include, but are not limited to: batch reactors, fed-batch reactors, continuous cultivation reactors, plug flow reactors, recycle bioreactor systems, immobilized cell reactors, biofilm reactors, packed bed reactors, fluidized-bed reactors, and dialysis cultivation reactors (Williams 2002). These reactor types can contain either mixed or pure cultures, and can stimulate heterotrophic and/or phototrophic cellular functions depending on the specific reactor design. Additionally, these reactor schemes can be used to produce products directly, or to harvest biomass or other products for downstream processes. Due to the complex nature of bioreactors, particularly anaerobic digesters, the use of metagenomics is helpful to understand the physiology of such systems.

Anaerobic digestion of organic wastes has been applied to treat wastewater with great success over the last 20 years, and is one of the more widely used strategies in industry (Talbot et al. 2008). Methanogenesis from organic material involves a series of symbiotic relationships between hydrolytic-fermentative bacteria, acetogenic bacteria, and methanogenic archaea (Shin et al. 2008; Talbot et al. 2008). This process releases energy in the form of methane gas, where methanogens metabolize low molecular weight carbon compounds such as CO2, acetate, and methylated compounds through the process of methanogenesis (Talbot et al. 2008; Krober et al. 2009; Zeng et al. 2010). Research in this area has distinguished the complex nature of anaerobic communities, including bioreactors (Talbot et al. 2008), emphasizing the need for further metagenomic research in anaerobic environments. Figure 4.1 is a photograph showing two 3,785 l continuous flow anaerobic sludge blanket reactors.

Metagenomics refers to culture-independent studies of diverse genomes from environmental samples, and in particular microbial consortiums. These include microbial

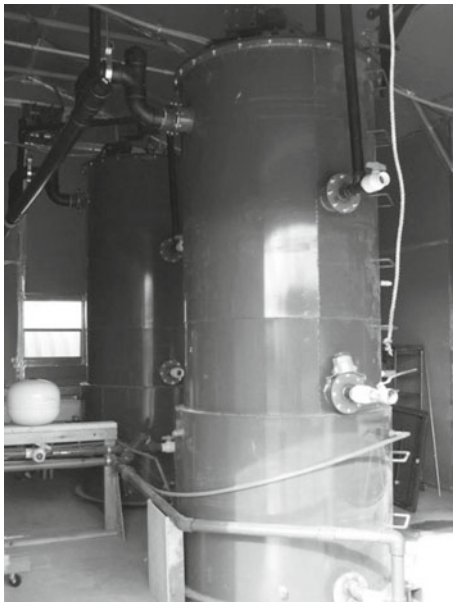


Fig. 4.1 Photograph of a pair of 3,785 l continuous flow anaerobic sludge blanket reactors

consortiums residing in plants, animals, other environmental niches, and even in controlled systems such as bioreactors. The development of next-generation sequencing has engaged large scale sequencing projects that were unimaginable several years ago (Petrosino et al. 2009). Prior to next-generation sequencing, understanding communities only went as far as sampling and microbial isolation; however, current technology has provided the means to understand not only the genetic composition of microbial communities, but also their specific functions. This capability has given rise to new discoveries and an overall greater understanding of biochemical and physiological interactions within microbial communities throughout the biosphere (Committee on Metagenomics 2007; Huang et al. 2009; Morales and Holben 2011).

It is estimated that only 0.1–1.0% of microbes can be cultivated and studied *in vitro*, further supporting the importance of 'omic' technologies (Committee on Metagenomics 2007). Metagenomic approaches eliminate the potential to impose intense selective pressure found in laboratory cultivation, which can introduce altered phenotypes compared to the original ecosystem (Morales and Holben 2011). Metagenomics, particularly pyrosequencing methods, have introduced a user-friendly approach where microbial diversity can be more extensively studied with greater laboratory efficiencies (Petrosino et al. 2009). Novel functions and discovery-based techniques have greatly improved throughout the last decade due to function-based screening. The ability to clone and express fragmented metagenomic DNA from environmental samples *in vitro* has overcome some of the limitations of culture dependent techniques (Straalen and Roelofs 2006).

Metatranscriptomics, metaproteomics, and metabolomics are also widely used in conjunction with metagenomic analyses. These 'omic' technologies can provide a more comprehensive understanding of the types of genes being expressed

by analyzing mRNA (metatranscriptomics), as well as the types of proteins (metaproteomics) and metabolites (metabolomics) being produced within microbial communities (Park et al. 2005; Kuystermans et al. 2007; Park and Lee 2008).

Although all 'omics' technologies produce highly valuable information, this chapter will focus mainly on metagenomics, which serves as a basis for performing and understanding genomics, and has contributed significantly towards providing a more comprehensive understanding of the microbial world.

4.2 Microbial Bioreactors, Bioproducts, Control and Optimization

There are several types of bioreactor systems that can be employed to produce bioproducts. Batch reactors are the most frequently used type of reactor in biotechnological productions, and can be stirred or static (Jenkins et al. 1992). Batch reactors are vessels that facilitate the growth of microorganisms and bioproduct production without supply of additional substrate after inoculation. Substrate is metabolized, and biomass is produced along with products during cultivation. Once products reach maximal concentration, the system is stopped and harvested. Fed-batch reactors use a similar approach; however concentrated substrate is typically added after inoculation, once high cell densities are achieved. Substrate can be added continuously, or step-wise, and the reactor can be stirred or static. Continuous cultivation reactors continuously add fresh media and withdraw broth that includes biomass and bioproducts at a constant volume. This process allows the supplementation of nutrients promoting growth, while inhibiting compounds (metabolites) are removed from the system (Ullmann 2007). Continuous cultivation is preferred to batch reactors because it can be automated, it has lower operating costs, and can operate for longer periods of time (Gómez et al. 2012). Similar systems are also described where cells or nutrients are recycled back into the reactor (Ullmann 2007). Plug flow reactors are characterized by a high length-to-width (aspect) ratio, and can be arranged as one long reactor or many short reactors in a tube bank. These reactors are used in continuous operation, with the source entering, and the product exiting the reactor, while concentrations of products increase in the direction of flow (Jenkins et al. 1992). There are many other types of bioreactor designs, such as sequential batch reactors, immobilized cell reactors (Jenkins et al. 1992; Ullmann 2007; Han et al. 2012), fluidized-bed reactors (Ullmann 2007; Sbizzaro et al. 2012), fixed-bed reactors (Kolios et al. 2000; Furuta et al. 2006), and dialysis cultivation reactors (Nakano et al. 1997; Ullmann 2007).

4.2.1 Bioproducts: Yield and Kinetics

The key to designing an operative bioreactor is to establish control, so that biological reactions are positively influenced through an in-depth understanding of the system.

Control and optimization of reaction yields through stoichiometry and thermodynamics, and reaction kinetics integrated with an optimized reactor design, will strengthen and enhance bioproduct production (Guillard and Tragardh 1999).

Bioproduct production in bioreactors is measured in terms of yield and productivity. Yield (Y) represents the amount of bioproduct normalized to some input parameter, such as carbon substrate, nutrient, or an environmental parameter such as oxygen. Common expressions for yield include mass of bioproduct produced per mass of carbon substrate utilized (Y p/s), mass of bioproduct per mass of nitrogen utilized (Y p/N), or mass of bioproduct per mass of oxygen utilized (Y p/O_2). Kinetics of bioproduct production are generally expressed in terms of bioproduct productivity, with units of bioproduct mass per reactor volume or surface area per time, e.g. mg/L-h or mg/m^2-h. Yield information can be combined with the limiting input parameter, e.g. substrate, to predict bioreactor productivity performance according to the following relationship:

$$Yield \cdot \ Substrate\ feed\ rate = bioproduct\ productivity$$

Therefore, an understanding of the theoretical maximum value for yield determined by community structure and gene expression parameters assists in modeling bioreactor performance, while genetic engineering tools may assist in improving yield values and therefore bioreactor performance. In addition to community structure and gene expression, the effects of mass transfer and environmental conditions within a bioreactor are critical for bioreactor performance. Mass transfer of electron donor (substrate), nutrients, and electron acceptor (e.g. oxygen) to the microorganism(s), and transfer of toxic byproducts away from the microorganism(s), affects the overall rate of biochemical reaction (Shuler and Kargi 2002). Mass transfer is accomplished by mixing intensity, and becomes more limiting as the bioreactor size (scale) is increased. While biochemical reaction rate is generally not limited at laboratory and bench-scale sizes because of the relative instantaneous transfer of reactants to the microorganisms and transfer of by-products away from the microbes, a switch in the limiting regime from reaction rate to mass transfer rate becomes limiting as the scale of the bioreactor is increased (Guillard and Tragardh 1999; Shuler and Kargi 2002). Increasing mass transfer has technical challenges due to the viscosity of the medium and potential cell shear stress for mammalian and some eukaryotic cells. Economic constraints related to exponential increases in the power required per unit volume of a bioreactor to accomplish rapid mass transfer are also a technical challenge (Shuler and Kargi 2002; Ullmann 2007).

Based on the information presented above, the understanding of bioproduct yield and productivity requires in-depth knowledge of community content and gene expression integrated with knowledge of bioreactor design, modeling, operation, and monitoring. Metagenomics can be utilized to discover new organisms, to discover new functions that can be expressed in more proficient systems operated in bioreactors, or to determine the microbial consortium within a bioreactor (i.e., anaerobic digesters). Optimization of bioproduct productivity will benefit from applying genetic engineering principles and tools to increase yield, and from efficient bioreactor design and operation regarding mass transfer and control of environmental conditions.

The integration of new metagenomic data related to microbial diversity, and potential for new bioreactor designs based on ecological principles for mixed cultures of multispecies of microbes, represents the greatest potential for developing new bioprocesses and bioproducts for large-scale implementation and production (Shuler and Kargi 2002).

4.2.2 Downstream Operations

Decreasing cost and increasing productivity generally involve optimizing bioproduct production and downstream processing operations. Bioproduct production has been addressed in the previous section. Innovative technologies in bioreactor design along with integration of both upstream and downstream processes that is focused on bioproduct production, have great potential for producing high value bioproducts more efficiently (Williams 2002). Downstream processes, which involve recovery, concentration, and purification of bioproducts produced in bioreactors, often account for as much as 40–50% of the final product cost. If downstream processes can be designed and operated to address upstream bioproduct and non-product processes used for separation, concentration, and purification, the integrated system will have potential to achieve the lowest cost bioproduct production system. Several factors impacting product recovery include temperature, specific cake resistance of cellular material, fluid viscosity, transmembrane pressure, cross-flow velocity, separation membrane type and surface area, membrane load, and the quality of upstream reactor conditions such as culture quality and homogenization. Downstream processing at small or laboratory scale should be optimized to determine the feasibility of large scale production (Lee and D'Amore 2011).

4.3 Metagenomics: Process Workflow and Bioinformatics

Metagenomic libraries can be screened for the discovery of novel genes and metabolic functions through bioinformatic tools (Mohapatra et al. 2011). Previously, microbial diversity within complex communities was determined by PCR amplifying community DNA, which was subsequently cloned into plasmid vectors. This method requires a substantial amount of cloning and sequencing to determine the microbial diversity. Current high-throughput sequencing technologies bypass cloning steps, and provide large quantities of sequence data. These technologies are commonly referred to as "sequencing by synthesis" (Mardis 2008) and they allow researchers to sequence more than 25 million nucleotides in a 4-h reaction (Binladen et al. 2007). This technology not only allows researchers to analyze large quantities of data efficiently, but also allows them to derive information through bioinformatics on microbial interactions, functions, and metabolism in a more resourceful manner (Morales and Holben 2011). Next-generation sequencing technologies have also

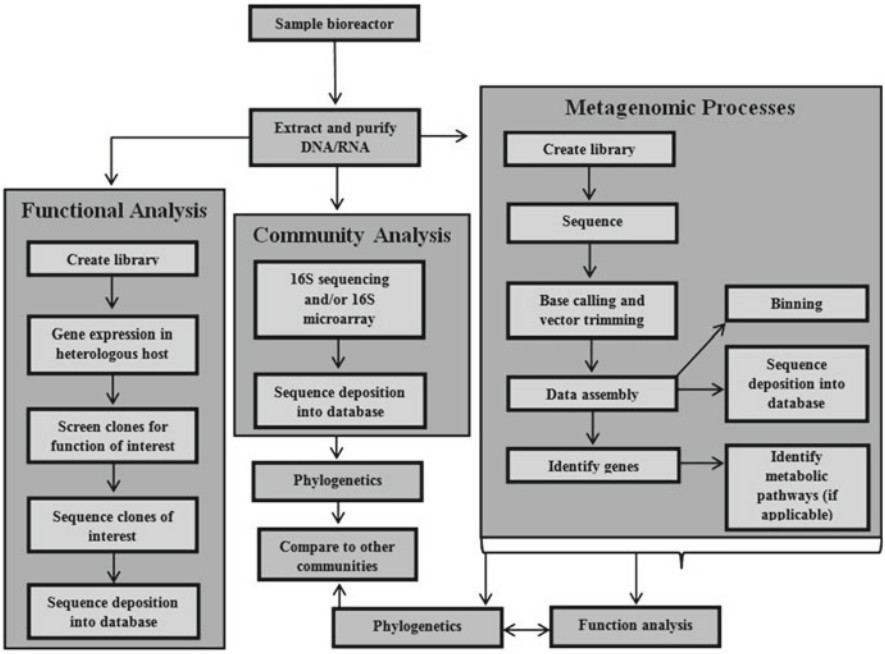


Fig. 4.2 Metagenomic analysis flowchart depicting the procedures involved in functional analysis, community analysis, and metagenomic process (Modified from Committee on Metagenomics 2007; Kunin et al. 2008)

introduced the ability to identify single genetic targets through primer based PCR incorporated with high-throughput technologies (Binladen et al. 2007; Petrosino et al. 2009), and have great potential for identifying microbial diversity and associated functions within bioreactors. The ability to identify single genetic targets using these technologies overrides cloning biases and sequence limitations exhibited by traditional clone libraries (Krober et al. 2009).

Despite advanced sequencing technologies, achieving comprehensive coverage of community samples is difficult due to the unevenness and complexity of these communities. According to Morales and Holben (2011), metagenomic sequence libraries must surpass the size of the metagenome of the community by 100–1,000 fold. This is especially true if minority population sequences are desired at the species level (Morales and Holben 2011). For decades, microbes have been studied as autonomous entities, and by their responses to chemicals in media. This paradigm has limited the thinking of microbiologists (Committee on Metagenomics 2007). Advances in metagenomics will allow researchers full understanding of microbes and their overall behaviors pertaining to relationships (whether mutualistic or antagonistic) and chemical responses within bioreactors (Handelsman 2004). Figure 4.2 provides a schematic overview of the major approaches and workflow used throughout metagenomic analyses.

4.3.1 Sampling and DNA Extraction

The starting material for sequencing environmental samples from bioreactors may contain a variety of organisms such as bacteria, archaea, eukaryotic, and/or viral species. It is important to acquire a homogenous sample from the bioreactor in order to extract DNA representing the consortium of interest. High quality DNA extraction and purification are considered the main bottleneck in metagenomics (Kunin et al. 2008; Wooley et al. 2010). Often times DNA sample collection can be limited by low biomass, small quantities of DNA, contaminated DNA, and inhibitory compounds that can interfere with DNA extraction and purification (Committee on Metagenomics 2007; Kunin et al. 2008). The presence of nucleases in environmental samples is an issue when extracting environmental DNA and RNA. Nucleic acids are initially obtained through cell lysis, either mechanical or enzymatic. Mechanical lysis with phenol is ideal since nucleases are destroyed, however protocols need to be adapted accordingly to represent community DNA accurately (Talbot et al. 2008). Small quantities of whole genomic DNA can be overcome by using whole genome amplification techniques (Angly et al. 2006; Kunin et al. 2008).

4.3.2 Shotgun Metagenomics

Whole-genome shotgun sequencing of metagenome samples is now possible due to advances in computational power and bioinformatics. This sequencing approach randomly shears DNA into many short overlapping sequences that can then be assembled into contigs based on overlapping sequences (Kaiser et al. 2003). This method can provide complete, or near complete, genome sequences. This data gives researchers the ability to determine phylogenetic diversity and metabolic potential within a complex community (Chen and Pachter 2005). Shotgun sequencing techniques along with fosmid libraries have been used to screen for and isolate novel enzymatic functions, both in bioreactors (Jiang et al. 2010), and in environmental samples (Jiang et al. 2012). Using metagenomics and gene cloning strategies together to study and screen for function has been reviewed in detail (Daniel 2004; Handelsman 2004; Streit and Schmitz 2004; Jiang et al. 2010, 2012).

4.3.3 High-Throughput Sequencing

The development of high throughput sequencing has allowed researchers to produce millions of sequences simultaneously. This technology has drastically lowered the cost of sequencing compared to standard dye-terminator methods (Sanger sequencing) while producing massive quantities of data in parallel (Moorthie et al. 2011). Several sequencing-by-synthesis platform technologies, considered high-throughput,

have been derived throughout the last decade. Platforms associated with these sequencing technologies include, but are not limited to: Roche 454 FLX Titanium sequencer (pyrosequencing), Illumina Genome Analyzer II, Applied Biosystems SOLiD sequencer (Mardis 2008), Lynx Therapeutics' Massively Parallel Signature sequencing, and Helioscope single molecule sequencing (http://www.helicosbio.com). The two most frequently used platforms to date are the Roche 454 FLX Titanium sequencer and the Illumina Genome Analyzer II. These two platforms provide a comparable view of microbial community composition with great efficiency when used in side-by-side studies (Luo et al. 2012).

Sequencing by synthesis platforms has been widely used to study metagenomics within environmental and bioreactor samples. Conventional PCR methods using 5′ tagged, or barcoded, primers can be used to generate homologous amplicons from multiple species, which are then utilized in high-throughput sequencing. This data can determine phylogenetics, population genetics, and comparative genomics in fermentative samples. Sequence data can be traced back to its source based on the 5′ tag (or barcode) sequence (Binladen et al. 2007). This method is particularly useful to determine microbial diversity, using conserved genes such as 16S rRNA and methyl coenzyme-M reductase (*mcrA*), or to detect functional genes of interest within a microbial community (Kanokratana et al. 2011; Sahl et al. 2011). Multiple tagged PCR products can be pooled together, each targeting different regions, and used in single sequencing reactions using high-throughput sequencing technologies (Binladen et al. 2007). Pooling tagged amplicons together reduces the cost of this sequencing technology and allows researchers to compile large quantities of data within complex environmental niches.

The workflow to produce next generation sequence libraries is straightforward. Figure 4.3 demonstrates the workflow for Roche 454 pyrosequencing. Sequencing is prepared by ligating specific adaptor oligos to both sides of DNA fragments. A pyrophosphate molecule released on nucleotide incorporation in an impulsion PCR step stimulates further downstream reactions that subsequently produce light from the oxidation of luciferin into oxyluciferin by luciferase (Mashayekhi and Ronaghi 2007; Mardis 2008). Further information on 454 pyrosequencing, as well as on other high-throughput technologies such as the Illumina genome analyzer and the Applied Biosystems SOLiD sequencer, has been reviewed in detail (Mardis 2008).

4.3.4 Bioinformatics for Community Analysis

Bioinformatics is the study and management of biological elements in the environment through computers, mathematics, and statistics (Luscombe et al. 2001). It is used mostly to organize existing information in databases and make it available to users, while allowing them to input data to expand the knowledge base. Bioinformatics has also established a variety of tools for data analysis such as assembly, diversity, alignment, phylogenetics, and function based software (Luscombe et al. 2001; Horner et al. 2010).

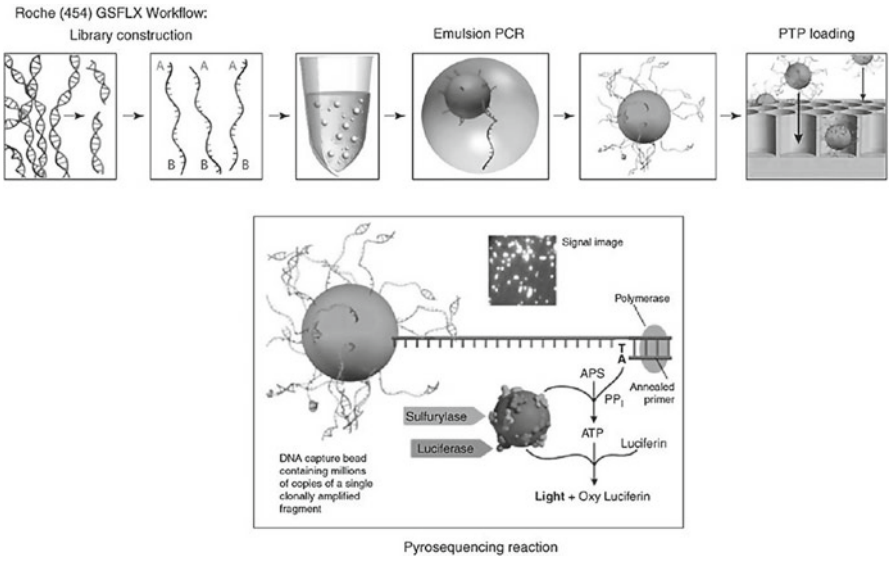


Fig. 4.3 Roche 454 workflow: Library construction ligates 454-specific adapters to DNA fragments, and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing. The beads are loaded into picotiter plate (PTP). The final step is the pyrosequencing reaction that occurs on nucleotide incorporation to report sequencing by synthesis (Figure is reproduced with permission from Mardis 2008)

4.3.4.1 Assembly

Assembly is a procedure that combines sequence reads into overlapping stretches of DNA called contigs, based on overlapping segments that represent consensus regions. They are finally combined fully to generate the whole genome (Kunin et al. 2008). This analysis allows the researcher to analyze open reading frames, operons, functional transcriptional units, and their related promoter elements and transcription factor binding sites. Mobile genetic elements, such as pathogenicity elements, can be evaluated once large fragments of the genome are assembled (Wooley et al. 2010).

There are many software programs available for assembly of metagenomic sequence reads. Genome sequences can be assembled using the Genome Sequencer *De Novo* Assembler Software (Roche Applied Science, Mannheim, Germany) (Schluter et al. 2008), the AMOS comparative assembler (Pop et al. 2004; Kunin et al. 2008), or TGICL software (Xie et al. 2011), for example. Software such as EULER and VELVET are also available and work well with assembling short contigs (Wooley et al. 2010). A variety of other tools for genome assembly have been reviewed in detail (Horner et al. 2010). It is important to inspect assembled data, as all assemblers to date tend to make numerous errors. This limitation can be addressed by analyzing data outputs from multiple assembly programs (Garcia Martin et al. 2006; Kunin et al. 2008). Final assembly products are subsequently submitted to various databases (NCBI and JGI). Considering the microheterogeneity of environmental samples and

mixed culture bioreactors, along with current limitations in metagenomics, complete assembly of metagenomic sequence data for dominant and minor organisms remains a challenge (Kunin et al. 2008; Wooley et al. 2010). However, with advances in technology, metagenomics will provide better coverage of genomes (Wooley et al. 2010), as well as providing improved software to assemble contigs with greater precision and speed.

4.3.4.2 Gene Prediction

Gene prediction is the process of identifying functional coding regions within an assembly. Gene content in bioreactor samples can be characterized by functionally annotating using the Clusters of Orthologous Groups of proteins database (COG) (Tatusov et al. 2001). COGs can be identified and assigned by comparing a BLASTx search against the COG database using specific parameters (Schluter et al. 2008). Homologous regions of other identified genes previously observed can be used to predict function centered on evidence-based methods. Programs or tools like CRITICA (Badger and Olsen 1999), and Orpheus (Frishman et al. 1998; Kunin et al. 2008) use this evidence-based approach. Tools are also available that can overcome frameshifts or chimeras due to poor assembly (Krause et al. 2006).

Functional annotation of assembled contigs is possible using a variety of software programs and databases. The GenDB genome annotation system can be used to assign function (Meyer et al. 2003; Schluter et al. 2008). Prediction of coding sequences can be accomplished using Gismo (Krause et al. 2007) and the Reganor Pipeline (Linke et al. 2006). Functional analysis of short contigs can be performed using BLAST tools, such as: BLAST2x vs. the KEGG database, BLAST2x vs. the COG protein database, BLAST2x vs. the SWISSPROT protein database, and the BLAST2x vs. the NCBI NR (non-redundant) protein database. In addition, the Sequence Analysis and Management System (SAMS) provides bioinformatic analysis of assembled contigs, sequencing libraries, and single reads from metagenome reads (Bekel et al. 2009). This system can analyze and functionally annotate short sequence fragments. Function prediction can be determined by inferring results from multiple bioinformatics databases, such as NCBI, InterPro (Mulder and Apweiler 2007; Mulder et al. 2007; Schluter et al. 2008), and the Integrated Microbial Genomes (IMG) data management and analysis systems (http://img.jgi.doe.gov/).

4.3.4.3 Novel Microbes and Functions

Discovery is an extremely valuable tool associated with metagenomics. Metagenomics has previously been demonstrated to show the presence of diverse and large sets of biocatalysts, such as cellulase and xylanase enzymes for hydrolysis of lignocellulosic material by microbes residing symbiotically within the hindgut paunch of wood-feeding termites (Warnecke et al. 2007). Specific targeted approaches through metagenomics can successfully identify natural and novel gene

products such as biocatalysts, antimicrobials, and sensory molecules (Morales and Holben 2011). Metagenomic approaches have enabled the understanding of entire microbial communities' genomic potential and have thus provided the ability to analyze organization, regulation, and function of unknown organisms and their genomes. Additionally, metatransciptomics can discover novel mRNA within a system, allowing the differentiation between expressed and nonexpressed genes (Shi et al. 2009; Morales and Holben 2011; Simon and Daniel 2011).

Detecting proteins of interest can be accomplished by searching all contigs from the metagenome data set from the bioreactor sample against a protein database containing similar protein sequences. This approach is typically considered sequence-driven screening (Straalen and Roelofs 2006). Discovering a cellulolytic enzyme, for example, can be accomplished by compiling protein sequences associated with these enzymes from particular genera from the NCBI database. Searching all contigs from the metagenome data set against the protein database, using the tBLASTn function, can identify matching reference proteins with high homology and *e*-values (Schluter et al. 2008).

Discovery of glycoside hydrolases from switchgrass-compost communities within a bioreactor was previously demonstrated (Allgaier et al. 2010). 454 pyrosequencing was used to detect 800 genes encoding glycoside hydrolase domains. A synthetic glycoside hydrolase was generated and successfully cloned and expressed in *Escherichia coli* (Allgaier et al. 2010). This example demonstrates the ability to identify full length genes of interest from metagenomic sequence data, and then synthesize and express the function *in vitro* (Allgaier et al. 2010).

Fosmid library construction and sequencing approaches using metagenomic DNA are valuable tools for screening and discovering novel functions. This approach is typically considered function-driven screening (Straalen and Roelofs 2006). Purified and fragmented metagenomic DNA is cloned into vectors and transformed into a suitable host. Once this is achieved, mining for clones and DNA sequences containing novel functions or functions of interest can be completed (Streit and Schmitz 2004). Jiang et al. (2010), used these approaches to screen β-glucosidase genes, which encode enzymes that hydrolyze β-glucosidic linkages in certain oligosaccharides and glycosides to form glucose, or shorter length oligosaccharides. A biogas community was sampled, and metagenomic DNA was used to screen for β-glucosidase activity using esculin hydrate and ferric ammonium citrate to detect transformed colonies containing this activity (Jiang et al. 2010). Detection of Calvin-Benson-Bassham cycle genes, reductive tricarboxylic acid genes, sulfur oxidation genes (Xie et al. 2011), esterase genes (Jiang et al. 2012), methyl coenzyme-M reductase genes (Hallam et al. 2003), polycyclic aromatic hydrocarbon degradation genes (Singleton et al. 2009), as well as a variety of cellulolytic and hemicellulolytic genes (Nacke et al. 2012) has been demonstrated using this approach.

Metatranscriptomics, which studies the mRNA within a community, can connect more thoroughly (while complementing metagenomic data) with the taxonomic make up of a community to its function, or *in situ* activity. By profiling transcripts and relating them to bioreactor conditions, researchers can differentiate between genes that are being expressed or nonexpressed (Simon and Daniel 2011).

Additionally, metatranscriptomic data can also be utilized to predict genes that are functional within an organism (Shi et al. 2009; Chistoserdova 2010; Morales and Holben 2011).

Another approach is to incorporate metaproteomic analysis along with metagenomic data. Proteomics is a powerful tool for analyzing whole cell proteins; however this approach does not provide a comprehensive understanding of all proteins present, due to unknown proteins and incomplete databases. Nevertheless, by establishing genomic information through metagenomics, a foundation for performing further functional genomic studies such as metaproteomics can be created with respect to genomic data (Park and Lee 2008).

A technique known as stable isotope probing can be employed to link specific substrates to certain functional groups while reducing the complexity of community analysis. Introducing a bioreactor with labeled substrate allows rapid detection of microbes involved in the metabolism of interest, such as sugar metabolism, methylotrophy, etc. This process uses substrates labeled with heavy isotopes, allowing for fractionation of the heavy communal DNA from microbes that actively metabolized the labeled substrate. The result is rapid screening of certain functions and offers reduced microbial complexity for metagenomic analysis (Friedrich 2006; Chistoserdova 2010).

4.3.4.4 Microbial Diversity and Phylogenetics

Microbial diversity can be determined by identifying metagenomic reads that contain, for example, 16S rRNA gene sequence homology. The process of determining microbial diversity from metagenomic data is called binning (Kembel et al. 2011), and can be determined by analyzing overall sequence similarity (DeLong et al. 2006), or single markers (Rusch et al. 2007). Stated simply, binning is the process of designating a particular sequence with an organism (Kembel et al. 2011).

Hundreds to thousands of sequences can be identified by conducting a homology search of the ribosomal database project (RDP10) database by means of BLASTn (http://blast.ncbi.nlm.nih.gov) analysis (Krober et al. 2009). Alignments can be filtered based on required *e*-values and alignment lengths to accommodate research needs and validity. Diversity can also be similarly analyzed by conducting high-throughput sequencing on 5′ tagged primer based PCR amplicons using highly conserved genes such as 16S rRNA. Output metagenome data can be filtered according to alignment, *e*-values, and length. Additionally, contigs and singletons assembly can be mapped to a database of known 16S rRNA sequences using the BLASTn algorithm (Xie et al. 2011). Either of these methods can provide large quantities of useful data for microbial phylogenetic analysis. Compiled sequencing data can be aligned using ClustalW (Luo et al. 2012), and phylogenetic trees can be generated using software such as MEGA (Tamura et al. 2007, 2011; Steinberg and Regan 2008; Krober et al. 2009). These phylogenetic trees allow researchers to determine the evolutionary relationship between organisms based on specific and conserved genetic markers. There are many other software programs available to construct and visualize phylogenetic trees, such as CLUSTER and TREEVIEW respectively (Xie et al. 2011).

Phylogenetic trees are constructed using either character-based methods, or distance-based methods. Popular character-based methods include maximum likelihood and maximum parsimony, however alternate methods such as the Bayesian-inference method are available (Huson et al. 2010). Maximum likelihood uses probabilistic models of evolution to produce trees (Bruno et al. 2000). This model may be experimental or parametric, and utilizes a strong statistical foundation for constructing trees (Sleator 2011). In other words, this method determines trees by maximizing the likelihood of producing a particular dataset, under a given evolutionary model (Huson et al. 2010). Maximum parsimony searches for phylogenetic trees based on observable substitutions or mutations among sequences (Sleator et al. 2008; Huson et al. 2010). Distance-based methods such as neighbor-joining use algorithms to compute distance matrices derived from amino acid substitutions, which are related back to the last common ancestor (Sleator 2011). This method has been said to produce the most accurate and consistent trees when given precise distances, or distances with small errors (Bruno et al. 2000). However, recent advances in phylogenetic software have made character- and distance-based methods highly efficient. Programs that produce trees typically use both methods, where a tree is constructed using a distance-based method and is used to evaluate parameters from a character-based method to increase consistency and confidence in phylogenetic trees (Sleator 2011). Bootstrapping is also an important tool to evaluate the strength at the nodes on phylogenetic trees regarding taxonomic relationships (Huson et al. 2010).

4.3.4.5 Diversity Indices

Statistical tools may be used to relate species abundance in a particular ecological niche to environmental conditions. Metagenomics can reveal which microbes live with whom and why, and statistical tools can provide valuable insight into the quality of data acquired (Talbot et al. 2008). Multivariate statistical analyses and diversity indices are widely used to monitor environmental restrictions controlling microbial changes within a community estimated from fingerprints (Bernhard et al. 2005; Mills et al. 2006; Miura et al. 2007; Talbot et al. 2008). Hierarchical clustering methods and ordination methods can be used to compare community fingerprints. Diversity indices, such as the Shannon index, can provide useful information on community adaptation with regard to environmental changes. The Shannon index, and other indices, can be calculated to determine biodiversity in a defined habitat or other bionetwork (Talbot et al. 2008; Wooley et al. 2010), or to examine the structural diversity of a controlled consortium of microbes within bioreactors (Zhang et al. 2010). Rarefaction curves are also desirable to assess and estimate the species coverage from sampling (Wooley et al. 2010).

4.3.4.6 Understanding Microbial Competition and Communication

Microbial community structure, function, and health are typically derived from competition for resources among community members. Mutualistic or antagonistic

relationships between organisms select for diverse survival mechanisms, and thus diverse microbial communities. The genes utilized for microbial competition (both mutualistic and antagonistic) have been previously described using mutant analysis (Handelsman 2004). These genes are hard to determine based on genomics since the phenotypic response is dependent on the nature of the system being analyzed. However, genomics allows researchers to develop hypotheses. These hypotheses from metagenomics, along with other 'omic' technologies, coupled with chemical studies, can provide informative answers on competition and communication molecules, such as antibiotics and other quorum sensing molecules (Handelsman 2004).

High throughput screens can be applied to bioreactor community DNA to identify novel antibiotics and other types of compounds that induce expression of genes controlled under quorum-sensing promoters. Metagenomic DNA can be incorporated in the same cell as the sensor for quorum sensing induction, comprising reporter genes for rapid screening and measurements on expression levels (Handelsman 2004). Metatranscriptomics, the study of all the transcripts within a community, can be employed to determine a profile of gene expression in relation to a condition of interest. This method can allow researchers to determine which genes are turned "on" or "off" during particular phases of the cell cycle and/or reactor conditions (Straalen and Roelofs 2006; Chistoserdova 2010). Additionally, metaproteomics and metabolomics can be used in conjunction to produce valuable data showing which proteins and metabolites are being produced within a system (Park et al. 2005). This information will greatly advance our understanding of the biological activity of these small molecules, regarding both competition and communication, within bioreactor environments.

4.3.4.7 Bioreactors and Metagenomics

Several examples of metagenomic libraries of bioreactor communities have been studied. Gene content regarding fermentative pathways in biogas-producing microbial communities has been previously demonstrated. 454-pyrosequencing technology was used, and COG (clusters of orthologous groups of proteins) categories, derived from single reads, revealed fermentative metabolic pathways. Additionally, the assembly of single reads demonstrated genes involved in sugar and amino acid metabolism from a variety of methanogens. A large number of clostridia species and their cellulolytic functions were also detected. These cellulolytic functions play a key role in the hydrolysis of organic matter, and are the initial step in anaerobic digestion (Schluter et al. 2008). Examples of detecting lignocellulolytic genes from bioreactors using metagenomics along with synthesizing, and expressing these genes, has been demonstrated. This particular bioreactor was established using switchgrass feedstock inoculated with green-waste compost. This system simulated thermophilic composting conditions to facilitate targeted discovery of glycoside hydrolases from a switchgrass adapted microbial community (Allgaier et al. 2010). There are other examples of using metagenomics to isolate and characterize alkali-thermostable lipases from fed-batch reactors

(Meilleur et al. 2009), to understand the microbial diversity and activity of dye degrading microbes from a two stage anoxic-oxic continuous reactor inoculated with dye effluent (Dafale et al. 2010), and to detect novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates (Chen et al. 2007), for example.

4.4 Engineering Microbes Based on 'Omics'

Discovery of novel or more efficient genes or pathways to produce bioproducts from metagenomic sequence data is shaping the future of science and bio-engineering. More efficient enzymes are constantly being discovered through metagenomics, which can be engineered into suitable hosts for highly efficient production of bioproducts. Integrated metagenomic, metatranscriptomic, metaproteomic, and metabolomic tools can be applied to cell and metabolic engineering strategies to identify improved or new phenotypes (Kuystermans et al. 2007). 'Omics' technologies can identify genetic targets that can be engineered, both metabolically and regulatory, to develop 100 % genetically defined strains with high performance suitable for industrial applications (Park and Lee 2008). Engineering microbes based on 'omics' technologies have been reviewed in detail (Park et al. 2005; Kuystermans et al. 2007; Park and Lee 2008).

4.5 Limitations of Metagenomics

Metagenomic technologies have radically influenced the future of microbial fingerprinting; however, several disadvantages are associated with this current technology. Cost and time associated with sequencing and data analysis can be overwhelming (Mohapatra et al. 2011). Sequence databases have been flooded with sequence data that has not been assigned taxonomically or assigned a function. Better bioinformatics tools are needed to analyze larger libraries rapidly and accurately, along with more affordable gene chips to continue rapid advances. Additionally, improvements in heterologous gene expression are also needed for advances in functional analysis (Handelsman 2004). Complete assembly of all microbial genomes within a sample is not currently possible. Sampling typically does not represent all of the species' genomes. Incomplete species information as well as poorly annotated sequence data within databases make it difficult to map reads to their function and/or species (Koonin and Wolf 2008; Wooley et al. 2010). High-throughput platforms currently do not provide adequate coverage within complex systems (Sleator et al. 2008). The potential for discovery and a comprehensive understanding of the microbial world from metagenomic techniques are strictly dependent on advances in methods, instrumentation, software, and accurate databases.

4.6 The Future of Metagenomics

Metagenomics has changed, and will continue to change, the way in which microbiologists approach many problems, while accelerating the rate of discovery, from genes to pathways to organisms (Handelsman 2004). Obtaining a comprehensive understanding through aggressive discovery will answer many questions and solve many issues pertaining to microbiology and bio-based engineering. Metagenomics will continue to stimulate and complement research on microbial genetics and microbial communities, while enhancing our understanding of function and overall physiology. This knowledge will allow us to address concerns regarding environmental and energy issues.

Soon, Gb-scale metagenomic projects will be possible, and sequencing data on microbial communities will eventually approach saturation. With improved resolution, sampling techniques, assembly tools, and increased data sharing and database infrastructure, researchers will have the knowledge and tools needed to incorporate the best functions for biotechnological processes (Sleator et al. 2008; Chistoserdova 2010). Metagenomics will not only aid the scientific community, but will assist in solving problems facing civilization such as health, agriculture, and sustainable energy production (Committee on Metagenomics 2007). The future demands the development of sustainable and domestic energy economies (Ellis et al. 2012), and metagenomics will provide information to engineer highly optimized microbes to address these issues.

4.7 Conclusion

Current technologies for the production of many high value products, such as antibiotics, vitamins, biofuels, enzymes, bioplastics, and various other primary and secondary metabolites, require an extensive knowledge of bioreactor design, modeling, operation, and monitoring capabilities integrated with a knowledge of metagenomics. Control and optimization of reaction yields and reaction kinetics, integrated with an optimized reactor design, will strengthen and enhance bioproduct production. Metagenomic approaches, along with metatranscriptomics, metaproteomics, and metabolomics, provide knowledge concerning the complex relationships that are present in microbial communities throughout the biosphere and throughout complex bioreactor systems. In addition, “omics” technologies are facilitating the discovery of more efficient and novel biological systems for increased bioproduct productivity. The integration of new metagenomic data with new and optimized bioreactor designs represents the greatest potential for developing new and efficient bioprocesses and bioproducts for industrial applications (Shuler and Kargi 2002).

Acknowledgements We thank the Utah State University Bioenergy program, the Utah Science Technology and Research (USTAR) initiative, the Sustainable Waste-to-Bioproducts Engineering Center (SWBEC), and the US Department of Energy (DOE) for financial support.

References

Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons BA, VanderGheynst JS, Hugenholtz P (2010) Targeted discovery of glycoside hydrolases from a switchgrass adapted compost community. PLoS One 5:e8812

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F (2006) The marine viromes of four oceanic regions. Plos Biol 4:2121–2131

Badger JH, Olsen GJ (1999) CRITICA: coding region identification tool invoking comparative analysis. Mol Biol Evol 16:512–524

Bekel T, Henckel K, Kuster H, Meyer F, Runte VM, Neuweger H, Paarmann D, Rupp O, Zakrzewski M, Puehler A, Stoye J, Goesmann A (2009) The sequence analysis and management system – SAMS-2.0: data management and sequence analysis adapted to changing requirements from traditional Sanger sequencing to ultrafast sequencing technologies. J Biotechnol 140:3–12

Bernhard AE, Colbert D, McManus J, Field KG (2005) Microbial community dynamics based on 16S rRNA gene profiles in a pacific northwest estuary and its tributaries. FEMS Microbiol Ecol 52:115–128

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:197–205

Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol 17:189–197

Chen K, Pachter L (2005) Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput Biol 1:106–112

Chen ZW, Liu YY, Wu JF, She Q, Jiang CY, Liu SJ (2007) Novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates. Appl Microbiol Biot 74:688–698

Chistoserdova L (2010) Recent progress and new challenges in metagenomics for biotechnology. Biotechnol Lett 32:1351–1359

Committee on Metagenomics (2007) The new science of metagenomics: revealing the secrets of our microbial planet. The National Academy Press, Washington, DC, pp 40–75

Dafale N, Agrawal L, Kapley A, Meshram S, Purohit H, Wate S (2010) Selection of indicator bacteria based on screening of 16 S rDNA metagenomic library from a two-stage anoxic-oxic bioreactor system degrading azo dyes. Bioresour Technol 101:476–484

Daniel R (2004) The soil metagenome – a rich resource for the discovery of novel natural products. Curr Opin Biotech 15:199–204

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311:496–503

Ellis JT, Hengge NN, Sims RC, Miller CD (2012) Acetone, butanol, and ethanol production from wastewater algae. Bioresour Technol 111:491–495

Friedrich MW (2006) Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. Curr Opin Biotech 17:59–66

Frishman D, Mironov A, Mewes HW, Gelfand M (1998) Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. Nucleic Acids Res 26:2941–2947

Furuta S, Matsuhashi H, Arata K (2006) Biodiesel fuel production with solid amorphous-zirconia catalysis in fixed bed reactor. Biomass Bioenerg 30:870–873

Garcia Martin H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol 24:1263–1269

Gómez E, Máximo MF, Montiel MC, Gómez M, Murcia MD, Ortega S (2012) Continuous tank reactors in series: an improved alternative in the removal of phenolic compounds with immobilized peroxidase. Environ Technol 33:103–111

Guillard F, Tragardh C (1999) Modeling of the performance of industrial bioreactors with a dynamic microenvironmental approach: a critical review. Chem Eng Technol 22:187–195
Hallam SJ, Girguis PR, Preston CM, Richardson PM, DeLong EF (2003) Identification of methyl coenzyme M reductase a (mcrA) genes associated with methane-oxidizing archaea. Appl Environ Microbiol 69:5483–5491
Han W, Wang B, Zhou Y, Wang DX, Wang Y, Yue LR, Li YF, Ren NQ (2012) Fermentative hydrogen production from molasses wastewater in a continuous mixed immobilized sludge reactor. Bioresour Technol 110:219–223
Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685
Horner DS, Pavesi G, Castrignano T, De Meo PD, Liuni S, Sammeth M, Picardi E, Pesole G (2010) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Br Bioinform 11:181–197
Huang Y, Gilna P, Li WZ (2009) Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics 25:1338–1340
Huson DH, Rupp R, Scornavacca C (2010) Phylogenetic networks: concepts, algorithms and applications. Cambridge University Press, Cambridge/New York, pp 23–30
Jenkins RO, Leach CK, Mijnbeek G (1992) Bioreactor design and product yield. Butterworth-Heinemann, Oxford/Boston, pp 71–140
Jiang C, Wu LL, Zhao GC, Shen PH, Jin K, Hao ZY, Li SX, Ma GF, Luo FF, Hu GQ, Kang WL, Qin XM, Bi YL, Tang XL, Wu B (2010) Identification and characterization of a novel fumarase gene by metagenome expression cloning from marine microorganisms. Microb Cell Fact 9:91
Jiang X, Xu X, Huo Y, Wu Y, Zhu X, Zhang X, Wu M (2012) Identification and characterization of novel esterases from a deep-sea sediment metagenome. Arch Microbiol 194:207–214
Kaiser O, Bartels D, Bekel T, Goesmann A, Kespohl S, Puhler A, Meyer F (2003) Whole genome shotgun sequencing guided-by bioinformatics pipelines – an optimized approach for an established technique. J Biotechnol 106:121–133
Kanokratana P, Uengwetwanit T, Rattanachomsri U, Bunterngsook B, Nimchua T, Tangphatsornruang S, Plengvidhya V, Champreda V, Eurwilaichitr L (2011) Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. Microb Ecol 61:518–528
Kembel SW, Eisen JA, Pollard KS, Green JL (2011) The phylogenetic diversity of metagenomes. PLoS One 6:e23214
Kolios G, Frauhammer J, Eigenberger G (2000) Autothermal fixed-bed reactor concepts. Chem Eng Sci 55:5945–5967
Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688–6719
Krause L, Diaz NN, Bartels D, Edwards RA, Puhler A, Rohwer F, Meyer F, Stoye J (2006) Finding novel genes in bacterial communities isolated from the environment. Bioinformatics 22:E281–E289
Krause L, McHardy AC, Nattkemper TW, Puhler A, Stoye J, Meyer F (2007) GISMO – gene identification using a support vector machine for ORF classification. Nucleic Acids Res 35:540–549
Krober M, Bekel T, Diaz NN, Goesmann A, Jaenicke S, Krause L, Miller D, Runte KJ, Viehover P, Puhler A, Schluter A (2009) Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. J Biotechnol 142:38–49
Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008) A bioinformatician's guide to metagenomics. Microbiol Mol Biol Rev 72:557–578
Kuystermans D, Krampe B, Swiderek H, Al-Rubeai M (2007) Using cell engineering and omic tools for the improvement of cell culture processes. Cytotechnology 53:3–22
Lee T, D'Amore T (2011) Membrane separation theoretical and applicable considerations for optimum industrial bioprocessing. J Bioprocess Biotechques 01. doi:10.4172/2155-9821.1000101e
Linke B, McHardy AC, Neuweger H, Krause L, Meyer F (2006) REGANOR: a gene prediction server for prokaryotic genomes and a database of high quality gene predictions for prokaryotes. Appl Bioinformatics 5:193–198

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparisons of illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7:e30087

Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. Method Inf Med 40:346–358

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141

Mashayekhi F, Ronaghi M (2007) Analysis of read length limiting factors in pyrosequencing chemistry. Anal Biochem 363:275–287

Meilleur C, Hupe JF, Juteau P, Shareck F (2009) Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. J Ind Microbiol Biotechnol 36:853–861

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Puhler A (2003) GenDB – an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31:2187–2195

Mills DK, Entry JA, Voss JD, Gillevet PM, Mathee K (2006) An assessment of the hypervariable domains of the 16 S rRNA genes for their value in determining microbial community diversity: the paradox of traditional ecological indices. FEMS Microbiol Ecol 57:496–503

Miura Y, Hiraiwa MN, Ito T, Itonaga T, Watanabe Y, Okabe S (2007) Bacterial community structures in MBRs treating municipal wastewater: relationship between community stability and reactor performance. Water Res 41:627–637

Mohapatra BR, Gould WD, Dinardo O, Koren DW (2011) Tracking the prokaryotic diversity in acid mine drainage-contaminated environments: a review of molecular methods. Miner Eng 24:709–718

Moorthie S, Mattocks CJ, Wright CF (2011) Review of massively parallel DNA sequencing technologies. HUGO J 5:1–12

Morales SE, Holben WE (2011) Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? FEMS Microbiol Ecol 75:2–16

Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Method Mol Biol 396:59–70

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJ, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2007) New developments in the InterPro database. Nucleic Acids Res 35:D224–D228

Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R (2012) Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett 34:663–675

Nakano K, Rischke M, Sato S, Markl H (1997) Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. Appl Microbiol Biot 48:597–601

Park JH, Lee SY (2008) Towards systems metabolic engineering of microorganisms for amino acid production. Curr Opin Biotech 19:454–460

Park SJ, Lee SY, Cho J, Kim TY, Lee JW, Park JH, Han MJ (2005) Global physiological understanding and metabolic engineering of microorganisms based on omics studies. Appl Microbiol Biot 68:567–579

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. Clin Chem 55:856–866

Pop M, Phillippy A, Delcher AL, Salzberg SL (2004) Comparative genome assembly. Brief Bioinform 5:237–248

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe

J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical pacific. Plos Biol 5:e77

Sahl JW, Gary MO, Harris JK, Spear JR (2011) A comparative molecular analysis of water-filled limestone sinkholes in north-eastern Mexico. Environ Microbiol 13:226–240

Sbizzaro D, Bota GB, Borri RB, Teran FJC (2012) Electrocoagulation/flotation followed by fluidized bed anaerobic reactor applied to tannery effluent treatment. Desalin Water Treat 37:359–363

Schluter A, Bekel T, Diaz NN, Dondrup M, Eichenlaub R, Gartemann KH, Krahn I, Krause L, Kromeke H, Kruse O, Mussgnug JH, Neuweger H, Niehaus K, Puhler A, Runte KJ, Szczepanowski R, Tauch A, Tilker A, Viehover P, Goesmann A (2008) The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. J Biotechnol 136:77–90

Shi Y, Tyson GW, DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature 459:266–269

Shin SG, Lee CS, Hwang K, Ahn JH, Hwang S (2008) Use of order-specific primers to investigate the methanogenic diversity in acetate enrichment system. J Ind Microbiol Biotechnol 35:1345–1352

Shuler ML, Kargi F (2002) Bioprocess engineering: basic concepts. Prentice Hall, Upper Saddle River, pp 1–576

Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. Appl Environ Microbiol 77:1153–1161

Singleton DR, Ramirez LG, Aitken MD (2009) Characterization of a polycyclic aromatic hydrocarbon degradation gene cluster in a phenanthrene-degrading acidovorax strain. Appl Environ Microbiol 75:2613–2620

Sleator RD (2011) Phylogenetics. Arch Microbiol 193:235–239

Sleator RD, Shortall C, Hill C (2008) Metagenomics. Lett Appl Microbiol 47:361–366

Steinberg LM, Regan JM (2008) Phylogenetic comparison of the methanogenic communities from an acidic, oligotrophic fen and an anaerobic digester treating municipal wastewater sludge. Appl Environ Microbiol 74:6663–6671

Straalen NMV, Roelofs D (2006) An introduction to ecological genomics. Oxford University Press, Oxford/New York, pp 11–146

Streit WR, Schmitz RA (2004) Metagenomics – the key to the uncultured microbes. Curr Opin Microbiol 7:492–498

Talbot G, Topp E, Palin MF, Masse DI (2008) Evaluation of molecular methods used for establishing the interactions and functions of microorganisms in anaerobic bioreactors. Water Res 42:513–537

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22–28

Ullmann F (2007) Ullmann's biotechnology and biochemical engineering. Wiley-VCH, Weinheim, pp 636–645

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG,

Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450:560–565

Williams JA (2002) Keys to bioreactor selections. Chem Eng Prog 98:34–41

Wooley JC, Godzik A, Friedberg I (2010) A primer on metagenomics. PLoS Comput Biol 6:e1000667

Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, Huang G, Li Y, Yan Q, Wu S, Wang X, Chen S, He G, Xiao X, Xu A (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J 5:414–426

Zeng SJ, Yuan XZ, Shi XS, Qiu YL (2010) Effect of inoculum/substrate ratio on methane yield and orthophosphate release from anaerobic digestion of Microcystis spp. J Hazard Mater 178:89–93

Zhang B, Sun BS, Ji M, Liu HN, Liu XH (2010) Quantification and comparison of ammonia-oxidizing bacterial communities in MBRs treating various types of wastewater. Bioresour Technol 101:3054–3059

Chapter 5
Synthetic Biology Triggers New Era of Antibiotics Development

Jianfeng Wang, Zhiqiang Xiong, Hailin Meng, Yiguang Wang, and Yong Wang

Abstract As a discipline to design and construct organisms with desired properties, synthetic biology has generated rapid progresses in the last decade. Combined synthetic biology with the traditional process, a new universal workflow for drug development has been becoming more and more attractive. The new methodology exhibits more efficient and inexpensive comparing to traditional methods in every aspect, such as new compounds discovery & screening, process design & drug manufacturing. This article reviews the application of synthetic biology in antibiotics development, including new drug discovery and screening, combinatorial biosynthesis to generate more analogues and heterologous expression of biosynthetic gene clusters with systematic engineering the recombinant microbial systems for large scale production.

Keywords Antibiotics screening • Environmental DNA • Heterologous biosynthesis • Synthetic biology • Systematic engineering

J. Wang. Z. Xiong • H. Meng • Y. Wang (✉)
Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China
e-mail: yongwang@sibs.ac.cn

Y. Wang
Key Laboratory of Biotechnology of Antibiotics, Ministry of Health, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_5,

Abbreviations

6dEB	6-deoxy-erythronolide B
ACP	acetyl carrier protein
AMPs	antimicrobial peptides
CHO	Chinese hamster overy
DMAPP	dimethylallyl pyrophosphate
DXP	1-deoxylulose 5-phosphate
eDNA	environmental DNA
FSEOF	flux scanning based on enforced objective flux
IPP	isopentenyl pyrophosphate
LK model	leu/lys model
MEP	methylerythritol-phospate
MOMA	minimization of metabolic adjustment
MRSA	methicillin-resistant *Staphylococcus aureus*
MVA	mevalonate
PSP	promoter strength predictive
ROOM	regulatory on/off minimization
TAR	transformation-associated recombination

5.1 Introduction

Since the introduction of penicillin in therapeutic use in 1940s, antibiotics have been chosen as the most effective option for anti-microbial agents. However, due to the rising and dispersing of antibiotic-resistance pathogens, such as the methicillin-resistant *Staphylococcus aureus* (MRSA) and newly emerging gram-negative multidrug-resistant NMD-1 (Devasahayam et al. 2010; Kumarasamy et al. 2010), its wide application was strongly restricted in the past decades. Meanwhile, the pace of discovering novel antibiotics with new scaffold was dramatically slowing down, because of the repetitive acquisition of known compounds by conventional bioactivity-based screening approaches, which were always time-consuming and laborious. Nevertheless, previous inspiring studies have demonstrated that the original sources still reserved amazing synthetic capacities for new compounds, just waiting to be unveiled by efficient screening strategies (Li and Vederas 2009). In recent years, renaissance of antibiotics discovery was accelerated, mainly represented by outstanding achievements in cryptic gene cluster mining (Wilkinson and Micklefield 2007) and marine natural product screening (Blunt et al. 2010). As antibiotics commonly generated during the complex secondary metabolism by microorganism (e.g. *Actinomycete*, fungi and *Psedomonas*) for some unknown purposes, microorganisms often produced antibiotics in very low quantities under natural conditions, which always made its supply costly or even unavailable to the demand.

To create new artificial organisms harboring most desirable properties, synthetic biology is to integrate different specific parts into a biological background by rational designing and engineering endeavor in a standardized pattern (Agapakis and Silver 2009; Na et al. 2010; Neumann and Neumann-Staubitz 2010). Of course, a rich subassembly wrapper containing various biological parts and well-equipped toolboxes to carry out our devises are two indispensable prerequisites. According to this concept, antibiotics was generally synthesized by iteratively assembling universal building blocks, such as acetyl-CoA, malonyl-CoA and methylmalonyl-CoA in polyketides, through stepwise condensation reactions catalyzed by highly organized enzymatic complex. Generally, the related coding genes clustered together on the genome in modularity, and each catalytic domain has certain substrate or product flexibility, which facilitates interchange of domains between different sources or pathways to generate a large natural product derived compound library for antibiotics screening. Other important technological progress, such as DNA synthesis (Bugl et al. 2007), DNA manipulation technology for pathway/whole genome assembly (Isaacs et al. 2011; Shao and Zhao 2009; Wingler and Cornish 2011; Zhang et al. 1998), and *in silico* microbial system simulation (Park et al. 2007; Price et al. 2003; Weber et al. 2009; Yadav et al. 2009), has led to encouraging achievements in production of fine chemicals (Keasling 2010), bio-fuels (Clomburg and Gonzalez 2010) and pharmaceuticals (Medema et al. 2011; Neumann and Neumann-Staubitz 2010). Moreover, the yield and productivity often could increase hundreds fold after multiple cycles of strains improvement. All these achievements have shown us a good promise of bio-industrial manufacturing as an alternative to petrochemical-based production or natural extraction. This review outlines the recent progresses of novel antibiotics screening and development in the context of synthetic biology, including synthetic biology-based therapy screening, rational design of antimicrobial antibiotics based on synthetic biological parts libraries, heterologous expression of antibiotics biosynthetic gene cluster for weakening silent gene clusters and environmental DNA (eDNA) and synthetic industrial microbe construction for large-scale production of antibiotics.

5.2 Synthetic Biology Based Therapy Screening

Validity of traditional bioactivity-based drug discovery methods were attenuated fast at the end of last century, due to the repeated acquisition of known chemical entities and great rise of multidrug-resistant strains. Target-based approaches, which always incorporated a potential drug target into an elaborate genetic circuit, were used for high-throughput and efficient new drug screenings. Artificial genetic circuit, its dynamic state modulated by the kinetic interaction between the specific small molecular and its receptor protein, leading to open or shut off the specific gene expression, was devised to control a physiological property in organisms (Haseltine and Arnold 2007; Sprinzak and Elowitz 2005). Several antibiotics-based

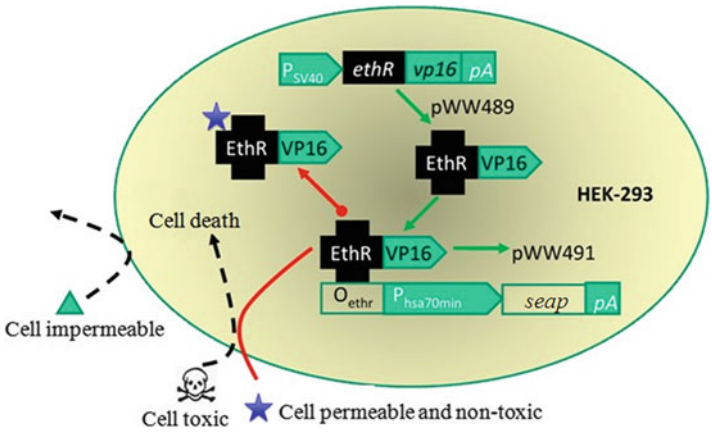


Fig. 5.1 Artificial genetic circuit designed for drug screening

gene regulation circuits were established for the controllable expression of transgenes in mammalian cells, such as streptogramin, tetracycline, and macrolides (Aubel et al. 2001; Fussenegger et al. 2000; Gossen and Bujard 1992; Weber et al. 2002). All of these circuits consist of two chimeric components. An artificial transactivator/repressor, which was a fusion protein containing an antibiotics binding domain (designated as A domain) and a transactivation/repression domain (designated as B domain), was always constitutively expressed from an independent plasmid. A hybrid responsive promoter containing two DNA regions responding to A and B domain, respectively, was engineered to control the expression of a reporter gene. Interaction between antibiotics and its specific binding protein thus can turn on or shut off the expression of reporter gene. Replacing antibiotics-binding protein with different drug target proteins in some diseases, these systems can be quickly engineered into synthetic drug screening platforms.

Ethionamide, as the current last line of defense in curing tuberculosis, plays its function through a Baeyer-Villiger oxygenation reaction to convert prodrug into active form. The enzyme catalyzed reaction is EthA, whose expression can be repressed by EthR through binding to *ethA* promoter. Therefore, the compounds, which can obstruct the binding of EthR and *ethA* promoter, could increase the antituberculosis efficacy of Ethionamide. The EthR-based gene circuit shown as Fig. 5.1, was thus designed by Weber et al. to screen the target potential drug candidates (Weber et al. 2008). EthR and VP16 transactivation domains were constitutively expressed in a fusion protein with no metabolic burden to the host human cells in one plasmid, and another plasmid contains a reporter gene SEAP constructed under the control of a chimeric promoter including the EthR-specific operator (O_{ethR}) and 5' of a minimal *Drosophila* heat shock protein 70 promoter. When exposing recombinant mammalian cells harboring this system to different compounds, EthR-binding compounds identified as drug candidates leading to obstruct the interaction of chimeric promoter will be easily screened according to the inhibition of SEAP expression.

Gonzalez-Nicolini et al. have developed a CHO (Chinese hamster overy)-$p27^{Kip1}$based synthetic evaluation model to imitate the development of cancer, and applied this model for high-throughput drug discovery processes (Gonzalez-Nicolini et al. 2004). $P27^{Kip1}$, a cyclin-dependent kinase inhibitor inducing mammalian cell-growth arrest, was placed under the control of a tetracycline responsive TET_{OFF} circuit consisting of regulation protein tTA and its target promoter. In the presence of tetracycline, circuit was OFF and synthetic mammalian model can grow as wide type. When tetracycline was removed, $p27^{Kip1}$ switched to express and induced a sustained G1-specific growth arrest. However, genetic instability of transgenic $p27^{Kip1}$ system caused its spontaneous losing and resulted in that a specific fraction of arrested cell would re-enter the cell cycle in the absent of tetracycline, which can precisely stimulate the development of cancer. Moreover, comparative proof-of-concept studies using licensed anticancer drug, such as 5-fluorouracil, doxorubicin, and etoposide, were tested to evaluate the validity of this model. Results showed that this CHO-$p27^{Kip1}$-based drug screening model was very efficient for high-throughput anticancer drug screening (Gonzalez-Nicolini et al. 2004).

Prokaryotic antibiotics response genetic circuits occurred in nature enable to detect antibiotics more sensitive than current Swiss and EC threshold values up to 2 magnitude orders, such as streptogramin. Novel antibiotic belonging to the same family of streptogramin can be discovered based on engineered these native circuits into a synthetic biosensor and incorporated it into mammalian cells. Aubel et al. established a CHO cell-based system using streptogramin interaction with transcriptional regulator to modulate reporter gene expression (Aubel et al. 2001). This system can be performed in a high-throughput 96-well plate assay procedure and effective to detect pristinamycin compounds. Weber et al. developed this approach into a class-specific antibiotics detection using broad-spectrum sensing proteins suitable for the classes of macrolide, tetracycline and streptogramin (Weber et al. 2005).

Except the target based genetic circuit screening approaches, Wither et al. have used the previously engineered synthetic *E. coli* strain capable of producing high level of terpene precursors as a host for expression of genomic DNA clones from the isoprene-producing bacterium *Bacillus subtilis*. The clones with a retrieved growth characterization caused by converting the excessive isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) precursors to non-toxic terpene compounds were identified as candidates coding for enzymes responsible for biosynthesis of novel terpene backbones (Withers et al. 2007). This approach provided an effective synthetic model for mining the biosynthetic capacity of cryptic genes.

5.3 Rational Design of Antimicrobial Natural Products

The modular and elegant programming characteristic of antibiotics synthetic process has provided us varieties of functional elements gathered from natural product biosynthetic pathways. Based on type of catalyzed reaction or specificity of substrate and product, all natural and artificial modified elements can be categorized in *synthetic*

biology space defined by Bromley et al. (2008), which endows us with an accessible part list to construct synthetic pathway or organism with lots of choices. Combinatorial biosynthesis originated in 1990s has facilitated us to highly diversify natural compounds by assembling hybrid pathway in random combinations. Its power for generating novel compounds has already been reviewed in numerous literatures (Baltz 2006; Rodriguez and McDaniel 2001; Weissman and Leadlay 2005). Compared to the combinatorial chemistry, combinatorial biosynthesis can incorporate alterations in every step during a whole biosynthetic process by core or tailoring enzyme manipulation. Recently, great efforts on novel catalytic elements exploration by DNA mining or protein engineering, systematic specification and clustering analysis of known part list have achieved. In addition, more directive information, such as inter-communication between different modules and protein structure elucidation of catalytic domains, have endowed combinatorial biosynthesis more rational devise (Prather and Martin 2008; Wilkinson and Micklefield 2007). Moreover, these artificial assembly pathways located in microbes easy to perform genetic modifications can be further engineered to achieve industrial production of the specific compounds, which were difficult to chemistry methods.

5.3.1 Novel Cationic Antimicrobial Peptides Devise

A new class of broad-spectrum antibiotics, cationic antimicrobial peptides (AMPs), which the antimicrobial mechanism of action is directly disrupted microorganism's membrane and cellular processes (Findlay et al. 2010; Powers and Hancock 2003), are potential alternative to combat with multidrug-resistant strains. The amino acids composition, number and sequence of an AMPs determine its antimicrobial efficacy and cytotoxicity in clinical use. Thus, numerous studies were carried out by amino acid replacement or rearrangement to generate improved AMPs (Park et al. 2003), such as commonly known leu/lys (Beven et al. 2003) or gly/leu (Sousa et al. 2009) rich peptides. The leu/lys model (referred as LK model) has strong capacity to form helix structure and was proved to be an excellent model for peptide development (Beven et al. 2003). Due to the important role of tryptophan for strengthening the interaction of peptide and target membrane, Kang et al. (2009) incorporated it into LK model peptide at the critical amphipathic interface between its hydrophobic and hydrophilic domains to develop more potent peptides. Different variants conformed to the formula of $L_lK_mW_2$ containing a core sequence of WLLKW or WLKKW were chemically synthesized and activity-structure analysis. The results exhibited that several variants have a good potential as antimicrobial drug candidates. Loose et al. developed a linguistic model to create novel peptides with limited sequence homology to natural scaffold (Loose et al. 2006). They treated the repeated amino acid sequence motifs in natural AMPs as a formal language. A set of regular grammars was built to describe this language which led to generate varieties of novel peptides. This method was identified as an effective way to search novel peptide with limited homology to natural peptide. Cloning and recombinant expression of

AMPs genes in microbes were also developed and reviewed, recently (Zorko and Jerala 2010). Theoretically, simplicity of AMPs biosynthetic process made heterologous production of large quantities of synthetic AMPs more convenient than polyketide, which can provide a rapid approach to screen the synthetic AMPs library and largely facilitate novel AMPs devise.

5.3.2 *Structure Diversification of Polyketides and Nonribosomal Peptides*

The magic and complexity of polyketides and nonribosomal peptides biosynthesis provide a tremendous space for combinatorial biosynthesis, which exhibit its power in generation of natural product derivatives, especially with the aid of synthetic biology. In rational reprogramming natural biosynthetic capacity, by engineered a native or de novo design of an assembly line can highly diversify the known compounds library (Baltz 2006). Earlier studies mainly using mix-and-match methods including domain swap and deletion/insertion always coupled with precursor supplement have been reviewed (Floss 2006; Zhang and Tang 2008). These methods are convenient to exert small modification on native structure, e.g., decoration on macrolide cyclic ring. However, traditional cloning technique with restriction enzyme was unavailable to perform genetic modification due to the lack of suitable enzyme restriction site. In order to investigate combinatorial constructs as more as possible, Menzella et al. (Menzella et al. 2005) developed a generic approach to design synthetic modules which can easily achieve various kinds of combinations with the repeated set of flanking restriction sites. 14 modules from eight polyketide clusters were synthesized and 154 bi-modular combinations were assembled, half of which mediate biosynthesis of triketide in *E. coli*, and demonstrated that these modules has a relative broad substrate specificity. A database of nonribosomal peptides (called Norine) with comprehensive information about monomer and biological activity has been developed (Caboche et al. 2008, 2009). This demonstrated a rule of certain monomer specificity related to a class of biological activities by analysis of monomer composition of 1,000 peptides. This was very instructive for amino acid residual replacement in native peptides. Inactivation of native acetyl carrier protein (ACP) domain and complementation with exogenous one in native producer or heterologous host, various unnatural peptides such as vancomycin and balhimyicn were generated which always identified with improved efficacy (Caboche et al. 2010).

Besides the above modifications on the scaffold, post-modifications, such as glycosylated, halogenated and methylated substituents in natural products, play an important role in their biological properties and can create new substituted patterns for combinatorial biosynthesis, which identified as a powerful tool to generate novel analogs (Olano et al. 2010; Rix et al. 2002). Sugar substituents on the core aglycone of natural product are important determinants for their bioactivity. Glycosyltransferase gene toolbox containing 70 bacterial glycosyltransferases and summary of their function, specificity and structure has reviewed (Erb et al. 2009). Borisova et al.

(2004) have demonstrated that a macrolide glycosyltransferase, DesVII from *S. venezuelae*, catalyzes the attachment of TDP-desosamine to 10-deoxymethynoli de or narbonolide. A thorough *in vitro* investigation on the substrate specificity of DesVII was proved that all of 12, 14, 16-memebered macrolides can be glycosylated by several different aminosugars to generate 19 novel macrolides (Jung et al. 2007). Using this broad substrate glycosyltransferase, different sugar units can be attached to different macrolactones to generate unnatural nature product library in engineered microbes. Williams et al. (2008) carried out a comprehensive program of saturation mutagenesis of hot spot amino acid residual in glycosyltransferase, *OleD* and successfully led to the improvement of catalyst proficiency and donor promiscuity of *OleD*. Besides glycosylation, halogenation widely occurring in marine derived compounds also attracted much attention for combinatorial biosynthesis. Present modification was mainly focused on the pyrrole-2-carboxylic acid moiety of aminocoumarins, e.g. clorobiocin (Freitag et al. 2006; Yin et al. 2010) and coumermycin A (Wolpert et al. 2008).

Currently, toolboxes of tailoring enzymes are more amenable to exploit, for their discrete operation on the aglycone with less consideration to intermediate transfer, so enzymatic assembly line for core structure always failed due to incompatibility caused by modification. In addition, the nature of tailoring enzymes made us more feasible to its specification in the framework of synthetic biology.

5.4 Production Improvement by Means of Synthetic Biology

The great structure diversity in natural products libraries and a wealth of structure analogs generated by genetic engineering had provided a prolific chemical pools for antibiotics development. However, natural products commonly accumulate in very low quantities in native producers and more and more novel natural products predicted by genome or eDNA mining await to be unraveled. Synthetic biology can not only provide a general approach to heterologous production of many known or predicted compounds by establishing artificial biosynthetic pathways in microbial systems with related enzymes (natural or engineered), and also is a foundational platform for further engineering of recombinant microbes to maximize the yield of desired compounds with multi-cycle genetic manipulations, seen as Fig. 5.2. In the past decade, various sources (plant, bacteria, fungi, eDNA, etc.) derived natural products categorized into several classes including polyketides (Pfeifer et al. 2001b), nonribosomal peptides (Gruenewald et al. 2004), hydrid polyketides-nonribosomal peptides, terpenoids (Withers and Keasling 2007), or alkaloids, were heterologous synthesized by engineered microbes. Exemplified by more than 1 g/L yield of 6-deoxy-erythronolide B (6dEB) in *E. coli* (Lau et al. 2004) and 40 g/L yield of amorphadiene in yeast (Westfall et al. 2012), commercialized production of some valuable natural products will be achieved in engineered microbes by the means of synthetic biology in the near future. Inevitably, synthetic biology will be an indispensable and general method for natural products production.

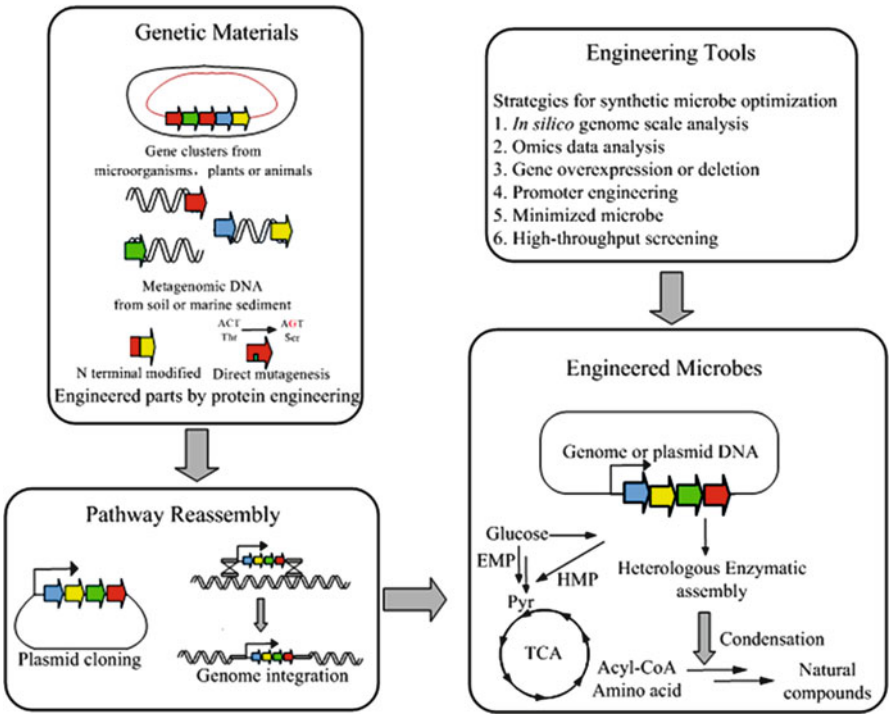


Fig. 5.2 Construction and optimization of a synthetic microbe for the natural product heterologous biosynthesis

5.4.1 Pathway Assembly and Host Redesign for Natural Products Heterologous Biosynthesis

To establish a commercially available microbial system for desired natural products, the first step was to select an appropriate host which must harbor several characteristics: (i) ease in DNA manipulation and transformation; (ii) support active and stable gene expression from different sources, (iii) supply sufficient precursors. If not, the selected microbe should possess a great potential for further development.

Several model microorganisms, such as *E. coli*, *S. cerevisiae* and *B. subtilis*, belonging to different genus, have already been developed into valuable hosts (Zhang et al. 2008). All of these hosts can satisfy the functional expression of many artificial pathways, but some problems severely restricted their further application, such as poor or inactive expression of genes derived from members of phyla distantly genus (*Streptomyces* or eDNA) caused by genetic codon bias or mismatching between host transcriptional machinery and native expression signals, insufficient supply of building blocks especially for natural products derived from *Streptomyces,* marine microbes or eDNA which have a great diversity in building units. Thus, different *Streptomyces,* genome minimized microbes, and marine microbes were developed

in recent years. Baltz (2010) has reviewed the recent studies of using *Streptomyces* and *Saccharopolyspora* as a host for cryptic gene cluster expression. By contrast to *E. coli and S. cerevisiae* with less secondary metabolism, the inherent rich secondary metabolism in these two strains can provide adequate precursors and cofactors for heterologous product biosynthesis.

Genome-minimized microorganisms, whose non-essential DNA regions were deleted including genes responsible for native secondary metabolism, were prospective biological chassis for heterologous biosynthesis. The advantages and applications of genome minimized hosts for heterologous biosynthesis have been reviewed by Gao et al. (2010) and Picataggio (2009). Genome minimized microbes can increase genetic stability, largely eliminate by-products formation to concentrate its metabolic potential into desired metabolism, and reduce consumption on cell growth and maintaining. Recently, the genome of *S. avermitilis* was reduced by large deletion of more than 1.4 Mb subtelomeric regions containing non-essential genes coding for its secondary metabolism. This minimized strain was successfully used for heterologous production of different classes of NPs with significant titers, e.g., aminoglycosides, nonribosomal peptides, polyketides and terpenes (Komatsu et al. 2010). Other related researches in *S. coelicolor* and *E. coli* further confirmed the superiority of genome minimized strains for various valuable compounds production (Flinspach et al. 2010; Mizoguchi et al. 2007). Further improvements will be focused on the construction of biotechnologically synthetic platform cell with a minimized genome and efficient regulatory structure, which can be easily customized for various pre-defined tasks from producing therapeutics to decontaminating waste streams (Foley and Shuler 2010).

Traditional molecular biological manipulation cannot meet the requirement for rapid cloning, reassembling, transferring the whole gene cluster with numerous element modification. Recently, Wenzel et al. (2005) have introduced Red/ET recombineering and transposition technique into large DNA fragment manipulation and transfer in *Pseudomonads*., complete gene cluster involved in myxochromide S biosynthesis, which was stitched from the BAC or Cosmid library of *Stigmatella aurantiaca*, and additional necessary elements for conjugation, integration, and expression in *Pseudomonads* were constructed into cosmid CMch37 in *E. coli*, and then the engineered biosynthetic cluster was transferred into *Pseudomonads* by conjugation. The results showed that the fermentation period time was reduced 3-fold (2 versus 6 days), while the yield was five-fold increase (40 versus 8 mg/L), compared with the native host producer. More importantly, the 30 kb myxochromide S biosynthetic gene cluster, and even ~60 kb biosynthetic gene cluster for epothilones, were efficiently chromosomal integrated into *Myxococcus xanthus* and *Pseudomonas* chromosomes by transposition (Wenzel et al. 2005), which would be difficult using electroporation due to the large size of DNA. Harnessing endonuclease-stimulated homologous recombination, Wingler et al. developed an user-friendly and high-yielding tool called Reiterative Recombination, for multigene pathway assembly in vivo with seamless and robustness stitch (Wingler and Cornish 2011). As the expanding of twenty-first-century genetic manipulation toolkit and advance in genome synthesis, the design of excellent cell chassis will be easier.

To create a synthetic microbe for successful production of a desired compound was the first step. The introduction of heterologous pathway generally breaks out the steady-state of host microbe's metabolism, always resulting in excessive intermediates accumulation and cell toxicity. Additionally, large metabolic discrepancy between heterologous and native host leads to scarcity of building blocks and cofactors for the heterologous compounds biosynthesis. Both of these two reasons caused the low yield of primitive engineered microbes, such as the production of 6-dEB (Pfeifer et al. 2001a) and anti-malarial drug precursor artemisinic acid in *E. coli* (Chang et al. 2007). Except for those reported or easily identified targets restricted to local network, systematic strategies were needed to identify and modify the genome-wide rate-limiting steps for further optimization of the synthetic cell factories, i.e., using *in silico* whole genome-wide stimulation and omics analysis. Computational analysis based on whole genome-scale flux balance analysis is an effective strategy to simulate the value of each single or double gene-knockout. Boghigian et al. (2009) identified several knockouts that theoretically improve the specific production of 6-dEB using the MOMA algorithm (Segre et al. 2002). However, this algorithm was only calculating double-gene-knockout simulations, owing to the tremendous time-consumption and the limit performance of a personal computer. These problems could be overcome by applying a novel algorithm (FDCA) based on FBA developed by our group. More targets including knockout, down-regulated and up-regulated sites for further experimental improvements were obtained by using this developed algorithm (Meng et al. 2011a). Compared to the rational methods, combinatorial gene targets knockout approaches (e.g., transposon mutagenesis) were also used to identify the interesting genetic targets for lycopene production in *E. coli*, resulting in several folds improvement on lycopene yield (Alper et al. 2005b).

In the last decade, strategies for fine tuning gene expression have attracted much attention, such as promoter replacement (Alper et al. 2005a) and RBS (ribosomal binding site) engineering (Pfleger et al. 2006). The methylerythritol-phospate (MEP) pathway was tightly regulated in native *E. coli*, thus a higher yield production of terpenoids compounds (e.g. amorphadiene and taxadiene) cannot be achieved due to an insufficient flux to the biosynthesis of terpene precursors (DMAPP and IPP). To solve this problem, apart from incorporating *S. cerevisiae* derived mevalonate pathway into *E. coli* to elevate terpenoid precursors supply (Martin et al. 2003), Ajikumar et al. extensively investigated the capability of MEP pathway by a systematic multivariate search approach. The results revealed that a copy of *dxs-idi-ispDF* operon integrated into chromosome has a higher yield of taxadiene than multi-copy expression by plasmids, further suggesting that fine-tuning expression of *dxs-idi-ispDF* operon could be obtained a better result (Ajikumar et al. 2010). Recently, Mey et al. have combined a promoter strength predictive (PSP) mode with promoter knock-in technique for rational modulation the expression level of target enzyme in appropriate levels (De Mey et al. 2010). This approach can be extended to other expression control modules, and a desired functional operon can be assembly in a defined pattern. Beside individually fine tuning these gene targets, Wang et al. have developed a multiplex genome engineering (MAGE) to rapidly modify many locations

on the chromosome to produce combinatorial genomic diversity. After 3 days cyclical evolution with 24 targets genes, the yield of lycopene in *E. coli* achieved five-fold increase by MAGE (Wang et al. 2009).

5.4.2 *In Silico Aided Strategies for Improving NPs Yield in Engineered Microbes*

The development and application of high-throughput techniques in studies have accumulated a large amount of omics data (Walter 2010). In parallel, *in silico* modelling and simulation approaches are being developed for quantitative analysis of cellular metabolism at the systems level (Lee et al. 2005b). Systems biology thus becomes the foundation for genome-scale synthetic biology (Barrett et al. 2006). Various models and tools developed for systems biology are suitable for serving as a guideline for the construction of synthetic biosystems and genetic devices as well. Palsson group have made great efforts on reconstructing the genome-scale gene-protein-reaction computational models for various kinds of organisms (Becker and Palsson 2005; Duarte et al. 2007, 2004; Feist et al. 2007; Forster et al. 2003; Herrgard et al. 2008; Lee et al. 2008; Oh et al. 2007; Reed et al. 2003; Thiele et al. 2005, 2009). These models, along with the constraint-based simulation methods (Becker et al. 2007), such as metabolic flux analysis (MFA) or flux balance analysis (FBA) (Edwards and Palsson 2000; Kauffman et al. 2003), dynamic flux balance analysis (DFBA) (Mahadevan et al. 2002), minimization of metabolic adjustment (MOMA) (Segre et al. 2002), and regulatory on/off minimization (ROOM) (Shlomi et al. 2005), are powerful tools to investigate the properties of different heterologous hosts and helpful for further *in silico* strain improvement for particular compound biosynthesis (Boghigian et al. 2009; Choi et al. 2010; Lee et al. 2005a; Park et al. 2007). Park et al. constructed a high-performance *E. coli* strain for L-valine production based on transcriptome analysis and gene knockout simulation of *in silico* genome-scale metabolic network (Park et al. 2007). Choi et al. reported a strategy called flux scanning based on enforced objective flux (FSEOF) to identify gene amplification targets. Combining FSEOF for gene amplification and MOMA for gene knockout simulation finally resulted in synergistic enhancement of lycopene production (Choi et al. 2010).

Besides gene target identification for strain improvement, *in silico* aided analysis strategies can also be employed for selecting superior heterologous host and precursor supply pathway for heterologous biosynthesis. Gonzalez-Lergier et al. (2006) analyzed the maximum theoretical yield of the erythromycin precursor, 6dEB produced by *E. coli*; Our group further investigated this yield in other organisms (e.g. *S. cerevisiae* and *B. subtilis*) in comparison with *E. coli*, and found that *E. coli* was the most efficient host for 6dEB biosynthesis with the highest yield. For terpenoids biosynthesis, the central precursor isopentenyl pyrophosphate (IPP) can be provided via two separate pathways, the 1-deoxylulose 5-phosphate (DXP) pathway and the mevalonate (MVA) pathway, in different organisms. Previous studies supported the

utilization of the MVA pathway for IPP over the DXP pathway (Anthony et al. 2009; Morrone et al. 2010; Tsuruta et al. 2009). However, we found that the DXP pathway owns more potential for IPP supply than MVA pathway through an *in silico* comparison of the maximum theoretical IPP yields and thermodynamic properties between the DXP and MVA pathways with different hosts and carbon sources. Following work on heterologous biosynthesis of Taxol's key intermediate taxadiene via DXP pathway has confirmed our viewpoint (Meng et al. 2011b).

5.4.3 *Activating the Potential of "Silent" Biosynthetic Gene Clusters*

Inspired by the explosive growth of genomic sequences data and comparative genomic analysis in the last decade, the extensive distribution of cryptic secondary metabolite biosynthetic gene clusters in microbial genomes were recognized. How to exploit these hiding synthetic capacities concealed in frequently touched microorganisms have attracted much attentions (Alper et al. 2005b). Hertweck et al. have summarized some powerful methods recently adopted, such as epigenomic manipulation, genomic-sequence based prediction combined with metabolome, and genetic modification on regulatory control (Scherlach and Hertweck 2009). Beside these options, cloned and heterologous expression was also identified as a more powerful tool to unravel these capacities. More examples of silent gene or gene clusters for terpene, PKSs and NRPSs synthase from bacteria, fungi and plants, were introduced into heterologous hosts and targeted compounds were isolated in anticipate. Blasiak et al. captured a promising biosynthetic gene cluster in *Pseudoalteromonas tunicata* by sequence-based searching for homologous ATP-grasp enzymes (Blasiak and Clardy 2009). Heterologous expression of this cluster in *E. coli* allowed for the production of two new 3-formyl tyrosine metabolites. Wyszynski et al. elucidated tunicamycin biosynthesis through a chemical-guided strategy for *in silico* genome mining and heterologous expression of a likely minimal gene set necessary for antibiotic production, which facilitate future redesign and generation of novel tunicamycin analogs for targeting of diverse carbohydrate-processing enzymes (Wyszynski et al. 2010).

5.4.4 *Novel Antibiotics via Heterologous Expression of eDNA*

Besides novel microorganisms isolation and culturing, eDNA library directly extracted from extremely complicated circumstance, especially in untouched terrestrial or marine source is also an invaluable treasure for antibiotics discovery (Banik and Brady 2010; Li and Qin 2005). Currently, heterologous expression of eDNA clones coupling with biological activity or sequence-based screening methods were

powerful approach for eDNA mining (Banik and Brady 2010). Although much attempts on heterologous expression of eDNA in *E. coli* have been done, only some compounds with simple biosynthetic diagram were isolated, such as indigo, violacein, turbomycin A and B, and patellamide (Long et al. 2005). These drawbacks were caused by the inability of acquisition of intact gene fragment and incompatibility of expression regulatory elements such as promoter in heterologous host. To increase the small-molecule biosynthetic diversity found in eDNA mining, Craig et al. have reported the phenotypic screening of broad-host-range environmental DNA libraries in six different proteobacteria: *Agrobacterium tumefaciens, Burkholderia graminis, Caulobacter vibrioides, E. coli, Pseudomonas putida, and Ralstonia metallidurans* (Craig et al. 2010). Advances in synthetic biology, including large DNA fragment assembly techniques and highly developed microbial chassis for a specific class of compound, largely facilitated the functional expression of entire eDNA cluster. Sequence-based screening using homologous PCR or clone hybridization was first employed for identification of the indispensible genetic elements for the intact cluster assembly (Singh et al. 2009). And then, these large discrete fragments were stitched together and transformed into appropriate host for functional expression with necessary expression elements exchanges. Using this strategy, more gene fragments coding partial of biosynthetic pathway of antibiotics were isolated. A general framework for recovery of large functional DNA fragments from cosmid clones and efficient re-assembly of these large fragments using transformation-associated recombination (TAR) in *S. cerevisiae* has been reported (Kim et al. 2010). They used this TAR method to reassemble two characterized eDNA clones (named as AB649 and AB185) into a complete Type II PKS biosynthetic pathway and heterologous expressed in *S. albus*. Results showed that four novel metabolites were detected, which were not seen in AB649 cultures alone (Feng et al. 2010). They also explored β-Proteobacterium *Ralstonia metallidurans* as an alternative host for the expression of eDNA library. Interestingly, cosmids conferred the production of novel pigmented/antibacterially-active metabolites in *R. metallidurans* clones were not functional in *E. co*li (Craig et al. 2009). Thus, *R. metallidurans*, a new model system, can be used to expand the collection of metabolites found in future metagenomic discovery efforts.

5.5 Conclusion

Along with the pace of technique advance, synthetic biology penetrates into all aspects of drug development process, and opens up an efficient way for antibiotics screening and manufacturing. Numerous materials, methods and experiences have accumulated since penicillin discovery in 1940s, synthetic biology will give them a new interpretations, such as the development of antibiotics no longer just relies on microbial sources; all kinds of natural products can be synthesized in general heterologous host; combinatorial biosynthesis no longer just depends on the similarity module replacement; arbitrary pathway assembly can be achieved and its content

will be extended unprecedentedly; antibiotic screening can also be achieved by designing more delicate models; large-scale industrial production of antibiotics will generally be performed in controllable *E. coli* and other host; and the yield, quality and industrial scale will be further improved. Clearly, synthetic biology will be exert a profound impact on the development and manufacturing of antibiotics. After all, synthetic biology is a new discipline, its theory, techniques and methods should be further developed, e.g., large-scale genome synthesis and stitching, the construction of artificial cell platforms will be the key to the development of antibiotics.

Acknowledgments This work was supported by the National Basic Research Program of China (973 Program, Grant no. 2012CB721104), the National Natural Science Foundation of China (Grants no. 31170101 and 31100073) and the major Projects of Knowledge Innovation Program of Chinese Academy of Sciences (Grant no. KSCX2-EW-J-12).

References

Agapakis CM, Silver PA (2009) Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. Mol Biosyst 5:704–713

Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science 330:70–74

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005a) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683

Alper H, Miyaoku K, Stephanopoulos G (2005b) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23:612–616

Anthony JR, Anthony LC, Nowroozi F, Kwon G, Newman JD, Keasling JD (2009) Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. Metab Eng 11:13–19

Aubel D, Morris R, Lennon B, Rimann M, Kaufmann H, Folcher M, Bailey JE, Thompson CJ, Fussenegger M (2001) Design of a novel mammalian screening system for the detection of bioavailable, non-cytotoxic streptogramin antibiotics. J Antibiot 54:44–55

Baltz RH (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. Nat Biotechnol 24:1533–1540

Baltz RH (2010) Streptomyces and saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol 37:759–772

Banik JJ, Brady SF (2010) Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. Curr Opin Microbiol 13:603–609

Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY (2006) Systems biology as a foundation for genome-scale synthetic biology. Curr Opin Biotechnol 17:488–492

Becker SA, Palsson BO (2005) Genome-scale reconstruction of the metabolic network in *staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. BMC Microbiol 5:8

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc 2:727–738

Beven L, Castano S, Dufourcq J, Wieslander A, Wroblewski H (2003) The antibiotic activity of cationic linear amphipathic peptides: lessons from the action of leucine/lysine copolymers on bacteria of the class mollicutes. Eur J Biochem 270:2207–2217

Blasiak LC, Clardy J (2009) Discovery of 3-formyl-tyrosine metabolites from *pseudoalteromonas tunicata* through heterologous expression. J Am Chem Soc 132:926–927

Blunt JW, Copp BR, Munro MH, Northcote PT, Prinsep MR (2010) Marine natural products. Nat Prod Rep 27:165–237

Boghigian BA, Lee K, Pfeifer BA (2009) Computational analysis of phenotypic space in heterologous polyketide biosynthesis – applications to *Escherichia coli, Bacillus subtilis*, and *Saccharomyces cerevisiae*. J Theor Biol 262:197–207

Borisova SA, Zhao L, Melancon IC, Kao CL, Liu HW (2004) Characterization of the glycosyltransferase activity of desVII: analysis of and implications for the biosynthesis of macrolide antibiotics. J Am Chem Soc 126:6534–6535

Bromley EH, Channon K, Moutevelis E, Woolfson DN (2008) Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems. ACS Chem Biol 3:38–50

Bugl H, Danner JP, Molinari RJ, Mulligan JT, Park H-O, Reichert B, Roth DA, Wagner R, Budowle B, Scripp RM, Smith JAL, Steele SJ, Church G, Endy D (2007) DNA synthesis and biological security. Nat Biotechnol 25:627–629

Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G (2008) NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36:D326–331

Caboche S, Pupin M, Leclere V, Jacques P, Kucherov G (2009) Structural pattern matching of nonribosomal peptides. BMC Struct Biol 9:15

Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J Bacteriol 192:5143–5150

Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD (2007) Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. Nat Chem Biol 3:274–277

Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76:3097–3105

Clomburg JM, Gonzalez R (2010) Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. Appl Microbiol Biotechnol 86:419–434

Craig JW, Chang F-Y, Brady SF (2009) Natural products from environmental DNA hosted in ralstonia metallidurans. ACS Chem Biol 4:23–28

Craig JW, Chang F-Y, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol 76:1633–1641

De Mey M, Maertens J, Boogmans S, Soetaert WK, Vandamme EJ, Cunin R, Foulquie-Moreno MR (2010) Promoter knock-in: a novel rational method for the fine tuning of genes. BMC Biotechnol 10:26

Devasahayam G, Scheld WM, Hoffman PS (2010) Newer antibacterial drugs for a new century. Expert Opin Investig Drugs 19:215–234

Duarte NC, Herrgard MJ, Palsson BO (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14:1298–1309

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A 104:1777–1782

Edwards JS, Palsson BO (2000) Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. BMC Bioinform 1:1

Erb A, Weiss H, Harle J, Bechthold A (2009) A bacterial glycosyltransferase gene toolbox: generation and applications. Phytochemistry 70:1812–1821

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121

Feng Z, Kim JH, Brady SF (2010) Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. J Am Chem Soc 132:11902–11903

Findlay B, Zhanel GG, Schweizer F (2010) Cationic amphiphiles, a new generation of antimicrobials inspired by the natural antimicrobial peptide scaffold. Antimicrob Agents Chemother 54:4049–4058

Flinspach K, Westrich L, Kaysser L, Siebenberg S, Gomez-Escribano JP, Bibb M, Gust B, Heide L (2010) Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. Biopolymers 93:823–832

Floss HG (2006) Combinatorial biosynthesis – potential and problems. J Biotechnol 124:242–257

Foley PL, Shuler ML (2010) Considerations for the design and construction of a synthetic platform cell for biotechnological applications. Biotechnol Bioeng 105:26–36

Forster J, Famili I, Fu P, Palsson BO, Nielsen J (2003) Genome-scale reconstruction of the *saccharomyces cerevisiae* metabolic network. Genome Res 13:244–253

Freitag A, Méndez C, Salas JA, Kammerer B, Li S-M, Heide L (2006) Metabolic engineering of the heterologous production of clorobiocin derivatives and elloramycin in *Streptomyces coelicolor* M512. Metab Eng 8:653–661

Fussenegger M, Morris RP, Fux C, Rimann M, von Stockar B, Thompson CJ, Bailey JE (2000) Streptogramin-based gene regulation systems for mammalian cells. Nat Biotechnol 18:1203–1208

Gao H, Zhuo Y, Ashforth E, Zhang L (2010) Engineering of a genome-reduced host: practical application of synthetic biology in the overproduction of desired secondary metabolites. Protein Cell 1:621–626

Gonzalez-Lergier J, Broadbelt LJ, Hatzimanikatis V (2006) Analysis of the maximum theoretical yield for the synthesis of erythromycin precursors in *Escherichia coli*. Biotechnol Bioeng 95:638–644

Gonzalez-Nicolini V, Fux C, Fussenegger M (2004) A novel mammalian cell-based approach for the discovery of anticancer drugs with reduced cytotoxicity on non-dividing cells. Invest New Drugs 22:253–262

Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A 89:5547–5551

Gruenewald S, Mootz HD, Stehmeier P, Stachelhaus T (2004) In vivo production of artificial non-ribosomal peptide products in the heterologous host *Escherichia coli*. Appl Environ Microbiol 70:3282–3291

Haseltine EL, Arnold FH (2007) Synthetic gene circuits: design with directed evolution. Annu Rev Biophys Biomol Struct 36:1–19

Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novere N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasic I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26:1155–1160

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333:348–353

Jung WS, Han AR, Hong JS, Park SR, Choi CY, Park JW, Yoon YJ (2007) Bioconversion of 12-, 14-, and 16-membered ring aglycones to glycosylated macrolides in an engineered strain of *Streptomyces venezuelae*. Appl Microbiol Biotechnol 76:1373–1381

Kang SJ, Won HS, Choi WS, Lee BJD (2009) Novo generation of antimicrobial LK peptides with a single tryptophan at the critical amphipathic interface. J Pept Sci 15:583–588

Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol 14:491–496

Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330:1355–1358

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF (2010) Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93:833–844

Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H (2010) Genome-minimized streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107:2646–2651

Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, Chaudhary U, Doumith M, Giske CG, Irfan S, Krishnan P, Kumar AV, Maharjan S, Mushtaq S, Noorie T, Paterson DL, Pearson A, Perry C, Pike R, Rao B, Ray U, Sarma JB, Sharma M, Sheridan E, Thirunarayan MA, Turton J, Upadhyay S, Warner M, Welfare W, Livermore DM, Woodford N (2010) Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis 10:597–602

Lau J, Tran C, Licari P, Galazzo J (2004) Development of a high cell-density fed-batch bioprocess for the heterologous production of 6-deoxyerythronolide B in *Escherichia coli*. J Biotechnol 110:95–103

Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY (2005a) Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl Environ Microbiol 71:7880–7887

Lee SY, Lee DY, Kim TY (2005b) Systems biotechnology for strain improvement. Trends Biotechnol 23:349–358

Lee J, Yun H, Feist AM, Palsson BO, Lee SY (2008) Genome-scale reconstruction and in silico analysis of the clostridium acetobutylicum ATCC 824 metabolic network. Appl Microbiol Biotechnol 80:849–862

Li X, Qin L (2005) Metagenomics-based drug discovery and marine microbial diversity. Trends Biotechnol 23:539–543

Li JW, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325:161–165

Long PF, Dunlap WC, Battershill CN, Jaspars M (2005) Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. ChemBioChem 6:1760–1765

Loose C, Jensen K, Rigoutsos I, Stephanopoulos G (2006) A linguistic model for the rational design of antimicrobial peptides. Nature 443:867–869

Mahadevan R, Edwards JS, Doyle FJ 3rd (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83:1331–1340

Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21:796–802

Medema MH, Breitling R, Bovenberg R, Takano E (2011) Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. Nat Rev Micro 9:131–137

Meng HL, Lu Z, Wang Y, Wang X, Zhang S (2011a) *In silico* improvement of heterologous biosynthesis of erythromycin precursor 6-deoxyerythronolide B in *Escherichia coli*. Biotechnol Bioprocess Eng 16:445–456

Meng HL, Wang Y, Hua Q, Zhang SL, Wang XN (2011b) *In silico* analysis and experimental improvement of taxadiene heterologous biosynthesis in *Escherichia coli*. Biotechnol Bioprocess Eng 16:205–215

Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat Biotechnol 23:1171–1176

Mizoguchi H, Mori H, Fujio T (2007) *Escherichia coli* minimum genome factory. Biotechnol Appl Biochem 46:157–167

Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ (2010) Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. Appl Microbiol Biotechnol 85:1893–1906

Na D, Kim TY, Lee SY (2010) Construction and optimization of synthetic pathways in metabolic engineering. Curr Opin Microbiol 13:363–370

Neumann H, Neumann-Staubitz P (2010) Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. Appl Microbiol Biotechnol 87:75–86

Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282:28791–28799

Olano C, Mendez C, Salas JA (2010) Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. Nat Prod Rep 27:571–616

Park Y, Lee DG, Jang SH, Woo ER, Jeong HG, Choi CH, Hahm KS (2003) A Leu-Lys-rich antimicrobial peptide: activity and mechanism. Biochim Biophys Acta 1645:172–182

Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci U S A 104:7797–7802

Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science 291:1790–1792

Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24:1027–1032

Picataggio S (2009) Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. Curr Opin Biotechnol 20:325–329

Powers J-PS, Hancock REW (2003) The relationship between peptide structure and antibacterial activity. Peptides 24:1681–1691

Prather KL, Martin CH (2008) De novo biosynthetic pathways: rational design of microbial chemical factories. Curr Opin Biotechnol 19:468–474

Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol 21:162–169

Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4:R54

Rix U, Fischer C, Remsing LL, Rohr J (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat Prod Rep 19:542–580

Rodriguez E, McDaniel R (2001) Combinatorial biosynthesis of antimicrobials and other natural products. Curr Opin Microbiol 4:526–534

Scherlach K, Hertweck C (2009) Triggering cryptic natural product biosynthesis in microorganisms. Org Biomol Chem 7:1753–1760

Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99:15112–15117

Shao Z, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:e16

Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci U S A 102:7695–7700

Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V, Batra N (2009) Metagenomics: concept, methodology, ecological inference and recent advances. Biotechnol J 4:480–494

Sousa JC, Berto RF, Gois EA, Fontenele-Cardi NC, Honorio JE Jr, Konno K, Richardson M, Rocha MF, Camargo AA, Pimenta DC, Cardi BA, Carvalho KM (2009) Leptoglycin: a new glycine/leucine-rich antimicrobial peptide isolated from the skin secretion of the South American frog leptodactylus pentadactylus (leptodactylidae). Toxicon 54:23–32

Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438:443–448

Thiele I, Vo TD, Price ND, Palsson BO (2005) Expanded metabolic reconstruction of helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. J Bacteriol 187:5818–5830

Thiele I, Jamshidi N, Fleming RM, Palsson BO (2009) Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput Biol 5:e1000312

Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD (2009) High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. PLoS One 4:e4489

Walter T (2010) Visualization of image data from cells to organisms. Nat Methods 7:S26–S40

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898

Weber W, Fux C, Daoud-el Baba M, Keller B, Weber CC, Kramer BP, Heinzen C, Aubel D, Bailey JE, Fussenegger M (2002) Macrolide-based transgene control in mammalian cells and mice. Nat Biotechnol 20:901–907

Weber CC, Link N, Fux C, Zisch AH, Weber W, Fussenegger M (2005) Broad-spectrum protein biosensors for class-specific detection of antibiotics. Biotechnol Bioeng 89:9–17

Weber W, Schoenmakers R, Keller B, Gitzinger M, Grau T, Daoud-El Baba M, Sander P, Fussenegger M (2008) A synthetic mammalian gene circuit reveals antituberculosis compounds. Proc Natl Acad Sci U S A 105:9994–9998

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, Wohlleben W (2009) CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol 140:13–17

Weissman KJ, Leadlay PF (2005) Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 3:925–936

Wenzel SC, Gross F, Zhang Y, Fu J, Stewart AF, Muller R (2005) Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. Chem Biol 12:349–356

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109:E111–E118

Wilkinson B, Micklefield J (2007) Mining and engineering natural-product biosynthetic pathways. Nat Chem Biol 3:379–386

Williams GJ, Goff RD, Zhang C, Thorson JS (2008) Optimizing glycosyltransferase specificity via "hot spot" saturation mutagenesis presents a catalyst for novobiocin glycorandomization. Chem Biol 15:393–401

Wingler LM, Cornish VW (2011) Reiterative recombination for the in vivo assembly of libraries of multigene pathways. Proc Natl Acad Sci U S A 108:15135–15140

Withers ST, Keasling JD (2007) Biosynthesis and engineering of isoprenoid small molecules. Appl Microbiol Biotechnol 73:980–990

Withers ST, Gottlieb SS, Lieu B, Newman JD, Keasling JD (2007) Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. Appl Environ Microbiol 73:6277–6283

Wolpert M, Heide L, Kammerer B, Gust B (2008) Assembly and heterologous expression of the coumermycin A1 gene cluster and production of new derivatives by genetic engineering. ChemBioChem 9:603–612

Wyszynski FJ, Hesketh AR, Bibb MJ, Davis BG (2010) Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster. Chem Sci 1:581–589

Yadav G, Gokhale RS, Mohanty D (2009) Towards prediction of metabolic products of polyketide synthases: an in silico analysis. PLoS Comput Biol 5:e1000351

Yin X, Chen Y, Zhang L, Wang Y, Zabriskie TM (2010) Enduracidin analogues with altered halogenation patterns produced by genetically engineered strains of *Streptomyces fungicidicus*. J Nat Prod 73:583–589

Zhang W, Tang Y (2008) Combinatorial biosynthesis of natural products. J Med Chem 51:2629–2633

Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet 20:123–128

Zhang H, Wang Y, Pfeifer BA (2008) Bacterial hosts for natural product production. Mol Pharm 5:212–225

Zorko M, Jerala R (2010) Production of recombinant antimicrobial peptides in bacteria. Method Mol Biol 618:61–76

Chapter 6
Cascades and Networks of Regulatory Genes That Control Antibiotic Biosynthesis

Juan F. Martín and Paloma Liras

Abstract Onset of the biosynthesis of bioactive secondary metabolites in batch cultures of actinomycetes occurs after the rapid growth phase, following a transition phase which involves complex metabolic changes. This transition is triggered by nutrient starvation or by other environmental stress signals. Expression of genes encoding bioactive secondary metabolites is governed by cascades of pathway specific regulators and networks of cross-talking global regulators. Pathway specific regulators such as *Streptomyce*s antibiotic regulatory proteins, LAL-type and LysR-type regulators respond to autoregulatory proteins that act in concert with their cognate ligands (e.g. γ-butyrolactone receptor proteins and their cognate γ-butyrolactone ligands). Global regulators such as PhoR-PhoP and other two component systems and orphan response regulators, such as GlnR, control set of genes affecting primary and secondary metabolism. GlnR and, therefore, nitrogen metabolism genes are under phosphate control exerted by binding of PhoP to PHO boxes located in the promoter region of GlnR. A few pleiotropic regulatory genes, such as *areB* (*ndgR*), *dmdR1* or *dasR* connect primary metabolism (amino acid biosynthesis, N-acetylglucosamine or iron levels) with antibiotic biosynthesis. Some atypical response regulators that require specific small ligands appear to be involved in feedback control of antibiotic production. All these mechanisms together modulate, in a coordinated manner, different aspects of *Streptomyces* metabolism as a real "protection net" that prevents drastic changes in metabolism that may be deleterious for cell survival.

Keywords Antibiotic biosynthesis pathways • Phosphate and nitrogen control • Pathway-specific regulators • *Streptomyces* regulatory mechanisms • Two component systems

J.F. Martín (✉) • P. Liras
Department of Molecular Biology, University of León, León 24071, Spain
e-mail: jf.martin@unileon.es; paloma.liras@unileon.es

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_6,

Abbreviations

ARE	autoregulatory element
ARR	atypical response regulators
BRP	butyrolactone receptor protein
CDA	calcium-dependent antibiotic
EMSA	electrophoretic mobility shift assay
GlcNac	N-acetyl-glucosamine
HTH	helix-turn-helix
i-TRAQ	isobaric tags for relative and absolute quantification
LAL	large ATP-binding regulators of the LuxR family
LC-MS/MS	liquid chromatography mass spectrometry
PHO box	sequence for PhoP binding
Q-RT-PCR	quantitative reverse polymerase chain reaction
SARP	*Streptomyces* antibiotic regulatory protein
TCS	two component systems

6.1 Secondary Metabolite Gene Clusters

Secondary metabolites of microorganisms include antibacterial, antifungal, antiviral and antitumor agents, insecticides, pigments, inmunomodulators (including potent inmunosupressors) and a variety of other bioactive products (Davies 2007; Yim et al. 2007; Demain 2009).

The genes encoding enzymes involved in the biosynthesis of secondary metabolites are usually clustered in "genome islands" extending for up to 100 kb, which are not conserved in the genomes of closely related actinomycetes. Expression of the genes encoding the biosynthesis of secondary metabolites is controlled by complex regulatory mechanisms. The onset of antibiotic production is triggered by a nutritional shift-down (see below). Broadly speaking we can divide the regulatory genes that control antibiotic biosynthesis in two types: (i) some of these genes are located in the antibiotic gene clusters together with genes for biosynthesis, secretion and resistance to the antibiotics (to avoid suicide of the producer bacteria) and are usually pathway-specific (Martín and Liras 1989; Cundliffe and Demain 2010); (ii) others, are global (wide domain) pleiotropic regulators that control several metabolic pathways and may not be linked to specific biosynthetic gene clusters (Martín and Liras 2010). These global regulators respond to a variety of nutritional or environmental stress signals, e.g. phosphate or nitrogen starvation, iron deficiency, the presence of chitin or N-acetylglucosamine in the medium, cell wall damage, heat shock, osmotic or pH stress, among others.

6.2 Transition from Primary to Secondary Metabolism

Streptomyces species and other filamentous microorganisms show a rapid growth phase that is not strictly exponential, because of the polar growth of the hyphae, followed by a transition phase that leads to a stationary phase, which in some species is rapidly followed by massive lysis of the culture while in others may be prolonged for a few days before lysis occurs (Kolter et al. 1993).

The transition phase from primary metabolism (trophophase or rapid growth phase) to secondary metabolism (idiophase or production phase) following nutrient starvation is commonly found in most actinomycetes (Nieselt et al. 2010), although the time period and speed of the transition phase depends upon the media composition and the individual strains. The transition phenomenon and the time span of proper working of the idiophase received attention for a long time (Martín and Demain 1980), but new details of the regulation of the "metabolic switch" during the transition phase have emerged in recent years (Hesketh et al. 2007, 2008; Huang et al. 2001; Lian et al. 2008; Nieselt et al. 2010; Manteca et al. 2010a, b; Thomas et al. 2012).

In solid cultures of *Streptomyces* there are two differentiated mycelial stages: an early compartmentalized vegetative mycelium (first mycelium) and a multinucleated reproductive mycelium (second mycelium) arising after programmed cell death processes (Manteca et al. 2010a). Detailed proteomics analysis of the distinct developmental stages of solid confluent *Streptomyces coelicolor* cultures using iTRAQ (isobaric tags for relative and absolute quantitation) labeling and LC-MS/MS revealed important protein changes during the switch from primary to secondary metabolism between the initial compartmentalized mycelium and the multinucleated hyphae. Similarly, studies of nonsporulating liquid cultures demonstrate a complex differentiation process comparable to that occurring in sporulating solid cultures (Manteca et al. 2010b).

For decades most metabolic switch studies in liquid cultures of *Streptomyces* species have been performed in shake flasks in which the control of environmental conditions (pH, temperature, oxygen transfer) is poor and the results are not easily scaled-up to large batch cultures in stirred fermenters. Recently these problems have been overcome in highly controlled fermenters by using a glutamate-based minimum medium which is phosphate limited and by strictly controlling the environmental growth conditions (Nieselt et al. 2010) that allowed robust transcriptomic and proteomic studies (Thomas et al. 2012; Martín et al. 2012b). These studies revealed drastic changes in the expression of about 50 genes following phosphate depletion in the fermenter culture (Martín et al. 2012b).

6.3 Pyramidal Control of Antibiotic Biosynthesis Versus Networks of Global Regulators

For many years evidence has been obtained suggesting that expression of antibiotics (and other bioactive metabolites) biosynthetic genes is controlled through cascades of regulators that in turn modulate the expression of pathway-associated regulatory genes

(Bibb 2005; Martín and Liras 2010). We will refer to the well-known streptomycin and cephamycin/clavulanic acid regulatory cascades as model systems, but the tylosin regulatory cascade and other similar cascades are also well known (Cundliffe 2008). In all these cases the top position in the regulatory cascade is occupied by: (i) a butyrolactone receptor protein (BRP) in coordination with its cognate butyrolactone molecule (Takano 2006), or (ii) a pleiotropic regulator (e.g. AreB/NdgR) that governs butyrolactone biosynthesis, and therefore, the butyrolactone-dependent cascade.

6.3.1 Autoregulatory Factors and Their Receptor Proteins

A class of important small molecules that affect the behavior of *Streptomyces* species are autoregulators that work as intercellular communication signals (Horinouchi 2007; Takano 2006). Autoregulatory molecules play important roles in the control of biosynthesis of secondary metabolites in numerous streptomycetes and other actinobacteria (Choi et al. 2003; Takano 2006). The best known autoregulators in *Streptomyces* belong to the γ-butyrolactone family although there are other autoregulators with different chemical structures (Yamada 1999; Recio et al. 2004; Xu et al. 2010). The γ-butyrolactones act at very low effective concentration (about 100 nM) and they control multiple cellular functions in specific growth phases (Horinouchi and Beppu 2007). About 15 signal molecules with γ-butyrolactone structure have been found in *Streptomyces* species; a few γ-butyrolactones have been purified from *Streptomyces* species including *S. griseus, S. virginiae, S. avermitilis* or *S. coelicolor* (Takano 2006; Kato et al. 2007; Kitani et al. 2011). On the basis of minor structural differences the γ-butyrolactones have been classified as: (i) A-factor type, with a 10-keto group in the alkyl side chain, (ii) Virginia-butanolides with a 10-α-hydroxyl group, and (iii) IM-2 type with a 10-ß-hydroxyl group. The pathway for γ-butyrolactone biosynthesis was elucidated a few years ago (Kato et al. 2007), although some doubts on the cyclization of the precursors to form the butyrolactone ring still remain.

In addition, signal molecules with other structures, such as the IP factor [2,3-diamino, 2,3-bis(hydroxymethyl)-1,4-butanediol) in *S. natalensis*, have been described to play a similar role to that of γ-butyrolactones, although little is still known about the regulation mechanisms of these alternative autoregulatory molecules (Recio et al. 2004; Vicente et al. 2009).

Autoregulatory factors exert multiple effects on cell metabolism. DNA microarray analysis of global gene regulation by A-factor in *Streptomyces griseus* showed that the expression of more than 150 genes is affected by the A factor (Hara et al. 2009). A few of these genes are primary targets of the autoregulators and respond immediately to the addition of A-factor whereas others are affected by cascade mechanisms. It is interesting that the effects of autoregulators on antibiotic production and morphogenesis are different from species to species. The butyrolactones in culture media are taken up by the cells and act as quorum sensing molecules by releasing their cognate γ-butyrolactone receptor proteins (BRP) from operator sites,

thus activating gene expression (Horinouchi 2007).γ-Butyrolactone receptor proteins are transcriptional regulators belonging to the TetR superfamily (Natsume et al. 2004). These BRP proteins possess a γ-butyrolactone-binding domain at the C-terminus and a characteristic DNA-binding helix-turn-helix (HTH) motif with high overall conservation at its N-terminus. The high conservation of the HTH motif in these proteins suggests that AT-rich DNA sequences conserved in the promoters of the BRP regulated genes (called ARE, for autoregulatory elements) are the targets recognised by these autoregulator receptors (Folcher et al. 2001; Natsume et al. 2004). Since the γ-butyrolactones are species specific, the protein sequence of their corresponding BRPs must have specificity, particularly in their ligand-binding domains.

The butyrolactone-BRP systems in response to external signals trigger regulatory genes encoding proteins, collectively known as SARP (Streptomyces antibiotic regulatory proteins), most of which act as positive regulators of the genes for antibiotic biosynthesis. However, in most cases the initial steps of the signal transduction pathway from nutritional or environmental conditions to the γ-butyrolactone synthases is not yet known (Kim et al. 2004; Santamarta et al. 2005).

An interesting question is "who regulates the regulator". In several cases the BRP-encoding genes are autoregulated. Moreover, the butyrolactone synthesizing genes are under the control of sigma factors or pleiotropic regulatory genes such as NdgR or PhoP (see below) that responds to nutrient starvation. In some *Streptomyces* species the *brp* gene has evolved in close association with the butyrolactone synthase gene. This is the case of the *arpA-afsA* genes in *S. griseus* (Nishida et al. 2007), but the validity of this association in general is unclear as more *brp* genes without an attached butyrolactone synthase gene are found (Salehi-Najafabadi et al. 2011).

6.3.2 *Pathway Specific Regulators: Intermediates in the Regulatory Cascades*

Most of the pathway-specific antibiotic regulatory genes encode activators with a low molecular weight (around 25 kDa). Many of these small pathway specific regulators belong to the SARP family. They contain, close to the N-terminus, a DNA binding domain (Wietzorrek and Bibb 1997) and a bacterial transcriptional activation domain (BTAD). The regulators ActII-ORF4, CdaR and RedD in *S. coelicolor*, controlling actinorhodin, the calcium-dependent-antibiotic (CDA) and undecylprodigiosin production, respectively, belong to this family; CcaR of *Streptomyces clavuligerus* that regulates cephamycin and clavulanic acid biosynthesis, and StrR controlling streptomycin production in *Streptomyces griseus* are also SARPs. The number of SARP proteins is continuously increasing (Martín and Liras 2010). Although they are pathway-specific, some of the so-called pathway-specific regulators sometimes control also unrelated biosynthetic pathways (Huang et al. 2005).

Deletion of the SARP-encoding genes results in impaired production or complete loss of antibiotic formation, and this ability is regained by complementation with the wild type gene. Comparative RT-PCR studies of the wild type strain and

SARP null mutants allows to identify the genes or operons activated by each SARP in the antibiotic biosynthesis clusters. Binding of SARPs to promoters upstream of antibiotic biosynthetic genes has been demonstrated in several gene clusters using electrophoretic mobility shift assays. Initial studies on binding specificity were made with ActII-ORF4 and DnrI (Arias et al. 1999; Sheldon et al. 2002) and with the CcaR of *S. clavuligerus* (Santamarta et al. 2002). The available evidence indicates that SARP activators regulate expression of key genes of an antibiotic pathway but not all of them.

Expression of SARP genes is frequently controlled by butyrolactone receptor proteins. These BRPs recognize specific 22 bp ARE sequences located upstream or close to the promoter of SARP-encoding genes, as shown by footprinting experiments (Stratigopoulos et al. 2002; Santamarta et al. 2005). Although many genes encoding BRPs have been identified in different antibiotic producing actinomycetes and several of them have been cloned and the corresponding proteins expressed in *E. coli*, the experimental evidence for binding of BRPs to SARPs promoters is still limited.

The complexity of the cascades of regulatory proteins is variable depending on the *Streptomyces* species. We will describe two examples of regulation of antibiotics with different structures.

6.3.2.1 Regulation of Streptomycin Biosynthesis Through the Pathway Specific strR Gene in *Streptomyces griseus*

A γ-butyrolactone receptor protein, named ArpA, binds a 22 palindromic specific ARE sequence upstream of the promoter of *adpA*, a gene encoding a transcriptional activator of the AraC/XylS family. Binding of ArpA results in the repression of *adpA* transcription. In the presence of nanomolar concentrations of A-factor, the A-factor–ArpA complex is released from the DNA and *adpA* transcription proceeds (Ohnishi et al. 1999). AdpA binds the promoter region of *strR*, the pathway-specific regulatory gene required to express streptomycin biosynthesis and resistance genes. In addition, AdpA activates genes of the *adpA*-regulon, including genes for morphogenesis and spore formation, metaloendopeptidases and serine proteases, and genes for grisazone biosynthesis and resistance.

6.3.2.2 Pathway-Specific Regulators of β-Lactam Biosynthesis in *Streptomyces clavuligerus*

S. clavuligerus produces several antibiotics, including the β-lactams cephamycin C and clavulanic acid, a β-lactamase inhibitor. A SARP type protein, CcaR, is required for the biosynthesis of cephamycin C and clavulanic acid (Pérez-Llarena et al. 1997). The clusters of genes for both compounds are located side by side in three different *Streptomyces* species (Ward and Hodgson 1993) and the *ccaR* gene is located in the cephamycin gene cluster.

Expression of *ccaR* is modulated by: (i) the probable binding of *S. clavuligerus* BRP to a ARE sequence located upstream of *ccaR,* what negatively affects antibiotic

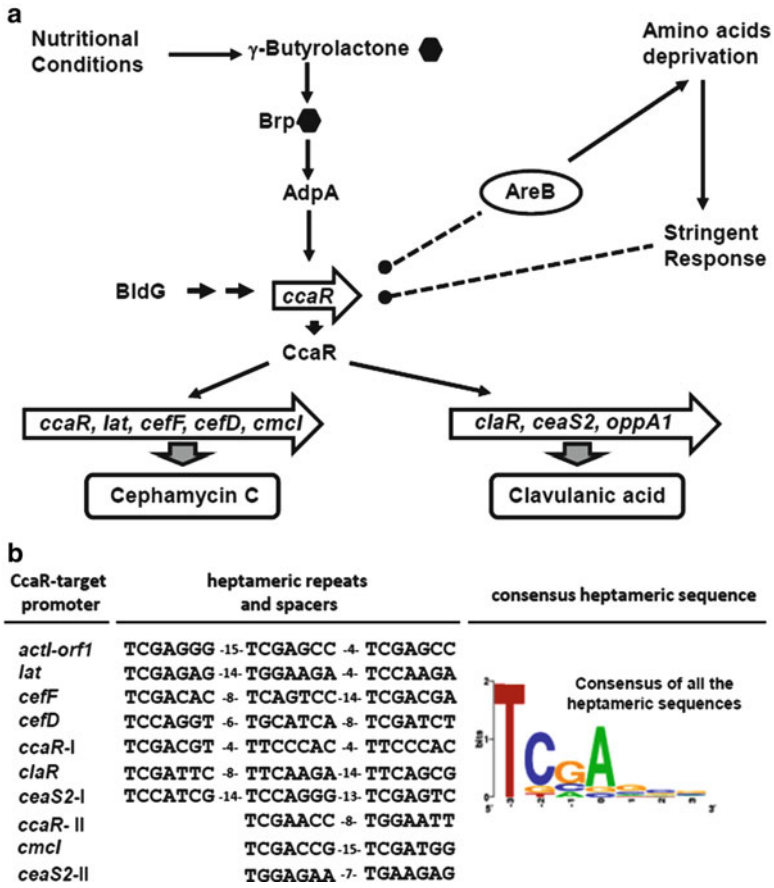


Fig. 6.1 (**a**) Scheme of the regulatory cascades that control cephamycin C and clavulanic acid biosynthesis in *S. clavuligerus*. Production of both compounds is controlled through the expression of *ccaR* by a cascade involving Brp and AdpA (Santamarta et al. 2005; López-García et al. 2010). It is modulated by the non-phosphorylated BldG (Bignell et al. 2005), AreB (Santamarta et al. 2007) and the stringent response (Gomez-Escribano et al. 2008). The presence of a butyrolactone (indicated with a pentagon) in *S. clavuligerus* has not been fully confirmed. Positive regulation is indicated by arrows, while negative regulation is indicated by *broken lines* and *solid circles*. (**b**) Heptameric sequences found upstream of cephamycin (*ccaR, lat, cefD, cmcI*) and clavulanic acid (*claR, ceaS2*) biosynthesis genes which are binding sites for CcaR. As reference the heptameric sequence found in the promoter of *actI-orf1* for actinorhodin biosynthesis is shown. The consensus heptameric sequence is indicated at the *right side*

production (Santamarta et al. 2005), (ii) the IclR-like regulator AreB, which also exerts pleiotropic effects on amino acids biosynthesis (particularly on expression of *leuCD*), fatty acid biosynthesis and utilization (Santamarta et al. 2007) (see below), and iii) the AdpA protein which increases *ccaR* expression seven-fold in the wild type strain in relation to a *S. clavuligerus adpA* mutant. The *adpA* gene itself contains an ARE sequence which is recognized by the BRP protein resulting in lower levels of expression of *adpA* (López-García et al. 2010) (Fig. 6.1a).

EMSA assays using rCcaR shows that CcaR binds the *ccaR* promoter controlling its own expression, the *lat* and *cefF* promoters and the bidirectional *cefD-cmcI* promoter region. In this way CcaR activates the expression of early, middle and late genes of the cephamycin pathway (Fig. 6.1).

In the clavulanic acid cluster the *claR* gene, encodes a LysR-type regulator which contains an HTH motif. Mutants disrupted in *claR* are unable to produce clavulanic acid but forms cephamycin. EMSA and Q-RT-PCR studies show that rCcaR binds the *ceaS* gene promoter, activating expression of the early genes for clavulanic acid formation. Moreover, binding of CcaR to the *claR* promoter controls expression of late clavulanic acid genes through a cascade mechanism. Footprinting and bioinformatic studies allowed the recognition of duplicated or triplicated heptameric boxes (TCGANNN) (Santamarta et al. 2011) for CcaR binding in all the recognized promoters (Fig. 6.1b). This box is in agreement with those proposed for SARP regulatory proteins of the actinorhodin, daunorubicin, valanimycin and fredericamycin gene clusters (Wietzorrek and Bibb 1997; Garg and Parry 2010).

6.3.2.3 Other Pathway Specific Regulators: The LAL and LysR-Type Regulators

Members of the LAL family of regulatory proteins, (originally typified by the regulator of the maltose regulon in *Escherichia coli* MalT) are frequent in some antibiotic gene clusters. These include *FkbN* for compound FK520 from *Streptomyces hygroscopicus* var. *ascomyceticus* (Wu et al. 2000). FkbN of *Streptomyces tsukubaensis* (Goranovič et al. 2010; Mo et al. 2012), PikD for pikromycin from *S. venezuelae* (Wilson et al. 2001), RapH for rapamycin from *Streptomyces hygroscopicus* (Aparicio et al. 1996; Molnár et al. 1996; Kuscer et al. 2007), Cvm7P which controls the formation of the 5S clavams (Tahlan et al. 2007) but not clavulanic acid formation in *S. clavuligerus,* NysRI/RIII for nystatin in *Streptomyces noursei* (Brautaset et al. 2000), AmphIV and PteF that regulate amphotericin and filipin of *Streptomyces nodosus* and *Streptomyces avermitilis,* respectively, and GdmRI/GdmRII for geldanamycin in *Streptomyces hygroscopicus* 17997 (He et al. 2008). The LAL regulatory proteins are relatively large in size (872–1159 amino acids) compared to the better studied SARPs (277–665 aa).

One of the best known examples of a LAL-regulated cascade is that of PimM in *Streptomyces natalensis.* The pimaricin gene cluster contains two genes, *pimM* and *pimR*, located back to back, encoding transcriptional activators (Anton et al. 2004, 2007). Both regulatory genes are required for pimaricin production. PimR encodes a SARP-type regulator with an additional C-terminal region homologous to guanilate cyclases. On the other hand PimM is a large protein containing an N-terminal PAS sensor domain and C-terminal LuxR-type HTH domain for DNA binding. Genes homologous to *pimM* are present in all known polyene biosynthesis gene clusters. GST-fused PimM (or GST-fused to the C terminal DNA-binding domain) gives DNA shift of 8 of the 12 promoters of the pimaricin cluster. PimM does not produce a band shift on the PimM-PimR bidirectional promote indicating a lack of

autoregulation. The 16 protected regions determined by footprinting reveals a 14 nt consensus sequence TVGGGAWWTCCCBA (where V is A, C or G; W is A or T; B is C, G or T), with dyad symmetry, in agreement with the recognition sequence of LuxR-type regulators, which overlaps with the −35 region of the bound promoter (Santos-Aberturas et al. 2011a). It seems that the transcriptional activation of LAL regulators is due to protein-protein contacts between the regulator and the RNA polymerase.

Genes homologous to PimM as well as 14 nt similar sequences for binding of the corresponding LAL regulator are present in other polyene biosynthesis gene clusters. These sequences are highly conserved. Indeed, a DNA fragment containing the 14 nt consensus sequence of an amphotericin biosynthesis gene of *S. nodosus* is retarded by PimM of *S. natalensis* in EMSA experiments. In addition the *pimM*-orthologous genes *amphIV, nysRIV* and *pteF* for regulation of amphotericin, nystatin and filipin in *S. nodosus, S. avermitilis* and *S. noursei,* complement a *pimM*-deleted mutant of *S. natalensis* (Santos-Aberturas et al. 2011b).

Members of the LysR-type family of transcriptional regulators have a size of about 310–325 amino acids, a C-terminal HTH (residues 1–60) and a well conserved substrate-binding domain (Maddocks and Oyston 2008). These regulators frequently require the presence of a small molecule (co-inducer). The DNA-binding domain of the LysR regulators is also involved in co-inducer recognition and/or response (Henikoff et al. 1988; Schell 1993). Some of the known LysR- regulators present in antibiotic gene clusters include ClaR, located in the clavulanic acid gene cluster of *S clavuligerus* (Pérez-Redondo et al. 1998), FkbR1 of *S. hygroscopicus* var. *ascomyceticus* (Wu et al. 2000) and its orthologous FkbR of *S. tsukubaensis* (Mo et al. 2012) or SAV_4790 in *S. avermitilis* (Omura et al. 2001).

6.4 A Connection Between Primary and Secondary Metabolism: AreB (NdgR)

Some pleiotropic regulators connect primary and secondary metabolism. This is the case of AreB of *S. clavuligerus* and its orthologue NdgR in *S. coelicolor* (Santamarta et al. 2005; Yang et al. 2009), and also of DasR.

AreB (ARE binding protein) was isolated by an affinity capture technique using as bait a DNA fragment located upstream of the SARP-type *ccaR* gene of *S. clavuligerus* that controls clavulanic acid and cephamycin C biosynthesis. This DNA fragment contains an ARE sequence that appears to be also the site of recognition by a butyrolactone-receptor protein.

The AreB regulator belongs to the IclR family of transcriptional regulators. A protein binding to the ARE sequence upstream of the regulatory *ccaR* gene of *S. clavuligerus* was isolated by DNA affinity binding. The *areB* gene, encoding this protein, is located upstream and in opposite orientation to the *leuCD* operon of *S. clavuligerus*; it encodes a 239-amino acid protein with a HTH motif at the

N-terminal region. An *areB*-deleted mutant of *S. clavuligerus* requires leucine for optimal growth in defined media. Expression of the *leuCD* operon is delayed in *S. clavuligerus areB*, because AreB binds the *areB-leuCD* intergenic region acting as a positive modulator. Clavulanic acid and cephamycin C production are improved in the *areB* mutant although no drastic difference in *ccaR* expression was observed in the *areB* mutant with respect to the parental strain.

Later, the AreB orthologous gene was identified in the doxorubicin producer *S. peucetius* and in the model actinomycete *S. coelicolor* (Yang et al. 2009) and named NdgR (for nitrogen source dependent growth regulator). *S. coelicolor* mutants deleted in *ndgR* showed reduced growth and defects in undecylprodigiosine production in minimal medium containing amino acids. No binding of NdgR to the promoter regions of *glnA, glnII* or *glnR* was shown, excluding a direct regulation of these nitrogen metabolism genes by NdgR. However, binding of NdgR to the promoter of the *scbR* gene (a butyrolactone synthesis regulator) suggests an indirect effect on nitrogen metabolism mediated by ScbR.

The NdgR protein binds to the intergenic region of *ndgR-leuCD* (as reported previously for AreB) and to the *scbR-scbA* region and results in many protein changes as shown by proteomic studies (Yang et al. 2009).

In summary AreB (NdgR) connects primary nitrogen metabolism with antibiotic production and differentiation. Interestingly this gene is well conserved in all *Streptomyces* species and many other related bacteria such as *Mycobacterium* and Corynebateria.

6.5 DasR: A Pleiotropic Regulator Controlling N-Acetylglucosamine Utilization and Antibiotic Biosynthesis in *S. coelicolor*

Another important pleiotropic regulator that connects primary and secondary metabolism in *S. coelicolor* is DasR (Rigali et al. 2008). DasR is primarily involved in the metabolism of N-acetylglucosamine (GlcNAc). This molecule is a favored carbon and nitrogen source for streptomycetes. Its intracellular catabolism requires the combined actions of the N-acetylglucosamine-6-phosphate deacetylase, NagA, and the glucosamine-6-phosphate deaminase/isomerase, NagB. GlcNAc acts as a signaling molecule in the DasR-mediated nutrient sensing system, activating development and antibiotic production under poor growth. Using *nagA, nagB,* and *nagK* mutants it was shown that the *nag* genes are part of the DasR regulon in *S. coelicolor,* what explains their transcriptional induction by GlcNAc. In *S. coelicolor*, GlcNAc is internalized as GlcNAc-6-P via the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS).

GlcNAc is an important signaling molecule for streptomycetes and a major decision point towards the onset of development and antibiotic production (Rigali et al. 2006, 2008). Under rich growth conditions, GlcNAc promotes growth; thereby blocking differentiation processes, while under poor nutritional conditions GlcNAc

promotes development and antibiotic production (Rigali et al. 2008; Colson et al. 2008; van Wezel and McDowall 2011). Under famine conditions, a complete signaling pathway was proposed, from GlcNAc uptake to the onset of antibiotic production, whereby the pleiotropic regulator DasR controls the genes for GlcNAc transport and metabolism and for antibiotic synthesis (Rigali et al. 2006, 2008). Glucosamine-6-phosphate is a key molecule in this signaling cascade and acts as an allosteric effector preventing the DNA binding ability of DasR. This results in the loss of transcriptional repression of specific activator genes for antibiotic production, including *actII*-ORF4, encoding the pathway-specific activator for actinorhodin production, and *redZ*, encoding the response regulator RedZ that activates undecylprodiginine production. DasR also represses the *cpk* gene cluster encoding the cryptic polyketide Cpk, but this effect is probably indirect (van Wezel et al. 2009).

6.6 The Divalent Metal Regulator DmdR1 Links Iron Limitation Stress with Primary and Secondary Metabolism

One of the connections between stress sensing and primary and secondary metabolism in which the signal transduction cascade is well known is that of iron deficiency or iron overload stress responses in *S. coelicolor* (Tunca et al. 2007). The bioavailability of iron is very low because salts of the oxidized ferric ion formed under normal oxic conditions are largely insoluble (Crosa 1997). Iron deficiency prevents efficient growth since iron is an essential element with very important metabolic roles, e.g., in cytochromes and Fe-S containing proteins. On the opposite side, iron overload is highly damaging to the cell because iron triggers the Fenton reaction resulting in the formation of reactive oxygen species (ROS) that react with DNA and proteins (Galaris and Pantopoulos 2008).

The signal transduction cascade in this case is mediated by a strict control of the biosynthesis of siderophores and siderophore-mediated iron transport; both processes are controlled by the DmdR1 regulator.

To take up iron from soil many microorganisms synthesize high-affinity iron chelators (siderophores) (Neilands 1995). *Streptomyces* produce different types of siderophores (Muller and Raymond 1984; Bunet et al. 2006). Desferrioxamines are nonpeptide hydroxamate siderophores composed of alternating dicarboxylic acid and diamine units linked by amide bonds. They are produced by many *Streptomyces* species (Imbert et al. 1995; Yamanaka et al. 2005; Shupp et al. 1987, 1988; Lautru et al. 2005; Dimkpa et al. 2008). The expression of genes involved in iron metabolism in actinomycetes is under the control of a divalent metal-dependent regulatory protein (DmdR) analogous to the diphtheria toxin repressor of *Corynebacterium diphtherieae*. *S. coelicolor* has two similar genes, *dmdR1* and *dmdR2*, encoding regulatory proteins of this family (Flores et al. 2005) of which DmdR1 is the main one.

Flores and Martín (2004) reported that *S. coelicolor* DmdR1 binds specific sequences (iron boxes) in the upstream region of the diphtheria toxin (*tox*) gene of *C. diphtheriae* and the *desA* gene of *Streptomyces pilosus*. At least ten putative iron boxes were found upstream of 10 different ORFs in the genome of *S. coelicolor* (Flores and Martín 2004). One of the iron boxes is located in the promoter of the *desABCD* gene cluster, which is responsible for desferrioxamine biosynthesis (Bentley et al. 2002; Barona-Gómez et al. 2004; Challis 2005). The first step in the desferrioxamine pathway is the conversion of L-lysine into cadaverine catalyzed by the enzyme lysine decarboxylase (Shupp et al. 1987, 1988) which, in *S. coelicolor* is encoded by *desA* (Tunca et al. 2007). Transcriptional analysis of the *desABCD* cluster showed that the genes involved in desferrioxamine production are under iron control mediated by the DmR1 transcriptional regulator. The transcription start point of the *desA* gene overlaps with the palindromic iron box. Binding of purified DmdR1 protein to the *desA* promoter region of both *S. coelicolor* and *S. pilosus* proved that iron control of the expression of the *desABCD* cluster is mediated by the DmdR1 regulator (Tunca et al. 2007), and by the product of a gene overlapping *dmdR1,* which is expressed in the opposite orientation to *dmdR1* (Tunca et al. 2009).

In addition to the *desABCD* cluster involved in desferrioxamines B and E biosynthesis two additional genes, *desE* and *cdtB,* have been proposed to recognize and transport ferrioxamines B and E (molecules carrying an iron atom), respectively (Tierrafria et al. 2011). Disruption of *desE* in *S. coelicolor* results in impaired growth and lack of sporulation whereas supplementation of the impaired mutant with desferrioxamine E triggers actinorhodin production.

Other iron boxes that appear to bind DmdR1 are located upstream of four different coelichelin biosynthesis genes, *cchA, cchE, cchJ* and *cchF.* Although sensing and transport of iron-loaded ferrioxamines seem to be exerted by the DesE and CdtB transporters, the molecular mechanisms by which desferrioxamines influence antibiotic biosynthesis is still unclear.

6.7 Network Connecting Wide-Domain Regulators

Different types of wide-domain (also named pleiotropic) regulators are known in *Streptomyces*. Some of them, including the paired two component systems, the orphan response regulators and the stringent response (Gómez-Escribano et al. 2008) have been reviewed elsewhere (Bibb 2005; Liras et al. 2008; Martín et al. 2012b) and therefore are only summarized in this article.

6.7.1 Two-Component Systems

The genome of *Streptomyces* species is very large (from 8 to 10.5 Mb). Sixty-seven paired two component systems (TCS), 13 orphan response regulators and 17 unpaired sensor kinases have been identified in the genome of *S. coelicolor*

(Bentley et al. 2002) and similar numbers of TCS occur in the genome of other *Streptomyces* species. The impressive array of these TCS has been reviewed recently (Hutchings et al. 2004; Martín et al. 2012b). However, the stress signals or other environmental imputs recognized by these systems are largely unknown. Two of the best known TCSs are the PhoR-PhoP and the AbsA1-AbsA2 systems. About 20 additional TCSs have been studied in some detail in different species of *Streptomyces*. They have some features in common with PhoR-PhoP and several of them have effect on antibiotic production and differentiation (Martín et al. 2012a).

The PhoR-PhoP system belongs to class IIIA of two-component systems (Hutchings et al. 2004). The PhoR protein is a standard membrane sensor kinase although it contains only one full transmembrane spanning domain and a short additional hydrophobic stretch. PhoP is a member of the OmpR family of DNA-binding response regulators. PhoP, as other TCS response regulators, contains the DNA-binding domain (DBD) in the C terminal region (amino acids 187–201) (Sola-Landa et al. 2005). Phosphate regulation of antibiotic biosynthesis in *Streptomyces lividans* is mediated by the two-component PhoR–PhoP system (Sola-Landa et al. 2003; Martín 2004; Ghorbel et al. 2006) that also controls the alkaline phosphatase gene (*phoA*) and other related genes (Apel et al. 2007). In *S. lividans* expression of *phoA* is induced by phosphorylated PhoP and mutants lacking phoP (or the phoR–phoP cluster) do not synthesize PhoA. These mutants overproduce large amounts of actinorhodin and undecylprodigiosin (Sola-Landa et al. 2003). However some differences exist between *Streptomyces* species. In liquid cultures *S. coelicolor* mutants deleted in *phoP* behave in a different manner with respect to *Streptomyces lividans*; although actinorhodin and undecylprodigiosin are under phosphate control, the *S. coelicolor phoP* mutants do not overproduce these pigmented metabolites (Santos-Beneit et al. 2009). The available information indicates that pigmented antibiotic biosynthesis in *S. lividans* is under a more strict phosphate control than in *S. coelicolor* and the derepressing effect observed in the *S. lividans PhoP* mutants is stronger than in *S. coelicolor*.

The response regulator PhoP binds to a large number of well-characterized promoters (Martín et al. 2012b) but not to the best known promoters of genes involved in actinorhodin or undecylprodigiosin biosynthesis. The PhoP-mediated regulation of the secondary metabolism takes place through the control of the expression of *afsS*, a small σ-factor-like protein, whose amplification in a high-copy number plasmid leads to an overproduction of actinorhodin and undecylprodigiosin in both *S. coelicolor* and *S. lividans* (Vögtli et al. 1994; Matsumoto et al. 1995; Floriano and Bibb 1996). PhoP downregulates the expression of *afsS* by direct competition with AfsR, since both PhoP and AfsR regulators recognize overlapping sequences in the *afsS* promoter (Santos-Beneit et al. 2009, 2011b). In addition, PhoP affects expression of the promoter of the *S. coelicolor* RNA polymerase omega factor *rpoZ*, required for normal biosynthesis of actinorhodin and undecylprodigiosin (Santos-Beneit et al. 2011a).

Usually PhoP binds to a PHO box around the −35 region of the PhoP-regulated promoters and stimulates transcription by helping to recruit the RNA polymerase and to stabilize the RNA-polymerase-promoter DNA complex. Although PhoP is usually a transcriptional activator, sometimes it may work as a repressor. This is the

case of PhoP repression of the *hrdA* and *rpoZ* genes, encoding, respectively, a σ−factor and the omega factor of the RNA polymerase (Sola-Landa et al. 2008; Santos-Beneit et al. 2011a). This repression is due to a 'road-block' effect when PhoP binds to the −10 to +1 region of the promoter thus interfering with RNA polymerase binding.

6.7.2 *Other Two-Component Systems Involved in Control of Secondary Metabolite Biosynthesis*

Another very interesting TCS is the *absA1-absA2* system encoding a sensor kinase with phosphatase activity, AbsA1, and the response regulator AbsA2. The *absA1-absA2* operon is located within the CDA biosynthetic cluster and has strong regulatory effects on CDA production but surprisingly, exerts also important effects on actinorhodin and undecylprodigiosin biosynthesis. Certain mutant alleles of *absA1* cause nearly complete inhibition of the biosynthesis of these three metabolites in *S. coelicolor*. It is known that the ratio of protein kinase/phosphatase activity of the sensor kinase is an important parameter in the action of a TCS (Stewart 2010). The cytoplasmic domain of AbsA1 can phosphorylate AbsA2 but also dephosphorylate it. Interestingly, a mutation that impairs the kinase activity of AbsA1 enhances antibiotic production *in vivo*, while those mutations that impair the phosphatase activity acting on AbsA2-P dramatically reduce antibiotic production (Sheeler et al. 2005). These results clearly indicate that the phosphorylated form of the response regulator AbsA2-P is a repressor of antibiotic production (Ryding et al. 2002).

Transcript analysis indicated that mutations in the *absA* locus strongly influence the expression of the CDA biosynthesis genes. An interesting question is if there is a direct interaction of AbsA2-P with its targets; McKenzie and Nodwell (2007) have used a chromatin immunoprecipitation (ChIP) assay to identify the targets of AbsA2-P. This protein interacts with the *actII-ORF4* and *redZ* promoter regions *in vivo* and *in vitro*, and the *in vitro* interactions are stimulated by AbsA2 phosphorylation. However, the precise AbsA2-P binding sequence has not been established yet.

6.8 Cross-Talk Between Global Regulators: Regulatory Networks

The discovery of cross-talk between global regulators in *Streptomyces* has highlighted the importance of networks of regulatory systems. An increasing number of cross-talk regulations are reported every year. Interesting examples are the cross-regulation of PhoP and GlnR and the cross-regulation of PhoP and AfsR resulting in the modulation of the antibiotic clusters associated genes *actII-orf4* and *redD* in *S. coelicolor* (Santos Beneit et al. 2009, 2012).

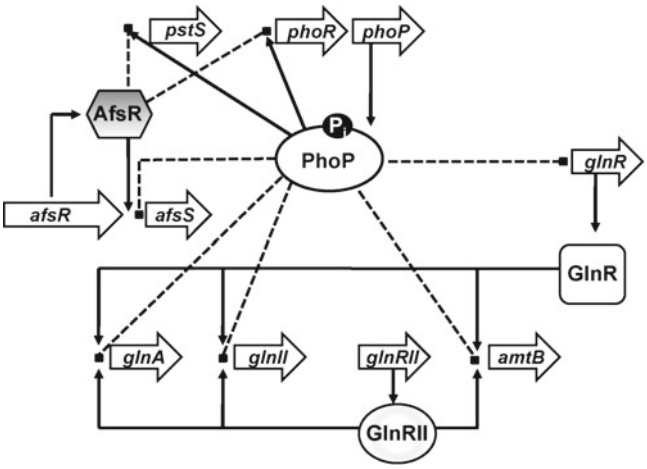


Fig. 6.2 Cross-regulation of PhoP and AfsR and control of *glnR* and *amtB* by PhoP in *S. coelicolor*. GlnRII binds to the same sequences as GlnR but expression of *glnRII* is not controlled by PhoP. Positive regulation at the different promotors is indicated by *continuous lines* and *arrows*. Negative regulation is indicated by *broken lines* and *solid squares*

6.8.1 Cross-Talk of PhoP and GlnR

In addition to its role on phosphate uptake and secondary metabolism PhoP also affects other nutritional regulators (Martín et al. 2011). An important interaction is that between PhoP and the genes involved in nitrogen metabolism. In this case the main regulator is GlnR that controls directly or indirectly the expression of numerous genes (Wray and Fisher 1993; Reuther and Wohlleben 2007; Tiffert et al. 2008, 2011; Pullan et al. 2011). A few years ago it was shown that some genes related to nitrogen metabolism are overexpressed in a Δ*phoP* mutant after a phosphate concentration downshift (Rodríguez-García et al. 2007). More recently it was shown that PhoP binds directly to the promoters of several genes involved in nitrogen assimilation, including that of glnR (Fig. 6.2) (Rodríguez-García et al. 2009).

Unbalanced primary metabolism caused by phosphate or nitrogen starvation is undesirable for the cell and triggers a stress response that leads to survival efforts by scavenging residual nutrients and to the production of secondary metabolites.

6.8.2 Cross-Regulation of PhoP and AfsR

Another important example of interactions between global regulators is the cross-talk of PhoP with AfsR, a member of the STAND (signal transduction ATPases with numerous domains) family. This protein contains an N-terminal SARP domain,

a central ATPase domain and a terminal tetratrico peptide repeat (TPR) domain. In AfsR only the SARP domain is strictly required to form a complex with the RNA polymerase and the promoter of the *afsS* gene; this interaction is essential for *afsS* transcription and, therefore, for antibiotic production (Fig. 6.2). Two repeated nine nucleotide sequences were reported to be protected by AfsR, covering the −40 to −15 region of the *afsS* sense strand (Tanaka et al. 2007). Surprisingly the AfsR binding sequence in the *afsS* promoter overlaps almost fully with a PhoP-binding sequence (PHO box) (Santos-Beneit et al. 2009). A fine study of the binding of PhoP and AfsR to this region allowed discriminating between the important nucleotides for binding of each of these two global regulators in the common nucleotide sequence (Santos-Beneit et al. 2011b).

6.9 Atypical Response Regulators with Ligands: Role in the Feedback Regulation of Secondary Metabolite Biosynthesis

In the response regulators five conserved amino acids are important for phosphorylation of these proteins at the aspartic acid site. However, a subset of the orphan response regulators lack two or three of these conserved amino acids (Hutchings 2007). These proteins have been named "atypical response regulators" (ARR) (Wang et al. 2009) and may require small ligands (Guthrie et al. 1998; Ainsa et al. 1999; Galperin 2006; Pflock et al. 2007; Tian et al. 2007; Wang et al. 2009). Although their overall structure is similar to that of the classical response regulators it is unclear if these ARRs are really phosphorylated *in vivo.* These ARRs control a variety of processes including virulence, cell differentiation, iron transport and secondary metabolism (Guthrie et al. 1998; Ulijasz et al. 2004; Pflock et al. 2007; Tian et al. 2007).

One of these ARRs in *S. coelicolor* is RedZ belonging to the NarL family. RedZ is an orphan response regulator encoded by a gene located in the *red* gene cluster for undecylprodigiosin biosynthesis. RedZ has been shown to regulate expression of the *red* gene cluster by binding to the promoter region of the *redD* gene, a SARP-type pathway-specific regulator. Binding of purified RedZ to the *redD* promoter was inhibited by undecylprodigiosin, whereas this compound alone was unable to bind directly to the DNA even at very high concentrations (Wang et al. 2009). By this mechanism *Streptomyces* species control expression of antibiotic biosynthetic genes using as a regulatory ligand, the final product of the pathway.

A similar mode of action was described for the atypical response regulator JadR1 of *S. venezuelae* that activates the transcription of jadomycin biosynthetic genes. Jadomycin B binds to the N-terminal region of JadR1 causing its release from the target promoter. In summary, antibiotic biosynthesis appears to be controlled by a feedback autoregulatory mechanism when high levels of jadomycin are accumulated (Wang et al. 2009).

6.10 Future Perspectives

The genomes of *Streptomyces* contain a large amount of genetics information (up to 10.5 Mb) that corresponds to their ability to utilize a variety of nutrients and to respond to drastic changes in environmental and nutritional conditions. Part of this genetic information is dedicated to the biosynthesis of a wealth of bioactive secondary metabolites with important cell to cell communication roles. Although the application of the "omics" tools to the molecular analysis of *Streptomyces* metabolism has allowed considerable progress in our understanding of antibiotic biosynthesis pathways, the knowledge of regulatory mechanisms is still sketchy: We have information on the mechanisms of some regulatory systems (e.g. phosphate control and nitrogen source regulation) but we lack a broad view of the overall regulatory mechanisms as required for systems biology. The arising information on cross-talk between global regulators provides a glimpse of the complexity of the regulatory interactions. Networks have been implemented by evolution as a safeguard of metabolism against drastic nutritional unbalances. These unbalances lead to the production of antibiotics and other bioactive metabolites as "alarm signals". Therefore, we need to know much better the regulatory mechanisms that control and limit antibiotic production for an adequate biotechnological application.

Acknowledgements This article was supported by Grants BIO2009-09820 to P. Liras and BIO2010-16094 to Juan F. Martín, of the CICYT, Spanish Ministry of Economy and Competitivity, and by the European ERA-IB2 Project PIM2010EE1-00677 "INMUNOTEC" to both authors.

References

Aínsa JA, Parry HD, Chater KF (1999) A response regulator-like protein that functions at an intermediate stage of sporulation in *Streptomyces coelicolor* A3(2). Mol Microbiol 34:607–619

Antón N, Mendes MV, Martín JF, Aparicio JF (2004) Identification of PimR as a positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. J Bacteriol 186:2567–2575

Antón N, Santos-Aberturas J, Mendes MV, Guerra SM, Martín JF, Aparicio JF (2007) PimM, a PAS domain positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. Microbiology 153:3174–3183

Aparicio JF, Molnár I, Schwecke T, König A, Haydock SF, Khaw LE, Staunton J, Leadlay PF (1996) Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene 169:9–16

Apel AK, Sola-Landa A, Rodríguez-García A, Martín JF (2007) Phosphate control of *phoA, phoC* and *phoD* gene expression in *Streptomyces coelicolor* reveals significant differences in binding of PhoP to their promoter regions. Microbiology 153:3527–3537

Arias P, Fernández-Moreno MA, Malpartida F (1999) Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2) as a DNA-binding protein. J Bacteriol 181:6958–6968

Barona-Gómez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J Am Chem Soc 126:16282–16283

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature 417:141–147

Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol 8:208–215

Bignell DR, Tahlan K, Colvin KR, Jensen SE, Leskiw BK (2005) Expression of ccaR, encoding the positive activator of cephamycin C and clavulanic acid production in *Streptomyces clavuligerus*, is dependent on bldG. Antimicrob Agents Chemother 49:1529–1541

Brautaset T, Sekurova ON, Sletta H, Ellingsen TE, StrŁm AR, Valla S, Zotchev SB (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. Chem Biol 7:395–403

Bunet R, Brock A, Rexer UH, Takano A (2006) Identification of genes involved in siderophore transport in *Streptomyces coelicolor* A3 (2). FEMS Microbiol Lett 262:57–64

Challis G (2005) A widely distributed bacterial pathway for siderophore biosynthesis independent of non ribosomal peptide. ChemBioChem 6:601–611

Choi SU, Lee CK, Hwang YI, Kinosita H, Nihira T (2003) γ-Butyrolactone autoregulators and receptor proteins in non-*Streptomyces* actinomycetes producing commercially important secondary metabolites. Arch Microbiol 180:303–307

Colson S, van Wezel GP, Craig M, Noens EE, Nothaft H, Mommaas AM, Titgemeyer F, Joris B, Rigali S (2008) The chitobiose-binding protein, DasA, acts as a link between chitin utilization and morphogenesis in *Streptomyces coelicolor*. Microbiol 154:373–382

Crosa JH (1997) Signal transduction and transcriptional and post-transcriptional control or iron-regulated genes in bacteria. Microbiol Mol Biol Rev 61:319–336

Cundliffe E (2008) Control of tylosin biosynthesis in *Streptomyces fradiae*. J Microbiol Biotechnol 18:1485–1491

Cundliffe E, Demain AL (2010) Avoidance of suicide in antibiotic-producing microbes. J Ind Microbiol Biotechnol 37:643–672

Davies J (2007) Small molecules: the lexicon of biodiversity. J Biotechnol 29:3–5

Demain AL (2009) Antibiotics: natural products essential to human health. Med Res Rev 29:821–842

Dimkpa C, Svatos A, Merten D, Büchel G, Kothe E (2008) Hydroxamate siderophores produced by *Streptomyces acidiscabies* E13 bind nickel and promote growth in cowpea (*Vigna unguiculata* L.) under nickel stress. Can J Microbiol 54:163–172

Flores FJ, Martín JF (2004) Iron-regulatory proteins DmdR1 and DmdR2 of *Streptomyces coelicolor* form two different DNA-protein complexes with iron boxes. Biochem J 380:497–503

Flores FJ, Barreiro C, Coque JJR, Martín JF (2005) Functional analysis of Two divalent metal-dependent genes *dmdR1* and *dmdR2* in *Streptomyces coelicolor* and proteome changes in deletion mutants. FEBS J 272:725–735

Floriano B, Bibb M (1996) *afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). Mol Microbiol 21:385–396

Folcher M, Gaillard H, Nguyen LT, Nguyen KT, Lacroix P, Bamas-Jacques N, Rinkel M, Thompson CJ (2001) Pleiotropic functions of a *Streptomyces pristinaespiralis* autoregulator receptor in development, antibiotic biosynthesis, and expression of a superoxide dismutase. J Biol Chem 276:44297–44306

Galaris D, Pantopoulos K (2008) Oxidative stress and iron homeostasis: mechanistic and health aspects. Crit Rev Clin Lab Sci 45:1–23

Galperin MY (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J Bacteriol 188:4169–4182

Garg RP, Parry RJ (2010) Regulation of valanimycin biosynthesis in *Streptomyces viridifaciens*: characterization of VlmI as a *Streptomyces* antibiotic regulatory protein (SARP). Microbiol 156:472–483

Ghorbel S, Kormanec J, Artus A, Virolle MJ (2006) Transcriptional studies and regulatory interactions between the phoR-phoP operon and the phoU, mtpA and ppk genes of *Streptomyces* lividans TK24. J Bacteriol 188:677–686

Gomez-Escribano JP, Martín JF, Hesketh A, Bibb MJ, Liras P (2008) *Streptomyces clavuligerus relA*-null mutants overproduce clavulanic acid and cephamycin C: negative regulation of secondary metabolism by (p)ppGpp. Microbiol 154:744–755

Goranovič D, Kosec G, Mrak P, Fujs S, Horvat J, Kuščer E, Kopitar G, Petković H (2010) Origin of the allyl group in FK506 biosynthesis. J Biol Chem 285:14292–14300

Guthrie EP, Flaxman CS, White J, Hodgson DA, Bibb MJ, Chater KF (1998) A response-regulator-like activator of antibiotic synthesis from *Streptomyces coelicolor* A3(2) with an amino-terminal domain that lacks a phosphorylation pocket. Microbiol 144:727–738

Hara H, Ohnishi Y, Horinouchi S (2009) DNA microarray analysis of global gene regulation by A-factor in *Streptomyces griseus*. Microbiol 155:2197–2210

He W, Lei J, Liu Y, Wang Y (2008) The LuxR family members GdmRI and GdmRII are positive regulators of geldanamycin biosynthesis in *Streptomyces hygroscopicus* 17997. Arch Microbiol 189:501–510

Henikoff S, Haughn GW, Calvo JM, Wallace JC (1988) A large family of bacterial activator proteins. Proc Natl Acad Sci USA 85:6602–6606

Hesketh A, Chen WJ, Ryding J, Chang S, Bibb M (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). Genome Biol 8:R161

Hesketh A, Bucca G, Laing E, Flett F, Hotchkiss G, Smith CP, Chater KF (2008) New pleiotropic effects of eliminating a rare tRNA from *Streptomyces coelicolor*, revealed by combined proteomic and transcriptomic analysis of liquid cultures. BMC Genomics 8:261

Horinouchi S (2007) Mining and polishing of the treasure trove in the bacterial genus *Streptomyces*. Biosci Biotechnol Biochem 71:283–299

Horinouchi S, Beppu T (2007) Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. Proc Jpn Acad 83:277–295

Huang J, Lih CJ, Pan KH, Cohen SN (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. Genes Dev 15:3183–3192

Huang J, Shi J, Molle V, Sohlberg B, Weaver D, Bibb MJ, Karoonuthaisiri N, Lih CJ, Kao CM, Buttner MJ, Cohen SN (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. Mol Microbiol 58:1276–1287

Hutchings MI (2007) Unusual two-component signal transduction pathways in the actinobacteria. Adv Appl Microbiol 61:1–26

Hutchings MI, Hoskisson PA, Chandra G, Buttner MJ (2004) Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of *Streptomyces coelicolor*A3(2). Microbiol 150:2795–2806

Imbert M, Béchet M, Blondeau R (1995) Comparison of the main siderophores produced by some species of *Streptomyces*. Curr Microbiol 31:129–133

Kato JY, Funa N, Watanabe H, Ohnishi Y, Horinouchi S (2007) Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. Proc Natl Acad Sci USA 104:2378–2383

Kim HS, Lee YJ, Lee CK, Choi SU, Yeo SH, Hwang YI, Yu TS, Kinoshita H, Nihira T (2004) Cloning and characterization of a gene encoding the gamma-butyrolactone autoregulator receptor from *Streptomyces clavuligerus*. Arch Microbiol 182:44–50

Kitani S, Miyamoto KT, Takamatsu S, Herawati E, Iguchi H, Nishitomi K, Uchida M, Nagamitsu T, Omura S, Ikeda H, Nihira T (2011) Avenolide, a Streptomyces hormone controlling antibiotic production in *Streptomyces avermitilis*. Proc Natl Acad Sci USA 108:16410–16415

Kolter R, Siegele DA, Tormo A (1993) The stationary phase of the bacterial life cycle. Annu Rev Microbiol 47:855–874

Kuscer E, Coates N, Challis I, Gregory M, Wilkinson B, Sheridan R, Petković H (2007) Roles of *rapH* and *rapG* in positive regulation of rapamycin biosynthesis in *Streptomyces hygroscopicus*. J Bacteriol 189:4756–4763

Lautru S, Deeth R, Bailey L, Challis G (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat Chem Biol 1:265–269

Lian W, Jayapal KP, Charaniya S, Mehra S, Glod F, Kyung YS, Sherman DH, Hu WS (2008) Genome-wide transcriptome analysis reveals that a pleiotropic antibiotic regulator, AfsS, modulates nutritional stress response in *Streptomyces coelicolor* A3(2). BMC Genomic 9:56

Liras P, Gomez-Escribano JP, Santamarta I (2008) Regulatory mechanisms controlling antibiotic production in *Streptomyces clavuligerus*. J Ind Microbiol Biotechnol 35:667–676

López-García MT, Santamarta I, Liras P (2010) Morphological differentiation and clavulanic acid formation are affected in a *Streptomyces clavuligerus adpA*-deleted mutant. Microbiol 156:2354–2365

Maddocks SE, Oyston PC (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiol 154:3609–3623

Manteca A, Sánchez J, Jung HR, Schwämmle V, Jensen ON (2010a) Quantitative proteomics analysis of *Streptomyces coelicolor* development demonstrates that onset of secondary metabolism coincides with hypha differentiation. Mol Cell Proteomics 9:1423–1436

Manteca A, Jung HR, Schwämmle V, Jensen ON, Sánchez J (2010b) Quantitative proteome analysis of *Streptomyces coelicolor* Nonsporulating liquid cultures demonstrates a complex differentiation process comparable to that occurring in sporulating solid cultures. J Proteome Res 3:4801–4811

Martín JF (2004) Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR–PhoP system: an unfinished story. J Bacteriol 186:5197–5201

Martín JF, Demain AL (1980) Control of antibiotic biosynthesis. Microbiol Rev 44:230–251

Martín JF, Liras P (1989) Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. Ann Rev Microbiol 43:173–206

Martín JF, Liras P (2010) Engineering of regulatory cascades and networks controlling antibiotic biosynthesis in *Streptomyces*. Curr Opin Microbiol 13:263–273

Martín JF, Sola-Landa A, Santos-Beneit F, Fernández-Martínez LT, Prieto C, Rodríguez-García A (2011) Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. Microb Biotechnol 4:165–174

Martín JF, Sola-Landa A, Rodríguez-García A (2012a) Two-component systems in *Streptomyces*. In: Gross R (ed) Two component systems in Bacteria. Horizon Sci Press, Norfolk, pp 315–331

Martín JF, Santos-Beneit F, Rodríguez-García A, Sola-Landa A, Smith MCM, Ellingsen TE, Nieselt K, Burroughs NJ, Wellington EMH (2012b) Transcriptomic studies of phosphate control of primary and secondary metabolism in *Streptomyces coelicolor*. Appl Microb Biotechnol 95:61–75

Matsumoto A, Ishizuka H, Beppu T, Horinouchi S (1995) Involvement of a small ORF downstream of the *afsR* gene in the regulation of secondary metabolism in *Streptomyces coelicolor* A3(2). Actinomycetologica 9:37–43

McKenzie NL, Nodwell JR (2007) Phosphorylated AbsA2 negatively regulates antibiotic production in *Streptomyces coelicolor* through interactions with pathway-specific regulatory gene promoters. J Bacteriol 189:5284–5292

Mo S, Yoo YJ, Ban YH, Lee SK, Kim E, Suh JW, Yoon YJ (2012) Roles of *fkbN* in Positive Regulation and tcs7 in Negative Regulation of FK506 Biosynthesis in *Streptomyces* sp. strain KCTC 11604BP. Appl Environ Microbiol 78:2249–2255

Molnár I, Aparicio JF, Haydock SF, Khaw LE, Schwecke T, König A, Staunton J, Leadlay PF (1996) Organisation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of genes flanking the polyketide synthase. Gene 169:1–7

Muller G, Raymond KN (1984) Specificity and mechanism of ferrioxamine-mediated iron transport in *Streptomyces pilosus*. J Bacteriol 160:304–312

Natsume R, Ohnishi Y, Senda T, Horinouchi S (2004) Crystal structure of a gamma-butyrolactone autoregulator receptor protein in *Streptomyces coelicolor* A3(2). J Mol Biol 336:409–419

Neilands JB (1995) Siderophores: structure and function of microbial iron-transport compounds. J Biol Chem 270:26723–26726

Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen OM, Sletta H, Alam MT, Merlo ME, Moore J et al (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC Genomic 11:10

Nishida H, Ohnishi Y, Beppu T, Horinouchi S (2007) Evolution of gamma-butyrolactone synthases and receptors in Streptomyces. Environ Microbiol 9:1986–1994

Ohnishi Y, Kameyama S, Onaka H, Horinouchi S (1999) The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: identification of a target gene of the A-factor receptor. Mol Microbiol 34:102–111

Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc Natl Acad Sci USA 98:12215–12220

Pérez-Llarena FJ, Liras P, Rodriguez-García A, Martín JF (1997) A regulatory gene (*ccaR*) required for cephamycin and clavulanic acid productionin *Streptomyces clavuligerus*: amplification results in overproduction of both beta-lactam compounds. J Bacteriol 179:2053–2059

Pérez-Redondo R, Rodríguez-García A, Martín JF, Liras P (1998) The *cla*R gene of *Streptomyces clavuligerus*, encoding a LysR-type regulatory protein controlling clavulanic acid biosynthesis, is linked to the clavulanate-9-aldehyde reductase (*car*) gene. Gene 211:311–321

Pflock M, Bathon M, Schär J, Müller S, Mollenkopf H, Meyer TF, Beier D (2007) The orphan response regulator HP1021 of *Helicobacter pylori* regulates transcription of a gene cluster presumably involved in acetone metabolism. J Bacteriol 189:2339–2349

Pullan ST, Chandra G, Bibb MJ, Merrick M (2011) Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. BMC Genomic 12:175–189

Recio E, Colinas A, Rumbero A, Aparicio JF, Martín JF (2004) PI factor, a novel type quorum-sensing inducer elicits pimaricin production in *Streptomyces natalensis*. J Biol Chem 279:41586–41593

Reuther J, Wohlleben W (2007) Nitrogen metabolism in *Streptomyces coelicolor*: transcriptional and post-translational regulation. J Mol Microbiol Biotechnol 12:139–146

Rigali S, Nothaft H, Noens EE, Schlicht M, Colson S, Müller M, Joris B, Koerten HK, Hopwood DA, Titgemeyer F, van Wezel GP (2006) The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. Mol Microbiol 61:1237–1251

Rigali S, Titgemeyer F, Barends S, Mulder S, Thomae AW, Hopwood DA, van Wezel GP (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. EMBO Rep 9:670–675

Rodríguez-García A, Barreiro C, Santos-Beneit F, Sola-Landa A, Martín JF (2007) Genome-wide transcriptomic and proteomic analysis of the primary response to phosphate limitation in *Streptomyces coelicolor* M145 and in a Δ*phoP* mutant. Proteomics 7:2410–2429

Rodríguez-García A, Sola-Landa A, Apel K, Santos-Beneit F, Martín JF (2009) Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of *glnR*, *glnA*, *glnII* and *amtB* expression by the response regulator PhoP. Nucleic Acids Res 37:3230–3242

Ryding NJ, Anderson TB, Champness WC (2002) Regulation of the *Streptomyces coelicolor* calcium-dependent antibiotic by *absA*, encoding a cluster-linked two-component system. J Bacteriol 184:794–805

Salehi-Najafabadi Z, Barreiro C, Martínez-Castro M, Solera E, Martín JF (2011) Characterization of a γ-butyrolactone receptor of *Streptomyces tacrolimicus*: effect on sporulation and tacrolimus biosynthesis. Appl Microbiol Biotechnol 92:971–984

Santamarta I, Rodríguez-Garcia A, Pérez-Redondo R, Martín JF, Liras P (2002) CcaR is an Autoregulatory Protein that binds to the *ccaR* and the *cefD-cmcI* promoters of the cephamycin C-clavulanic acid cluster in *Streptomyces clavuligerus*. J Bacteriol 184:3106–3113

Santamarta I, Pérez-Redondo R, Lorenzana LM, Martín JF, Liras P (2005) Different proteins bind to the butyrolactone receptor protein ARE sequence located upstream of the regulatory ccaR gene of *Streptomyces clavuligerus*. Mol Microbiol 56:824–835

Santamarta I, López-García MT, Pérez-Redondo R, Koekman B, Martín JF, Liras P (2007) Connecting primary and secondary metabolism: AreB, an IclR-like protein, binds the ARE(*ccaR*) sequence of *S. clavuligerus* and modulates leucine biosynthesis and cephamycin C and clavulanic acid production. Mol Microbiol 66:511–524

Santamarta I, López-García MT, Kurt A, Nárdiz N, Alvarez-Álvarez R, Pérez-Redondo R, Martín JF, Liras P (2011) Characterization of DNA-binding sequences for CcaR in the cephamycin-clavulanic acid supercluster of *Streptomyces clavuligerus*. Mol Microbiol 81:968–981

Santos-Aberturas J, Payero TD, Vicente CM, Guerra SM, Cañibano C, Martín JF, Aparicio JF (2011a) Functional conservation of PAS-LuxR transcriptional regulators in polyene macrolide biosynthesis. Metab Eng 13:756–767

Santos-Aberturas J, Vicente CM, Guerra SM, Payero TD, Martín JF, Aparicio JF (2011b) Molecular control of polyene macrolide biosynthesis: direct binding of the regulator PimM to eight promoters of pimaricin genes and identification of binding boxes. J Biol Chem 286:9150–9161

Santos-Beneit F, Rodríguez-García A, Sola-Landa A, Martín JF (2009) Cross-talk between two global regulators in *Streptomyces*: PhoP and AfsR interact in the control of *afsS, pstS* and *phoRP* transcription. Mol Microbiol 72:53–68

Santos-Beneit F, Barriuso-Iglesias M, Fernández-Martínez LT, Martínez-Castro M, Sola-Landa A, Rodríguez-García A, Martín JF (2011a) The RNA polymerase omega factor RpoZ is regulated by PhoP and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. Appl Environ Microbiol 77:7586–7594

Santos-Beneit F, Rodríguez-García A, Martín JF (2011b) Complex transcriptional control of the antibiotic regulator *afsS* in *Streptomyces*: PhoP and AfsR are overlapping, competitive activators. J Bacteriol 193:2242–2251

Santos-Beneit F, Rodríguez-García A, Martín JF (2012) Overlapping binding of PhoP and AfsR to the promoter region of *glnR* in *Streptomyces coelicolor.* Microbiol Res (in press)

Schell MA (1993) Molecular biology of the LysR family of transcriptional regulators. Annu Rev Microbiol 47:597–626

Schupp T, Waldmeier U, Divers M (1987) Biosynthesis of desferrioxamine B in *Streptomyces pilosus*: evidence for the involvement of lysine decarboxylase. FEMS Microbiol Lett 42:135–139

Schupp T, Toupet C, Divers M (1988) Cloning and expression of two genes of *Streptomyces pilosus* involved in the biosynthesis of siderophore desferroxamine B. Gene 64:179–188

Sheeler NL, MacMillan SV, Nodwell JR (2005) Biochemical activities of the *absA* two-component system of *Streptomyces coelicolor*. J Bacteriol 187:687–696

Sheldon PJ, Busarow SB, Hutchinson CR (2002) Mapping the DNA-binding domain and target sequences of the *Streptomyces peucetius* daunorubicin biosynthesis regulatory protein. DnrI Mol Microbiol 44:449–460

Sola-Landa A, Moura RS, Martín JF (2003) The two-component PhoR–PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. Proc Natl Acad Sci U S A 100:6133–6318

Sola-Landa A, Rodríguez-García A, Franco-Domínguez E, Martín JF (2005) Binding of PhoP to promoters of phosphate regulated genes in *Streptomyces coelicolor:* identification of PHO boxes. Mol Microbiol 56:1373–1385

Sola-Landa A, Rodríguez-García A, Apel AK, Martín JF (2008) Target genes and structure of the direct repeats in the DNA binding sequences of the response regulator PhoP in *Streptomyces coelicolor*. Nucleic Acids Res 36:1358–1368

Stewart RC (2010) Protein histidine kinases: assembly of active sites and their 981 regulation in signaling pathways. Curr Opin Microbiol 13:133–141

Stratigopoulos G, Gandecha AR, Cundliffe E (2002) Regulation of tylosin production and morphological differentiation in *Streptomyces fradiae* by TylP, a deduced gamma-butyrolactone receptor. Mol Microbiol 45:735–744

Tahlan K, Anders C, Wong A, Mosher RH, Beatty PH, Brumlik MJ, Griffin A, Hughes C, Griffin J, Barton B, Jensen SE (2007) 5 S clavam biosynthetic genes are located in both the clavam and paralog gene clusters in *Streptomyces clavuligerus*. Chem Biol 14:131–142

Takano E (2006) γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol 9:287–294

Tanaka A, Takano Y, Ohnishi Y, Horinouchi S (2007) AfsR recruits RNA polymerase to the *afsS* promoter: a model for transcriptional activation by SARPs. J Mol Biol 369:322–333

Thomas L, Hodgson DA, Wentzel A, Nieselt K, Ellingsen TE, Moore J, Morrissey ER, Legaie R, STREAM Consortium, Wohlleben W, Rodríguez-García A, Martín JF, Burroughs NJ, Wellington EM, Smith MC (2012) Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and phoP mutant grown in batch culture. Mol Cell Proteomic 11:M111.013797

Tian Y, Fowler K, Findlay K, Tan H, Chater KF (2007) An unusual response regulator influences sporulation at early and late stages in *Streptomyces coelicolor*. J Bacteriol 189:2873–2885

Tierrafría VH, Ramos-Aboites HE, Gosset G, Barona-Gómez F (2011) Disruption of the siderophore-binding *desE* receptor gene in *Streptomyces coelicolor* A3(2) results in impaired growth in spite of multiple iron-siderophore transport systems. Microb Biotechnol 4:275–285

Tiffert Y, Supra P, Wurm R, Wohlleben W, Wagner R, Reuther J (2008) The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. Mol Microbiol 67:861–880

Tiffert Y, Franz-Wachtel M, Fladerer C, Nordheim A, Reuther J, Wohlleben W, Mast Y (2011) Proteomic analysis of the GlnR-mediated response to nitrogen limitation in *Streptomyces coelicolor* M145. Appl Microbiol Biotechnol 89:1149–1159

Tunca S, Barreiro C, Sola-Landa A, Coque JJ, Martín JF (2007) Transcriptional regulation of the desferrioxamine gene cluster of *Streptomyces coelicolor* is mediated by binding of DmdR1 to an iron box in the promoter of the *desA* gene. FEBS J 274:1110–1122

Tunca S, Barreiro C, Coque JJ, Martín JF (2009) Two overlapping antiparallel genes encoding the iron regulator DmdR1 and the Adm proteins control siderophore and antibiotic biosynthesis in *Streptomyces coelicolor* A3(2). FEBS J 276:4814–4827

Ulijasz AT, Andes DR, Glasner JD, Weisblum B (2004) Regulation of iron 1007 transport in *Streptococcus pneumoniae* by RitR, an orphan response regulator. J Bacteriol 186:8123–8136

van Wezel GP, McDowall KJ (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. Nat Prod Rep 28:1311–1333

van Wezel GP, McKenzie NL, Nodwell JR (2009) Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics. Methods Enzymol 458:117–141

Vicente CM, Santos-Aberturas J, Guerra SM, Payero TD, Martín JF, Aparicio JF (2009) PimT: an amino acid exporter controls polyene production via secretion of the quorum sensing pimaricin-inducer PI-factor in *Streptomyces natalensis*. Microb Cell Fact 8:8–33

Vögtli M, Chang PC, Cohen SN (1994) *afsR2*: a previously undetected gene encoding a 63-amino acid protein that stimulates antibiotic production in *Streptomyces lividans*. Mol Microbiol 14:643–653

Wang L, Tian X, Wang J, Yang H, Fan K, Xu G, Yang K, Tan H (2009) Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. Proc Natl Acad Sci U S A 106:8617–8622

Ward JM, Hodgson JE (1993) The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. FEMS Microbiol Lett 110:239–242

Wietzorrek A, Bibb M (1997) A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol Microbiol 25:1181–1184

Wilson DJ, Xue Y, Reynolds KA, Sherman DH (2001) Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of *Streptomyces venezuelae*. J Bacteriol 183:3468–3475

Wray LV, Fisher SH (1993) The *Streptomyces coelicolorglnR* gene encodes a protein similar to other bacterial response regulators. Gene 130:145–150

Wu K, Chung L, Revill WP, Katz L, Reeves CD (2000) The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251:81–90

Xu G, Wang J, Wang L, Tian X, Yang H, Fan K, Yang K, Tan H (2010) "Pseudo" γ-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. J Biol Chem 285:27440–27448

Yamada Y (1999) Autoregulatory factors and regulation of antibiotic production in *Streptomyces*. In: England R, Hobbs G, Bainton N, McRoberts DL (eds) Microbial signalling and communication society for general microbiology. Cambridge University Press, Cambridge, pp 177–196

Yamanaka K, Oikawa H, Ogawa HO, Hosono K, Shinmachi F, Takano H, Sakuda S, Beppu T, Ueda K (2005) Desferrioxamine E produced by *Streptomyces griseus* stimulates growth and development of *Streptomyces tanashiensis*. Microbiology 151:2899–2905

Yang YH, Song E, Kim EJ, Lee K, Kim WS, Park SS, Hahn JS, Kim BG (2009) NdgR, an IclR-like regulator involved in amino-acid-dependent growth, quorum sensing, and antibiotic production in *Streptomyces coelicolor*. Appl Microbiol Biotechnol 82:501–511

Yim G, Wang HH, Davies J (2007) Antibiotics as signalling molecules. Philos Trans R Soc Lond B Biol Sci 362:1195–1200

Chapter 7
Systems Analysis of Microbial Adaptations to Simultaneous Stresses

Ross P. Carlson, Olusegun J. Oshota, and Reed L. Taffs

Abstract Microbes live in multi-factorial environments and have evolved under a variety of concurrent stresses including resource scarcity. Their metabolic organization is a reflection of their evolutionary histories and, in spite of decades of research, there is still a need for improved theoretical tools to explain fundamental aspects of microbial physiology. Using ecological and economic concepts, this chapter explores a resource-ratio based theory to elucidate microbial strategies for extracting and channeling mass and energy. The theory assumes cellular fitness is maximized by allocating scarce resources in appropriate proportions to multiple stress responses. Presented case studies deconstruct metabolic networks into a complete set of minimal biochemical pathways known as elementary flux modes. An economic analysis of the elementary flux modes tabulates enzyme atomic synthesis requirements from amino acid sequences and pathway operating costs from catabolic efficiencies, permitting characterization of inherent tradeoffs between resource investment and phenotype. A set of elementary flux modes with competitive tradeoffs properties can be mathematically projected onto experimental fluxomics datasets to decompose measured phenotypes into metabolic adaptations, interpreted as cellular responses proportional to the experienced culturing stresses. The resource-ratio based method describes the experimental phenotypes with greater accuracy than other contemporary approaches and further analysis suggests the results are both statistically and biologically significant. The insight into metabolic network design principles including tradeoffs associated with concurrent stress adaptation provides a foundation for interpreting physiology as well as for rational control and engineering of medically, environmentally, and industrially relevant microbes.

R.P. Carlson (✉) • O.J. Oshota • R.L. Taffs
Chemical and Biological Engineering Department, Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717-3920, USA
e-mail: rossc@coe.montana.edu

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_7,

Keywords Elementary flux mode • Evolution of metabolism • Nutrient limitation • Resource-ratio theory • Stoichiogenomics

Abbreviations

^{13}C	carbon 13
$C_{fumarate}$	concentration of fumarate
C_{inv}	carbon investment
Cmole	carbon mole
Cmol glc	Cmol glc, carbon moles of glucose
Cmol X	Cmol X, carbon moles of biomass
$C_{op,X}$	carbon operating cost for growth
CμM	micromoles of carbon per liter of cytosol
$\boldsymbol{E}$	matrix containing ecologically important elementary modes
[E]	enzyme concentration
EFM	elementary flux mode
EFMA	elementary flux mode analysis
FBA	flux balance analysis
k_{cat}	catalytic turnover number
K_m	half-saturation constant
MFA	metabolic flux analysis
N_{inv}	nitrogen investment
$N_{inv,X,1:1}$	nitrogen investment for growth, minimalist flux-to-enzyme approach
$O_{2,op,X}$	oxygen operating cost for growth
$\boldsymbol{v}$	vector containing fluxes
v_{max}	maximum enzyme-catalyzed reaction rate
$\boldsymbol{w}$	vector containing weighting factors

7.1 Introduction: Resources and Life

Life is driven by mass and free energy extracted from the environment. The immense diversity of life highlights the variety of these mass and energy forms as well as the array of successful strategies for exploiting them. Metabolic networks channel these resources under an array of environmental constraints and stresses using competitive sets of gene products (Elser et al. 2007). The discipline of ecology has explicitly and implicitly studied mass and free energy flows on a continuum of size scales for more than 100 years. Central themes of many ecological analyses include resource investment strategies, energetic efficiencies of different physiologies and the tradeoffs between the two. These theories, developed and tested on macroscale systems, are useful for decoding the rapidly accumulating omics databases representing microbial-scale systems.

Scarcity of resources over evolutionary time is thought to have influenced the elemental make-up of microbes (Dekel and Alon 2005; Elser et al. 2011; Makino et al. 2003; Sterner and Elser 2002; Zinn et al. 2004). For instance, elemental availability is believed to have influenced the evolution of microbial DNA and amino acid sequences (Bragg and Hyder 2004; Bragg and Wagner 2009). The chromosome sequence of *Pelagibacter* is biased toward low nitrogen containing codons, presumably as a response to the extreme nitrogen scarcity in oceanic ecosystems (Giovannoni et al. 2005). Proteins involved with carbon, nitrogen, and sulfur acquisition contain less of the respective element than average proteins in both prokaryotic and eukaryotic microorganisms (Baudouin-Cornu et al. 2001; Elser et al. 2011). Highly expressed proteins are also thought to be influenced by 'elemental sparing' where the relative material synthesis costs of limiting resources are reduced (Bragg et al. 2012). Strategic use of limiting resources also extends to enzyme cofactors. Respiratory chains require large investments of iron; microbes native to low iron environments often shift from iron containing enzymes to non-iron requiring enzymes such as from ferredoxin to flavodoxin under conditions of iron scarcity (Erdner and Anderson 1999).

It is hypothesized that not only are macromolecule sequences and cofactor requirements influenced by resource scarcity, but also the structure and regulation of entire biochemical networks. Selective pressures have eliminated microbes that fail to allocate limiting resources to cellular functions providing the most competitive return on investment (Carlson 2007, 2009; Dhurjati et al. 1985; Molenaar et al. 2009; Wessely et al. 2011). The return on investment can reflect a multitude of fitness properties including toxin production or resistance, attachment to surfaces or chemotaxis, osmolyte synthesis or salvage, and high-yielding or high-rate extraction of free energy. The role of high-yielding, metabolic efficiency on cellular physiology has been established both theoretically and experimentally (Carlson and Srienc 2004b; Varma and Palsson 1993; Westerhoff et al. 1983).

This chapter details a theory for decoding the multiplicity of microbial strategies to extract and channel mass and energy flows under simultaneous environmental stresses, with a focus on resource scarcity. Elucidating tradeoffs between resource availability and microbial physiology provides a theoretical basis for systemic analyses of omics data, and a rational basis for controlling microbes in medical, environmental and industrial applications.

7.2 *In Silico* Analysis of Metabolic Systems

Notable advances have been made in understanding the basis for metabolic functioning, but there is still need for guiding network design principles (Huang 2000; Papp et al. 2009). These principles need to be investigated through computer-based analyses integrating the numerous metabolic components and their interactions into testable models. For instance, any broad search of the interconnected fluxes of atoms, electrons, and energy in large-scale reaction networks is simply beyond manual feasibility.

Applications of biochemical kinetic models, typically systems of coupled ordinary differential equations describing the evolution of concentrations in time, have illuminated microbial processes for decades (e.g. Kargi and Weissman 1987; Nielsen and Villadsen 1992; Smallbone et al. 2010; Steuer et al. 2006; Straight and Ramkrishna 1994; Varner 2000). Unfortunately, these approaches when applied to metabolic networks are typically either simplistic in terms of the number of processes considered, or in terms of the mechanistic model, or require extensive sets of condition-specific kinetic parameters. Published kinetic parameters, as compiled by databases like BRENDA (Chang et al. 2008; Schomburg et al. 2004), can vary over several orders of magnitude for the same enzyme, requiring expert knowledge for appropriate incorporation into biological models and interpretation of results (Teusink et al. 2000).

An alternative class of successful *in silico* methods, circumventing requirements for large condition-sensitive parameter sets, is known as stoichiometric modeling. Stoichiometric models extract systemic information from conservation relationships and molecular-level network structure, frequently deduced from omics datasets (Reed and Palsson 2003; Trinh et al. 2009). There are three major stoichiometric modeling approaches: metabolic flux analysis (MFA), linear programming methods (also known as flux balance analysis or FBA), and metabolic pathway analysis (including elementary flux mode analysis or EFMA). All three methods define a high-dimension solution space of physiologically permissible metabolic flux distributions based on a stoichiometric matrix specifying system conservation relationships. This solution space is often represented as a pointed convex cone, although the shape can vary depending on the reversibility properties of the reactions (Llaneras and Picó 2010; Wagner and Urbanczik 2005). The three approaches differ in how metabolic flux distributions are selected from the solution space. Detailed descriptions and comparisons of these methods can be found elsewhere (Blank and Kuepfer 2010; Reed and Palsson 2003; Schilling et al. 1999; Trinh et al. 2009).

The continuous solution space defined by stoichiometric methods such as FBA or EFMA contains an effectively infinite number of mathematical possibilities. Identifying biologically and ecologically relevant flux distributions in this continuum requires selection criteria. Maximization of biomass yield on substrate is a widely assumed basis for identifying a cellular physiology. The measure has an appealing simplicity, reflecting an ecologically reasonable structure-function relationship for metabolic networks. This criterion has been utilized successfully to predict and interpret microbial behaviors, including *Escherichia coli* grown in glucose-limited chemostats at modest dilution rates (Carlson and Srienc 2004b; Fong et al. 2003; Fong and Palsson 2004; Varma et al. 1993; Varma and Palsson 1994). However, biomass yield maximization does not always adequately describe metabolic behaviors such as batch growth, or chemostat growth under conditions of nitrogen limitation (Carlson 2007, 2009; Papp et al. 2009; Schuetz et al. 2007; Schuster et al. 2008, 2011).

Because of its role in the current chapter, EFMA will be discussed briefly here. A more detailed description can be found in Chap. 2 of this book which is dedicated to EFMA. EFMA begins with the construction of a stoichiometric matrix representing

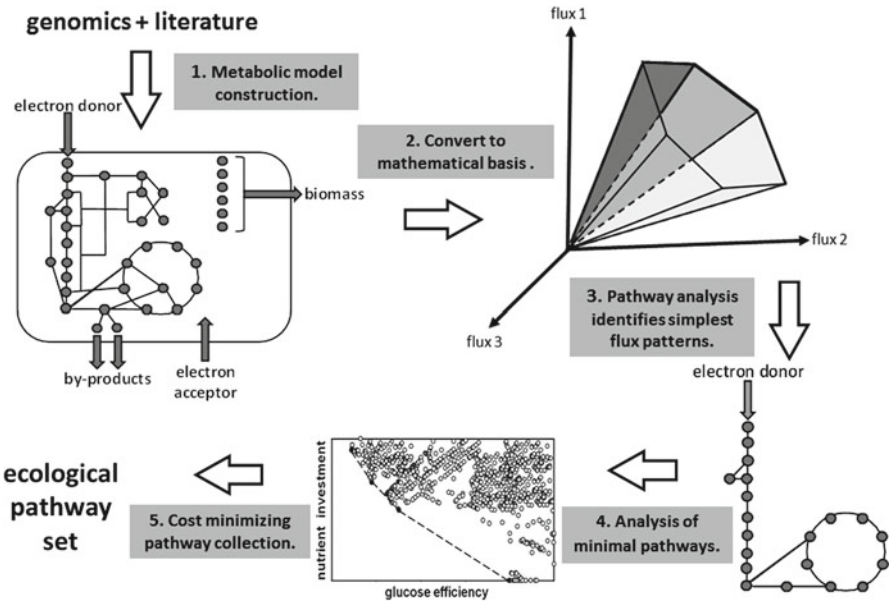


Fig. 7.1 Schematic representation of elementary flux mode analysis (EFMA) model building and cost-benefit analysis. EFMA model building and cost-benefit analyses involve the following steps: (1) metabolic model construction using genomic information and literature resources; (2) model conversion into a mathematical basis, whereby the reaction stoichiometries and reversibilities constrain steady-state cellular phenotypes to flux distributions within a space known as a flux cone; (3) EFMA decomposition of a reaction network into its simplest steady-state pathways, called elementary flux modes (EFMs); (4) cost-benefit analysis translates the EFMs into resource costs based on biochemical and genomic information; (5) minimizing combinations of these resource costs identifies competitive EFMs, collected as ecologically relevant pathway sets

the set of biochemical reactions to be considered. These are frequently compiled based on enzymes annotated from a genomic or metagenomic dataset (Fig. 7.1). The mathematical representation of the model encompasses all thermodynamically relevant system flux distributions. The complete set of enzymatically unique, minimal steady-state pathways spanning this permissible space is known as the elementary flux modes (EFMs) (Gagneur and Klamt 2004; Klamt et al. 2005; Schuster and Hilgetag 1994; Schuster et al. 2000). EFMs allow straightforward investigation of a network's metabolic potential from the bottom up (Klamt and Stelling 2003; Llaneras and Picó 2010; Papin et al. 2004; Trinh et al. 2009). These simplest pathways represent building blocks for steady-state metabolism, and their linear combinations allow system-level analyses of individual microbes, pure cultures, and consortia (e.g. Carlson et al. 2002, 2005; Carlson 2007, 2009; Klamt et al. 2008; Poolman et al. 2004; Taffs et al. 2009; Trinh et al. 2006, 2008; Wlaschin et al. 2006; Zhao and Kurata 2009).

7.3 Analyzing Metabolic Networks with Molecular Level Resource-Ratio Theory

The shortcomings of the often used 'biomass maximization on substrate' criterion highlight the need for additional theoretical frameworks for explaining microbial behaviors. The ecological fields are rich in theoretical approaches for analyzing fitness in various environments. In addition, decades of economic research have highlighted the importance of strategic resource allocation. These theoretical frameworks are well-suited for describing and interpreting many levels of biological organization, from molecular-level metabolic systems to entire ecosystems (Carlson and Taffs 2010). Resource-ratio theory, from the discipline of ecology, is a unifying approach for understanding shared resource competition and investment (Tilman 1980). A recent meta-analysis found that experimental tests supported predictions based on resource-ratio theory 75% of the time, making it one of the most successful theories in ecology (de Mazancourt and Schwartz 2010; Miller et al. 2005).

An approach similar to resource-ratio theory has been applied to *in silico* metabolic networks to quantify the fitness of the genome encoded metabolic potential subjected to pairs of scarce resources (Carlson 2007, 2009). It is hypothesized that evolutionary selection has favored regulation schemes directing resource investment toward effective, stress-specific metabolic pathways. The methodology examines tradeoffs between possible genome-encoded physiologies, represented as EFMs. Cells with regulation schemes permitting effective use of genomic potential would gain a fitness advantage under the relevant stress. To date, the approach has been applied to combinations of two limiting resources which can be either catabolic or anabolic in nature (Carlson and Srienc 2004b; Carlson 2007, 2009). This metabolism-focused application is a departure from traditional resource-ratio theory, which analyzes competition between different organisms for shared resources, rather than comparing potential phenotypes available to a single organism. The approach also shares conceptual similarities with the efficient frontier curve traditionally utilized in economic risk-return analysis and recently applied to biological systems (Kitano 2010).

Case studies in molecular level resource-ratio analysis decomposed *in silico* stoichiometric models of *E. coli* into EFMs (Carlson and Srienc 2004b; Carlson 2007, 2009). As mentioned previously, EFMs are minimal biochemical pathways comprised of metabolite transport and chemical reactions; the enzyme-based steps require an investment of anabolic resources like nitrogen or iron. EFMs, in this context, represent theoretical proteomes for which investment requirements can be tabulated. Calculation of investment requirements necessitates an assumed relationship between fluxes and enzyme concentrations. This relationship varies depending on specific enzyme properties, as well as the chemical environment. Two scenarios are proposed as bounds on the pathway-level flux-to-enzyme concentration relationships. The first approach is a minimalist relationship between pathway enzymes: the concentration ratio of every enzyme pair is set to one. This scenario would utilize varying metabolite pool size and activity modulation including allosteric regulation to achieve a specified flux distribution (this flux-to-enzyme model

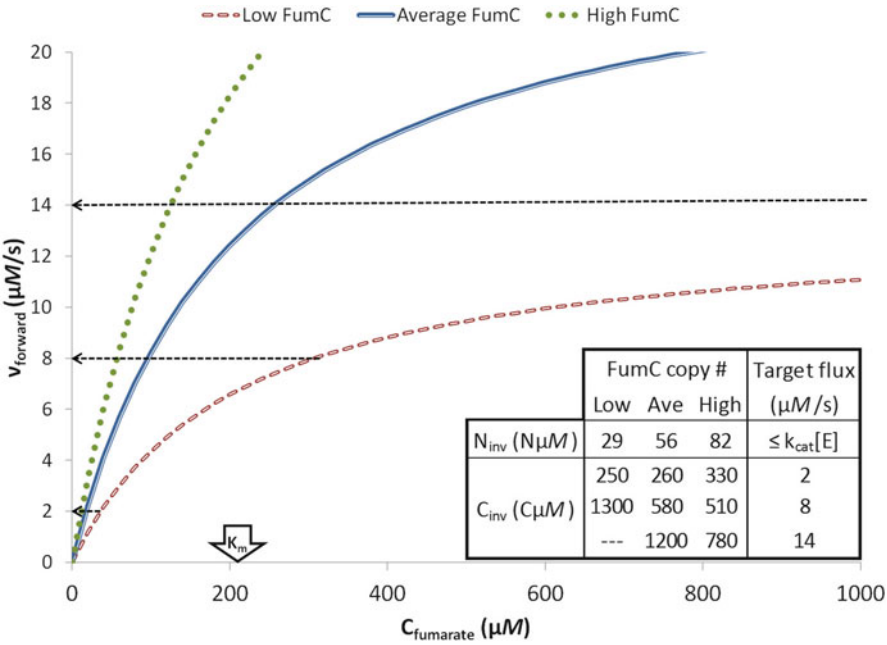


Fig. 7.2 Forward flux through the *E. coli* fumarase FumC (ordinate) as a function of fumarate concentration (abscissa), for three enzyme concentrations (*curved traces*). Each of the dashed *horizontal lines* represents a target flux, i.e. a fitness objective. The intersections of these lines with the various traces represent alternative investment strategies equally capable of achieving the fitness objective. The carbon and nitrogen investments (N_{inv} and C_{inv}) are quantified in the inset table; they represent a sum of both enzyme and metabolite pool investments. Michaelis-Menten kinetics were assumed, and the reverse reaction was neglected. Kinetic parameters for *E. coli* FumC were taken from the literature (Estévez et al. 2002), and the substrate concentration associated with half-saturation of FumC is indicated on the x-axis with an *arrow*. The plotted range of fumarate concentrations is well within intracellular measurements of *E. coli* from the literature (Bennett et al. 2009). The 'low', 'average', and 'high' enzyme concentrations were estimated based on intracellular copy numbers measured in *E. coli* (Taniguchi et al. 2010), assuming a cylindrical cell 2 μm long with a 1 μm diameter

was previously termed the first order method). The second flux-to-enzyme approach assumes a flux proportional relationship. Enzyme concentration ratios for any pathway enzyme pair are proportional to flux ratios through the corresponding reactions. The proportional scenario would use differences in enzyme concentration as the major controller of flux (this model was previously termed the zeroth order method). A comparison of the minimal and proportional flux-to-enzyme approaches to experimental fluxomic data suggests the minimalist approach is a better approximation for *E. coli*, although many additional, potentially relevant, relationships are imaginable (Carlson 2009).

The two flux-to-enzyme approaches are compared in Fig. 7.2 using experimental data for the *E. coli* fumarase enzyme FumC. A flux, representing a fitness objective, can be realized by a number of different proteomes, depending on the associated

metabolome. Two central concepts are illustrated in Fig. 7.2 by the intersections of the fitness objective fluxes (dashed lines) and the enzyme kinetics traces. First, each target flux can be driven by multiple combinations of enzyme and substrate concentrations. Second, low enzyme concentration may not support a high target flux, no matter how high the substrate concentration. These ideas illuminate tradeoffs between alternative investment strategies for achieving the same phenotype.

A straightforward tradeoff exists between investment of substrate into enzymes and into metabolites. Steady state metabolite pools are an investment because stationary concentrations represent substrate that is unavailable for other processes. The optimal strategy in terms of this tradeoff isn't universal: it depends in complex fashion on kinetic parameters, enzyme and substrate stoichiometry, and the target flux. Comparing the second and third lines of the inset table from Fig. 7.2 shows that it takes less carbon investment to drive very small fluxes through FumC with a large fumarate pool, but larger fluxes can be driven more economically with a large enzyme pool. This effect, which may seem counterintuitive initially, is a result of extreme differences in investment per molecule of enzyme versus per molecule of substrate.

Tradeoffs are environment specific and investment of an abundant resource into metabolite pools may not represent a fitness burden. The first line from the Fig. 7.2 inset table illustrates this point: if nitrogen limitation is the only relevant stress, it may be competitive to express the smallest FumC concentration capable of supporting a given target flux, even if that requires a heavy investment of carbon into metabolite pool. Additional environmentally-dependent tradeoffs are also noteworthy. In an environment that changes rapidly, the target flux may increase faster than a microbe can reorganize its proteome. In those environments, it may be competitive to overbuild enzymes for the current condition. As an example, the second line of the inset table from Fig. 7.2 implies that, in a carbon-limited environment with a low flux target, it is effective to drive that flux via the metabolite pool ($C_{inv} = 250$ vs. 330 CμM). This conclusion, however, rests on the assumption that the flux target won't suddenly and drastically increase. If it did, the final line of the inset table from Fig. 7.2 demonstrates that the new flux target might exceed the capacity of the small enzyme pool, providing a competitive advantage to the 'overbuilt' phenotype.

7.4 Constructing Resource Pair Tradeoff Curves

EFM anabolic resource investment requirements can be analyzed in concert with a catabolic physiological fitness metric termed an operating cost. EFM operating costs are the amount of substrate, typically an electron donor or acceptor, required to synthesize a specified amount of cellular product, such as one Cmole of biomass or one mole of ATP: equivalent to inverted yields. The operating cost reflects the net conversion efficiency of the pathway. The relationship of investment requirements against operating costs ranks EFMs efficiently based on combined resource use minimization.

Minimizing combined EFM resource costs for the same cellular product identifies a tradeoff curve, also known as a 'cost minimization envelope' (Carlson and Srienc 2004a, b). Figure 7.3 illustrates two tradeoff curves for the *E. coli* central metabolism. Figure 7.3a is an electron donor (glucose) vs. electron acceptor (oxygen) tradeoff curve while Fig. 7.3b is an electron donor (glucose) vs. anabolic resource (nitrogen) investment cost tradeoff curve. The plots are graphical illustrations of resource-ratio planes defined by the genome encoded metabolic potential. Each circle represents one distinct EFM. The EFM coordinates are the costs in terms of the two plotted resources to synthesize one Cmole of biomass. Physiologies that minimize requirements for resources are hypothesized to represent ecologically and economically competitive strategies. The leftmost EFM along the electron donor axis (abscissa in Fig. 7.3a, b) minimizes the cost of biomass synthesis on glucose; however, this reduction is offset by a higher requirement for the resource plotted on the ordinate. High tradeoff curve resource-ratios (paired resource/electron donor) permit increasingly efficient extraction of substrate free energy, up to a system constrained maximum. Moving right along a tradeoff curve maps competitive physiologies as the resource-ratio of either oxygen or nitrogen decreases relative to glucose. Lowering the resource-ratio increases glucose operating costs (Fig. 7.3).

Tradeoff curves represent combinations of enzymatic steps that provide the highest possible free energy yield based on the interaction between network stoichiometry and resource availability ratios. When the ratio of oxygen or nitrogen to carbon is small, tradeoff analysis indicates it is competitive for *E. coli* to down regulate the citric acid cycle, extracting easily accessible free energy from glycolysis while secreting partially oxidized metabolic intermediates (e.g. acetic or lactic acid). Retaining the 'lost' material and energy through further oxidation would represent a poor investment of the scarce resource. When the same resource ratios are high, the EFMs populating the left extreme of the tradeoff curve indicate it becomes ecologically competitive to completely oxidize glucose. The slopes of a tradeoff curve formalize the diminishing return of free energy extraction from substrate relative to efficient use of the second resource. These specific results are based on the physiological potential of *E. coli*, and general results depend on an organism's genome-encoded potential.

Physiological responses consistent with resource-ratio theory can have tradeoff properties that seem initially contradictory. For instance, the phenomena of microbial overflow metabolisms have been reported under a variety of conditions, including both nutrient excess and scarcity. The widespread occurrence of this metabolic strategy suggests there are fundamental adaptive principles guiding microbial responses (El-Mansi and Holms 1989; El-Mansi 2004; Majewski and Domach 1990; Neijssel et al. 1996; Straight and Ramkrishna 1994; Teixeira de Mattos and Neijssel 1997). The presented resource-ratio approach provides a single theory to explain overflow metabolism for both conditions based on small resource availability ratios. Under conditions of unrestricted environmental resources, the cellular resource ratio is constrained by the ratio of the corresponding transporter capacities (V_{max}). An imbalance in transporter capacity can create nutrient limitation stress responses even when

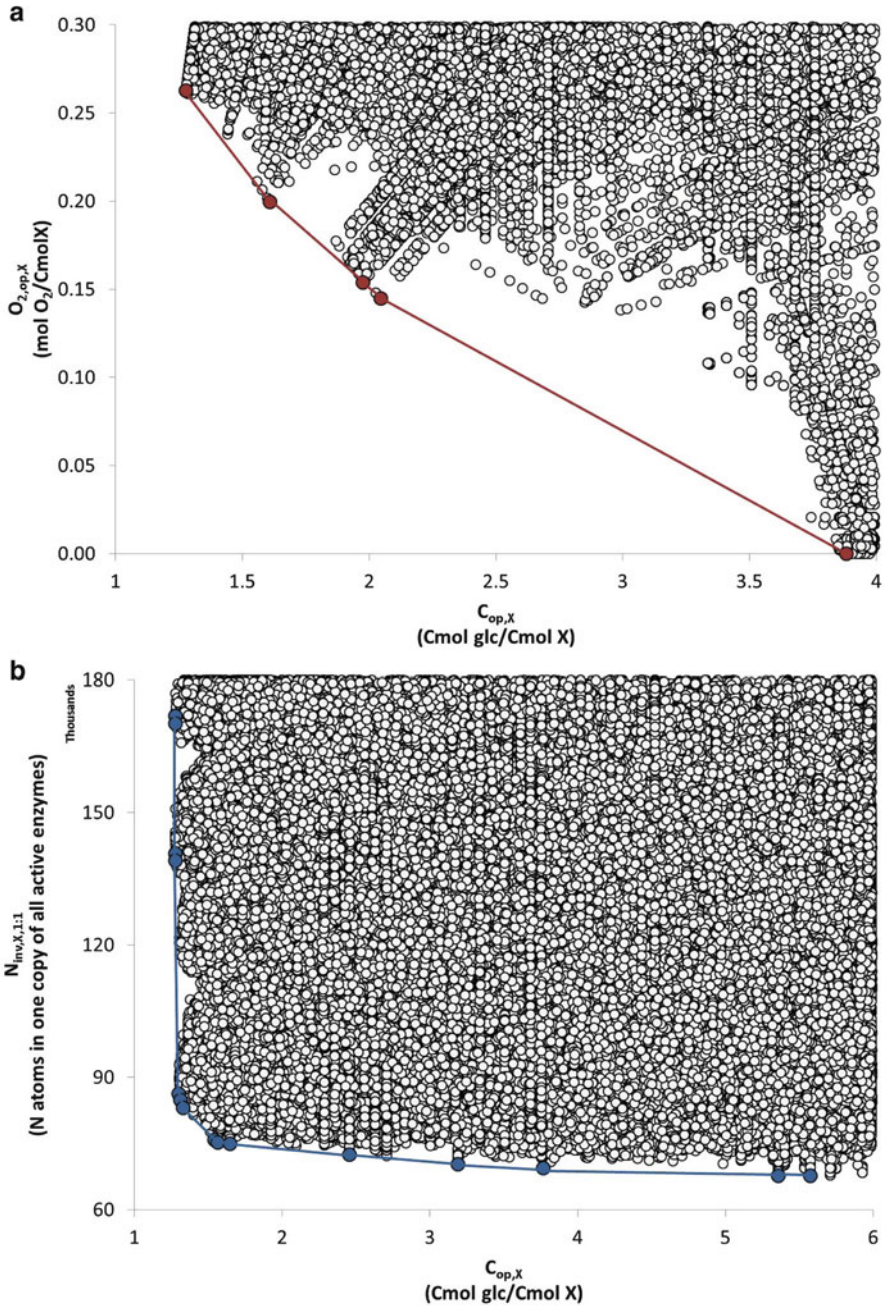


Fig. 7.3 Tradeoff curves for biomass synthesis in *E. coli* MG1655. Each *circle* represents an elementary flux mode (EFM) producing one carbon mole of biomass; *dark filled circles* and *line segments* represent flux distributions minimizing the combined costs plotted on the respective axes. Both abscissae represent glucose operation costs ($C_{op,x}$): the amount of glucose required to

the limiting resource is present at high levels, relative to enzyme half-saturation values. Under conditions of nutrient scarcity, however, availability depends on environmental supply and transporter affinity (K_m). The resource-ratio-based analysis of metabolic potential predicts a physiology consistent with an overflow metabolism in both cases.

Additional *in silico* studies have investigated alternative enzyme investment proxies, including the minimization of total flux or number of biochemical steps (de Figueiredo et al. 2009; Hoffmann et al. 2006; Holzhütter 2004; Poolman et al. 2004; Stelling et al. 2002), a summation of enzyme set molecular weights or volume (Beg et al. 2007; Vazquez et al. 2008a, b; Wessely et al. 2011), and a number of alternatives (Schuetz et al. 2007). While the approaches don't explicitly consider resource-ratio theory, they do consider allocation of resources, and sometimes arrive at similar predicted physiologies. An explicit consideration of resource investment has advantages when the resource is not uniformly distributed across proteins like the cofactor iron. Recent simulations and experiments in our laboratory demonstrated that the iron-limited physiological response is distinct from nitrogen limitation in *E. coli*, and that these differences map well to a resource-ratio interpretation.

7.5 Decomposing Fluxomics Data into Simultaneous Stress Adaptations

Microbes live in complex environments and are exposed frequently to simultaneous demands (Elser et al. 2007). Adaptations to maximize fitness require allocating limiting resources in optimal proportions between multiple cellular responses. These adaptations are proposed to be predictable using economic, evolutionary, and ecological theory (Bloom et al. 1985; Carlson and Taffs 2010; Kitano 2010; Papp et al. 2009; Perrin and Sibly 1993). In fact, dynamic modeling methods have been applied for decades to study simultaneous environmental pressures (Bader 1978; Kooijman 2000; Molenaar et al. 2009; Straight and Ramkrishna 1994), but these models are not yet able to fully capitalize on the omics revolution. Continuing advances in stoichiometry-based network models, however, are enabling systems-wide modeling efforts to understand microbial resource allocation.

A recent study used the stoichiometric modeling approach EFMA to decode potential *E. coli* adaptations to multiple stresses (Carlson 2009). EFMs from the *E. coli* model were translated into resource costs, and a set of ecologically relevant EFMs were assembled from cost minimizing tradeoff curves (Fig. 7.3). Each individual

Fig. 7.3 (continued) build one Cmole of biomass using a pre-existing proteome. The ordinate in subfigure (**a**) represents oxygen operation cost ($O_{2,op,x}$): a similar metric for oxygen. The ordinate in subfigure (**b**) represents nitrogen investment cost ($N_{inv,x}$): the amount of nitrogen contained in the proteome supporting each EFM flux distribution. The plotted investment costs were calculated using the minimalist flux-to-enzyme method

EFM consists of a unique enzyme pattern, providing a basis for flux fingerprinting. The flux distributions were projected onto experimental ^{13}C fluxomics datasets (Schuetz et al. 2007) to decompose measured phenotypes into metabolic responses, interpreted as proportional to the experienced culturing stresses.

The mathematics behind this analysis used experimentally measured ^{13}C-based flux distributions ($\vec{v}$) and the set of ecologically competitive elementary modes (E) identified through paired cost minimizations. The fluxomic data described the *E. coli* central carbon metabolism using ten degrees of freedom; the EFMs were projected into the same ten dimensional space to permit comparison. The proportional contribution of each individual EFM to the overall phenotype is given by a weighting vector $\vec{w}$ such that:

$$\vec{v} = E\vec{w} \tag{7.1}$$

where E is a matrix containing the elementary modes as column vectors. The vector $\vec{w}$ was determined using nonnegative least squares analysis to solve the following problem:

$$\text{Min}_{\vec{w}} \left\| E\vec{w} - \vec{v} \right\|; \forall i : w_i > 0 \tag{7.2}$$

Only positive elements were considered for $\vec{w}$ to limit considerations to biologically meaningful solutions; biomass- and ATP-synthesizing EFMs are not reversible. The magnitude of each weighting element represents the proportional contribution of the corresponding EFM to the mathematical description of measured cellular physiology.

In the described study (Carlson 2009), the *E. coli* metabolic network contained 197,018 distinct EFMs; the set E consisted of 38 distinct pathways identified from tradeoff curve analysis. Only three or four of these pathways were needed to best describe each of the four experimentally measured metabolic phenotypes ($\vec{v}$). These phenotypes comprised unrestricted batch growth, two carbon-limited chemostats (dilution rates of 0.2 and 0.4 h^{-1}), and a nitrogen-limited chemostat (dilution rate of 0.2 h^{-1}). The best fit of $\vec{v}$ for each of the four experimental distributions included EFMs from multiple tradeoff curves, suggesting that every culture investigated was responding to simultaneous stresses. For instance, the batch growth flux distribution was best described as a combination of metabolic strategies likely to be competitive for oxygen and nitrogen limitation, as well as an optimal biomass synthesis pathway. The description is attractive considering the often-reported overflow metabolism associated with *E. coli* unrestricted batch growth (e.g. Meadows et al. 2010; Xu et al. 1999).

In addition to its promising qualitative properties, the accuracy of the *in silico* description was compared quantitatively to an extensive collection of FBA-based flux distributions using the Euclidean distance metric. The set of ecologically selected EFMs described the experimental data with better accuracy than any of the tested FBA-based distributions. The best FBA descriptions required different objective functions for different culturing conditions. The most accurate reported FBA-based

description for the experimental batch growth data utilized the maximization of ATP yield per flux unit objective function (Schuetz et al. 2007).

The nonnegative least squares-based predictions were tested for statistical and biological significance by comparing the ability of 10,000 randomly selected EFM sets (of the same size) to describe the experimentally measured phenotypes. Not one random set described the four tested experimental datasets as accurately as the EFM set selected by minimization of combined costs. A perturbation analysis was performed to characterize the sensitivity of the approach to experimental uncertainty in the measured flux distributions. The predictions were remarkably stable, with the magnitude of the weighting factors falling into a tight distribution and no more than 6% of 10,000 simulations selecting EFMs different than the initial batch growth analysis (Carlson 2009).

7.6 Future Directions

An important consideration in stoichiometric modeling of metabolism is the robustness of phenotype. For every metabolism our group has investigated, the resource-ratio space very near the tradeoff surface includes many additional EFMs (Fig. 7.3). Laboratory chemostat experiments have shown that yield differences as small as 0.5% are selectable (Dykhuizen and Hartl 1980), but selection pressures vary in both magnitude and kind over evolutionary time, suggesting that even greater margins in 'optimal' metabolic behavior are likely fixed in nature. While perhaps not 'competitive' in strictly-controlled, non-varying environments, versatility and resilience matter in the long game, and tradeoffs between robustness and optimal performance are well known in nature as well as engineering (Csete and Doyle 2002; Kitano 2010). Our laboratory is experimenting with different margins of optimality and clustering approaches to find an appropriate balance between the concepts of metabolic robustness, resource requirements and cellular performance for selecting important metabolic phenotypes.

Not all investments produce a benefit that is easy to compare objectively between EFMs. For example, extensive theoretical and experimental work supports the notion that rate-based strategies can outcompete yield-based strategies in certain environments (MacLean 2008; Schuster et al. 2008). Unfortunately, quantifying the relative benefit of rate differences between pathways would obviate the main benefit of stoichiometric modeling: wide applicability without extensive parameterization. Similarly, some investments can be difficult to tabulate; cofactor requirements for enzymes are frequently unknown and potentially flexible with likely fitness consequences, causing difficulties in analysis of trace element scarcity. Differential investments in non-metabolic functions (e.g. chaperones, sensory proteins, or even ribosomal synthesis) can require special effort to be integrated into the tradeoff curve framework, although simple dynamic models are being developed for that purpose (Molenaar et al. 2009). With eukaryotic systems, organelles and their associated biochemical pathways have additional investment requirements that

should be accounted for to adapt the presented methodology. Questions remain: for instance, what additional membrane investment cost should be considered for organelle enzymes, or is this investment negligible when compared to the outer cellular membrane synthesis costs? Detailed ecological models exist for cellular resource budgets, accounting for uptake and growth machinery, storage compounds, and baseline physiologic functions. These are important considerations: while the metabolic enzyme collection is flexible in microbial generalists, the resources invested into other processes are critical throughout microbiology and presumably non-uniform. In addition microbes often express enzymes that are not required for their current circumstances. In a form of 'hedging metabolic bets', microbes may divert a fraction of their resources to a backup plan if the environment suddenly changes. Economic models associated with hedge funds surely will be useful for interpreting what fraction of a resource pool should be invested into risk-management strategies. For these reasons, development of suitable proxies for various types of investments and benefits will continue as an open area in the field of metabolic analysis. The best proxies will be calculable without extensive laboratory work, but effective in selecting a set of modes that is responsive to a specific stress.

Finally, microbial communities can be highly regulated and stable structures based on extensive trading of resources (Miller et al. 2010; Pfeiffer and Bonhoeffer 2004; van der Meer et al. 2005; Wintermute and Silver 2010). Stoichiometric analysis of community metabolism is a growing field (Dias et al. 2008; Stolyar et al. 2007; Taffs et al. 2009; Zhuang et al. 2011; Zomorrodi and Maranas 2012), and aspects of the presented analysis have been extended to coevolved microbial consortia to examine whether members forgo individual optimality for community-related benefits (Taffs et al. 2009). Resource-ratio theory has also recently been expanded to interpret resource exchanges between species (de Mazancourt and Schwartz 2010). Refining this theory to allow stoichiometry-based modeling of exchanges between and relative abundance of interacting microbes will play an important role in deciphering geochemical cycles. Highly efficient natural systems will also provide design blueprints for robust artificial consortia in bioprocess applications (Bernstein et al. 2012).

7.7 Conclusion

Stoichiometric network models are important theoretical tools, facilitating systems-wide modeling efforts for interpreting how microbes allocate resources in response to different environmental demands. This chapter explored microbial phenotypes that maximize fitness in multi-factorial environments through an economic analysis of metabolism. The economic analysis identified costs associated with every genome-encoded phenotype, which were used with ecological resource-ratio theory to characterize phenotypic tradeoff surfaces. These metabolic strategies support the notion of fundamental adaptive principles to the regulation of metabolism. For instance under the seemingly opposite scenarios of unrestricted and scarce resources, the resource-ratio based analysis predicts an overflow metabolism is ecologically

competitive. The presented schemata, developed primarily to understand metabolic network structure and function, provide a rational basis for investigations in environmental ecology, as well as for control of microbial processes in medical and industrial applications.

References

Bader FG (1978) Analysis of double-substrate limited growth. Biotechnol Bioeng 20:183–202

Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D (2001) Molecular evolution of protein atomic composition. Science 293:297–300

Beg QK, Vazquez A, Ernst J, De Menezes MA, Bar-Joseph Z, Barabási AL, Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc Natl Acad Sci U S A 104:12663–12668

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol 5:593–599

Bernstein HC, Paulson SD, Carlson RP (2012) Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. J Biotechnol 157:159–166

Blank LM, Kuepfer L (2010) Metabolic flux distributions: genetic information, computational predictions, and experimental validation. Appl Microbiol Biotechnol 86:1243–1255

Bloom AJ, Chapin FS, Mooney HA (1985) Resource limitation in plants – an economic analogy. Annu Rev Ecol Syst 16:363–392

Bragg JG, Hyder CL (2004) Nitrogen versus carbon use in prokaryotic genomes and proteomes. Proc R Soc Lond Ser B 271:S374–S377

Bragg JG, Wagner A (2009) Protein material costs: single atoms can make an evolutionary difference. Trends Genet 25:5–8

Bragg JG, Quigg A, Raven JA, Wagner A (2012) Protein elemental sparing and codon usage bias are correlated among bacteria. Mol Ecol 21:2480–2487

Carlson RP (2007) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics 23:1258–1264

Carlson RP (2009) Decomposition of complex microbial behaviors into resource-based stress responses. Bioinformatics 25:90–97

Carlson R, Srienc F (2004a) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. Biotechnol Bioeng 85:1–19

Carlson R, Srienc F (2004b) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. Biotechnol Bioeng 86:149–162

Carlson RP, Taffs RL (2010) Molecular-level tradeoffs and metabolic adaptation to simultaneous stressors. Curr Opin Biotechnol 21:670–676

Carlson RP, Fell DA, Srienc F (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol Bioeng 79:121–134

Carlson R, Wlaschin A, Srienc F (2005) Kinetic studies and biochemical pathway analysis of anaerobic poly-(R)-3-hydroxybutyric acid synthesis in *Escherichia coli*. Appl Environ Microbiol 71:713–720

Chang A, Scheer M, Grote A, Schomburg I, Schomburg D (2008) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res 37:D588–D592

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. Science 295:1664–1669

De Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25:3158–3165

De Mazancourt C, Schwartz MW (2010) A resource ratio theory of cooperation. Ecol Lett 13:349–359
Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592
Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27:1–9
Dias JML, Oehmen A, Serafim LS, Lemos PC, Reis MAM, Oliveira R (2008) Metabolic modelling of polyhydroxyalkanoate copolymers production by mixed microbial cultures. BMC Syst Biol 2:59
Dykhuizen DE, Hartl DL (1980) Selective neutrality of 6pgd allozymes in *Escherichia coli* and the effects of genetic background. Genetics 96:801–817
El-Mansi M (2004) Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications. J Ind Microbiol Biotechnol 31:295–300
El-Mansi EM, Holms WH (1989) Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. J Gen Microbiol 135:2875–2883
Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett 10:1135–1142
Elser JJ, Acquisti C, Kumar S (2011) Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. Trends Ecol Evol 26:38–44
Erdner DL, Anderson DM (1999) Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnol Oceanogr 44:1609–1615
Estévez M, Skarda J, Spencer J, Banaszak L, Weaver TM (2002) X-ray crystallographic and kinetic correlation of a clinically observed human fumarase mutation. Protein Sci 11:1552–1557
Fong SS, Palsson BØ (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat Genet 36:1056–1058
Fong SS, Marciniak JY, Palsson BØ (2003) Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. J Bacteriol 185:6400–6408
Gagneur J, Klamt S (2004) Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinform 5:175
Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245
Hoffmann S, Hoppe A, Holzhütter HG (2006) Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). Genome Inf 17:195–207
Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem 271:2905–2922
Huang S (2000) Complexity: the practical problems of post-genomic biology. Nat Biotechnol 18:471–472
Kargi F, Weissman JG (1987) Kinetic parameter estimation in microbial desulfurization of coal. Biotechnol Bioeng 30:1063–1066
Kitano H (2010) Violations of robustness trade-offs. Mol Syst Biol 6:384
Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis? Trends Biotechnol 21:64–69
Klamt S, Gagneur J, Von Kamp A (2005) Algorithmic approaches for computing elementary modes in large biochemical reaction networks. Syst Biol (Stevenage) 152:249–255
Klamt S, Grammel H, Straube R, Ghosh R, Gilles ED (2008) Modeling the electron transport chain of purple non-sulfur bacteria. Mol Syst Biol 4:156
Kooijman SALM (2000) Dynamic energy and mass budgets in biological systems. Cambridge University Press, Cambridge
Llaneras F, Picó J (2010) Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. J Biomed Biotechnol 753904

Maclean RC (2008) The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. Heredity 100:471–477
Majewski RA, Domach MM (1990) Simple constrained-optimization view of acetate overflow in *Escherichia coli*. Biotechnol Bioeng 35:732–738
Makino W, Cotner JB, Sterner RW, Elser JJ (2003) Are bacteria more like plants or animals? Growth rate and resource dependence of bacterial C:N:P stoichiometry. Funct Ecol 17:121–130
Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B (2010) Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab Eng 12:150–160
Miller TE, Burns JH, Munguia P, Walters EL, Kneitel JM, Richards PM, Mouquet N, Buckley HL (2005) A critical review of twenty years' use of the resource-ratio theory. Am Nat 165:439–448
Miller LD, Mosher JJ, Venkateswaran A, Yang ZK, Palumbo AV, Phelps TJ, Podar M, Schadt CW, Keller M (2010) Establishment and metabolic analysis of a model microbial community for understanding trophic and electron accepting interactions of subsurface anaerobic environments. BMC Microbiol 10:149
Molenaar D, Van Berlo R, De Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol 5:323
Neijssel OM, Teixeira De Mattos MJ, Tempest DW (1996) Growth yield and energy distribution. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella*: Cellular and molecular biology American Society for Microbiology. Washington DC, pp 1683–1692
Nielsen J, Villadsen J (1992) Modeling of microbial kinetics. Chem Eng Sci 47:4225–4270
Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BØ (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405
Papp B, Teusink B, Notebaart RA (2009) A critical view of metabolic network adaptations. HFSP J 3:24–35
Perrin N, Sibly RM (1993) Dynamic models of energy allocation and investment. Annu Rev Ecol Syst 24:379–410
Pfeiffer T, Bonhoeffer S (2004) Evolution of cross-feeding in microbial populations. Am Nat 163:E126–E135
Poolman MG, Venkatesh KV, Pidcock MK, Fell DA (2004) A method for the determination of flux in elementary modes, and its application to lactobacillus rhamnosus. Biotechnol Bioeng 88:601–612
Reed JL, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 185:2692–2699
Schilling CH, Schuster S, Palsson BØ, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Prog 15:296–303
Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32:D431–D433
Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119
Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2:165–182
Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332
Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252:497–504
Schuster S, De Figueiredo LF, Schroeter A, Kaleta C (2011) Combining metabolic pathway analysis with evolutionary game theory: explaining the occurrence of low-yield pathways by an analytic optimization approach. Biosystems 105:147–153
Smallbone K, Simeonidis E, Swainston N, Mendes P (2010) Towards a genome-scale kinetic model of cellular metabolism. BMC Syst Biol 4:6
Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193
Sterner RW, Elser JJ (2002) Ecological stoichiometry: the biology of elements from molecules to the biosphere. Princeton University Press, Princeton

Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. Proc Natl Acad Sci U S A 103:11868–11873

Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA (2007) Metabolic modeling of a mutualistic microbial community. Mol Syst Biol 3:92

Straight JV, Ramkrishna D (1994) Modeling of bacterial growth under multiply-limiting conditions: experiments under carbon- or/and nitrogen-limiting conditions. Biotechnol Prog 10:-588–605

Taffs R, Aston JE, Brileya K, Jay Z, Klatt CG, Mcglynn S, Mallette N, Montross S, Gerlach R, Inskeep WP, Ward DM, Carlson RP (2009) In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. BMC Syst Biol 3:114

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS, Quantifying E (2010) *Coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329:533–538

Teixeira De Mattos MJ, Neijssel OM (1997) Bioenergetic consequences of microbial adaptation to low-nutrient environments. J Biotechnol 59:117–126

Teusink B, Passarge J, Reijenga CA, Esgalhado E, Van Der Weijden CC, Schepper M, Walsh MC, Bakker BM, Van Dam K, Westerhoff HV, Snoep JL (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur J Biochem 267:5313–5329

Tilman D (1980) Resources: a graphical-mechanistic approach to competition and predation. Am Nat 116:362–393

Trinh CT, Carlson R, Wlaschin A, Srienc F (2006) Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab Eng 8:628–638

Trinh CT, Unrean P, Srienc F (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74:3634–3643

Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826

Van Der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M, De Leeuw JW, Damste JSS, Ward DM (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone national park. Appl Environ Microbiol 71:3978–3986

Varma A, Palsson BØ (1993) Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. J Theor Biol 165:503–522

Varma A, Palsson BØ (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60:3724–3731

Varma A, Boesch BW, Palsson BØ (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59:2465–2473

Varner JD (2000) Large-scale prediction of phenotype: concept. Biotechnol Bioeng 69:664–678

Vazquez A, Beg QK, De Menezes MA, Ernst J, Bar-Joseph Z, Barabási AL, Boros LG, Oltvai ZN (2008a) Impact of the solvent capacity constraint on E. coli metabolism. BMC Syst Biol 2:7

Vazquez A, De Menezes MA, Barabási AL, Oltvai ZN (2008b) Impact of limited solvent capacity on metabolic rate, enzyme activities, and metabolite concentrations of S. cerevisiae glycolysis. PLoS Comput Biol 4:6

Wagner C, Urbanczik R (2005) The geometry of the flux cone of a metabolic network. Biophys J 89:3837–3845

Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Mol Syst Biol 7:515

Westerhoff HV, Hellingwerf KJ, Van Dam K (1983) Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. Proc Natl Acad Sci U S A 80:305–309

Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol 6:407

Wlaschin AP, Trinh CT, Carlson R, Srienc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. Metab Eng 8:338–352

Xu B, Jahic M, Enfors SO (1999) Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. Biotechnol Prog 15:81–90

Zhao Q, Kurata H (2009) Maximum entropy decomposition of flux distribution at steady state to elementary modes. J Biosci Bioeng 107:84–89

Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R, Lovley DR (2011) Genome-scale dynamic modeling of the competition between rhodoferax and geobacter in anoxic subsurface environments. ISME J 5:305–316

Zinn M, Witholt B, Egli T (2004) Dual nutrient limited growth: models, experimental observations, and applications. J Biotechnol 113:263–279

Zomorrodi AR, Maranas CD (2012) OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 8:e1002363

Chapter 8
Metabolic Reprogramming Under Microaerobic and Anaerobic Conditions in Bacteria

Yue Shan*, Yong Lai*, and Aixin Yan

Abstract Oxygen has a great impact on the metabolism and physiology of microorganisms. It serves as the most efficient terminal electron acceptor to drive the energy conservation process of cellular respiration and is required in many biosynthetic reactions. Bacteria encounter oxygen fluctuation and limitation during their growth in both natural ecological niches and in laboratory vessels. In response to oxygen limitation, facultative bacteria undergo substantial metabolic reprogramming to switch from the aerobic respiration to either anaerobic respiration, fermentation, or photosynthesis. Two key factors determine the metabolic pathways bacteria adopt under oxygen deprived microaerobic and anaerobic conditions: maximal energy conservation and redox homeostasis. In this chapter, we first describe how the fulfillment of these two key factors governs the metabolic reprogramming of facultative bacteria and how the process is tightly controlled by several global regulatory factors: FNR, ArcBA, as well as NarL and NarP. We then utilizes fermentation of glycerol, a large surplus byproduct of biodiesel industry, as an example to illustrate how environment, process, and strain based approaches can be exploited to manipulate and engineer the anaerobic metabolic pathways so that desirable fermentation products can be achieved with optimal yield.

Keywords Aerobic respiration • Anaerobic respiration • Fermentation • Glycerol fermentation • Redox homeostasis

*Yue Shan and Yong Lai are contributed equally

Y. Shan • Y. Lai • A. Yan (✉)
School of Biological Sciences, The University of Hong Kong,
Pokfulam Road, Hong Kong, China
e-mail: ayan8@hku.hk

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_8,

Abbreviations

Arc	aerobic respiratory control
ATP	adenosine-5′-triphosphate
ETC	electron transport chain
FNR	fumarate and nitrite reduction
NADH	nicotinamide adenine dinucleotide (reduced form)
TCA cycle	tricarboxylic acid cycle

8.1 Introduction

Oxygen (O_2) as an atmospheric component has a great impact on the life style of living organisms on earth. It serves as the most efficient terminal electron acceptor to drive the energy conservation process of cellular respiration and is required in many biosynthetic reactions. Owing to its substantial effect on the growth and metabolism of microorganisms, how these organisms respond to the presence and levels of O_2 has been an important scenario to categorize microbial species. From this regard, microorganisms are classified as obligate aerobes, facultative anaerobes, aerotolerant anaerobes, microaerophiles, and obligate anaerobes. Facultative anaerobes are those that can grow both in the presence (aerobic) and absence (anaerobic) of O_2 and have been an important class of organisms exploited in biotechnology owing to their broad growth conditions and metabolic diversities. In order to grow both in the presence and absence of O_2 in their ecological niches, these microorganisms have evolved complex strategies to optimize their metabolic efficiencies in response to the changes of oxygen levels. In laboratory cultivation of bacteria for research or biotechnological purposes, fluctuation of oxygen levels in growth vessels likewise has a significant impact on the physiological status of bacteria and consequently the types of metabolic products they generate. Understanding this metabolic reprogramming process of bacteria in response to oxygen availabilities thus is essential for the design and optimization of bacterial cultivation systems for research and biotechnological applications. In this chapter, we first summarize the fundamental principles that govern the metabolic reprogramming of facultative anaerobic bacteria during their switch from aerobic to microaerobic or anaerobic conditions, and then discuss various factors and approaches that can be utilized to optimize the metabolic reprogramming processes in biotechnological applications.

8.2 Key Factors That Govern the Metabolic Reprogramming of Bacteria in Response to O_2

Two major requirements govern the metabolic reprogramming for facultative bacteria in response to different O_2 levels: to achieve optimal energy conservation and to maintain the cellular redox homeostasis. Under aerobic growth conditions, facultative

anaerobic bacteria conserve energy mainly by aerobic respiration, during which organic carbon sources are converted to CO_2 and the resulting reducing equivalents (NADH and $FADH_2$) are oxidized by the terminal electron acceptor O_2 through the aerobic electron transport chain (ETC). ATP is simultaneously generated during this process by means of oxidative phosphorylation. Under anaerobic conditions bacteria must adjust their strategies for energy conservation and redox homeostasis owing to the lack of O_2 as the terminal electron acceptor of the ETC. Under this condition the energy generation pathways adopted by bacteria include anaerobic respiration, fermentation, or photosynthesis. For chemoorganotrophic and chemolithotrophic bacteria, ATP can be yielded from anaerobic respiration or fermentation. For facultative phototrophic bacteria, such as purple non sulfur bacteria, usually photosynthesis is adopted upon the reduction of oxygen concentration. Concomitant with the energy production, NADH generated during the breakdown of the reduced carbon sources are recycled to NAD^+ in either of the processes of anaerobic respiration, fermentation, or CO_2 fixation depending on the species utilized and the availability of external electron acceptors (McEwan 1994). Thus, cellular redox balance is achieved. Since chemoorganotrophic bacteria represent the majority of the species utilized in biotechnology, we next use this class of microorganisms as an example to illustrate the metabolic reprogramming of bacteria in response to O_2 availabilities.

8.3 Aerobic Metabolism of Facultative Anaerobic Bacteria

Under aerobic growth conditions with sufficient carbon sources, aerobic respiration is adopted by facultative anaerobes to conserve energy owing to the highest yield of ATP through this process. The process of aerobic respiration can be divided into carbon flow and electron transport through the electron transport chain. Bacteria will first convert carbon sources such as monosaccharides, amino acids, short chain alcohols or acids to pyruvate. This process usually generates a limited number of ATP and reducing equivalents, often in the form of NADH. A typical example of this conversion is the degradation of glucose to yield pyruvate, referred as glycolytic pathways or glycolysis (Fig. 8.1), which is found in all major groups of microorganisms and functions both in the presence or absence of O_2. During aerobic respiration, pyruvate is converted to acetyl-CoA and CO_2 by the pyruvate dehydrogenase complex (PDHC). Acetyl-CoA then enters the tricarboxylic acid (TCA) cycle and be oxidized to CO_2. The various intermediates in the TCA cycle can be supplied to anabolic pathways for biosynthesis (Fig. 8.1a). In addition to the glycolytic pathways, certain carbon sources such as fatty acids and specific amino acids are directly converted to acetyl-CoA and subsequently enter the TCA cycle. Carbon sources of some short chain organic acids such as succinate, malate and citrate can directly enter the TCA cycle through the various points of the pathway during aerobic metabolism (Fig. 8.1a) (Gennis and Stewart 1996). Through these pathways and reactions, various of complex organic carbon sources are fully oxidized and converted to CO_2. This process represents the carbon flow during the aerobic metabolism.

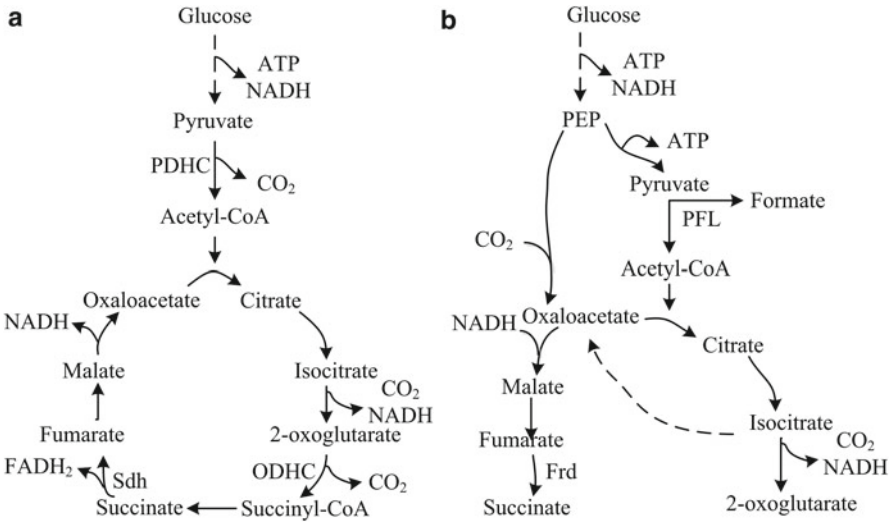


Fig. 8.1 The carbon flow during the aerobic and anaerobic catabolism of glucose. (**a**) Under aerobic condition, glucose is catabolized through glycolytic pathways into pyruvate and subsequently enters the aerobic TCA cycle through which reducing equivalents (in the form of NADH and $FADH_2$) and various carbon intermediates are generated. (**b**) During anaerobic metabolism, the glycolytic pathways are still active, but the TCA cycle becomes branched to generate less NADH. It functions mainly to supply intermediates to biosynthesis under anaerobic conditions. *Broken lines* represent multiple reaction steps. Abbreviations used are: *Frd* fumarate reductase, *ODHC* 2-oxoglutarate dehydrogenase complex, *PFL* pyruvate formate lyase, *PDHC* pyruvate dehydrogenase complex, *PEP* phosphoenolpyruvate, *Sdh* succinate dehydrogenase

Aerobic metabolism also includes the process of electron transport chain which is also the key step for redox homeostasis and energy production. The reducing power in the organic carbon sources is transferred to the electron carrier NADH (with a small portion being transferred to $FADH_2$) through the reactions in the glycolytic pathways and TCA cycle. NADH is the major electron donor during the aerobic respiration and it donates its electrons to the electron transport chain (ETC) located in the cytoplasmic membrane of bacteria. During this process, NADH is oxidized to NAD^+ and electrons released is transferred through a series of electron carriers in ETC and at the end through the terminal reductases to O_2 which is reduced to H_2O following accepting electrons. The electron transport through the ETC generates proton and electrical gradients across the cytoplastic membrane of bacteria which then drives ATP synthesis or be directly utilized to perform other cellular work. In aerobic ETC, the NADH dehydrogenases and the terminal reductases are linked by ubiquinones (Q) as electron carriers (Fig. 8.2a). Depending on the high or low cellular O_2 levels, either cytochrome *o* (Cyo) or cytochrome *d* (Cyd) terminal reductase complex is expressed. While cytochrome *o* complex operates under high oxygen concentration, cytochrome *d* complex is activated and functions under low oxygen concentration such as stationary growth phase owing to its high binding affinity to O_2 but lower activity comparing with the cytochrome *o* complex.

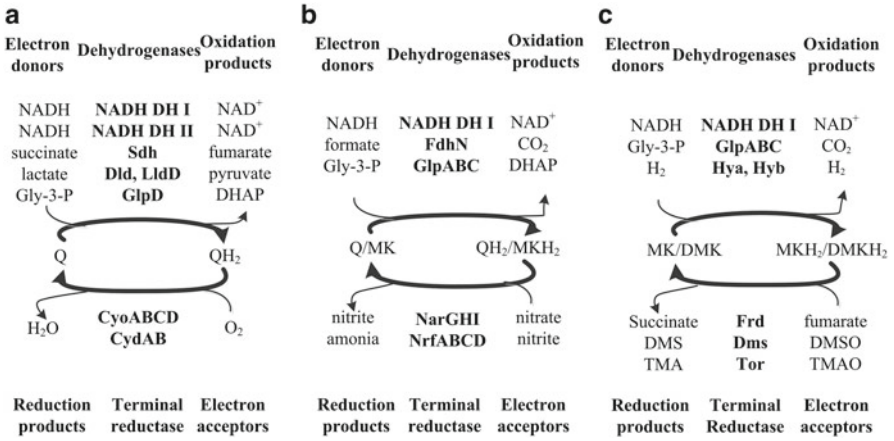


Fig. 8.2 Aerobic or anaerobic electron transport chains (ETC) with different electron carriers, dehydrogenases, and terminal reductases. (**a**) Aerobic electron transport chain. O_2 serves as the terminal electron acceptor of the ETC and either cytochrome *o* or cytochrome *d* complex is the terminal reductase. Ubiquinones serve as electron carriers. NADH, succinate, lactate, or glycerol-3-phosphate can serve as electron donors with the presence of specific dehydrogenases. (**b**) Anaerobic electron transport chain using nitrate or nitrite as electron acceptors. Ubiquinones (UQ) or menaquinone (MK) often serves as electron carriers when nitrate is utilized as the terminal electron acceptor. (**c**) Anaerobic electron transport chain using fumarate, DMSO or TMAO as electron acceptors. Menaquinone (MK) or demethylmenaquinone (DMK) often serves as electron carriers when cells utilize DMSO, TMAO or fumarate as electron acceptors. Abbreviations used are: *CydAB* cytochrome *d* oxidase, *CyoABCD* cytochrome *o* oxidases, *Dld* D-lactate dehydrogenase, *Dms* DMSO reductase, *FdhN* formate dehydrogenase-N, *Frd* fumarate reductase, *GlpD* aerobic glycerol-3-phosphate dehydrogenase, *GlpABC* anaerobic glycerol-3-phosphate dehydrogenase, *Hya* hydrogenase 1, *Hyb* hydrogenase 2, *LldD* L-lactate dehydrogenase, *NADH DH* NADH dehydrogenases, *NarGHI* nitrate reductase A, *NrfABCD* nitrite reductase, *Sdn* succinate dehydrogenase, *Tor* TMAO reductase

Through this ETC coupled ATP synthesis, both energy conservation and redox homeostasis is achieved which is essential to the physiological stability of bacteria.

In addition to NADH, several other compounds can also serve as electron donors to the aerobic ETC which is often linked to the special carbon sources utilized by bacteria. For instance, when glycerol is the sole carbon source available, glycerol-3-phosphate (Gly-3-P), which can be readily generated from glycerol by the enzyme glycerol kinase or derived from degradation of phospholipids and triacylglycerol, is a commonly used electron donor in the ETC (Gennis and Stewart 1996). In this case, glycerol-3-phosphate dehydrogenase (GlpD) is expressed and replaces the NADH dehydrogenase in the ETC which allows glycerol-3-phosphate donates its electrons to the electron carriers in the ETC and be oxidized to form dihydroxyacetone phosphate (DHAP) (Gennis and Stewart 1996). Owing to its less efficiency of energy conservation comparing with that using glucose as the carbon source and NADH as the electron donor, the expression of *glpD* is subject to the control of catabolite repression (Gennis and Stewart 1996). In addition, D-lactate and L-lactate

can also serve as the electron donors of ETC and be oxidized to pyruvate by D-lactate dehydrongenase (Dld) and L-lactate dehydrogenase (LldD) respectively. Some amino acids, such as proline, were also found to be able to serve as electron donors to the ETC in the presence of specific dehydrogenases during the aerobic growth of *E. coli* and *S. typhimurium* (Gennis and Stewart 1996).

8.4 Anaerobic Metabolism of Facultative Anaerobic Bacteria

During aerobic metabolism, the major biological process for redox homeostasis and energy conservation is the cellular respiration using O_2 as the terminal electron acceptor. The large difference of reduction potentials between NADH and O_2 (~1.14 V) ensures the highest yield of energy conservation in bacterial cells. Under anaerobic conditions when O_2 is absent, bacteria must adopt alternative means to oxidize the reducing equivalents NADH and generate ATP as aerobic respiration is no longer operated. Chemoorganotrophs utilize anaerobic respiration or fermentation to support their growth and maintain the cellular redox homeostasis under this condition. Since the compounds which can be utilized to replace O_2 in anaerobic ETC or to oxidize NADH in fermentation all have more positive reduction potentials than O_2, the yield and generation rate of ATP during anaerobic respiration and fermentation is slower than aerobic respiration. As a result of this insufficient oxidizing power, carbon flow is also adjusted under anaerobic growth to maintain redox homeostasis of bacterial cells.

Using glucose catabolism as an example, under anaerobic conditions, it is still converted to pyruvate by glycolysis (Fig. 8.1b). Different from aerobic metabolism, in the next step pyruvate is lysed to yield acetyl-CoA and formate by the pyruvate formate lyase (Pfl) during anaerobic growth. Owing to insufficient oxidizing power under this condition, the TCA cycle which produces majority of NADH during the aerobic metabolism becomes branched and incomplete so that less NADH is produced and the resulting intermediates are mainly used as precursors for biosynthesis. This metabolic reprogramming is achieved by repression of 2-oxoglutarate dehydrogenase complex (ODHC) and replacement of succinate dehydrogenase (Sdh) with fumarate reductase (Frd) in the pathway (Spiro and Guest 1991). Meanwhile, oxaloacetate is replenished from phosphoenolpyruvate (PEP) or isocitrate to maintain the balanced carbon flow in the incomplete TCA cycle (Fig. 8.1b) (Bock and Sawyers 1996; Cornan and Laporte 1996). As a result, less NADH will be generated comparing to aerobic carbon flow in order to cooperate with the lower efficiency of oxidation under this condition (Spiro and Guest 1991).

Under anaerobic conditions, NADH generated during glycolysis and other reactions still must be oxidized to maintain the redox balance and to obtain cellular energy. How NADH is re-oxidized determines which means of anaerobic metabolism, i.e. anaerobic respiration or fermentation, is adopted by chemoorganotrophic bacteria. If bacteria are supplied with exogenous electron acceptors such as nitrate, sulfate, trimethyl-amine-N-oxide (TMAO) *etc.*, they will utilize anaerobic respiration

for energy conservation as ATP generation from the electron transport chain is more efficient than fermentation. If exogenous electron acceptors are not available, bacteria have to utilize endogenous components, i.e. intermediates during catabolic reactions, as electron acceptors to re-oxidize NADH and this process is called fermentation. Environmental factors, such as the availabilities of alternative electron acceptors and fermentable carbon sources, determine which mode of anaerobic metabolism is adopted by bacteria.

8.4.1 Anaerobic Respiration

8.4.1.1 Terminal Electron Acceptors in the Anaerobic Respiration

During anaerobic respiration, the electron donor NADH still transfers its electrons to the quinone electron carriers in the anaerobic ETC via the NADH dehydrogenases complex. Quinones then transfer the electrons to accessible electron acceptors via specific terminal reductases. Different species can utilize different compounds as electron acceptors which often are related to the characteristics of their ecological niches. Common electron acceptors utilized by *E. coli* include nitrate, nitrite, trimethyl-amine-N-oxide (TMAO), dimethyle-sulphoxide (DMSO), and fumarate. Conceivably bacteria prefer electron acceptors with more positive reduction potentials, hence for *E. coli* the order of preferred electron acceptors is oxygen > nitrate > nitrite > DMSO > TMAO > fumarate (Gunsalus and Park 1994). For each electron acceptor utilized in anaerobic respiration, specific terminal reductases are needed (Fig. 8.2b). To achieve optimal energy conservation, the expression of the ETC dehydrogenases and terminal reductases is regulated by both the absence of oxygen and the availability of the corresponding terminal electron acceptors. In *E. coli* under aerobic conditions, the genes involved in the utilization of all other electron acceptors are repressed. Under anaerobic conditions if nitrate is present, DMSO, TMAO, and fumarate respiration is repressed. Only in the absence of oxygen and nitrate, these alternative terminal reductases are expressed (Gennis and Stewart 1996). Different from the aerobic ETC which always utilizes ubiquinones as the electron carriers, anaerobic electron transport chain utilizes different electron carriers depending on the electron acceptors utilized and the presence of corresponding terminal reductases. Instead of ubiquinones, menaquinone (MK) and demethylmenaquinone (DMK) often serves as electron carriers when cells utilize DMSO, TMAO or fumarate as electron acceptors while ubiquinones (UQ) and menaquinone (MK) serves as the electron carriers when nitrate is utilized as the terminal electron acceptor (Unden and Bongaerts 1997). In addition to the common terminal electron acceptors, certain species such as *Geobacter spp.* and *Shewanella oneidensis* can utilize insoluble Fe^{3+} as the terminal electron acceptor (Weber et al. 2006). Some bacteria, especially obligate anaerobes can utilize sulfate and CO_2 as the terminal electron acceptors although they are not common in facultative bacteria.

8.4.1.2 Electron Donors in the Anaerobic Respiration

Similar as in the case of aerobic respiration, some compounds other than NADH can serve as the electron donors to the anaerobic ETC. Examples include several reduced carbon compounds generated from the branched TCA cycle and formate and hydrogen which are generated from pyruvate cleavage by the pyruvate formate lyase. Anaerobic utilization of formate and hydrogen as electron donors requires the expression of formate dehydrogenase (Fdh) and hydrogenase (Hyb) respectively (Gennis and Stewart 1996). Glycerol-3-phosphate can also be utilized as the electron donor under anaerobic conditions with the presence of anaerobic glycerol-3-phosphate dehydrogenase (GlpABC) (Gennis and Stewart 1996). Pyruvate can be oxidized to acetate and carbon dioxide by the pyruvate oxidase (PoxB) and thus can also serve as the electron donor in *E. coli* under anaerobic conditions. However, this pathway is only activated during stationary phase culture and is tightly regulated. The oxidation of pyruvate is suggested to provide as the energy source under microaerobic conditions (Chang et al. 1994).

8.4.1.3 Other Factors Affect the Efficiency of Anaerobic Respiration

In addition to the reduction potentials of different terminal electron acceptors utilized in the anaerobic ETC, presence of other compounds has also found to be essential for the efficiency of anaerobic respiration. One important factor is metal ion homeostasis. This is because many enzymes in the anaerobic respiratory pathways are metal containing enzymes. It was reported that expression of operons containing enzymes in the anaerobic respiration pathway is highly reduced upon iron limitation under anaerobic conditions while the aerobic respiration pathways are only modestly affected in *E. coli* (Cotter et al. 1992). Moreover, induction of DMSO respiration pathways is reported to be dependent on the molybdate uptake (McNicholas et al. 1998).

It is notable that utilization of electron acceptors other than O_2 often produces metabolic intermediates and by-products that differ from aerobic respiration. Some of the intermediates during anaerobic respiration are harmful to bacteria and hence various means to counteract these effects are developed by bacteria when utilize these compounds as electron acceptors. Examples include enzymatic conversion of various nitrosative by-products to catabolizable carbon and nitrogen sources (Rankin et al. 2008) or export them through efflux pumps during the process of nitrate respiration (Zhang et al. 2011). It is also noteworthy that although in most cases the energy conservation mode using nitrate as the electron acceptor occurs only under anaerobic conditions, it may also be adopted under aerobic conditions by some bacterial species. It was reported that *Paracoccus pantothrpus* synthesizes a periplasmic nitrate reductase (NAP) that can conduct aerobic nitrate respiration. NAP is found to be induced under aerobic condition only when highly reduced organic compounds serve as carbon sources. It was proposed that the major purpose of aerobic nitrate respiration is for oxidation of excessive reducing equivalents and to maintain the redox homeostasis of bacterial cells (Sears et al. 2000).

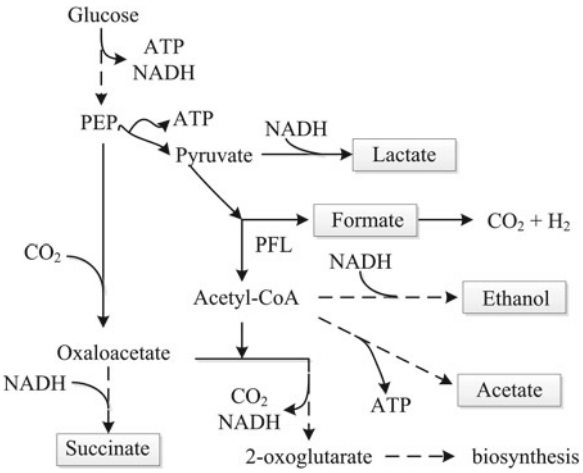


Fig. 8.3 Mixed acid fermentation of glucose by *E. coli*. During anaerobic growth in the absence of electron acceptors, glucose is fermented to form a mixture of products (*boxed*) including ethanol, acetate, lactate, succinate and formate by the model organism *E. coli* in order to regenerated NAD^+ and maintain the redox homeostasis of the cells. Abbreviations used are: *PFL* pyruvate formate lyase, *PEP* phosphoenolpyruvate. *Broken lines* represent multiple reaction steps

8.4.2 Fermentation

Under the conditions that both oxygen and exogenous electron acceptors are not accessible, fermentation will be adopted by facultative anaerobes to maintain their redox homeostasis, during which NADH is oxidized by endogenous intermediates and ATP is only generated by substrate level phosphorylation. Hence, fermentation requires supportable carbon sources that can generate intermediates to be utilized as electron acceptors.

E. coli often conducts a mixed acid fermentation when glucose is the carbon source during anaerobic growth (Fig. 8.3). The final products of mixed acid fermentation include ethanol, acetate, lactate, formate and succinate, which are usually excreted as ethanol, acetic acid, lactic acid, formic acid and succinic acid (Bock and Sawyers 1996). Another common type of fermentation is butanediol fermentation, which is usually adopted by *Enterobacter* (Bock and Sawyers 1996). Different from the aerobic metabolism in which glucose is almost exclusively oxidized to CO_2, the final products of fermentation are diverse and they are largely dependent on the carbon sources utilized and culture conditions. For example, *E. coli* can not use glycerol for fermentation in the absence of other electron acceptors at neutral PH, but under acidic conditions *E. coli* can ferment glycerol and produce ethanol and succinic acid (Dharmadi et al. 2006). In addition, different species may have different capabilities to ferment specific carbon sources. For example, *E. coli* could not grow with citrate as the sole carbon source, whereas *Salmonella typhimurium* can grow with citrate as the carbon source during which energy is derived from citrate fermentation to form acetate, formate and CO_2 (Bott 1997).

Although fermentation has the lowest energy conservation yield among the three metabolic processes, this mode of metabolism has significant biotechnological implications owing to the diversity of the products it can generate depending on the various genetic or environmental factors which provides a robust platform for metabolic engineering. It is notable that accumulation of certain fermentation products may have inhibition effects on cell growth. For example, ethanol, butanol and other alcohols can cause cellular toxicity by interrupting cell membranes (Huffer et al. 2011). Other products such as organic acids affect the pH homeostasis of bacterial cells. In natural microbial communities, these products are often excreted and can be utilized as carbon sources by other species in the same community. However in laboratory or biotechnological growth in a closed vessel, toxic products accumulation in the medium can dramatically inhibit cell growth as well as the yield of desirable products. One common solution to solve this problem is to use chemostat to maintain the physiological state of bacterial cells. Alternatively, genetically modification of bacterial strains to improve their tolerance to target products can be utilized (Ingram 1986) and it represents an active research area in modern microbial biotechnology.

8.5 Regulation of the Switch from Aerobic to Anaerobic Metabolism

8.5.1 Global Transcription Factors FNR and ArcBA

Owing to the significant diversity of anaerobic metabolic pathways and their impact on the global carbon and energy flow in bacterial cells, the transition from the aerobic to anaerobic lifestyle is tightly controlled in facultative bacteria in response to the availability of O_2, electron acceptors, as well as carbon sources. This regulatory network ensures the expression of essential proteins required in specific metabolic pathways, repression of unnecessary pathways, as well as optimal energy conservation under any specific conditions. In *E. coli* the metabolic switch between the aerobic to anaerobic growth is primarily controlled by two global transcription regulatory systems: FNR (fumarate and nitrate reduction) and the ArcBA (aerobic respiratory control) system. Both FNR and AcrA are transcription factors that can bind to specific DNA promoter regions and activate or repress gene transcription (Kiley and Beinert 2003; Gunsalus and Park 1994). The two systems have been found highly conserved among different Gram negative bacterial species, and an FNR-like protein has also been discovered in the Gram positive bacterium *Baccillus subtitlis*. It is interesting that most bacteria which have a FNR-like protein have been found to have a second oxygen or redox responsive protein (Sawers 1999), such as ArcA in the case of *E. coli*, suggesting the necessity of two global regulatory systems to cooperate the metabolic reprogramming in response to O_2 availability.

FNR is a global transcription factor that utilizes an $[4Fe\text{-}4S]^{2+}$ cluster to directly sense and respond to oxygen limitation. Under anaerobic conditions, FNR is synthesized as an $[4Fe\text{-}4S]^{2+}$ cluster containing protein. Assembly of the $[4Fe\text{-}4S]^{2+}$ cluster causes spontaneous FNR dimerization, enabling its specific DNA binding and transcription regulation. Upon being exposed to O_2, the $[4Fe\text{-}4S]^{2+}$ cluster is rapidly oxidized to an $[2Fe\text{-}2S]^{2+}$ cluster which causes the dissociation of FNR dimer and subsequent loss of specific DNA binding and transcription regulation. The O_2 concentration that fully activates FNR is reported to be below 1–5 mbar range of oxygen tensions (Sawers 1999). Upon being activated, FNR activates the expression of enzymes involved in anaerobic respiration and fermentation and represses the expression of those genes in aerobic respiration pathways (Gunsalus and Park 1994). Recent microarray studies reveal that in addition to the genes involved in the carbon and energy flow, FNR regulates more than two hundreds genes in *E. coli* genome and the function of many FNR regulated genes remains unknown (Kang et al. 2005).

While FNR directly senses oxygen through its $[4Fe\text{-}4S]^{2+}$ cluster and regulates gene expression, the ArcBA two component system mainly senses the redox potential in bacterial cells and thus respond to O_2 deprivation in an indirect manner. The system is activated under microaerobic to anaerobic conditions (10–5 mbar oxygen tensions) and functions mainly to repress the enzymes in aerobic metabolic pathways although it also activates genes involved in the microaerobic and anaerobic growth (Sawers 1999). The ArcBA system is a two component system including a membrane sensor protein ArcB and a cytoplasmic response regulator ArcA. Instead of directly sensing O_2 molecules, the transmembrane domain of ArcB senses membrane redox potential, i.e. reduced quinones, and subsequently undergoes phosphorylation at His^{292} and Asp^{576} amino acids at its C-terminal region. The phosphate group is then transferred to ArcA, enabling its capability to bind specific DNA sequences and regulate gene transcription (Rolfe et al. 2011).

Several genes are subject to dual regulation by both FNR and ArcBA. However, the two regulators may have additive or opposite effect on the transcription of the regulated genes. For instance, expression of *pfl* gene, which encodes pyruvate formate lyase that catalyses pyruvate cleavage to acetyl-CoA and formate under anaerobic conditions, is activated by both FNR and ArcBA, whereas the expression of *cydAB* gene which encodes cytochrome *d* oxidase and has optimal expression under microaerobic conditions, is activated by AcrBA but is repressed by FNR with the decrease of oxygen concentration. ArcA is activated first upon the reduction of O_2 level which in turn activates the expression of cytochrome *d* oxidase to optimize cellular respiration under microaerobic conditions. When oxygen concentration is further reduced in the growth environment, FNR will be activated and upon activation it represses the expression of *cydAB*, resulting in the shutdown of aerobic respiration and adoption of anaerobic metabolism under this condition (Tseng et al. 1996). It is conceivable that this cooperative regulation contributes to the fine-tuning of metabolic pathways under a range of different oxygen concentrations (Shalel-Levanon et al. 2005; Rolfe et al. 2011).

8.5.2 *Other Regulatory Factors*

In addition to FNR and ArcBA, NarL and NarP, two regulators that respond to the presence of nitrate and nitrite, often functions cooperatively with FNR and ArcBA and plays an important role in the anaerobic lifestyle of many bacteria. These two regulators activate genes involved in nitrate respiration and repress genes involved in anaerobic respiration of other alternative electron acceptors. Genes regulated by NarL and NarP are simultaneously regulated by FNR in response to the absence of oxygen availability and presence of nitrate. Through cooperative regulation by NarL, NarP and FNR, bacteria can optimize the anaerobic respiration pathways under different environmental conditions such that the carbon and electron acceptors will be utilized in a hierarchical manner to ensure optimal energy conservation under any given conditions (Constantinidou et al. 2006). It is noteworthy that in addition to these global transcription factors, recently a number of small RNAs, such as FnrS and ArcZ, were also identified to participate the regulation of genes involved in the anaerobic switch. The sRNAs FnrS and ArcZ are the components of FNR and ArcA regulon respectively and they function to fine-tune the expression of a subset of genes involved in the anaerobic metabolic reprogramming and enhance the efficiency of the switch (Mandin and Gottesman 2010; Durand and Storz 2010).

It is not unexpected that genetically modification of these transcription regulators and small RNAs can alter the metabolic pathways adopted by bacteria and consequently the metabolic products being generated. Thus it provides a useful tool to engineer the anaerobic metabolic processes. For instance deletion of *fnr* or *arcA* alters the ratio of different fermentation products in *E. coli* (Levanon et al. 2005). Deletion of *arcA* increased ethanol production during glucose fermentation and deletion of both *fnr* and *arcA* caused a high yield of lactate as the major fermentation product presumably due to the inefficiency of pyruvate formate lyase in the mutant strain (Levanon et al. 2005). These approaches will be further discussed in the following section.

8.6 Application and Manipulation of Anaerobic Metabolic Reprogramming in Biotechnology

Comparing with aerobic metabolism, anaerobic metabolism especially fermentation can lead to the production of a broad range of valuable chemicals which have great biotechnological implications. To apply and manipulate bacterial metabolic pathways towards the maximal production of the desirable products, the fundamental principles of maintaining the redox and energy homeostasis of bacteria must be fulfilled. On the basis of these fundamental principles, several environment based, process based, as well as strain based approaches can be applied to manipulate and engineer the metabolic pathways. Environment based approaches include control of environmental factors, such as the availability of carbon, nitrogen, and electron acceptors, as well as other factors such as pH, pressure, temperature *etc*. Process based

approaches refer to those that alter the mechanical and physical properties of the reaction systems, and strain based approaches include genetically engineering bacterial strains to overexpress target enzymes or block pathways competing for the same substrate. In this section we use glycerol fermentation as a paradigm to illustrate these strategies and applications.

8.6.1 Fermentative Metabolism of Glycerol

Fermentation of sugars to generate various chemicals such as ethanol, CO_2, and lactic acid has widely been utilized in industrial applications such as alcoholic beverage, food, and biofuels. The fermentative conversion of 6-carbon sugars derived from corn, sugarcane or sugar beet into the transportation biofuel ethanol has especially represented a major biotechnological development in the last decade owing to its potential to replace the fossil fuels. However, a major problem that hampers its economic viability is the co-production of a large amount of glycerol, which has resulted in a dramatic surplus of this by-product and a more than ten-fold decrease of the price of crude glycerol (Yazdani and Gonzalez 2007). How to convert this low price by-product to higher value products has become both an urgent need and a "target of opportunity" to increase the economic viability of biofuels industry (Yazdani and Gonzalez 2007).

Compare with glucose, glycerol is a carbon source with a higher degree of reduction. Several species of the *Enterobacteriaceae* family, such as *Klebsiella*, *Citrobacter*, and *Enterobacter*, and the genera of *Clostridium*, *Lactobacillus*, and *Bacillus* which contain the enzyme of 1,3-propanediol dehydrogenase (1,3-PDODH), can ferment glycerol to produce 1,3-propanediol (Yazdani and Gonzalez 2007). *E. coli*, the workhorse of modern biotechnology, was considered to be unable to ferment glycerol owing to its high reduction state. However, Dharmadi et al. (2006) have studied the feasibility of *E. coli* K-12 strain to ferment glycerol and found that *E. coli* MG1655 is able to ferment glycerol anaerobically in a pH dependent manner and the reaction requires the enzyme formate hydrogen lyase (FHL) (Fig. 8.4). Using different bacterial species and through metabolic engineering, glycerol fermentation can be utilized to produce several valuable chemicals, such as 1,3-propanediol, ethanol, succinic acid, propionic acid, and dihydroxyacetone *etc*.(da Silva et al. 2009). We next summarize the various strategies utilized to promote the production of each of these chemicals through metabolic engineering of the glycerol fermentation pathways.

8.6.2 1,3-Propanediol

Production of 1,3-propanediol (1,3-PDO) has attracted considerable interest in the past several years due to its promising applications in cosmetic and polymer industries (Saxena et al. 2009). Fermentative production of 1,3-PDO by bacterial species such as *Klebsiella* and *Citrobacter* requires the enzyme 1,3-PDODH to convert 3-hydroxypropionaldehyde, the dehydrating product of glycerol, to the 1,3-PDO

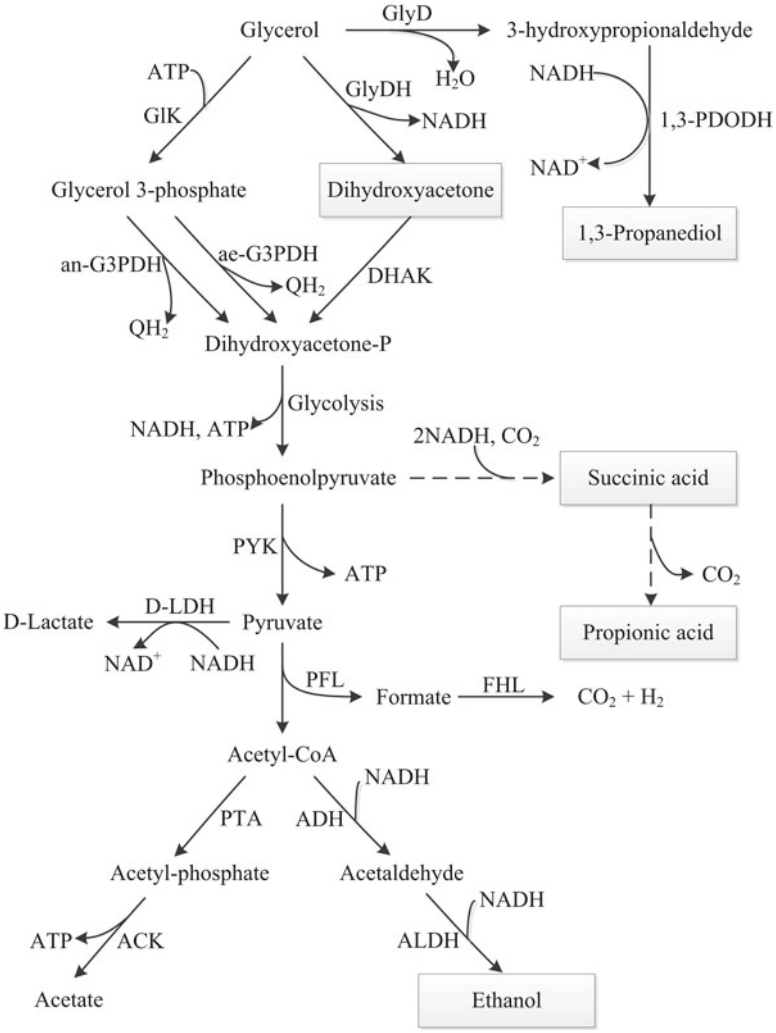


Fig. 8.4 Glycerol dissimilation and the synthesis of metabolites in bacteria under anaerobic and microaerobic conditions. In the species of *Klebsiella*, *Citrobacter*, *Enterobacter*, *Clostridium*, *Lactobacillus*, and *Bacillus,* glycerol fermentation can be divided into oxidative and reductive pathways. In the oxidative pathway, glycerol is converted to glycolytic intermediate dihydroxyacetone-P which is then degraded to diverse metabolites with the production of reducing equivalents. The reductive pathway, represented by the production of 1,3-propanediol, acts as a sink for the reducing equivalents generated in the oxidative pathway to maintain the redox balance during the metabolism. Important metabolites are boxed. *Broken lines* represent multiple reaction steps. Abbreviations used are: *ACK* acetate kinase, *ADH* alcohol dehydrogenase, *ALDH* aldehyde dehydrogenase, *ae-G3PDH* aerobic glycerol-3-phosphate dehydrogenase, *an-G3PDH* anaerobic glycerol-3-phosphate dehydrogenase, *DHAK* dihydroxyacetone kinase, *FHL* formate hydrogen-lyase, *GlK* glycerol kinase, *GlyD* glycerol dehydratase, *GlyDH* glycerol dehydrogenase, *LDH* lactate dehydrogenase, *PFL* pyruvate formate-lyase, *PTA* phosphoacetyl transferase, *PYK* pyruvate kinase; *1,3-PDODH* 1,3-propanediol dehydrogenase

product (Fig. 8.4). This step is essential and is referred as the reductive pathway of the process as it consumes NADH generated during glycolysis of glycerol to pyruvate (referred as the oxidative pathway of the process) and regenerates NAD^+. Both process based and strain based strategies were developed to optimize the production of 1,3-PDO. For example, a major problem of using glycerol as the sole carbon source is the slow growth and low biomass in the fermentor. To overcome this, a two-stage fermentation process was developed by Boenigk et al. (1993) in which sufficient biomass was produced in the first fermentor under low glycerol concentration (250 mM) but supplemented with 0.02% yeast extract, and production of 1,3-PDO was initiated and significantly increased by the addition of glycerol (625 mM) in the second fermentor following the accumulation of the biomass. The second approach to improve cell growth involves changing the anaerobic condition to microaerobic conditions (Chen et al. 2003). This modification increases the productivity of 1,3-PDO from 0.8 to 1.57 $g\,l^{-1}\,h^{-1}$. Improving cell growth can also be achieved by supplement of fumarate (5 mM), an electron acceptor for anaerobic respiration, to bacterial cultures such as *Klebsiella pneumoniae*. Through this approach, both the rate of glycerol consumption and 1,3-PDO production increased 35% compared with that in the absence of fumarate (Lin et al. 2005). This is presumably because fumarate can enhance the activities of several key enzymes in the pathway and decreases the NAD^+/NADH ratio resulting in more available reducing power to convert 3-HPA to 1,3-PDO. Production of 1,3-PDO can also be improved through strain based approaches. For example, to eliminate ethanol, a major competitor in the production of 1,3-PDO, aldehyde dehydrogenase gene (*aldA*) which plays a significant role in ethanol formation in *Klebsiella pneumoniae* was deleted and this enhanced the productivity and yield of 1,3-PDO to 14.05 $mmol\,L^{-1}\,h^{-1}$ and 0.699 $mol\,mol^{-1}$ respectively (Zhang et al. 2006). To eliminate the toxic effect of organic acid byproduct lactate, Yang *et. al* deleted lactate dehydrogenase gene (*ldhA*) of *Klebsiella oxytoca* and found that gene deletion led to 58 and 21% increase of the concentration and productivity of 1,3-PDO respectively grown under microaerobic conditions (Yang et al. 2007). Finally, it was reported that using *Clostridium butyricum* VPI 3266, the best natural 1,3-PDO producer identified so far, the productivity of 1,3-PDO was achieved to the highest level (10.3 $g\,l^{-1}\,h^{-1}$) at dilution rate of 0.30 h^{-1} and 60 $g\,l^{-1}$ of feed glycerol fermentation pattern (Gonzalez-Pajuelo et al. 2005).

8.6.3 Ethanol

Dharmadi et al. (2006) studied glycerol fermentation by *E. coli* at acidic pH and found that ethanol accounted for 86% of all products in the fermentation broth. A small percentage of succinic acid (7%) was also identified, this is because both ethanol and succinic acid production through glycerol fermentation is redox-balanced process (Fig. 8.4). However, compare with succinic acid production pathway, the reactions led to ethanol production have a net gain of one ATP per each molecule of

glycerol being converted, thus glycerol fermentation by *E. coli* is promising to be utilized to produce ethanol. Several means have been found to be effective to further improve the yield of ethanol production. These include using high concentration of glycerol to overcome the low affinity of glycerol dehydrogenase to glycerol substrate (Gonzalez et al. 2008); using microaerobic conditions to replace strict anaerobic conditions which greatly improve bacterial growth and eliminate the need for rich supplements such as tryptone or yeast extract (Durnin et al. 2009); as well as lower the concentration of potassium and phosphate which has a negative effect on the redox balance of the fermentation process (Gonzalez et al. 2008). Furthermore, *Paenibacillus macerans*, a species in the genus of *Paenibacillus*, which has broader metabolic capabilities than *E. coli* was exploited to ferment glycerol by Gupta et al. (2009) and it was found that the maximum specific growth rate of *P. macerans* during the log phase can achieve $0.40 \pm 0.03\ h^{-1}$, about ten-fold of that reported for *E. coli,* and the rate of ethanol production was 12 fold higher than that in *E. coli*.

8.6.4 Succinic Acid

Succinic acid is a biorefinery that has broad industrial values. Owing to its reduced state, succinic acid production from glucose fermentation has been limited (Sawers and Clark 2004), whereas fermentation of glycerol, a carbon source with higher degree of reduction, provides a promising means for its production (Fig. 8.4) (Dharmadi et al. 2006). Although glycerol fermentation by *E. coli* led to only ~7% succinic acid, its production has been found can be dramatically increased to 22% by simply conducting the fermentation in the presence of CO_2 (Dharmadi et al. 2006). Another species that can ferment glycerol to produce succinic acid is *Anaerobiospirillum succiniciproducens* and fermentation of glycerol by *A. succiniciproducens* led to a much higher yield of succinic acid (133%, g-succinic acid/g-glycerol) than fermentation of glucose (86%) (Lee et al. 2001). Cell growth, succinic acid yield, and productivity were found to be significantly enhanced when the culturing system of *A. succiniciproducens* was supplemented with hydrogen as an external electron donor in the form of H_2/CO_2 mixture (5:95, v/v) (Lee et al. 1999). Supplement with specific components have also found to be effective to increase succinic acid production by *Enterococcus flavescens* through glycerol fermentation. It was reported that a seven-fold increase of productivity was achieved when the culture medium was supplied with 3% sucrose, 15 mM $MgCO_3$, 1:0.5 ratio of tryptone and ammonium hydrogen phosphate at pH 6.5, and incubated at 39 °C (Agarwal et al. 2007).

8.6.5 Propionic Acid

Comparing with glucose, glycerol is a more optimal carbon source for propionic acid production owing to its low price, sufficient availability, higher yield, and reduced cost of purification (Yazdani and Gonzalez 2007; Barbirato et al. 1997). It was reported that production of propionic acid can be improved by co-fermentation of glycerol

(mainly used for propionic acid production) and glucose (provide both reducing equivalents and ATP for biomass) by the species of *Propionibacterium acidipropionici* (Liu et al. 2011). Comparing with the process using glycerol as the sole carbon source, a mixture of glycerol/glucose with 4/1 (mol/mol) ratio has led to 20 and 21% increase of the yield and productivity of propionic acid respectively. Feeding glycerol at a constant rate (fed-batch fermentation) provides another means to improve both the yield and productivity of propionic acid since this can help to eliminate the cell growth inhibition caused by propionic acid (Liang et al. 2012; Feng et al. 2010; Zhu et al. 2010). Strain based approach was also investigated to improve the yield of propionic acid. During glycerol fermentation by *Propionibacterium acidipropionici*, it was found that deletion of acetate kinase gene which eliminates the formation of competitive product acetate can increase the concentration of propionic acid in fed-batch fermentation broth to 100 $g\,l^{-1}$, which is much higher than that produced by the wild-type strain (Zhang and Yang 2009).

8.6.6 Dihydroxyacetone

Glycerol fermentation can also be utilized to produce dihydroxyacetone, an important intermediate for various chemicals in industry (Fig. 8.4). Two approaches have been investigated that can increase the productivity of dihydroxyacetone from glycerol fermentation. A strain based optimization revealed that overexpression of glycerol dehydrogenase gene can improve both bacterial growth and production of dihydroxyacetone by *Gluconobacter oxydans* (Gätgens et al. 2007). In addition, process based approach such as using repeated-fed-batch mode of fermentation, which eliminates the inhibition of both the substrates and products on bacterial growth, caused 75% increase of the productivity of dihydroxyacetone (Hekmat et al. 2003).

8.7 Concluding Remarks

It is noteworthy that in addition to the glucose and glycerol fermentation mentioned here, microbial anaerobic metabolism has much broader applications in various areas of industry, environment and medicine. Examples include bio-hydrogen production by the photosynthetic non-sulfur bacteria (PNSB) which eventually convert solar energy to usable hydrogen fuels (McKinlay and Harwood 2010; Argun and Kargi 2011; Basak and Das 2007; Yokoi et al. 2002; Asada et al. 2006), application of anaerobic sulfate reduction to remove heavy metal contamination (Muyzer and Stams 2008) and anaerobic ammonium oxidation (anammox) to promote the global nitrogen cycle (van der Star et al. 2007), as well as the recent exploitation of hypoxia targeted bacterial strains *Salmonella*, *Clostridium*, and *Bifidobacterium* for cancer treatment (Arrach et al. 2008; Forbes 2010; Ryan et al. 2009). While designing and manipulation of these systems must fulfill the fundamental principles of

energy and redox homeostasis of bacterial physiology, optimizing various environmental, process, and strain based factors can be explicitly determined by experiments. There is no doubt that with our increasing understanding of the physiology and biochemistry of microorganisms, identification and sequencing of more bacterial species, as well as the development of system-wide approaches, exploitation of microbial metabolism will have greater contribution to many areas of human life and have significant impact on human heath in the future.

References

Agarwal L, Isar J, Meghwanshi GK, Saxena RK (2007) Influence of environmental and nutritional factors on succinic acid production and enzymes of reverse tricarboxylic acid cycle from *Enterococcus flavescens*. Enzyme Microb Technol 40:629–636

Argun H, Kargi F (2011) Bio-hydrogen production by different operational modes of dark and photo-fermentation: an overview. Int J Hydrog Energ 36:7443–7459

Arrach N, Zhao M, Porwollik S, Hoffman RM, McClelland M (2008) Salmonella promoters preferentially activated inside tumors. Cancer Res 68:4827–4832

Asada Y, Tokumoto M, Aihara Y, Oku M, Ishimi K, Wakayama T, Miyake J, Tomiyama M, Kohno H (2006) Hydrogen production by co-cultures of Lactobacillus and a photosynthetic bacterium, *Rhodobacter sphaeroides* RV. Int J Hydrog Energ 31:1509–1513

Barbirato F, Chedaille D, Bories A (1997) Propionic acid fermentation from glycerol: comparison with conventional substrates. Appl Microbiol Biotechnol 47:441–446

Basak N, Das D (2007) The prospect of purple non-sulfur (PNS) photosynthetic bacteria for hydrogen production: the present state of the art. World J Microbiol Biotechnol 23:31–42

Bock A, Sawyers G (1996) Fermentation. In: FC N (ed) *Escherichia coli* and Salmonella: cellular and molecular biology, vol 1. ASM Press, Washington, DC, pp 262–282

Boenigk R, Bowien S, Gottschalk G (1993) Fermentation of glycerol to 1,3-propanediol in continuous cultures of Citrobacter-freundii. Appl Microbiol Biotechnol 38:453–457

Bott M (1997) Anaerobic citrate metabolism and its regulation in enterobacteria. Arch Microbiol 167:78–88

Chang YY, Wang AY, Cronan JE Jr (1994) Expression of *Escherichia coli* pyruvate oxidase (PoxB) depends on the sigma factor encoded by the rpoS(katF) gene. Mol Microbiol 11:1019–1028

Chen X, Zhang DJ, Qi WT, Gao SJ, Xiu ZL, Xu P (2003) Microbial fed-batch production of 1,3-propanediol by Klebsiella pneumoniae under micro-aerobic conditions. Appl Microbiol Biotechnol 63:143–146

Constantinidou C, Hobman JL, Griffiths L, Patel MD, Penn CW, Cole JA, Overton TW (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. J Biol Chem 281:4802–4815

Cornan JE Jr, Laporte D (1996) Tricarboxylic acid cycle and glyoxylate bypass. In: FC N (ed) *Escherichia coli* and Salmonella: cellular and molecular biology, vol 1. ASM Press, Washington, DC, pp 206–216

Cotter PA, Darie S, Gunsalus RP (1992) The effect of iron limitation on expression of the aerobic and anaerobic electron transport pathway genes in *Escherichia coli*. FEMS Microbiol Lett 79: 227–232

da Silva GP, Mack M, Contiero J (2009) Glycerol: a promising and abundant carbon source for industrial microbiology. Biotechnol Adv 27:30–39

Dharmadi Y, Murarka A, Gonzalez R (2006) Anaerobic fermentation of glycerol by *Escherichia coli*: a new platform for metabolic engineering. Biotechnol Bioeng 94:821–829

Durand S, Storz G (2010) Reprogramming of anaerobic metabolism by the FnrS small RNA. Mol Microbiol 75:1215–1231

Durnin G, Clomburg J, Yeates Z, Alvarez PJ, Zygourakis K, Campbell P, Gonzalez R (2009) Understanding and harnessing the microaerobic metabolism of glycerol in *Escherichia coli*. Biotechnol Bioeng 103:148–161

Feng XH, Chen F, Xu H, Wu B, Yao J, Ying HJ, Ouyang PK (2010) Propionic acid fermentation by *Propionibacterium freudenreichii* CCTCC M207015 with a fibrous-bed bioreactor. Bioprocess Biosyst Eng 24:1077–1085

Forbes NS (2010) Engineering the perfect (bacterial) cancer therapy. Nat Rev Cancer 10:785–794

Gätgens C, Degner U, Bringer-Meyer S, Herrmann U (2007) Biotransformation of glycerol to dihydroxyacetone by recombinant Gluconobacter oxydans DSM 2343. Appl Microbiol Biotechnol 76:553–559

Gennis RB, Stewart V (1996) Respiration. In: FC N (ed) *Escherichia coli* and Salmonella: cellular and molecular biology, vol 1. ASM Press, Washington, DC, pp 217–261

Gonzalez R, Murarka A, Dharmadi Y, Yazdani SS (2008) A new model for the anaerobic fermentation of glycerol in enteric bacteria: trunk and auxiliary pathways in *Escherichia coli*. Metab Eng 10:234–245

Gonzalez-Pajuelo M, Andrade JC, Vasconcelos I (2005) Production of 1,3-propanediol by Clostridium butyricum VPI 3266 in continuous cultures with high yield and productivity. J Ind Microbiol Biotechnol 32:391–396

Gunsalus RP, Park SJ (1994) Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. Res Microbiol 145:437–450

Gupta A, Murarka A, Campbell P, Gonzalez R (2009) Anaerobic fermentation of glycerol in Paenibacillus macerans: metabolic pathways and environmental determinants. Appl Environ Microbiol 75:5871–5883

Hekmat D, Bauer R, Fricke J (2003) Optimization of the microbial synthesis of dihydroxyacetone from glycerol with Gluconobacter oxydans. Bioprocess Biosyst Eng 26:109–116

Huffer S, Clark ME, Ning JC, Blanch HW, Clark DS (2011) Role of alcohols in growth, lipid composition, and membrane fluidity of yeasts, bacteria, and archaea. Appl Environ Microbiol 77:6400–6408

Ingram L (1986) Microbial tolerance to alcohols: role of the cell membrane. Trends Biotechnol 4:40–44

Kang Y, Weber KD, Qiu Y, Kiley PJ, Blattner FR (2005) Genome-wide expression analysis indicates that FNR of *Escherichia coli* K-12 regulates a large number of genes of unknown function. J Bacteriol 187:1135–1160

Kiley PJ, Beinert H (2003) The role of Fe-S proteins in sensing and regulation in bacteria. Curr Opin Microbiol 6:181–185

Lee PC, Lee WG, Kwon S, Lee SY, Chang HN (1999) Succinic acid production by *Anaerobiospirillum succiniciproducens*: effects of the H_2/CO_2 supply and glucose concentration. Enzyme Microb Technol 24:549–554

Lee PC, Lee WG, Lee SY, Chang HN (2001) Succinic acid production with reduced by-product formation in the fermentation of *Anaerobiospirillum succiniciproducens* using glycerol as a carbon source. Biotechnol Bioeng 72:41–48

Levanon SS, San KY, Bennett GN (2005) Effect of oxygen on the *Escherichia coli* ArcA and FNR regulation systems and metabolic responses. Biotechnol Bioeng 89:556–564

Liang ZX, Li L, Li S, Cai YH, Yang ST, Wang JF (2012) Enhanced propionic acid production from Jerusalem artichoke hydrolysate by immobilized *Propionibacterium acidipropionici* in a fibrous-bed bioreactor. Bioprocess Biosyst Eng 35:915–921

Lin R, Liu H, Hao J, Cheng K, Liu D (2005) Enhancement of 1,3-propanediol production by Klebsiella pneumoniae with fumarate addition. Biotechnol Lett 27:1755–1759

Liu Y, Zhang YG, Zhang RB, Zhang F, Zhu J (2011) Glycerol/glucose co-fermentation: one more proficient process to produce propionic acid by *Propionibacterium acidipropionici*. Curr Microbiol 62:152–158

Mandin P, Gottesman S (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J 29:3094–3107
McEwan AG (1994) Photosyntheticelectron transport and anaerobic metabolism in purple non-sulfur phototrophic bacteria. Antonie Van Leeuwenhoek 66:151–164
McKinlay JB, Harwood CS (2010) Photobiological production of hydrogen gas as a biofuel. Curr Opin Biotechnol 21:244–251
McNicholas PM, Chiang RC, Gunsalus RP (1998) Anaerobic regulation of the *Escherichia coli* dmsABC operon requires the molybdate-responsive regulator ModE. Mol Microbiol 27:197–208
Muyzer G, Stams AJ (2008) The ecology and biotechnology of sulphate-reducing bacteria. Nat Rev Microbiol 6:441–454
Rankin LD, Bodenmiller DM, Partridge JD, Nishino SF, Spain JC, Spiro S (2008) *Escherichia coli* NsrR regulates a pathway for the oxidation of 3-nitrotyramine to 4-hydroxy-3-nitrophenylacetate. J Bacteriol 190:6170–6177
Rolfe MD, Ter Beek A, Graham AI, Trotter EW, Asif HM, Sanguinetti G, de Mattos JT, Poole RK, Green J (2011) Transcript profiling and inference of *Escherichia coli* K-12 ArcA activity across the range of physiologically relevant oxygen concentrations. J Biol Chem 286:10147–10154
Ryan RM, Green J, Williams PJ, Tazzyman S, Hunt S, Harmey JH, Kehoe SC, Lewis CE (2009) Bacterial delivery of a novel cytolysin to hypoxic areas of solid tumors. Gene Ther 16:329–339
Sawers G (1999) The aerobic/anaerobic interface. Curr Opin Microbiol 2:181–187
Sawers R, Clark D (2004) Fermentative pyruvate and acetyl-coenzyme a metabolism. *Escherichia coli* and Salmonella: cellular and molecular biology, web edn. ASM Press, Washington, DC. Available at http://www.ecosal.org
Saxena RK, Anand P, Saran S, Isar J (2009) Microbial production of 1,3-propanediol: recent developments and emerging opportunities. Biotechnol Adv 27:895–913
Sears HJ, Sawers G, Berks BC, Ferguson SJ, Richardson DJ (2000) Control of periplasmic nitrate reductase gene expression (napEDABC) from Paracoccus pantotrophus in response to oxygen and carbon substrates. Microbiology 146(Pt 11):2977–2985
Shalel-Levanon S, San KY, Bennett GN (2005) Effect of ArcA and FNR on the expression of genes related to the oxygen regulation and the glycolysis pathway in *Escherichia coli* under microaerobic growth conditions. Biotechnol Bioeng 92:147–159
Spiro S, Guest JR (1991) Adaptive responses to oxygen limitation in *Escherichia coli*. Trends Biochem Sci 16:310–314
Tseng CP, Albrecht J, Gunsalus RP (1996) Effect of microaerophilic cell growth conditions on expression of the aerobic (cyoABCDE and cydAB) and anaerobic (narGHJI, frdABCD, and dmsABC) respiratory pathway genes in *Escherichia coli*. J Bacteriol 178:1094–1098
Unden G, Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. Biochim Biophys Acta 1320:217–234
van der Star WR, Abma WR, Blommers D, Mulder JW, Tokutomi T, Strous M, Picioreanu C, van Loosdrecht MC (2007) Startup of reactors for anoxic ammonium oxidation: experiences from the first full-scale anammox reactor in Rotterdam. Water Res 41:4149–4163
Weber KA, Achenbach LA, Coates JD (2006) Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. Nat Rev Microbiol 4:752–764
Yang G, Tian JS, Li JL (2007) Fermentation of 1,3-propanediol by a lactate deficient mutant of Klebsiella oxytoca under microaerobic conditions. Appl Microbiol Biotechnol 73:1017–1024
Yazdani SS, Gonzalez R (2007) Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. Curr Opin Biotechnol 18:213–219
Yokoi H, Maki R, Hirose J, Hayashi S (2002) Microbial production of hydrogen from starch-manufacturing wastes. Biomass Bioenerg 22:389–395
Zhang A, Yang ST (2009) Engineering *Propionibacterium acidipropionici* for enhanced propionic acid tolerance and fermentation. Biotechnol Bioeng 104:766–773
Zhang Y, Li Y, Du C, Liu M, Cao Z (2006) Inactivation of aldehyde dehydrogenase: a key factor for engineering 1,3-propanediol production by Klebsiella pneumoniae. Metab Eng 8:578–586

Zhang Y, Xiao M, Horiyama T, Li X, Nishino K, Yan A (2011) The multidrug efflux pump MdtEF protects against nitrosative damage during the anaerobic respiration in *Escherichia coli*. J Biol Chem 286:26576–26584

Zhu Y, Li J, Tan M, Liu L, Jiang L, Sun J, Lee P, Du G, Chen J (2010) Optimization and scale-up of propionic acid production by propionic acid-tolerant *Propionibacterium acidipropionici* with glycerol as the carbon source. Bioresour Technol 101:8902–8906

Chapter 9
Tunable Promoters in Synthetic and Systems Biology

Tore Dehli, Christian Solem, and Peter Ruhdal Jensen

Abstract Synthetic and systems biologists need standardized, modular and orthogonal tools yielding predictable functions in vivo. In systems biology such tools are needed to quantitatively analyze the behavior of biological systems while the efficient engineering of artificial gene networks is central in synthetic biology. A number of tools exist to manipulate the steps in between gene sequence and functional protein in living cells, but out of these the most straight-forward approach is to alter the gene expression level by manipulating the promoter sequence. Some of the promoter tuning tools available for accomplishing such altered gene expression levels are discussed here along with examples of their use, and ideas for new tools are described. The road ahead looks very promising for synthetic and systems biologists as tools to achieve just about anything in terms of tuning and timing multiple gene expression levels using libraries of synthetic promoters now exist.

Keywords Metabolic engineering • Standardization, orthogonality and modularity • Synthetic and systems biology • Synthetic gene networks • Synthetic promoter libraries

Abbreviations

ATc	anhydrotetracycline
DXP	1-deoxy-D-xylulose-5-phosphate
EGFP	enhanced green fluorescent protein

T. Dehli • C. Solem • P.R. Jensen (✉)
Center for Systems Microbiology, Department of Systems Biology,
Technical University of Denmark, Matematiktorvet 301/242,
2800 Lyngby, Denmark
e-mail: cso@bio.dtu.dk; tid@bio.dtu.dk; prj@bio.dtu.dk

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_9,

GFP	green fluorescent protein
GOI	gene of interest
gTME	global transcription machinery engineering
GusA	β-glucuronidase
KO	knock-out
MAGE	multiplex automated genome engineering
MCA	metabolic control analysis
ORF	open reading frame
Pfk	phosphofructokinase
PLS-R	partial least squares regression
PSP	promoter strength predictive
RBS	ribosome binding site
SOE PCR	splicing by overlap extension PCR
SPL	synthetic promoter library
TAL	transcription activator-like
TALOR	transcription activator-like orthogonal repressor
yEGFP	yeast enhanced green fluorescent protein

9.1 Introduction

Synthetic biology and systems biology are related but distinct disciplines. While both use modeling and simulation as core tools, synthetic biology focuses on the rational engineering of biological systems to display functions that do not exist in nature whereas systems biology aims to quantitatively understand the behavior of biological systems. Tools to modulate gene expression are central in both disciplines, a fact that has been and is still overlooked in many studies. Changing gene expression in small steps is essential for metabolic control analysis (MCA) in systems biology (Jensen and Hammer 1998a; Koebmann et al. 2005; Solem and Jensen 2002) and in synthetic biology, achieving appropriate gene expression levels is pivotal for redirecting fluxes to manipulate cell function (Bond-Watts et al. 2011). There are many examples of studies where functions have been added by genetic manipulation, but without achieving the specific quantitative adjustment of expression required for optimal metabolic change (Carothers et al. 2009). The promoters used in such studies have simply been characterized as ‘weak’ or ‘strong’ and this terminology is still frequently used.

A long-term goal in synthetic biology is to reengineer the entire cell function; a highly time-consuming process. To solve this challenge fast-track synthetic biology is required in an approach that involves: (A) Standardization—avoid the need for tailor-making new tools or extensive adjustment of old ones for each new purpose (Ellis et al. 2011); (B) Modularity—the ability of a system or component to function independently of context (Khalil and Collins 2010); (C) Orthogonality—separation of the modulated network from the chassis cell’s own regulatory networks (Blount et al. 2012a). An extremely useful tool in this respect is libraries of synthetic

promoters as their high resolution of outputs enables fine-tuning of gene expression levels to optimize network function (Blount et al. 2012a). Such promoters should be either inducible or constitutive. Constitutive promoters should exhibit constant gene expression levels through the different growth phases of the organism, whereas inducible promoters should exhibit tight control over gene expression and linear dependence to the amount of signal present (Keasling 1999). In the production of e.g. biofuels and biochemicals, where the goal is to re-route a major part of the metabolic flux to a desired product, the accuracy of gene expression is key. Such accuracy requires finely tuned gene expression levels (either using constitutive promoters or promoters inducible to known well-defined levels) and balancing of metabolic fluxes, metabolites and co-factors, along with expression of heterologous proteins (Blount et al. 2012a).

Synthetic promoters fulfill the above requirements whereas natural promoters do not. Using endogenous promoters for multiple genetic changes in a single cell is both a laborious process and can lead to genetic instability if identical or similar promoters are reused, especially in DNA recombination prone organisms. This problem is even more pertinent in eukaryotic systems than in prokaryotic ones as the lack of operons in the former requires the engineer to come up with a new promoter for every single gene in a pathway. A derivative of the important dairy strain *Lactococcus lactis* with all glycolytic genes up-regulated approximately two-fold using an SPL (Synthetic Promoter Library) approach has been generated without compromising genetic stability (Solem and Jensen, unpublished data). An additional drawback of wild-type promoters is that they are more prone to pleiotropic effects than synthetic ones (Blount et al. 2012a, b).

There are many ways in which gene expression can be changed; in this review we focus on approaches that create tunable promoters compatible with synthetic and systems biology studies.

9.2 Methods for Tuning Gene Expression Levels

Tools enabling fine-tuning of cell function are desirable in both basic and applied research (e.g. MCA and metabolic engineering, respectively). Many aspects of cell function are tunable, but the direct and often proportional link between promoter strength and protein concentration makes promoters an interesting starting point for fine-tuning of cell function through fine-tuning of gene expression levels. Historically changes made to promoters altering their strength have been changes in the length of the spacer region between the −35 and −10 box (Rossi et al. 1983), changes around and in the −35 box (Horwitz and Loeb 1986, 1988), changes to the −10 box (Jaurin et al. 1982) or a combination thereof (Horwitz and Loeb 1986; Jaurin et al. 1982). However as these approaches tinker with the most important properties of a promoter, the manipulations often result in an all or nothing response in terms of promoter strength rather than fine-tuning.

9.2.1 A Set of Insulated Promoters

Davis et al. created a set of insulated bacterial promoters to facilitate promoter transfer from ORF (Open Reading Frame) to ORF without modifying the relative expression level within the promoter set, i.e. promoter one should always be A times stronger than promoter two in different genetic backgrounds. To accomplish this a minimal 51 bp promoter sequence was identified and used as reference. This minimal promoter sequence was insulated up- and downstream with sequences designed to protect against transcription initiation and promoter escape (Davis et al. 2011).

To achieve differential expression Davis et al. varied sequences within the −10 and −35 regions manually introducing a number of cloning steps. Only the −35 and −10 regions were changed to keep the promoter and insulation sequences as constant between promoters as possible; nonetheless promoters varying two orders of magnitude in expressing levels were achieved. Four promoters were characterized and compared with four uninsulated minimal promoters. The promoter insulation proved to be very efficient compared to the minimal promoters when either the sequence up- or downstream of the insulated promoter was changed, but the insulated promoters did not have an improved stability towards changes in genetic context when comparing with the uninsulated promoters. This can be due to a number of factors affecting transcription efficiency (Bai Flagfeldt et al. 2009; Davis et al. 2011; Pruss and Drlica 1989; Pérez-Martín et al. 1994).

The method developed by Davis et al. was an attempt to improve the classical way of changing gene expression levels by enabling the scientist to once and for all make a library of promoters that is predictable in relative gene expression levels across genetic contexts. Although some significant improvements were made Davis et al. show exactly how difficult it is to keep expression levels of a fixed promoter set constant in changing genetic contexts. Instead it is favorable to insert libraries of promoters in different chromosomal locations and then choose those promoters appropriate to achieve a given cell function in that specific genetic context (see below).

9.2.2 Error-Prone PCR Based Randomization

In 2005 Alper et al. developed a method for creating promoter libraries by exposing the constitutive bacteriophage P_L-λ promoter to error-prone PCR followed by cloning into a reporter plasmid upstream of the GFP (Green Fluorescent Protein) gene. This yielded a large number of functional clones of which 200 were selected for further characterization on the single cell level. Many of the selected clones showed fluorescence instability and were discarded. Twenty-two clones were chosen for the functional promoter library and characterized in great detail by measuring GFP fluorescence, quantifying mRNA levels and cloning the library in front of the *cat* gene conferring resistance to chloramphenicol (Alper et al. 2005).

Amongst other things this method showed that elements regulating promoter activity can be far away (−120 bp) from the ORF (Fischer et al. 2006; Jensen et al. 2006).

Furthermore Alper et al. and Nevoigt et al. showed the broad applicability of promoter library technologies as ubiquitous tools, when they used the libraries in eukaryotic hosts. This was also shown for the SPL technology by Jeppsson et al. and others (Alper et al. 2005; Jensen and Hammer 1998c; Jeppsson et al. 2003; Nevoigt et al. 2006; Tornoe et al. 2002).

Qin et al. used the error-prone PCR based technique developed by Alper et al. to generate a library of P_{GAP} promoters for tunable expression in the yeast *Pichia pastoris*. Nearly 30,000 promoter mutants were screened to generate a library of 33 mutants spanning a 296-fold range of fluorescence intensity (Alper et al. 2005; Qin et al. 2011). This work clearly demonstrated the need for fine-tuning gene expression levels, as the overproduction of SAM (*S*-adenosylmethionine) using a promoter from the library to drive MAT gene expression was superior to simple over-expression (Qin et al. 2011).

9.2.3 Synthetic Promoter Libraries (SPL)

In 1998 Jensen and Hammer showed the importance of the context in which the −35 and −10 boxes are embedded by randomizing the DNA sequence of spacers between consensus sequences in *E. coli*, *L. lactis* and *S. cerevisiae* (Jensen and Hammer 1998a, c). In this first approach to generate SPLs a mixture of single-stranded oligonucleotides was synthesized based on the consensus sequence design approach (Fig. 9.1). The oligonucleotides were made double-stranded using a 3′-specific primer and DNA polymerase. The double stranded fragments were then cloned in front of the *lacLM* operon (encoding β-galactosidase) of plasmid pAK80 (Israelsen et al. 1995; Jensen and Hammer 1998a).

The plasmids were transformed into *E. coli* and then introduced in *L. lactis* MG1363 resulting in different blue coloring of the two organisms when plated on LB (*E. coli*) or GM17 (*L. lactis*) containing X-gal. Enzyme activity measurements were performed in liquid culture and resulted in activity (measured in Miller units) ranging from 0.3 to 2,050 in *L. lactis* and 0.2 to 528 in *E. coli* thus covering three orders of magnitude of activity in small steps. This promoter library was designed for *L. lactis* and higher *E. coli* activities would probably have been achieved using a reporter gene and an SPL specific for *E. coli* (Jensen and Hammer 1998a).

In 2002, Solem and Jensen simplified the use of SPLs by integrating the SPL in the forward primer used to amplify the gene of interest (GOI). The reporter gene *gusA* was used in this study and a linear correlation between GusA (β-glucuronidase) and Pfk (phosphofructokinase) activities was shown, i.e. GusA activity can be used as an indicator of Pfk activity. This linear correlation between reporter gene and gene or phenotype of interest has also been shown elsewhere (Hansen et al. 2009; Koebmann et al. 2002a, b).

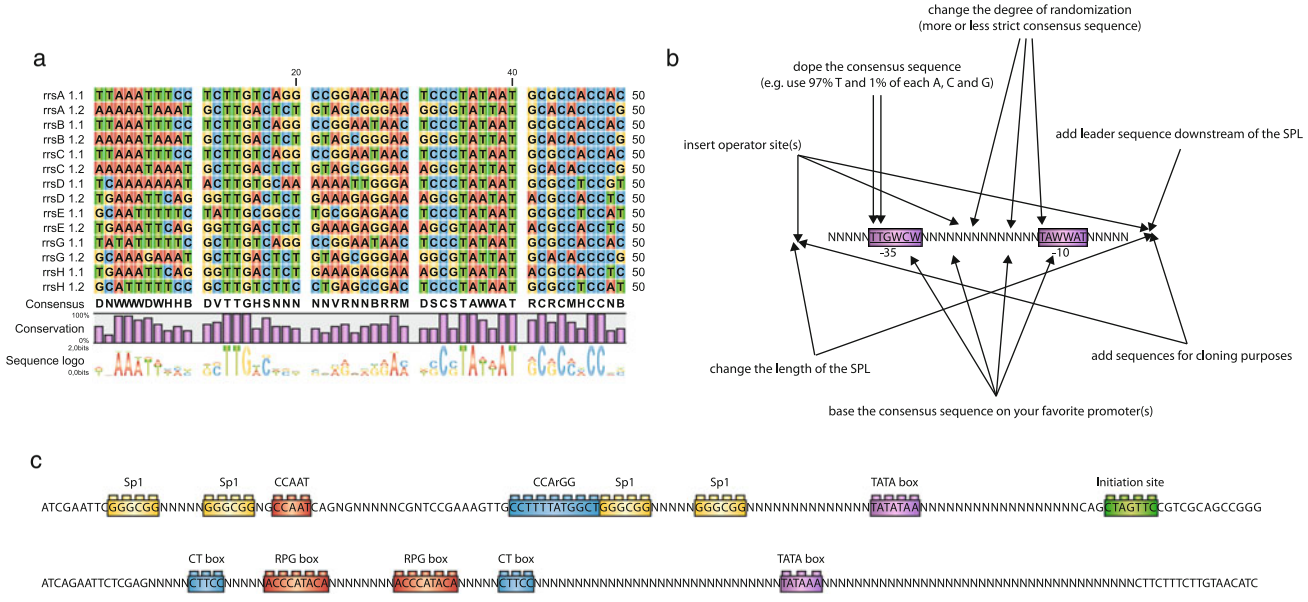


Fig. 9.1 Design strategies for synthetic promoter libraries. Part **a** shows an alignment of 50 bp promoter sequences from *E. coli* rRNA promoters. Ribosomal RNA promoters are typically very strong and thus an SPL based on such a consensus sequence is likely to yield promoters covering three orders of magnitude of gene expression levels. For data from libraries based on rRNA sequences see Rud et al. (2006). The consensus sequence for the SPL could equally well have been based on natural or synthetic constitutive or controllable promoter sequences. The alignment was made in CLC Main Workbench 6 (CLC bio, Aarhus, Denmark) and recolored. Part **b** shows a starting point for an *E. coli* SPL based on the rRNA promoter alignment in Part **a**. The −35 and −10 boxes are highlighted in *purple*. The *text* and *arrows* suggest ways to customize SPLs, e.g. by adding a weak, moderate or strong natural or synthetic leader sequence downstream of the SPL (Solem and Jensen 2002). The SPL can be inserted in a primer or used as an independent single- or double-stranded oligo. SPLs are compatible with any cloning method and can be maintained in cells either on the chromosome or extrachromosomally. Part **c** shows SPLs for eukaryotic promoters that were designed using a LEGO brick principle (high-lighted by the *colored boxes*). The *upper* sequence is the JeTr2 promoter used in mammalian cells (Tornoe et al. 2002) and the *bottom* sequence is a yeast synthetic promoter (Jeppsson et al. 2003). Besides the randomization introduced by the Ns, also the RPG and CT boxes were doped 1% (see Part **b**). A detailed protocol for experimental SPL work can be seen in Ellis (2009)

Further advantages for the SPL technology is that food-grade bacteria modified with SPLs remain food-grade (if either no antibiotic marker or a food-grade antibiotic marker is present), the gene expression level during growth is very stable, which is often not the case for inducible systems and gene expression levels higher than those for induced systems can be achieved (Jensen and Hammer 1998a; Rud et al. 2006). As the use of SPL technology is readily standardized, the 2010 DTU iGEM team '*Bi[o]stable switch*' developed BBF RCP 63, the '*DTU Synthetic Promoter Library Standard*' (DTU iGEM team 2010).

The SPL approach can be combined with changes to the −10 and −35 boxes (Lemire et al., unpublished data) and altered RBS (Ribosome Binding Site) sequences to achieve larger spans of promoter activity. It has been published that the approach suffers from some limitations (Meynial-Salles et al. 2005), one being (A) that a cloning step is needed, the other being (B) that only promoter replacement is allowed which limits the maximal level of expression of a gene present in only one copy (e.g. the RBS is not enhanced and therefore lower expression levels are obtained). (A) A cloning step is not needed as procedures analogous to the one used by Meynial-Salles et al. can be done with the SPL technology, e.g. SOE-PCR (Splicing by Overlap Extension PCR), recombineering approaches or *de novo* oligo synthesis (Horton et al. 1989; Meynial-Salles et al. 2005; Wang et al. 2009; Dehli T. et al., unpublished data). (B) One can either respect the original RBS (i.e. promoter replacement only) when designing an SPL or incorporate another. We routinely use the aldolase leader of *L. lactis* in front of SPLs for cloning of heterologous genes due to its mRNA stability and well-defined RBS (Solem et al. 2008b). SPLs can be designed based on the strongest promoters in a particular genome, e.g. rRNA promoters (Fig. 9.1). This coupled with an optimized RBS gives the highest expression level one can achieve in a single chromosomal copy. The randomization of the base pairs then yields different gene expression levels in small steps between the strongest promoters possible and the weakest possible (*de facto* gene knock-outs (KO) as the weakest promoters show no gene expression). Also when using degenerate RBS sequences to manipulate gene expression levels in operons one should be aware that only the gene closest to the operon promoter is manipulated unless the RBS sequences upstream of all genes in the operon are changed accordingly. Changing the promoter typically changes the individual gene expression levels in an operon equally (Solem and Jensen 2002).

Jeppsson et al. showed how the SPL technology can be applied in yeast (Fig. 9.1). Here, the level of glucose-6-phosphate dehydrogenase (G6PDH) was down-regulated using a promoter from an SPL which modulated the gene expression 0.01–1.79 fold around the wild-type level. Three selected promoters were used to replace the native *ZMF1* promoter. Compared to the *ZWF1*-KO strain, the weakest of these promoters (6% of wild-type *ZWF1*-level) resulted in a 5.1-fold increased xylose consumption rate (77% of wild-type level), and a growth rate similar to that of the wild-type in a lignocellulosic hydrolysate. This illustrates how the SPL technology can also be used to fine-tune gene expression below wild-type levels (Jeppsson et al. 2003).

Tornoe et al. designed a small synthetic highly active promoter using a LEGO brick principle (Fig. 9.1). The spacer sequence of the JeT promoter was randomized to create an SPL which was tested with the reporter genes EGFP (Enhanced GFP)

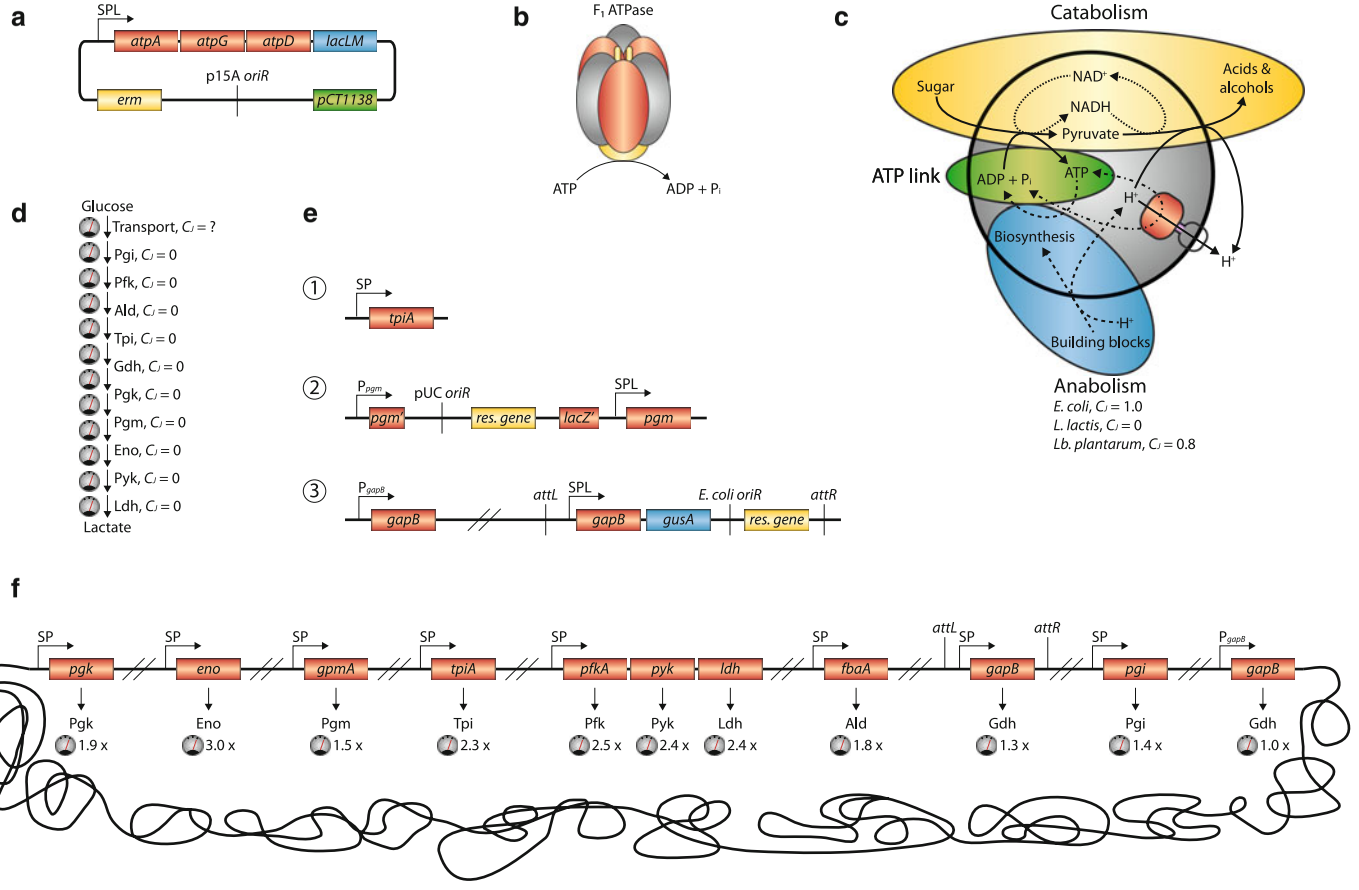

a
SPL
atpA
atpG
atpD
lacLM
p15A oriR
erm
pCT1138
b
F1 ATPase
ATP
ADP + Pi
c
Catabolism
Sugar
NAD+
NADH
Pyruvate
Acids & alcohols
ATP link
ADP + Pi
ATP
H+
Biosynthesis
H+
H+
Building blocks
Anabolism
E. coli, Cj = 1.0
L. lactis, Cj = 0
Lb. plantarum, Cj = 0.8
d
Glucose
Transport, Cj = ?
Pgi, Cj = 0
Pfk, Cj = 0
Ald, Cj = 0
Tpi, Cj = 0
Gdh, Cj = 0
Pgk, Cj = 0
Pgm, Cj = 0
Eno, Cj = 0
Pyk, Cj = 0
Ldh, Cj = 0
Lactate
e
1
SP
tpiA
2
Ppgm
pUC oriR
SPL
pgm'
res. gene
lacZ'
pgm
3
PgapB
attL
SPL
E. coli oriR
attR
gapB
gapB
gusA
res. gene
f
SP
pgk
Pgk
1.9 x
SP
eno
Eno
3.0 x
SP
gpmA
Pgm
1.5 x
SP
tpiA
Tpi
2.3 x
SP
pfkA
Pfk
2.5 x
pyk
Pyk
2.4 x
ldh
Ldh
2.4 x
SP
fbaA
Ald
1.8 x
attL
SP
gapB
Gdh
1.3 x
attR
SP
pgi
Pgi
1.4 x
PgapB
gapB
Gdh
1.0 x

Fig. 9.2 MCA and control by demand. Part **a** of the figure shows the general structure of the vectors used to investigate how much control the anabolic reactions of *E. coli*, *L. lactis* and *Lb. plantarum* have over the glycolytic flux in these organisms (Koebmann et al. 2002a, b; Rud et al. 2008). *LacLM* encodes β-galactosidase, *erm* confers resistance to erythromycin, p15A and pCT1138 are replicons for *E. coli* and *L. lactis*, respectively. *atpAGD* encode the F_1 part of the ATP synthase (F_1 ATPase), which is shown in Part **b**. The *atpAGD* operon is transcriptionally coupled to *lacLM* for screening purposes. Part **b** shows the F_1 ATPase that catalyzes ATP to ADP (Koebmann et al. 2002c). The *red*, *grey* and *yellow* subunits are called α, β and γ and are encoded by *atpA*, *atpD* and *atpG*, respectively. Part **c** shows free energy metabolism in bacteria (Koebmann et al. 2002c). The F_1 ATPase is shown in *red* together with the F_O part encoding the transmembrane region (*purple*). Catabolism, anabolism and the ATP link between the former two is shown in *yellow*, *blue* and *green*, respectively. *Dotted*, *full*, *dashed* and *semi-dashed lines* symbolize redox metabolism, energy generation, energy consumption and synthesis/hydrolysis of ATP, respectively. In strains expressing the plasmid from Part **a**, the ATP degradation is uncoupled from biomass formation to different well-defined levels using an SPL and the control exerted by the anabolic reactions of an organism on glycolysis can thus be estimated using MCA. The flux control coefficient (C_J) is calculated as $(dJ/J)/(dE/E)$, where E is enzyme concentration and J is the flux. The control exerted by the anabolic reactions on glycolysis varies greatly from organism to organism. In *E. coli*, *Lb. plantarum* and *L. lactis* 100, 80 and 0% of the control over glycolysis resided in ATP consuming reactions. The diagrams in (**a**), (**b**) and (**c**) are adapted from the original publications. Part **d** shows the result after all glycolytic enzymes of *L. lactis* as well as *ldh* have been subjected to MCA (see text for references). The gene expression levels and thereby the enzyme activities have been tuned through the three general approaches shown in Part **e**. Pgi, phosphoglucose isomerase; Pfk, phosphofructokinase; Ald, aldolase; Tpi, triosephosphate isomerase; Gdh, glyceraldehyde 3-phosphate dehydrogenase; Pgk, phosphoglycerate kinase; Pgm, phosphoglycerate mutase; Eno, enolase; Pyk, pyruvate kinase; Ldh, lactate dehydrogenase. In Part **e** the three general strategies used to manipulate all the genes in Part **d** are shown. (1) The native promoter is substituted with a synthetic promoter (SP) isolated from an SPL in a double crossover event; (2) The wild type promoter is replaced by an SPL in a single crossover event; (3) An additional copy of the gene of interest is placed in the *attB* phage integration site of *L. lactis*. In Part **f** an *L. lactis* strain with all glycolytic enzymes and *ldh* up-regulated app. two-fold is shown. Each gene has been screened with its own SPL separately to find the appropriate expression level for at two-fold overexpression of the corresponding enzyme activity. That particular promoter has then been cloned into another vector to allow replacement of the wild type promoter on the chromosome in a double crossover event. This 2× strain shows no signs of an elevated glycolytic flux (Solem and Jensen, unpublished data)

and SEAP (Secreted Alkaline Phosphatase) in HiB5, HEK 293, ARPE-19 and CHO cells. These results are promising for steady-state fine-tuning of gene expression levels in future human gene therapies thereby eliminating the need for fluctuating inducible systems (Tornoe et al. 2002).

SPLs have been used extensively to perform MCA, especially in *L. lactis* (Fig. 9.2) (Andersen et al. 2001a, b; Jørgensen et al. 2004; Koebmann et al. 2002a, b, 2005, 2006; Rud et al. 2008; Solem and Jensen 2002; Solem et al. 2003, 2007, 2008a, b, 2010).

The versatility of the SPL method and the ease of using it across species (from bacteria to yeast to mammalian cells) is an advantage that separates it from other published methods. Also one automatically obtains the relevant range of gene expression levels as promoters too weak or strong for the given purpose are counter-selected (Jeppsson et al. 2002; Meynial-Salles et al. 2005; Solem and Jensen 2002; Tornoe et al. 2002).

9.2.4 Comparing Error-Prone PCR and the SPL Technology

From a distance promoter libraries generated using either error-prone PCR or the SPL technology look similar as both generate promoter variation covering three orders of magnitude of promoter activity and yield high numbers of functional clones (35–90%) (Fischer et al. 2006; Hammer et al. 2006). However there are a number of important differences between the two when considering the steps needed to obtain the library and the usability of the library *in vivo*: (A) An SPL can be introduced using fewer cloning steps (Alper et al. 2005; Nevoigt et al. 2006; Solem and Jensen 2002); (B) The error-prone PCR based library is biased towards promoters of similar or equal strength to the wild-type promoter, as single mutations are introduced randomly and with low frequency. These single changes only affect promoter strength significantly when occurring in promoter strength hotspots, e.g. the −10 and −35 boxes of prokaryotic promoters. Assuming a single base in the typical 17 bp spacer between the −10 and −35 regions is changed, chances are high that little or no effect in promoter strength will be observed. Therefore, when using error-prone PCR for generating a library of promoters, one will have many 'false-positives', i.e. promoters that are mutated but have similar or equal promoter strength compared to the promoter used as template in error-prone PCR. This necessitates a laborious post-cloning screening process to find a suitable library to be used further on. One might even need to characterize the promoter library independently of the desired gene one wants to modulate and then subsequently clone relevant promoters in front of genes of interest (Alper et al. 2005; Nevoigt et al. 2006). (C) Due to (B) one always needs a good reporter system when using error-prone PCR as screening of the whole, or even sub-sets of the, library can be very inefficient for the modulated GOI. With the SPL technology one achieves a widespread level of activity covering two to three orders of magnitude of gene expression levels with as little as 20 clones. Therefore one do not necessarily need a reporter gene for the SPL technology as long as an assay or an antibody for the protein of interest exists (Solem and Jensen 2002); (D) For multiple

use the SPL technology provides more genetically stable strains as the error-prone based method features long stretches (>400 bp) of highly repetitive sequences (>95% homology). E.g. wild-type *S. cerevisiae* or *E. coli* strains used in recombineering need as little as 40 and 90 bp, respectively, for efficient recombination of two or more identical DNA sequences. This is highly important in the future of synthetic biology where whole genomes are likely to be manipulated simultaneously (Alper et al. 2005; Blount et al. 2012a, b; Nevoigt et al. 2006; Solem and Jensen 2002; Wang et al. 2009; Solem and Jensen, unpublished data); (E) If the promoter chosen for error-prone PCR is endogenous to the host cell there is a higher risk of pleiotropic effects affecting the promoter library, as only single mutations are introduced (Alper et al. 2005; Blount et al. 2012a).

9.2.5 Inducible SPLs

The expression level of single promoters in standard SPLs is considered constitutive (Rud et al. 2006) potentially leading to problems if the product of the GOI expressed is toxic to the cell resulting in undesirable mutations either in the promoter or the GOI complicating later analysis. This can be circumvented using inducible SPLs. Also the sometimes observed phenomenon of all-or-nothing induction of inducible promoters can effectively be avoided with inducible SPLs (Keasling 1999; Khlebnikov et al. 2000; Siegele and Hu 1997).

SPLs that respond to the general amino acid level, the arginine level and heat-shock have been developed previously in our laboratory (Jensen and Hammer 1998c) and other inducible systems are currently being developed (Rytter et al., manuscript in preparation). The strategy to regulate cell function by making SPLs based on natural controllable promoters should be applicable to e.g. elevated osmolarity, low pH, temperature, etc. as well thereby significantly expanding the toolbox for both basic and applied research.

The 2011 DTU iGEM team '*The Universal Tool for Gene Silencing*' developed an inducible SPL based on the arabinose promoter P_{BAD} (BioBrick part BBa_I0500). These promoters were not active in the absence of arabinose, but were induced to different well-defined levels in the presence of 0.2% arabinose. This brings together the best of both worlds of constitutive and inducible promoters to obtain genetic constructs that can be repressed until needed and then expressed to well-defined levels covering a wide range of expression levels in small steps (Ellis et al. 2009; DTU iGEM team 2011).

Miksch et al. used a restricted SPL technology approach to design a library of promoters that were only expressed in *E. coli* stationary phase. Due to the regulation of stationary promoters the consensus sequences were kept strict and based on the design only 4,608 different promoters were theoretically possible compared to the $>10^{15}$ possibilities using a standard SPL (Miksch et al. 2005a; Solem and Jensen 2002). Nonetheless twentyfold different expression levels were obtained with the strongest promoters being 3–4-fold stronger than known *E. coli* stationary phase

promoters. Also a difference in induction time upon stationary phase entry was observed, i.e. some promoters were induced in the late exponential phase while others were not induced until >2.5 h into stationary phase. The window of induction time was −1.5 to 2.7 relative to stationary phase entry (0) covered in small steps by the SPL. The natural stationary phase promoters were induced in late exponential phase or upon entry to the stationary phase (relative induction time −0.7 to 0.3) (Miksch et al. 2005a, 2005b). The induction fold from non-induced (exponential phase) to induced (stationary phase) was 4–6,800 for the SPL compared with 3–4 for natural stationary phase promoters. Promoters of similar strength showed different induction folds and degree of leakiness (i.e. expression during exponential phase). The low induction fold synthetic promoters were most leaky during exponential growth (Miksch et al. 2005b).

This strategy allows the biotech engineer to modulate and fine-tune fermentation processes in great detail using a single SPL. Thorough investigation of a library will enable optimal promoter choice in a three-dimensional space comprising optimal gene expression level; time of induction onset; and induction fold upon stationary phase entry and leakiness during exponential growth to obtain the best production characteristics for a certain product (Miksch et al. 2005a, b). These studies nicely show that stationary phase promoters stronger than those known today can be designed and the use of these promoters is very promising for the many applications that prefer or require induction only upon stationary phase entry (Miksch et al. 2005a).

Ellis et al. showed how inducible SPLs can be used to generate controllable genetic networks (Fig. 9.3). In this paper the Jensen-Hammer method (Jensen and Hammer 1998a) was extended to incorporate the TetR- and LacI-based repressor systems in the promoter sequence thereby equipping scientists with synthetic orthogonal promoters that can be induced to specific levels spanning a 22-fold range of gene expression levels in small steps. As only app. 3% of the obtained library (clones showing microtiter plate detectable yEGTP (yeast EGFP) expression levels only in presence of ATc (Anhydratetracycline)) was screened, the range of promoter expression levels is perhaps even higher yielding a more diverse library. The SPL consensus sequence was based on the *GAL1* promoter with an UAS (Upstream Activating Signal) 5′ of the SPL (Ellis et al. 2009). The TetR and LacI regulated SPL constructs and data obtained from screening the promoter libraries were used to create and model a type II negative feed-forward loop network (Mangan and Alon 2003) and a genetic toggle switch (Gardner et al. 2000); the behavior of both could be predicted qualitatively and quantitatively. This plug-and-play use of the SPL technology coupled with mathematical modeling to yield predictable functions *in vivo* is very promising for the future use of synthetic gene networks in all aspects of biotechnology (Ellis et al. 2009).

The reporter gene yEGTP was controlled by TetR, which was expressed on the same genome-integrated vector by the strong *TEF1* promoter. TetR binds the two $tetO_2$-boxes incorporated in the synthetic promoters ensuring low levels of gene expression without TetR inhibitor ATc (Ellis et al. 2009). The randomization of sequences close to the $tetO_2$- and *LacO*-repressors also affected the binding affinity of TetR and LacI thereby leading to both more and less leaky promoters (from 0.4

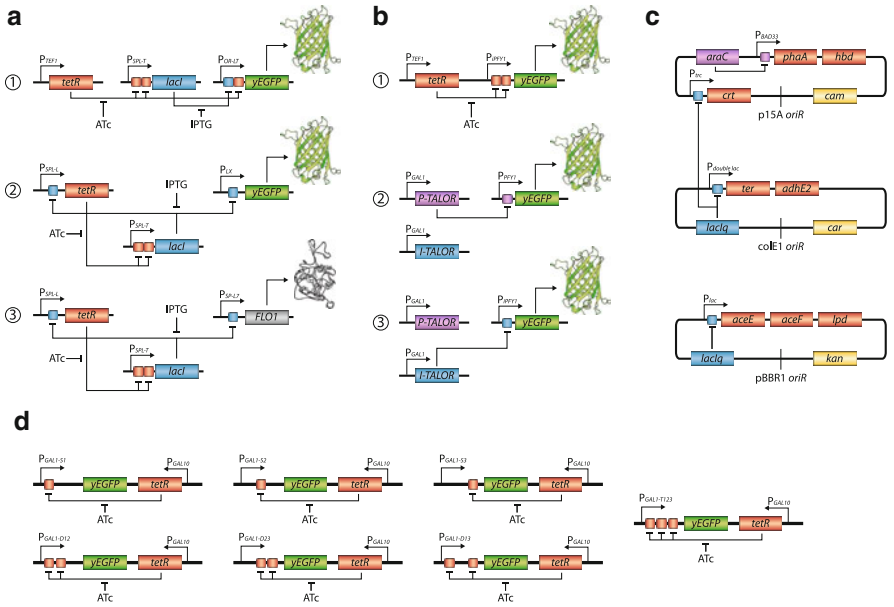


Fig. 9.3 Gene networks with predicted functions. Part **a** (Ellis et al. 2009). (1) Negative feed-forward loop controlling expression of *yEGFP*. The behavior of the negative feed-forward loop was predicted using a mathematical model and verified in ATc-IPTG dose–response studies monitoring *yEGFP* fluorescence; (2) and (3) Genetic toggle switches constructed from two promoter libraries acting as predictable genetic timer networks in either reset timing of *yEGFP* expression or *FLO1*-regulated yeast sedimentation. P_{TEF1}, constitutive promoter; P_{SPL-T}, promoter from the TetR-regulated SPL; P_{OR-LT}, TetR- and LacI-regulated OR-gate promoter; P_{SPL-L}, promoter from the LacI-regulated SPL; P_{LX}, strong LacI-regulated control promoter; P_{SP-L7}, promoter L7 from the LacI-regulated SPL; $tetO_2$, *red square*; *lacO*, *blue square*. Part **b** (Blount et al. 2012b). (1) Rational engineering of TetR-regulation into P_{PFY1} creating the inverter device P_{iPFY1} that carries two $tetO_2$-sites (*red squares*); (2) and (3) Engineered TALORs only bind and repress a well-defined stretch of DNA and show no crosstalk even though *P-TALOR* (*purple square*) and *I-TALOR* (*blue square*) operator sites are very similar (9 bp homologous region); P_{GAL1}, galactose inducible promoter; *P-TALOR*, TALOR protein recognizing a 17 bp region in the *PFY1* promoter region; *I-TALOR*, TALOR protein recognizing a 16 bp region in the *iPFY1* promoter region. Part **c** shows a three-plasmid system used for butanol production in *E. coli* (Bond-Watts et al. 2011). When IPTG and L-arabinose are added all genes (*red*) involved in butanol production are transcribed. Bond-Watts et al. needed to test a number of different promoters to obtain high butanol titers underlining the importance of tools effective in fine-tuning gene expression. P_{BAD33}, AraC (*purple*) regulated promoter; P_{trc}, P_{lac} and $P_{double\ lac}$, LacIq (*blue*) regulated promoters; *phaA*, acetoacetyl-CoA thiolase/synthase; *hbd*, 3-hydroxybutyryl-CoA dehydrogenase; *crt*, crotonase; *ter*, trans-enoyl-CoA reductase; *adhE2*, bifunctional butyraldehyde and butanol dehydrogenase; *aceEF-lpd*, pyruvate dehydrogenase complex (PDHc); *cam*, *car* and *kan*, chloramphenicol, carbenicillin and kanamycin resistance markers, respectively. Part **d** shows a combinatorial promoter design to study the effects of various operator site combinations on gene expression output variables such as dose–response curves and gene expression noise (Murphy et al. 2007). P_{GAL1} and P_{GAL10}, galactose inducible promoter; S, D and T means single, double or triple $tetO_2$ operators (*red squares*); the numbers *1*, *2* and *3* refer to the positioning of $tetO_2$ operator sites (All figures are adapted from the original publications)

to 4 fold around the control (leakiness measure) for TetR and 0.3–12 fold for LacI). However since the induction ratio was 4–129 fold for TetR and 2–55 for LacI over the range of promoters there was a clear-cut distinction between on and off in this study (Ellis et al. 2009).

The above methods involve well-characterized activator or repressor systems that can be used only once without crosstalk which is less attractive for synthetic biology purposes where rapid and repeated reengineering is central. TAL (Transcription Activator-Like) effector technology provides a solution as these specific DNA binding proteins can readily be altered for target site recognition (Bogdanove and Voytas 2011; Cermak et al. 2011). Blount et al. developed a system were TALORs (TAL Orthogonal Repressor) inhibiting natural host function could be readily designed and coupled with the SPL technology to allow tailor-making of both regulatory sites and promoter strength; properties highly desirable for multiple synthetic biology purposes (Fig. 9.3). This strategy can effectively remove the current limits to build complex synthetic networks (Blount et al. 2012b).

9.3 Predicting Phenotype from Genotype

An important goal of synthetic biology is to bring about reliable, higher-order functionality by wiring together basic circuit modules that can sense and process signals, perform logic operations and actuate biological responses (Khalil and Collins 2010). To this end Murphy et al. developed a combinatorial promoter design strategy to elucidate how the gene expression level and noise was affected when the architecture of yeast *GAL1* promoter was manipulated using $tetO_2$ sites (Murphy et al. 2007). The effect of operator positions within the *GAL1* promoter of *S. cerevisiae* was studied through the combinatorial set of seven promoters carrying one, two or three $tetO_2$ operator sites (Fig. 9.3). A generic computational model that described the dose–response curves and gene expression noise dependent on the promoter architecture was also developed. In this case the system predictability decreases when the number of operator sites is elevated underlining that bottom-up approaches using well-characterized parts might not be suitable for predicting the behavior of complex synthetic gene networks (Murphy et al. 2007).

Cox et al. created a very diverse library spanning five orders of magnitude of activity when combining the activators AraC and LuxR and the repressors LacI and TetR in a random fashion (up to three regulatory elements in each construct). The promoter sequence was divided into *distal*, *core* and *proximal* regions representing the 45 bp upstream of the −35 box, the region between the −35 and −10 box and 30 bp downstream of the −10 box, respectively (Cox et al. 2007).

The study showed that expression levels are restricted by an activation ceiling i.e. activation is limited by the absolute expression level (the higher the unregulated activity the less induction can be achieved with an activator). Thus in general activator operated promoters show low unregulated activity while repressor operated promoters show high unregulated activity (Cox et al. 2007).

Activation functions only in the *distal* region, but more surprisingly activation operators placed in the *core* and *proximal* regions resulted in neither inducible activation nor repression. Repression functions in both *distal*, *core* and *proximal* regions although repressors bound at the *core* region were more effective than *proximal* repressors that again were more effective than *distal* repressors. It was also shown that repressor positioning was more important than multiplicity. This can be used to design promoter libraries with few operator sites and still allow effective and predictable expression levels. Also in systems with more than one repressor, the dominant one can be chosen. When comparing activation and repression, repression dominated activation. These findings led to heuristic rules for promoter design (Cox et al. 2007).

In order for fine-tuning of gene expression to be truly synthetic one should be able to design promoters with a desired strength *in silico* and then use these to obtain a certain phenotype (De Mey et al. 2007; Ellis et al. 2009). To this end Jensen et al. created a model that could predict whether a certain promoter in a library was weak or strong and identify critical sites in the promoters that led to either weak or strong activity. This qualitative model was a first step to link promoter sequence to promoter strength (Jensen et al. 2006).

Gertz et al. investigated a library of synthetic promoters to construct a thermodynamic model that could accurately predict gene expression levels from promoter sequence in yeast. A thermodynamic model was built under the assumptions that gene regulation is controlled completely by DNA-protein and protein-protein interactions. The model calculates the probability that RNA polymerase binds to the promoter and then assumes that this results in gene transcription and protein production (Gertz et al. 2009).

The model was tested for seven different libraries and could explain 44–59% of the variance in these. This shows that a large part of the gene expression level can be predicted based on a simple thermodynamic DNA-protein and protein-protein model, but other factors can be considered in a more advanced model to increase the fraction of variance explained (Gertz et al. 2009; Pruss and Drlica 1989; Pérez-Martín et al. 1994). Also when applying one of the models based on the libraries of synthetic promoters to genome promoters, Gertz et al. were able to identify a number of genes not previously known to be regulated by the yeast transcription factor Mig1 (Gertz et al. 2009). This study illustrates how models based on investigation of synthetic parts can provide valuable information about natural systems.

De Mey et al. showed that promoter strengths in *E. coli* cannot be predicted from gene sequence based on (A) promoter anomalies as also shown earlier by Jensen and Hammer (De Mey et al. 2007; Jensen and Hammer 1998c); (B) the DNA sequence of individual promoters and spacers, i.e. no clear relationship between promoter strength and degree of alignment was observed; (C) the presence or absence of strong or weak nucleotide substrings (De Mey et al. 2007).

Instead promoter strength was predicted quantitatively based on a PLS-R(Partial Least Squares Regression) model named PSP (Promoter Strength Predictive) using nucleotide sequences of 49 promoters from an *E. coli* SPL. 42 of the promoters were used as training set in a leave-one-out cross-validated model determining the

number of latent variables using Wold's R criterion. The seven remaining promoters were used to test the model (De Mey et al. 2007, 2010; Wold 1978). The PSP model predicted promoter strength reasonably accurately and is generic as it can be used to determine the unregulated strength of natural *E. coli* promoters as well (De Mey et al. 2007, 2010). These findings are surprising as only 42 promoters used as a training set could seemingly predict the strength of six out of seven promoters taken from a library theoretically encoding $>10^{16}$ different sequences. Other attempts to predict gene expression levels from gene sequence have not been as successful, but perhaps the ability of PLS-R to predict latent structures in multivariate space has been the missing link in previous studies (Beer and Tavazoie 2004; Gertz et al. 2009). De Mey et al. verified the PSP model (De Mey et al. 2007) by accurately estimating the relative expression level of the native *E. coli ppc* gene and two synthetic promoters thus illustrating that the PSP model can be used instead of a time-consuming screening step when analyzing the promoter strength in a library (De Mey et al. 2010).

The use of PLS-R to estimate promoter strength *in silico* opens up for true synthetic biology by eliminating the need for characterization of created promoter libraries. Which promoter strength to choose to obtain a certain phenotype can be predicted using mathematical modeling as shown by Ellis et al. enabling rational design (De Mey et al. 2007, 2010; Ellis et al. 2009).

9.4 Conclusion

The experimental breakthrough that enabled quantitative measurements of gene expression levels of different promoter architectures was the invention of combinatorial promoter libraries such as the SPL technology (Mukherji and Van Oudenaarden 2009). Promoter libraries with a wide range of transcription activity can be made, investigated and integrated in *in silico* models and then utilized to develop synthetic gene networks that are both predictable and require no significant post-hoc adjustments (Lu et al. 2009).

Three main ways for producing component libraries for synthetic biology exists: (A) combinatorial shuffling (Cox et al. 2007; Gertz et al. 2009; Ligr et al. 2006; Murphy et al. 2007); (B) multi-round mutation/selection of an existing component (Alper et al. 2005, 2006) and (C) diversity-inherent component synthesis (Jensen and Hammer 1998a, b). Of these the third method appears most compatible with synthetic biology as it is rapid and generic as well as enables orthogonality, modularity, standardization and rational design of new components (Ellis 2009; Khalil and Collins 2010; Lu et al. 2009). Furthermore, the predictable behavior of constitutive SPLs makes them highly suitable for systems biology studies such as MCA.

With the steps taken by Ellis et al. (2009) one is now able to predict phenotype from measurements of promoter strength in simple systems using simple models. In the future this work should be taken further in order to accurately predict more complex behavior in e.g. real-time fermentations. Also steps have been taken to

develop tools for *in silico* prediction of the strength of promoters and someday we might not even have to measure promoter strength, but can just go straight to the laboratory and obtain the cell function of interest.

9.5 Outlook

A very interesting technology developed recently is MAGE (Multiplex Automated Genome Engineering) from the laboratory of George M. Church at Harvard Medical School. In essence MAGE allows the scientist to optimize multiple genetic loci in a given cell (or cell population) simultaneously. In the proof-of-concept paper Wang et al. modified 24 genetic components in the DXP (1-deoxy-D-xylulose-5-phosphate) pathway simultaneously to obtain a five-fold increase in lycopene production in only 3 days (Wang et al. 2009).

The approach introduced either inactivation mutations (double nonsense) or translation efficiency mutations (degenerated RBSs) into genomes using recombineering. Detailed *a priori* knowledge was required to pinpoint genes targeted for inactivation and genes targeted for RBS changes thereby possible narrowing the solution space by adding constraints. SPLs generate 1,000-fold different expression levels ranging from a *de facto* gene KO to gross overexpression while ensuring a stable genome without repetitive sequences and could thus be used as an unbiased tool to KO or overexpress genes in MAGE.

Alper et al. created another interesting technology, gTME (global Transcription Machinery Engineering), where sigma factors were altered using error-prone PCR. This led to genome-scale changes in expression levels and desired properties emerging from the new transcription profile of the organism studied could then be screened or selected for. To minimize pleiotropic effects, the mutated sigma factors were expressed in strains carrying wild-type sigma factors. Using this approach Alper et al. obtained multiple interesting phenotypes in both *S. cerevisiae* and *E. coli* suggesting that the method is well suited to obtain complex cell functions that are not readily accessible with traditional metabolic engineering approaches (Alper et al. 2006; Alper and Stephanopoulos 2007).

The ability to generate such tremendous diversity immediately calls for highly efficient screening or selection procedures. Wang et al. conveniently used the red color of lycopene producing strains to screen for colonies of interest (Wang et al. 2009) and Alper et al. used selection pressure to find ethanol tolerant strains after application of gTME (Alper et al. 2006; Alper and Stephanopoulos 2007). However it is far more difficult to screen for the production of e.g. ethanol or butanol in a huge population. Assuming that only four genes are modulated simultaneously using a MAGE-SPL approach that generate 100 different levels of expression for each gene one obtains a cell population containing up to 100 million unique gene expression profiles. With the best microfluidic techniques available today, it would take 10 h to screen such a library (Agresti et al. 2010). If one instead modulates six genes it would take more than 11 years to analyze the full size library. Therefore

future developments should focus on the ability to effectively select and screen for functionalities (e.g. catalysis) and production of metabolites (e.g. biofuels and biochemicals) in huge and diverse cell populations. Currently our abilities to generate sequence variation by far outperform our abilities to analyze the variation created.

References

Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proc Natl Acad Sci U S A 107:4004–4009

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9:258–267

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565–1568

Andersen HW, Pedersen MB, Hammer K, Jensen PR (2001a) Lactate dehydrogenase has no control on lactate production but has a strong negative control on formate production in *Lactococcus lactis*. Eur J Biochem 268:6379–6389

Andersen HW, Solem C, Hammer K, Jensen PR (2001b) Twofold reduction of phosphofructokinase activity in *Lactococcus lactis* results in strong decreases in growth rate and in glycolytic flux. J Bacteriol 183:3458–3467

Bai Flagfeldt D, Siewers V, Huang L, Nielsen J (2009) Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. Yeast 26:545–551

Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. Cell 117:185–198

Blount BA, Weenink T, Ellis T (2012a) Construction of synthetic regulatory networks in yeast. FEBS Lett 586:2112–2121

Blount BA, Weenink T, Vasylechko S, Ellis T (2012b) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One 7:e33279 1–e33279 11

Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333:1843–1846

Bond-Watts BB, Bellerose RJ, Chang MCY (2011) Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat Chem Biol 7:222–227

Carothers JM, Goler JA, Keasling JD (2009) Chemical synthesis using synthetic biology. Curr Opin Biotechnol 20:498–503

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82 1–e82 11

Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Syst Biol 3:1–11

Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 39:1131–1141

De Mey M, Maertens J, Lequeux G, Soetaert W, Vandamme E (2007) Construction and model-based analysis of a promoter library for *E. coli*: An indispensable tool for metabolic engineering. BMC Biotechnol 7:1–14

De Mey M, Maertens J, Boogmans S, Soetaert W, Vandamme E, Cunin R, Foulquié-Moreno M (2010) Promoter knock-in: a novel rational method for the fine tuning of genes. BMC Biotechnol 10:1–9

DTU iGEM team (2010) Bi[o]stable switch. http://2010.igem.org/Team:DTU-Denmark. Accessed 7 Apr 2012
DTU iGEM team (2011) The universal tool for gene silencing. http://2011.igem.org/Team:DTU-Denmark. Accessed 7 Apr 2012
Ellis T (2009) Synthesis and screening of regulatory component libraries for synthetic biology. Nat Protoc Exchange http://www.nature.com/protocolexchange/protocols/561
Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol 27:465–471
Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol 3:109–118
Fischer CR, Alper H, Nevoigt E, Jensen KL, Stephanopoulos G (2006) Response to Hammer et al.: tuning genetic control—importance of thorough promoter characterization versus generating promoter diversity. Trends Biotechnol 24:55–56
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342
Gertz J, Siggia ED, Cohen BA (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature 457:215–218
Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries-tuning of gene expression. Trends Biotechnol 24:53–55
Hansen ME, Wangari R, Hansen EB, Mijakovic I, Jensen PR (2009) Engineering of *Bacillus subtilis* 168 for increased nisin resistance. Appl Environ Microbiol 75:6688–6695
Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68
Horwitz M, Loeb LA (1986) Promoters selected from random DNA sequences. Proc Natl Acad Sci U S A 83:7405–7409
Horwitz M, Loeb LA (1988) DNA sequences of random origin as probes of *Escherichia coli* promoter architecture. J Biol Chem 263:14724–14731
Israelsen H, Madsen SM, Vrang A, Hansen EB, Johansen E (1995) Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-lacZ integrants with the new promoter probe vector, pAK80. Appl Environ Microbiol 61:2540–2547
Jaurin B, Grundström T, Normark S (1982) Sequence elements determining ampC promoter strength in *E. coli*. EMBO J 1:875–881
Jensen PR, Hammer K (1998a) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64:82–87
Jensen PR, Hammer K (1998b) Artificial promoters for metabolic optimization. Biotechnol Bioeng 58:191–195
Jensen PR, Hammer K (1998c) Artificial promoter libraries for selected organisms and promoters derived from such libraries. Patent WO 98/07846 1–90
Jensen K, Alper H, Fischer C, Stephanopoulos G (2006) Identifying functionally important mutations from phenotypically diverse sequence data. Appl Environ Microbiol 72:3696–3701
Jeppsson M, Johansson B, Hahn-Hägerdal B, Gorwa-Grauslund MF (2002) Reduced oxidative pentose phosphate pathway flux in recombinant xylose-utilizing *Saccharomyces cerevisiae* strains improves the ethanol yield from xylose. Appl Environ Microbiol 68:1604–1609
Jeppsson M, Johansson B, Jensen PR, Hahn-Hägerdal B, Gorwa-Grauslund MF (2003) The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. Yeast 20:1263–1272
Jørgensen CM, Hammer K, Jensen PR, Martinussen J (2004) Expression of the pyrG gene determines the pool sizes of CTP and dCTP in *Lactococcus lactis*. Eur J Biochem 271:2438–2445
Keasling JD (1999) Gene-expression tools for the metabolic engineering of bacteria. Trends Biotechnol 17:452–460
Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11:367–379

Khlebnikov A, Risa Ø, Skaug T, Carrier TA, Keasling J (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. J Bacteriol 182:7029–7034

Koebmann BJ, Westerhoff HV, Snoep JL, Nilsson D, Jensen PR (2002a) The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. J Bacteriol 184:3909–3916

Koebmann BJ, Solem C, Pedersen MB, Nilsson D, Jensen PR (2002b) Expression of genes encoding F1-ATPase results in uncoupling of glycolysis from biomass production in *Lactococcus lactis*. Appl Environ Microbiol 68:4274–4282

Koebmann BJ, Andersen HW, Solem C, Jensen PR (2002c) Experimental determination of control of glycolysis in *Lactococcus lactis*. Antonie Van Leeuwenhoek 82:237–248

Koebmann B, Solem C, Jensen PR (2005) Control analysis as a tool to understand the formation of the las operon in *Lactococcus lactis*. FEBS J 272:2292–2303

Koebmann B, Solem C, Jensen PR (2006) Control analysis of the importance of phosphorglycerateenolase for metabolic fluxes in *Lactococcus lactis* subsp. *lactis* IL1403. IEE Proc Syst Biol 153:346–349

Ligr M, Siddharthan R, Cross FR, Siggia ED (2006) Gene expression from random libraries of yeast promoters. Genetics 172:2113–2122

Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. Nat Biotechnol 27:1139–1150

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100:11980–11985

Meynial-Salles I, Cervin MA, Soucaille P (2005) New tool for metabolic pathway engineering in *Escherichia coli*: one-step method to modulate expression of chromosomal genes. Appl Environ Microbiol 71:2140–2144

Miksch G, Bettenworth F, Friehs K, Flaschel E (2005a) The sequence upstream of the −10 consensus sequence modulates the strength and induction time of stationary-phase promoters in *Escherichia coli*. Appl Microbiol Biotechnol 69:312–320

Miksch G, Bettenworth F, Friehs K, Flaschel E, Saalbach A, Twellmann T, Nattkemper TW (2005b) Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in *Escherichia coli*. J Biotechnol 120:25–37

Mukherji S, Van Oudenaarden A (2009) Synthetic biology: understanding biological design from synthetic circuits. Nat Rev Genet 10:859–871

Murphy KF, Balázsi G, Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. Proc Natl Acad Sci U S A 104:12726–12731

Nevoigt E, Kohnke J, Fischer CR, Alper H, Stahl U, Stephanopoulos G (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl Environ Microbiol 72:5266–5273

Pérez-Martín J, Rojo F, De Lorenzo V (1994) Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol Rev 58:268–290

Pruss GJ, Drlica K (1989) DNA supercoiling and prokaryotic transcription. Cell 56:521–523

Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol 77:3600–3608

Rossi JJ, Soberon X, Marumoto Y, McMahon J, Itakura K (1983) Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives. Proc Natl Acad Sci U S A 80:3203–3207

Rud I, Jensen PR, Naterstad K, Axelsson LA (2006) Synthetic promoter library for constitutive gene expression in *Lactobacillus plantarum*. Microbiology 152:1011–1019

Rud I, Solem C, Jensen PR, Axelsson L, Naterstad K (2008) Co-factor engineering in lactobacilli: effects of uncoupled ATPase activity on metabolic fluxes in *Lactobacillus (L.) plantarum* and *L. sakei*. Metab Eng 10:207–215

Rytter JV, Helmark S, Solem C, Lezyk MJ, Jensen PR (manuscript in preparation) Synthetic promoter libraries for metabolic engineering of *Corynebacterium glutamicum*

Siegele DA, Hu JC (1997) Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proc Natl Acad Sci U S A 94:8168–8172

Solem C, Jensen PR (2002) Modulation of gene expression made easy. Appl Environ Microbiol 68:2397–2403

Solem C, Koebmann BJ, Jensen PR (2003) Glyceraldehyde-3-phosphate dehydrogenase has no control over glycolytic flux in *Lactococcus lactis* MG1363. J Bacteriol 185:1564–1571

Solem C, Koebmann B, Yang F, Jensen PR (2007) The las enzymes control pyruvate metabolism in *Lactococcus lactis* during growth on maltose. J Bacteriol 189:6727–6730

Solem C, Koebmann B, Jensen PR (2008a) The extent of co-metabolism of glucose and galactose by *Lactococcus lactis* changes with the expression of the lacSZ operon from *Streptococcus thermophilus*. Biotechnol Appl Biochem 50:35–40

Solem C, Koebmann B, Jensen PR (2008b) Control analysis of the role of triosephosphate isomerase in glucose metabolism in *Lactococcus lactis.* IET Syst Biol 2:64–72

Solem C, Petranovic D, Koebmann B, Mijakovic I, Jensen PR (2010) Phosphoglycerate mutase is a highly efficient enzyme without flux control in *Lactococcus lactis*. J Mol Microbiol Biotechnol 18:174–180

Tornoe J, Kusk P, Johansen TE, Jensen PR (2002) Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297:21–32

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898

Wold S (1978) Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20:397–405

Chapter 10
Analysis of *Corynebacterium glutamicum* Promoters and Their Applications

Jan Nešvera, Jiří Holátko, and Miroslav Pátek

Abstract Promoters are DNA sequences which function as regulatory signals of transcription initiation catalyzed by RNA polymerase. Since promoters substantially influence levels of gene expression, they have become powerful tools in metabolic engineering. Methods for their localization used in *Corynebacterium glutamicum* and techniques for the analysis of their function are described in this review. *C. glutamicum* promoters can be classified according to the respective σ factors which direct RNA polymerase to these structures. *C. glutamicum* promoters are recognized by holo-RNA polymerase formed by subunits $\alpha_2\beta\beta'\omega+\sigma$. *C. glutamicum* codes for seven different sigma factors: the principal sigma factor σ^A and alternative sigma factors σ^B, σ^C, σ^D, σ^E, σ^H and σ^M, which recognize various classes of promoters. The promoters of housekeeping genes recognized by σ^A, which are active during the exponential growth, form the largest described group. These promoters and their mutant derivatives are the most frequently used elements in modulation of gene expression in *C. glutamicum*. Promoters recognized by alternative sigma factors and their consensus sequences are gradually emerging.

Keywords *Corynebacterium glutamicum* • Gene expression • Promoter • Promoter consensus sequence • Sigma factor

Abbreviations

RACE	rapid amplification of cDNA-ends
RNAP	RNA polymerase
ROMA	run-off transcription-microarray analysis

J. Nešvera • J. Holátko • M. Pátek (✉)
Institute of Microbiology, Academy of Sciences of the Czech Republic,
Vídeňská 1083, CZ-14220 Prague 4, Czech Republic
e-mail: patek@biomed.cas.cz

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_10,

TSP transcriptional start point
WT wild-type

10.1 Introduction

Transcription initiation is the major regulatory step in gene expression in bacteria. Promoters are the specific sequences within the noncoding regulatory regions of genes which are recognized by RNA polymerase (RNAP) and determine thus the location of the transcriptional start points. The sequence of a promoter can be crucial for the frequency of transcription initiation and consequently for the level of gene expression. Since a promoter is formed by a short DNA sequence, whose regulatory functions can be described in detail, it is an attractive and efficient element in the armory of tools for metabolic engineering. The choice of constitutive or inducible promoters still widens the possibilities for a strategy of purposeful controlled gene expression. Knowledge of promoters and their regulation in a particular organism is necessary for rational modulation of gene expression and for the application of these powerful tools in metabolic engineering. Promoters of *C. glutamicum*, a widely used industrial producer of amino acids, play an important role in molecular breeding of new producing strains.

10.2 Screening and Analysis of *Corynebacterium glutamicum* Promoters

10.2.1 Promoter-Probe Vectors

The promoter-probe plasmid vectors carrying various promoterless reporter genes coding for products, which can be easily detected and quantified, are the basic tools for screening and analysis of promoters. The insertions of promoter-containing DNA fragments upstream of the reporter genes form transcriptional fusions, which drive expression of the reporter genes. Construction of efficient promoter-probe vectors also requires the insertion of transcriptional terminator(s) upstream of the cloning sites to prevent readthrough from the vector to the reporter gene.

Promoter-probe plasmid vectors applied for analysis of *C. glutamicum* promoters have recently been listed (Nešvera and Pátek 2008). The reporter genes used in these vectors encode the following enzymes active in *C. glutamicum*: *lacZ* coding for *Escherichia coli* β-galactosidase (Eikmanns et al. 1991; Nishimura et al. 2008), *cat* coding for chloramphenicol acetyltransferase from Tn*9* (Eikmanns et al. 1991; Vašicová et al. 1998), *aph* encoding aminoglycoside phosphotransferase from Tn*5* (Barák et al. 1990; Cadenas et al. 1991), *uidA* coding for *E. coli* β-glucuronidase (Bardonnet and Blanco 1991) and *amy* encoding α-amylase from *Streptomyces griseus* (Cadenas et al. 1996). In addition, the pigment melanine from *Streptomyces*

glaucescens (Adham et al. 2003) and the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* (Letek et al. 2006; Knoppová et al. 2007; Kusumoto et al. 2011) are further reporter proteins easily detectable in *C. glutamicum* cells. The use of the *gfp* reporter gene ensures testing promoter activity by estimating the fluorescence intensity in living cells and it is suitable for measuring promoter activity during growth of bacterial cultures.

Examples of three autonomously replicating promoter-probe vectors used for analysis of *C. glutamicum* promoters are shown in Fig. 10.1. The promoter-probe vector pET2 carrying the promoterless *cat* reporter gene (Vašicová et al. 1998) (Fig. 10.1a) has been frequently employed for analysis of the transcriptional regulation of various *C. glutamicum* genes and operons, e.g. genes of phosphate (Ishige et al. 2003) and acetate metabolism (Gerstmeir et al. 2004), the ATPase operon (Barriuso-Iglesias et al. 2006), the genes involved in ethanol utilization (Arndt and Eikmanns 2007; Auchter et al. 2009) and glycogen synthesis (Seibold et al. 2010, 2011), the genes coding for enzymes of tricarboxylic acid cycle (Bussmann et al. 2009; Emer et al. 2009; van Ooyen et al. 2011) and the genes involved in uptake of dicarboxylates (Youn et al. 2008) and inorganic phosphate (Panhorst et al. 2011).

The replacement vector pEMel-1, transferable to *C. glutamicum* cells by conjugation, carries the *melC1-melC2* genes from *Streptomyces glaucescens* involved in melanine production under the P-*kan* promoter from Tn5 on the EcoRI-NdeI fragment (Fig. 10.1c). This plasmid can be used for testing the activity of promoters present on the EcoRI-NdeI fragments replacing the original P-*kan* fragment. The use of melanine as reporter protein ensures the visual, inexpensive and non-selective screening of thousands of promoter-containing clones (Adham et al. 2003). The promoter-probe vector pEMel-1 was applied for analysis of transcriptional regulation of *C. glutamicum* genes involved in gluconate utilization (Letek et al. 2006) and of the *ftsI* gene necessary for cell division (Valbuena et al. 2006).

The promoter-probe vector pEPR1 carrying the *gfp*uv reporter gene (Knoppová et al. 2007) (Fig. 10.1d) has been recently used for analysis of the transcriptional regulation of genes coding for resuscitation promoting factor (Jungwirth et al. 2008), glutamate dehydrogenase (Hänssler et al. 2009), glutamate synthase (Hasselt et al. 2011), malic enzyme (Krause et al. 2012) and of the operon involved in zinc metabolism (Schröder et al. 2010).

The vector pRAG5 carrying the reference gene *rfp* (coding for red fluorescence protein from the coral *Discosoma* sp.) under the strong constitutive *C. glutamicum* promoter P-45, in addition to the promoterless *gfp*uv reporter gene, was constructed on the basis of pEPR1 (Knoppová et al. 2007). The vector pRAG5 is convenient for normalized measurements of promoter activities during the growth of bacterial batch cultures.

To ensure the transcriptional fusion of a tested promoter to a promoterless reporter gene in a single copy in the cell, the integrative promoter-probe vectors pRIM2 carrying the *cat* reporter gene (Fig. 10.1b) (Vašicová et al. 1998) and pCRA741 carrying the *lacZ* reporter gene (Inui et al. 2007) were constructed. These vectors contain replicons of *E. coli* plasmids non-replicating in *C. glutamicum* and sequences (a non-coding sequence downstream of the *ppc* gene coding

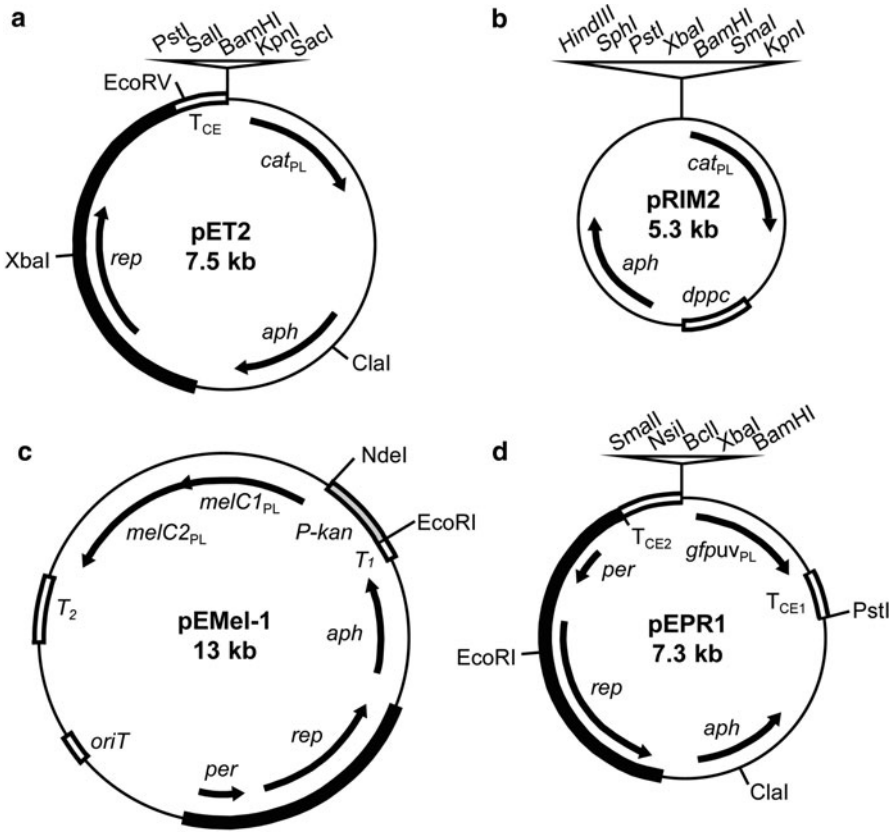


Fig. 10.1 Maps of the promoter-probe plasmid vectors. (**a**) *C. glutamicum/E. coli* shuttle promoter-probe vector pET2; (**b**) *E. coli* promoter-probe vector pRIM2 for insertion of a promoter and the reporter gene into the *C. glutamicum* chromosome; (**c**) *C. glutamicum/E. coli* shuttle promoter-probe vector pEMel-1; (**d**) *C. glutamicum/E. coli* shuttle promoter-probe vector pEPR1. The regions coming from plasmids of corynebacteria (pBL1 in the vector pET2, pCG1 in the vectors pEMel-1 and pEPR1) are shown as the *thick black lines*. The genes are shown inside the maps as *arrows*. *aph*, kanamycin resistance determinant; cat_{PL}, promoterless reporter gene coding for chloramphenicol acetyltransferase; $melC1_{PL}$-$melC2_{PL}$, promoterless reporter genes of melanine operon; $gfpuv_{PL}$, promoterless reporter gene coding for green fluorescent protein; *rep*, initiator of plasmid replication; *per*, positive effector of plasmid replication; T_1 and T_2, transcriptional terminators; T_{CE}, T_{CE1} and T_{CE2}, tandems of transcriptional terminators; *dppc*, non-coding sequence downstream of the *C. glutamicum ppc* gene; P-*kan*, promoter of kanamycin resistance determinant; *oriT*, origin of conjugative transfer

for phosphoenolpyruvate carboxylase in pRIM2 and a part of the non-essential strain-specific SSI7 island of *C. glutamicum* R in pCRA741) which allow their integration into the *C. glutamicum* chromosome via homologous recombination without any effect on the viability of recombinant strains.

10.2.2 Transcription Start Determination

The transcriptional start point (TSP), defines the 5′- end of the mRNA (transcript). Although the prediction of promoters from the DNA sequence is frequently used to define putative key motifs (−10 and −35 regions), experimental evidence is usually necessary for reliable localization of a functional promoter. Since the bacterial promoters are located closely upstream of the respective TSPs (designated as +1 position), determination of these TSPs indicates location of the key elements of the promoter. TSP may be determined by the primer extension technique, 5′-RACE (rapid amplification of cDNA-ends) or S1 nuclease protection mapping. In *C. glutamicum*, many TSPs were determined by radioactive (Pátek et al. 1996; Suda et al. 2008; Tanaka et al. 2008) or non-radioactive primer extension technique (Barreiro et al. 2004; Holátko et al. 2009). Several TSPs of a gene (and therefore several promoters) can be found by this analysis in a single experiment. The individual transcripts defined by the TSPs can be indirectly quantified by evaluation of the signal strength (Barreiro et al. 2004) produced by primer extension (Fig. 10.2). The 5′-RACE technique was also applied in determination of a number of *C. glutamicum* TSPs (Larisch et al. 2007; Nakunst et al. 2007; Schröder et al. 2010). The S1 nuclease protection mapping was used for TSP determination in a few cases (von der Osten et al. 1989; Han et al. 1990) but it seems that this is not currently enough reliable method for this purpose. The S1 mapping provided in comparison with primer extension less precise and less efficient procedure although the results were comparable (Pátek et al. 2003a). RNA sequencing methods have recently provided new approach to transcript analysis. TSPs could be deduced from the sequences of a large number of transcripts analyzed by high-throughput sequencing in various bacteria (Oliver et al. 2009). In addition, the results can be evaluated quantitatively and provide thus assessment of transcript levels in the cell. Such procedure has already been used also for the description of the *C. glutamicum* transcriptome (J. Kalinowski 2012, personal communication).

10.2.3 In Vitro Transcription

The analyses of promoter-RNAP interactions, functions of transcriptional regulators and characterization of promoters of different classes depend largely on the *in vitro* transcription assays, which use purified components of transcription machinery. *In vitro* transcription has been extensively used for investigation of properties of RNAP from *E. coli* (Ross and Gourse 2009) and *Bacillus subtilis* (Fujita 1999) and transcriptional regulation in these bacteria. *In vitro* transcription system also provided data on promoters and their regulation in human pathogen *Mycobacterium tuberculosis* (Rodrigue et al. 2006), related to corynebacteria. The *in vitro* transcription system for *C. glutamicum* has recently been developed (Holátko et al. 2012). The *C. glutamicum* strain carrying the gene encoding modified *rpoC* gene (coding

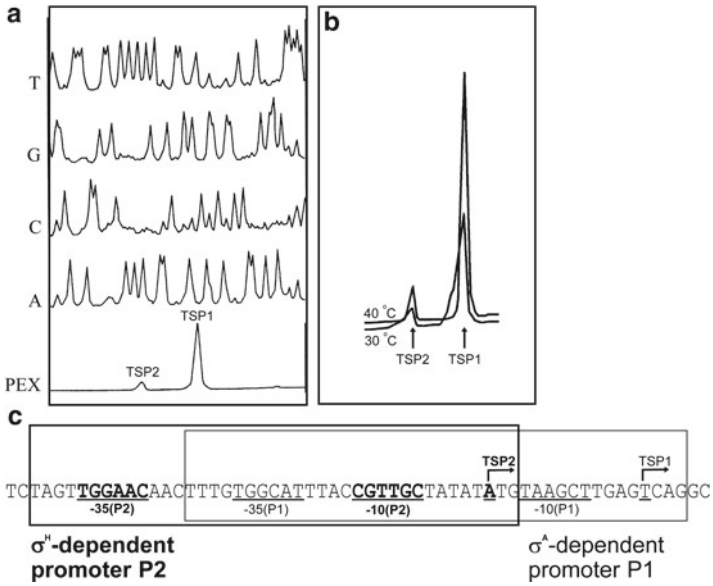


Fig. 10.2 Determination of transcriptional start points (TSPs) and its use for defining the promoter regions. TSPs of the *C. glutamicum dnaK* gene inducible by heat shock (Barreiro et al. 2004) are used as examples. (**a**) Determination of two *dnaK* TSPs by non-radioactive primer extension technique. The *bottom peaks* (PEX) represent cDNA synthesized in reverse transcription (primer extension) using RNA from *C. glutamicum*. The peaks generated by the automatic sequencer represent the products of sequencing reactions (T, G, C, A) performed with the same fluorescein-labeled primer as that used for primer extension. The nucleotides in TSPs were determined by a comparison of the positions of the primer extension products and the sequencing signal; (**b**) Quantitative comparison of the reverse *dnaK* transcripts synthesized on the RNA templates isolated from the cells cultivated under various conditions (30°C or 40°C); (**c**) Defining the promoter regions of two *dnaK* promoters. The key sequences are underlined, the sequences representing the σ^H-dependent promoter (P2) are shown in *bold letters* and the sequences of the σ^A-dependent promoter (P1) are in *standard letters*. The *boxes* represent the approximate regions of the P1 promoter (*thin line*) and the P2 promoter (*thick line*)

for β′ subunit of RNAP) within the chromosome was constructed using the gene-replacement technique (Schäfer et al. 1994) and the *C. glutamicum* core-RNAP was isolated in the native form (as the $\alpha_2\beta\beta'\omega$-complex) via affinity chromatography. The *sigA* (coding for the principal sigma factor in *C. glutamicum*), *sigB*, *sigE* and *sigH* (coding for alternative *C. glutamicum* sigma factors involved in stress responses) genes were cloned in the expression plasmid vector in *E. coli* and the sigma factor proteins were isolated by affinity purification after induced overexpression. Holo-RNAP (core-RNAP + specific sigma factor) was reconstituted and used for *in vitro* transcription from the plasmid templates carrying σ^A-, σ^B-, σ^H-, and σ^E-dependent promoters, respectively. The specific transcripts were formed in all cases, confirming the functionality of the *in vitro* transcription system and its usefulness in determination of sigma specificity in recognizing particular promoters. *In vitro* transcription may be used to identify promoters of all classes (i.e. recognized by any of

7 sigma factors of *C. glutamicum*), to prove that some promoters are recognized by two or more sigma factors and to define the roles of various transcriptional factors in regulation of activity of particular promoters. Run-off transcription of the *C. glutamicum* genome followed by microarray analysis of the resulting *in vitro*-synthesized transcriptome provided the new technique called ROMA (Cao et al. 2002). This technique may facilitate identification of the whole regulons controlled by specific sigma factors (sigmulons).

10.3 Classes of *C. glutamicum* Promoters

10.3.1 Promoters of Housekeeping Genes

Bacterial promoters of housekeeping genes were first described in *E. coli* and their DNA sequences form currently a large group of well-described regulatory signals. On the basis of the alignment of the promoter sequences, statistical evaluation of the conserved DNA regions and mutations affecting the promoter activity, two conserved hexamers located approximately 35 and 10 nucleotides (nt) upstream of the TSP (+1 position) were defined in *E. coli* promoters. The consensus sequence includes −35 regions TTGACA and −10 region TATAAT with a spacer of 17 ± 1 nt. Similar structures of promoters were found in most eubacterial species for the housekeeping genes, which are considered to be under the control of principal sigma factor in each organism. Statistical consensus of the key motifs of 159 *C. glutamicum* housekeeping promoters with experimentally determined TSPs was defined (Pátek and Nešvera 2011). It consists of −35 region TTGNCA and the extended −10 region GN**TA**NAN**T**NG (nucleotides in bold are present in more than 80% of the sequences, the other nucleotides appear in more than 35% of sequences; core hexamers are underlined). In general, the nucleotides in most positions are weakly conserved. In many *C. glutamicum* housekeeping promoters, the −35 sequences could not be defined and rather arbitrary hexamers are presented. The positions T1, A2 and T6 within the −10 core hexamer are reasonably conserved (Fig. 10.3).

This statistical consensus sequence may not represent the strongest promoter, since the promoter sequences evolved to fit the physiological functions rather than achieve the highest transcriptional activity. Moreover, function of many promoters is affected by one or more transcriptional protein regulators (activators or repressors), which bind within the promoters or in their vicinity in dependence on environmental conditions and physiological status of the cells and modulate their activity (Pátek et al. 2003b). Mutational studies of the nucleotides within and outside of the conserved regions may reveal the nucleotide positions which generally contribute to the activity of the promoters although these positions are not conserved. In *E. coli* a TG dimer located one nt upstream of the −10 hexamer (Burr et al. 2000) and an AT-rich UP element residing upstream of the −35 hexamer (Ross et al. 2001), which are not conserved, are present in some promoters and considerably enhance the promoter strength. In a few mutational studies of C. *glutamicum* promoter sequences

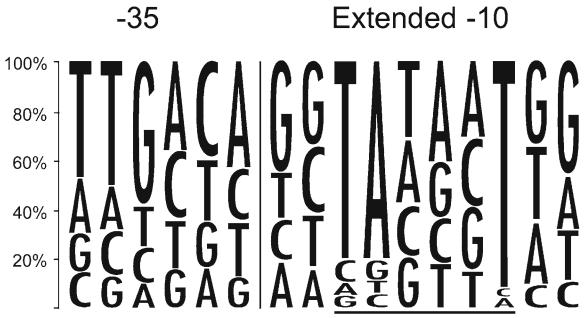


Fig. 10.3 Frequency of nucleotide occurrence within the two key motifs of the *C. glutamicum* housekeeping promoters. The distribution is based on the alignment of 159 promoter sequences and represents the levels of nucleotide conservation in particular positions in −35 and −10 regions. The key −10 hexamer is *underlined*

of strong or weak promoters were constructed (Vašicová et al. 1999; Asakura et al. 2007; Hänssler et al. 2009; Holátko et al. 2009). The TG dimer positioned 1 nt upstream of the −10 hexamer was also found to substantially enhance activity of *C. glutamicum* promoters. Mutations within the −35 hexamer resulting in sequences TTGACA or TTGCCA also provided stronger promoters (Asakura et al. 2007), although the −35 sequence is only loosely conserved in *C. glutamicum* promoters. As a conclusion of these studies analyzing up- and down-mutations of promoters, the statistical consensus may be modified into the functional consensus −35 region TTG$^{A}/_{C}$CA and extended −10 region TGNTATAATNG (key hexamers are underlined). It is not yet clear if the AT-rich regions found in some *C. glutamicum* promoters may function as UP elements.

In *C. glutamicum*, the principal sigma factor σ^A is believed to be responsible for the transcription of most of the housekeeping genes. The expression of the *sigA* gene is high during the exponential growth of *C. glutamicum* cells and is reduced when the cultures reach transition phase between the exponential and stationary growth phase (Larisch et al. 2007). It is therefore presumed that the promoters which are active during the exponential phase are recognized by RNAP+σ^A. However, it cannot be excluded that some of the promoters of housekeeping genes can also or exclusively be recognized by other sigma factors. The genes coding for several enzymes of glucose catabolism (*pfkA, fba, tpi, gapA, pgk, eno, ppc, fum* and *pqo*) were found to be under the control of σ^B even during the exponential growth phase (Ehira et al. 2008). These results indicate that σ^B may function as another sigma factor regulating transcription of housekeeping genes. In *E. coli*, many genes expressed during exponential growth were also found to be under control of alternative sigma factor σ^{38} (designated also σ^S) (Dong et al. 2008), which had been believed to function only during the transition to stationary growth phase. The −35 and −10 sequences of the *E. coli* σ^{70} -dependent and σ^{38} -dependent genes are very similar (Weber et al. 2005). The sequences of the key motifs of the *C. glutamicum* housekeeping promoters, which are recognized by RNAP+σ^A and RNAP+σ^B, respectively,

could not be distinguished, either. In agreement with this finding, a transcription from the promoter with the sequence TATAAT in −10 region was driven by both *C. glutamicum* σ^A and σ^B under the same conditions in the *in vitro* transcription study (Holátko et al. 2012). The activity of a number of constitutive and inducible housekeeping *C. glutamicum* promoters was analyzed in detail. These promoters represent a range of transcriptional signals with various strength and regulatory patterns, which can be used for efficient gene expression in the industrial *C. glutamicum* strains (see Sect. 10.4).

10.3.2 Promoters Recognized by SigB

SigB-dependent genes were revealed by the genome-wide transcription profiling of *C. glutamicum* WT strain in comparison to the *sigB* deletion strain (Larisch et al. 2007; Ehira et al. 2008). The nonessential principal-like (Group 2) sigma factor σ^B is mainly involved in transcription during the transition to the stationary phase and in response to various environmental stresses in *C. glutamicum*. Expression of *sigB* is lower than expression of *sigA* during the exponential growth (Oguiza et al. 1997) and the level of *sigB* expression increases when the cultures reach the transition to stationary phase. Increased expression of many *C. glutamicum* genes in the transition phase was shown to depend on the presence of σ^B (Larisch et al. 2007). Expression of the *sigB* gene also increases in response to ethanol, acid, salt, heat and cold stresses (Halgasova et al. 2002; Barreiro et al. 2009). In agreement with the increased level of σ^B after such stresses, expression of the genes involved in stress defense mechanisms was found to depend on σ^B. Ehira et al. (2008) showed that transcription of the genes involved in glucose consumption under the conditions of oxygen deprivation as well as during the aerobic exponential growth are also regulated by σ^B. Determination of the respective σ^B-dependent TSPs and localization of σ^B-specific promoters provided the sequences of the key promoter motifs. It was found that the σ^B-specific promoters contain sequences in the −10 region which are very similar to the −10 motifs in supposed σ^A-dependent promoters. In the set of 13 σ^B-dependent promoters (Larisch et al. 2007; Ehira et al. 2008) with the experimentally determined TSPs, the consensus of the sequence of the −10 regions can be defined as TA^{A}/$_{C}$AATTGA (−10 core hexamer underlined). The sequence in the −35 region is very weakly conserved and can be defined as GNGNCN (Ehira et al. 2008). These authors also hypothesized that the different −35 regions within the σ^B- and σ^A- dependent promoters affect the specific recognition of these sequences by the two sigma factors. In *E. coli*, the consensus sequences of σ^{70}-dependent promoters (involved in transcription of housekeeping genes) and σ^{38}-dependent promoters (involved in transcription of stress-response genes) are also very similar (Weber et al. 2005). Many *E. coli* promoters of the genes involved in glucose utilization are recognized by both RNAP containing σ^{70} or σ^{38} (Olvera et al. 2009). In *C. glutamicum*, σ^B inactivation partially reduced the transcription from the σ^B-dependent promoters of the genes involved in glucose metabolism but did not eliminate their activity.

This suggests that σ^A may also substitute the function of σ^B in *C. glutamicum* (Ehira et al. 2008). The particular physiological conditions and the potential regulatory elements (protein transcription factors or small effector molecules) which shift the preferential recognition of such dual promoters from σ^A to σ^B in *C. glutamicum* still have to be revealed. The promoters, which would be confirmed as σ^B-dependent, might be used in metabolic engineering of *C. glutamicum* strains to boost expression of genes under the stress conditions or in the stationary growth phase.

10.3.3 Promoters Recognized by SigH

A considerable number (about 45) of genes under the control of σ^H (a sigma factor of Group 4) was found in *C. glutamicum* using the gene expression profiling with *sigH*-deleted or *sigH*-overexpressed strains (Barreiro et al. 2009; Ehira et al. 2009). Another approach to detect genes of σ^H regulon preferred the genome profiling of the strain in which the gene of the specific anti-SigH factor (*rshA*) was deleted (Kalinowski, personal communication). Approximately 30 σ^H-specific promoters have been localized by determination of the respective TSPs (Halgasova et al. 2001; Engels et al. 2004; Zemanová et al. 2008; Ehira et al. 2009; our unpublished results). These promoters control particularly the genes involved in heat-shock response.

The *sigH* gene itself is transcribed in a relatively constant level during exponential growth and its transcription increases after oxidative stress and in stationary phase according to the results of the reporter gene fusions (Kim et al. 2005). However, no significant increase of the *sigH* transcript level was observed after heat shock (Barreiro et al. 2009), or in the transition phase (Larisch et al. 2007) by the genome profiling. Most of the results suggest that *sigH* is expressed from several housekeeping promoters (our unpublished results) and σ^H is deactivated during the exponential growth by interaction with the anti-SigH factor (Kalinowski, personal communication).

The key sequence motifs of the σ^H-specific promoters are highly conserved. The core of the consensus sequence is formed by the sequences **GGAA** – $N_{18–21}$ – **GTT**, which were also described in the promoters of stress genes in mycobacteria (Rodrigue et al. 2006) and streptomycetes (Paget et al. 1998). These motifs can be extended to the consensus sequence (–35) $^{G}/_{T}$**GGAA**TA and (–10) $^{C}/_{T}$**GTT**GAA (nucleotides in bold are present in more than 80% of the sequences, the others in more than 40% of sequences). Deletion of the *sigH* gene resulted in a decrease of the activity of some σ^H-dependent promoters but not in its complete elimination. This indicates that still other sigma factors may be involved in the transcription from these promoters in *C. glutamicum.* In agreement with this hypothesis, *in vitro* transcription proceeded with both RNAP + σ^H and RNAP + σ^E when some promoters were used as templates (our unpublished results). The σ^H-dependent promoters may be applied in biotechnology for overexpression of the chosen genes under the conditions of their induction by heat pulse.

10.3.4 Promoters Recognized by SigM, SigE, SigC and SigD

A few promoters were proposed to depend on the other sigma factors. Four genes *trxB, trxB1, trxC* and *sufR* involved in the oxidative stress response were defined as σ^M-dependent (Nakunst et al. 2007) and σ^H-dependent (Ehira et al. 2009). Since the *sigM* promoter is most probably recognized by RNAP + σ^H, the dependence of these genes on σ^H might be indirect. The *sigM* promoter may be also autoregulated according to our results. Therefore RNAP + σ^M and RNAP + σ^H might recognize the same or similar promoters. This is another hint of possible overlap of sigma-dependent regulons in *C. glutamicum*. There are some suggestions of genes regulated by σ^E-dependent promoters (Park et al. 2008), but these promoters have not yet been localized. The sigma factor σ^E is expressed after heat shock, surface stress and in the stationary growth phase and it probably mainly regulates genes involved in the surface stress response (Park et al. 2008). The consensus sequence of mycobacterial σ^E-dependent promoters is GGAAC – $N_{18\text{-}19}$ – GTT which is similar to the consensus of σ^H-dependent promoters in *C. glutamicum*. The P-*sigB* promoter is recognized by both σ^H and σ^E in *M. tuberculosis*. These data suggest that an overlap of functions of σ^H and σ^E may also exist in *C. glutamicum*. Our results of *in vitro* transcription from the promoters P-*dnaK* and P-*sigB* support such hypothesis. The promoters dependent on σ^C or σ^D have not as yet been described and the function of these sigma factors is not known in *C. glutamicum*.

10.4 Application of *C. glutamicum* Promoters in Metabolic Engineering

The knowledge of structure and function of promoters serves as a basis for their applications in the purposeful modulation of gene expression, an important tool used in metabolic engineering. Until now, most applications of controlled expression of *C. glutamicum* genes, cloned in various expression plasmid vectors, used well-described *E. coli* inducible promoters P-*lac*, P-*tac* and P-*trc* induced by isopropyl-β-D-thiogalactopyranoside (IPTG). The heat-induced P_RP_L promoters of phage λ (Tsuchiya and Morinaga 1988) and the *E. coli* P-*araBAD* promoter (Ben-Samoun et al. 1999) are further heterologous tightly regulated promoters used for controlled gene expression in *C. glutamicum*. The *C. glutamicum* expression plasmid vectors have been listed and described in detail in our previous review article (Nešvera and Pátek 2008).

The application of native inducible *C. glutamicum* promoters is still rather limited. During studies on regulation of primary metabolism in *C. glutamicum*, several promoters induced by various carbon sources were discovered. The promoters of the *aceA* and *aceB* genes coding for enzymes of the glyoxalate cycle were found to be induced by acetate (Gerstmeir et al. 2004). Gluconate was found to strongly induce the promoters of the *gntP* and *gntK* genes, which were used for the regulated expression of the *divIVA* genes involved in cell division in *C. glutamicum*

(Letek et al. 2006). The other gluconate-inducible promoter P-*git1* was applied, in addition to the maltose-inducible promoter P-*malE1*, for controlled expression of the *xynA* gene (coding for xylanase) from *Clostridium cellulovorans* in *C. glutamicum* (Okibe et al. 2010). Highly increased expression of the *prpDBC2* genes coding for the enzymes of 2-methylcitrate cycle was observed in the presence of propionate (Hüser et al. 2003). The strong induction of the P-*prpD2* by propionate requires the PrpR activator and the respective operator sequence within the P-*prpD2* promoter region (Plassmeier et al. 2012).

The constitutive promoter of the *cspB* gene, coding for the main *C. glutamicum* secreted surface-layer protein PS2, was found to be suitable for the expression of genes coding for secreted proteins in *C. glutamicum*. The *C. glutamicum-E. coli* expression vector pCC containing P-*cspB* (Tateno et al. 2007) was used as a basis for the construction of specialized vectors ensuring secretion or cell surface display of the products of the cloned genes (Tateno et al. 2009). Strong constitutive *C. glutamicum* promoters were also used for overexpression of selected genes in *C. glutamicum.* Cloning the *iolT1* and *iolT2* genes under the constitutive promoter of the *gapA* gene, coding for glyceraldehyde 3-phosphate dehydrogenase, in the multi-copy vector pCS299P caused their overexpression and proved their role in glucose uptake (Ikeda et al. 2011). Strong constitutive promoters were also inserted upstream of the selected genes in *C. glutamicum* chromosome to obtain stable and efficient plasmidless amino acid-producing strains. The strong promoter of the *C. glutamicum sod* gene coding for superoxide dismutase was inserted upstream of the genes *dld* (D-lactate dehydrogenase), *pyc* (pyruvate carboxylase) and *malE* (malic enzyme) within the *C. glutamicum* chromosome, which resulted in a *C. glutamicum* strain producing L-lysine during growth on lactate (Neuner and Heinzle 2011). The same promoter was also used for overexpression of the chromosomal genes *pyc, dapB* (dihydrodipicolinate reductase), *lysC* (aspartate kinase) and *tkt* (transketolase) in the L-lysine producing strain constructed with the use of data from metabolic flux distribution analysis (Becker et al. 2011). In addition, the promoter of the *fbt* gene (fructose 1,6-biphosphatase) was replaced by the strong *eftu* promoter within the same strain, which resulted in 15-fold higher specific enzyme activity and further improvement of L-lysine production.

The knowledge of the statistical and functional consensus sequences of *C. glutamicum* housekeeping promoters and of the significance of individual positions within their −10 and −35 hexamers has been a basis for purposeful modulation of promoter activities by nucleotide alterations in the specific positions of the consensus sequence. The set of *C. glutamicum* promoters mutated in the specific nucleotides within the −35 hexamers and the extended −10 regions is listed in the Table 10.1. Site-directed mutagenesis of the *dapA* promoter (dihydrodipicolinate synthase gene) provided a promoter set containing both strong and very weak promoters (Vašicová et al. 1999). The introduction of a *dapA* gene copy with the strong mutant P-*dapA*MC20 or P-*dapA*MA16 promoter (Table 10.1) into the chromosome enhanced L-lysine synthesis in the *C. glutamicum* production strain (Pfefferle et al. 2003). On the other hand, the very weak mutant P-*dapA*B6 promoter (Table 10.1) was used in experiments aimed to fine-tuning of *argF* (ornithine transcarbamoylase) expression

Table 10.1 Mutations in C. glutamicum promoters and their effects on promoter activity

Promoter	Gene product	–35	Extended – 10	Up/Down effect of mutation	References
P-*dapA* (WT)	Dihydrodipicolinate synthase	TAACCC	AGGTAACCTTG	–	Vašicová et al. (1999)
P-*dapA*MA16		TAACCC	AGGTA**TAA**TTG	Up	Vašicová et al. (1999), Pfefferle et al. (2003)
P-*dapA*MC20		TAACCC	**T**GGTAACCTTG	Up	Vašicová et al. (1999), Pfefferle et al. (2003)
P-*dapA*B6		TAACCC	AGG**C**AACC**A**TG	Down	Vašicová et al. (1999), Schneider et al. (2012)
P-*gdh* (WT)	Glutamate dehydrogenase	TGGTCA	TGCCATAATTG	–	Börmann et al. (1992)
P-*gdh*2		TGGTCA	TGC**T**ATAATTG	Up	Asakura et al. (2007), Hänssler et al. (2009)
P-*gdh*3		TT**G**A**C**A	TGC**T**ATAATTG	Up	Asakura et al. (2007)
P-*gdh*4		TT**G**TCA	TGC**T**ATAATTG	Up	Asakura et al. (2007)
P-*gdh*7		TT**GC**CA	TGC**T**ATAATTG	Up	Asakura et al. (2007)
P-*gdh*527_2		TGGTCA	TGCCATAA**A**TG	Down	Hänssler et al. (2009)
P-*gdh*527_3		TGGTCA	**C**CCCATAATTG	Down	Hänssler et al. (2009)
P-*gdh*527_4		TGGTCA	**C**CCCATAA**A**TG	Down	Hänssler et al. (2009)
P-*ilvD* (WT)	Dihydroxyacid dehydratase	GTGATA	AGCACTAGAGTGT	–	Holátko et al. (2009)
P-*ilvD*M7		GTGATA	**TGTG**CTA**T**AGTGT	Up	Holátko et al. (2009)
P-*ilvD*M14		GTGATA	AGCACT**GTG**GTA**T**	Up	Holátko et al. (2009)
P-*ilvE* (WT)	Transaminase	GTGTAT	AGGTGTACCTTAA	–	Holátko et al. (2009)
P-*ilvE*M6		GTGTAT	**TGTG**GTACC**A**TAA	Up	Holátko et al. (2009)
P-*ilvA* (WT)	Threonine deaminase	TAGGTG	GATTACACTAG	–	Pátek et al. (1996)
P-*ilvA*M1CG		TAGGTG	GAT**C**ACA**G**TAG	Down	Holátko et al. (2009)
P-*ilvA*M1CTG		TAGGTG	GAT**C**AC**TG**TAG	Down	Holátko et al. (2009)
P-*leuA* (WT)	Isopropylmalate synthase	TACCCA	TTGTATGCTTC	–	Pátek et al. (2003b)
P-*leuA*M3A		TACCCA	TTGTATGC**A**TC	Down	Holátko et al. (2009)
P-*leuA*M2TCG		TACCCA	TT**TC**A**G**GCTTC	Down	Holátko et al. (2009)
P-*leuA*M2C		TACCCA	TTG**C**ATGCTTC	Down	Holátko et al. (2009)

through modifications of the translation start codon and/or ribosome binding site. Among the obtained *C. glutamicum* strains, carrying the stably maintained plasmids with individual modifications and showing thus different levels of ornithine transcarbamoylase, the strain producing so far the highest amount of putrescine was found (Schneider et al. 2012).

Sequence modifications aimed at creating a stronger promoter were also constructed in the P-*gdh* promoter (glutamate dehydrogenase gene) (Table 10.1) (Asakura et al. 2007; Hänssler et al. 2009). Specific mutations within the −10 region enhanced glutamate dehydrogenase activity as much as 4.5-fold and the mutations in both −10 region and −35 hexamer resulted in further increase of activity of this enzyme (7-fold). These promoter modifications lead to an increase in L-glutamate production by the *odhA*-deficient *C. glutamicum* strain (lacking 2-oxo-glutarate dehydrogenase activity) (Asakura et al. 2007). Mutagenesis of native promoters within the chromosome, which affected the metabolism of valine, isoleucine and leucine, was used to improve L-valine production by *C. glutamicum* strains (Holátko et al. 2009). Up-mutations within the promoters of the *ilvD* and *ilvE* genes (Table 10.1) increased the activity of the respective biosynthesis enzymes. On the other hand, down-mutations were constructed in the promoters of the *ilvA* and *leuA* genes (Table 10.1), coding for enzymes which channel the flux of metabolites to the unwanted side-products (isoleucine and leucine). Combinations of particular promoter mutations resulted in a plasmidless strain exhibiting increased L-valine production (Holátko et al. 2009).

Although a considerable progress in studies of *C. glutamicum* promoters and in their use in metabolic engineering was achieved, still more reliable, strong, constitutive or inducible promoters are needed. Efficient corynebacterial promoters which can easily be induced by a cheap substrate without disturbing the biosynthesis of required products (e.g. amino acids) and biotechnological requirements are of particular interest. More detailed knowledge of stress-responsive promoters is necessary before including them into the toolbox for genetic manipulations in *C. glutamicum*. To construct an efficient and stable production strain, modifications of the pathways responsible for forming the precursors, drainage of intermediates, energy metabolism, product degradation, and excretion of the product are necessary. Detailed studies of the metabolism and regulation of expression of the involved genes should result in definitions of the targets for engineering in the studied strain. Such aims could be achieved particularly using genome-wide transcriptional profiling, metabolome analysis and knowledge of the metabolic flux distribution. In addition to use of various promoters, different approaches, like optimization of codon usage, engineering translational signals, introduction of genes responsible for uptake and utilization of cheap and useful carbon sources, elimination of by-products and optimization of growth medium can lead to the construction of powerful strains which can be used in biotechnological processes economic enough to surpass alternative procedures (e.g. chemical synthesis).

Acknowledgements Work in the authors' laboratory was supported by grants 204/09/J015 and P302/12/P633 from the Scientific Council of the Czech Republic and by project RVO 61388971.

References

Adham SA, Rodriguez S, Ramos A, Santamaria RI, Gil JA (2003) Improved vectors for transcriptional/translational signal screening in corynebacteria using the *melC* operon from *Streptomyces glaucescens* as reporter. Arch Microbiol 180:53–59

Arndt A, Eikmanns BJ (2007) The alcohol dehydrogenase gene *adhA* in *Corynebacterium glutamicum* is subject to carbon catabolite repression. J Bacteriol 189:7408–7416

Asakura Y, Kimura E, Usuda Y, Kawahara Y, Matsui K, Osumi T, Nakamatsu T (2007) Altered metabolic flux due to deletion of *odhA* causes L-glutamate overproduction in *Corynebacterium glutamicum*. Appl Environ Microbiol 73:1308–1319

Auchter M, Arndt A, Eikmanns BJ (2009) Dual transcriptional control of the acetaldehyde dehydrogenase gene *ald* of *Corynebacterium glutamicum* by RamA and RamB. J Biotechnol 140:84–91

Barák I, Koptides M, Jucovič M, Šišova M, Timko J (1990) Construction of a promoter-probe shuttle vector for *Escherichia coli* and brevibacteria. Gene 95:133–135

Bardonnet N, Blanco C (1991) Improved vectors for transcriptional signal screening in corynebacteria. FEMS Microbiol Lett 68:97–102

Barreiro C, Gonzalez-Lavado E, Pátek M, Martin JF (2004) Transcriptional analysis of the *groES-groEL1*, *groEL2*, and *dnaK* genes in *Corynebacterium glutamicum*: characterization of heat shock-induced promoters. J Bacteriol 186:4813–4817

Barreiro C, Nakunst D, Hüser AT, de Paz HD, Kalinowski J, Martin JF (2009) Microarray studies reveal a 'differential response' to moderate or severe heat shock of the HrcA- and HspR-dependent systems *in Corynebacterium glutamicum*. Microbiology 155:359–372

Barriuso-Iglesias M, Barreiro C, Flechoso F, Martin JF (2006) Transcriptional analysis of the F0F1 ATPase operon of *Corynebacterium glutamicum* ATCC 13032 reveals strong induction by alkaline pH. Microbiology 152:11–21

Becker J, Zelder O, Hafner S, Schröder H, Wittmann C (2011) From zero to hero-design-based systems metabolic engineering of *Corynebacterium glutamicum* for l-lysine production. Metab Eng 13:159–168

Ben-Samoun K, Leblon G, Reyes O (1999) Positively regulated expression of the *Escherichia coli araBAD* promoter in *Corynebacterium glutamicum*. FEMS Microbiol Lett 174:125–130

Börmann ER, Eikmanns BJ, Sahm H (1992) Molecular analysis of the *Corynebacterium glutamicum gdh* gene encoding glutamate dehydrogenase. Mol Microbiol 6:317–326

Burr T, Mitchell J, Kolb A, Minchin S, Busby S (2000) DNA sequence elements located immediately upstream of the −10 hexamer in *Escherichia coli* promoters: a systematic study. Nucleic Acids Res 28:1864–1870

Bussmann M, Emer D, Hasenbein S, Degraf S, Eikmanns BJ, Bott M (2009) Transcriptional control of the succinate dehydrogenase operon *sdhCAB* of *Corynebacterium glutamicum* by the cAMP-dependent regulator GlxR and the LuxR-type regulator RamA. J Biotechnol 143:173–182

Cadenas RF, Martin JF, Gil JA (1991) Construction and characterization of promoter-probe vectors for Corynebacteria using the kanamycin-resistance reporter gene. Gene 98:117–121

Cadenas RF, Fernandez-Gonzalez C, Martin JF, Gil JA (1996) Construction of new cloning vectors for *Brevibacterium lactofermentum*. FEMS Microbiol Lett 137:63–68

Cao M, Kobel PA, Morshedi MM, Wu MF, Paddon C, Helmann JD (2002) Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches. J Mol Biol 316:443–457

Dong T, Kirchhof MG, Schellhorn HE (2008) RpoS regulation of gene expression during exponential growth of *Escherichia coli* K12. Mol Genet Genomics 279:267–277

Ehira S, Shirai T, Teramoto H, Inui M, Yukawa H (2008) Group 2 sigma factor SigB of *Corynebacterium glutamicum* positively regulates glucose metabolism under conditions of oxygen deprivation. Appl Environ Microbiol 74:5146–5152

Ehira S, Teramoto H, Inui M, Yukawa H (2009) Regulation of *Corynebacterium glutamicum* heat shock response by the extracytoplasmic-function sigma factor SigH and transcriptional regulators HspR and HrcA. J Bacteriol 191:2964–2972

Eikmanns BJ, Kleinertz E, Liebl W, Sahm H (1991) A family of *Corynebacterium glutamicum/ Escherichia coli* shuttle vectors for cloning, controlled gene expression, and promoter probing. Gene 102:93–98

Emer D, Krug A, Eikmanns BJ, Bott M (2009) Complex expression control of the *Corynebacterium glutamicum* aconitase gene: identification of RamA as a third transcriptional regulator besides AcnR and RipA. J Biotechnol 140:92–98

Engels S, Schweitzer JE, Ludwig C, Bott M, Schaffer S (2004) *clpC* and *clpP1P2* gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor sigmaH. Mol Microbiol 52:285–302

Fujita M (1999) Identification of new sigma K-dependent promoters using an *in vitro* transcription system derived from *Bacillus subtilis*. Gene 237:45–52

Gerstmeir R, Cramer A, Dangel P, Schaffer S, Eikmanns BJ (2004) RamB, a novel transcriptional regulator of genes involved in acetate metabolism of *Corynebacterium glutamicum*. J Bacteriol 186:2798–2809

Halgasova N, Bukovska G, Timko J, Kormanec J (2001) Cloning and transcriptional characterization of two sigma factor genes, *sigA* and *sigB*, from *Brevibacterium flavum*. Curr Microbiol 43:249–254

Halgasova N, Bukovska G, Ugorcakova J, Timko J, Kormanec J (2002) The *Brevibacterium flavum* sigma factor SigB has a role in the environmental stress response. FEMS Microbiol Lett 216:77–84

Han KS, Archer JA, Sinskey AJ (1990) The molecular structure of the *Corynebacterium glutamicum* threonine synthase gene. Mol Microbiol 4:1693–1702

Hänssler E, Müller T, Palumbo K, Pátek M, Brocker M, Krämer R, Burkovski A (2009) A game with many players: control of *gdh* transcription in *Corynebacterium glutamicum*. J Biotechnol 142:114–122

Hasselt K, Rankl S, Worsch S, Burkovski A (2011) Adaptation of AmtR-controlled gene expression by modulation of AmtR binding activity in *Corynebacterium glutamicum*. J Biotechnol 154:156–162

Holátko J, Elišáková V, Prouza M, Sobotka M, Nešvera J, Pátek M (2009) Metabolic engineering of the L-valine biosynthesis pathway in *Corynebacterium glutamicum* using promoter activity modulation. J Biotechnol 139:203–210

Holátko J, Šilar R, Rabatinová A, Šanderová H, Halada P, Nešvera J, Krásný L, Pátek M (2012) Construction of in vitro transcription system for *Corynebacterium glutamicum* and its use in the recognition of promoters of different classes. Appl Microbiol Biotechnol. doi:10.1007/s00253-012-4336-1

Hüser AT, Becker A, Brune I, Dondrup M, Kalinowski J, Plassmeier J, Pühler A, Wiegrabe I, Tauch A (2003) Development of a *Corynebacterium glutamicum* DNA microarray and validation by genome-wide expression profiling during growth with propionate as carbon source. J Biotechnol 106:269–286

Ikeda M, Mizuno Y, Awane S, Hayashi M, Mitsuhashi S, Takeno S (2011) Identification and application of a different glucose uptake system that functions as an alternative to the phosphotransferase system in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90:1443–1451

Inui M, Suda M, Okino S, Nonaka H, Puskas LG, Vertes AA, Yukawa H (2007) Transcriptional profiling of *Corynebacterium glutamicum* metabolism during organic acid production under oxygen deprivation conditions. Microbiology 153:2491–2504

Ishige T, Krause M, Bott M, Wendisch VF, Sahm H (2003) The phosphate starvation stimulon of *Corynebacterium glutamicum* determined by DNA microarray analyses. J Bacteriol 185: 4519–4529

Jungwirth B, Emer D, Brune I, Hansmeier N, Pühler A, Eikmanns BJ, Tauch A (2008) Triple transcriptional control of the resuscitation promoting factor 2 (rpf2) gene of *Corynebacterium glutamicum* by the regulators of acetate metabolism RamA and RamB and the cAMP-dependent regulator GlxR. FEMS Microbiol Lett 281:190–197

Kim TH, Kim HJ, Park JS, Kim Y, Kim P, Lee HS (2005) Functional analysis of *sigH* expression in *Corynebacterium glutamicum*. Biochem Biophys Res Commun 331:1542–1547

Knoppová M, Phensaijai M, Veselý M, Zemanová M, Nešvera J, Pátek M (2007) Plasmid vectors for testing in vivo promoter activities in *Corynebacterium glutamicum* and *Rhodococcus erythropolis*. Curr Microbiol 55:234–239

Krause JP, Polen T, Youn JW, Emer D, Eikmanns BJ, Wendisch VF (2012) Regulation of the malic enzyme gene *malE* by the transcriptional regulator MalR in *Corynebacterium glutamicum*. J Biotechnol 159:204–215

Kusumoto T, Aoyagi M, Iwai H, Kabashima Y, Sakamoto J (2011) Monitoring enzyme expression of a branched respiratory chain of *Corynebacterium glutamicum* using an EGFP reporter gene. J Bioenerg Biomembr 43:257–266

Larisch C, Nakunst D, Hüser AT, Tauch A, Kalinowski J (2007) The alternative sigma factor SigB of *Corynebacterium glutamicum* modulates global gene expression during transition from exponential growth to stationary phase. BMC Genomics 8:4

Letek M, Valbuena N, Ramos A, Ordonez E, Gil JA, Mateos LM (2006) Characterization and use of catabolite-repressed promoters from gluconate genes in *Corynebacterium glutamicum*. J Bacteriol 188:409–423

Nakunst D, Larisch C, Hüser AT, Tauch A, Pühler A, Kalinowski J (2007) The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. J Bacteriol 189:4696–4707

Nešvera J, Pátek M (2008) Plasmids and promoters in corynebacteria and their applications. In: Burkovski A (ed) Corynebacteria. Genomics and molecular biology. Caister Academic Press, Norfolk, pp 113–154

Neuner A, Heinzle E (2011) Mixed glucose and lactate uptake by *Corynebacterium glutamicum* through metabolic engineering. Biotechnol J 6:318–329

Nishimura T, Teramoto H, Vertes AA, Inui M, Yukawa H (2008) ArnR, a novel transcriptional regulator, represses expression of the *narKGHJI* operon in *Corynebacterium glutamicum*. J Bacteriol 190:3264–3273

Oguiza JA, Marcos AT, Martin JF (1997) Transcriptional analysis of the *sigA* and *sigB* genes of *Brevibacterium lactofermentum*. FEMS Microbiol Lett 153:111–117

Okibe N, Suzuki N, Inui M, Yukawa H (2010) Isolation, evaluation and use of two strong, carbon source-inducible promoters from *Corynebacterium glutamicum*. Lett Appl Microbiol 50:173–180

Oliver HF, Orsi RH, Ponnala L, Keich U, Wang W, Sun Q, Cartinhour SW, Filiatrault MJ, Wiedmann M, Boor KJ (2009) Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. BMC Genomics 10:641

Olvera L, Mendoza-Vargas A, Flores N, Olvera M, Sigala JC, Gosset G, Morett E, Bolivar F (2009) Transcription analysis of central metabolism genes in *Escherichia coli*. Possible roles of sigma38 in their expression, as a response to carbon limitation. PLoS One 4:e7466

Paget MS, Kang JG, Roe JH, Buttner MJ (1998) SigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). EMBO J 17:5776–5782

Panhorst M, Sorger-Herrmann U, Wendisch VF (2011) The *pstSCAB* operon for phosphate uptake is regulated by the global regulator GlxR in *Corynebacterium glutamicum*. J Biotechnol 154:149–155

Park SD, Youn JW, Kim YJ, Lee SM, Kim Y, Lee HS (2008) *Corynebacterium glutamicum* sigmaE is involved in responses to cell surface stresses and its activity is controlled by the anti-sigma factor CseE. Microbiology 154:915–923

Pátek M, Nešvera J (2011) Sigma factors and promoters in *Corynebacterium glutamicum*. J Biotechnol 154:101–113

Pátek M, Eikmanns BJ, Pátek J, Sahm H (1996) Promoters from *Corynebacterium glutamicum*: cloning, molecular analysis and search for a consensus motif. Microbiology 142:1297–1309

Pátek M, Muth G, Wohlleben W (2003a) Function of *Corynebacterium glutamicum* promoters in *Escherichia coli*, *Streptomyces lividans*, and *Bacillus subtilis*. J Biotechnol 104:325–334

Pátek M, Nešvera J, Guyonvarch A, Reyes O, Leblon G (2003b) Promoters of *Corynebacterium glutamicum*. J Biotechnol 104:311–323

Pfefferle W, Möckel B, Bathe B, Marx A (2003) Biotechnological manufacture of lysine. Adv Biochem Eng Biotechnol 79:59–112

Plassmeier J, Persicke M, Pühler A, Sterthoff C, Rückert C, Kalinowski J (2012) Molecular characterization of PrpR, the transcriptional activator of propionate catabolism in *Corynebacterium glutamicum*. J Biotechnol 159:1–11

Rodrigue S, Provvedi R, Jacques PE, Gaudreau L, Manganelli R (2006) The sigma factors of *Mycobacterium tuberculosis*. FEMS Microbiol Rev 30:926–941

Ross W, Gourse RL (2009) Analysis of RNA polymerase-promoter complex formation. Methods 47:13–24

Ross W, Ernst A, Gourse RL (2001) Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. Genes Dev 15:491–506

Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, Pühler A (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145:69–73

Schneider J, Eberhardt D, Wendisch VF (2012) Improving putrescine production by *Corynebacterium glutamicum* by fine-tuning ornithine transcarbamoylase activity using a plasmid addiction system. Appl Microbiol Biotechnol 95:169–178

Schröder J, Jochmann N, Rodionov DA, Tauch A (2010) The Zur regulon of *Corynebacterium glutamicum* ATCC 13032. BMC Genomics 11:12

Seibold GM, Hagmann CT, Schietzel M, Emer D, Auchter M, Schreiner J, Eikmanns BJ (2010) The transcriptional regulators RamA and RamB are involved in the regulation of glycogen synthesis in *Corynebacterium glutamicum*. Microbiology 156:1256–1263

Seibold GM, Breitinger KJ, Kempkes R, Both L, Krämer M, Dempf S, Eikmanns BJ (2011) The *glgB*-encoded glycogen branching enzyme is essential for glycogen accumulation in *Corynebacterium glutamicum*. Microbiology 157:3243–3251

Suda M, Teramoto H, Imamiya T, Inui M, Yukawa H (2008) Transcriptional regulation of *Corynebacterium glutamicum* methionine biosynthesis genes in response to methionine supplementation under oxygen deprivation. Appl Microbiol Biotechnol 81:505–513

Tanaka Y, Okai N, Teramoto H, Inui M, Yukawa H (2008) Regulation of the expression of phosphoenolpyruvate: carbohydrate phosphotransferase system (PTS) genes in *Corynebacterium glutamicum* R. Microbiology 154:264–274

Tateno T, Fukuda H, Kondo A (2007) Direct production of L-lysine from raw corn starch by *Corynebacterium glutamicum* secreting *Streptococcus bovis* alpha-amylase using *cspB* promoter and signal sequence. Appl Microbiol Biotechnol 77:533–541

Tateno T, Okada Y, Tsuchidate T, Tanaka T, Fukuda H, Kondo A (2009) Direct production of cadaverine from soluble starch using *Corynebacterium glutamicum* coexpressing alpha-amylase and lysine decarboxylase. Appl Microbiol Biotechnol 82:115–121

Tsuchiya M, Morinaga Y (1988) Genetic control systems of *Escherichia coli* can confer inducible expression of cloned genes in coryneform bacteria. Bio/Technology 6:428–430

Valbuena N, Letek M, Ramos A, Ayala J, Nakunst D, Kalinowski J, Mateos LM, Gil JA (2006) Morphological changes and proteome response of *Corynebacterium glutamicum* to a partial depletion of FtsI. Microbiology 152:2491–2503

van Ooyen J, Emer D, Bussmann M, Bott M, Eikmanns BJ, Eggeling L (2011) Citrate synthase in *Corynebacterium glutamicum* is encoded by two *gltA* transcripts which are controlled by RamA, RamB, and GlxR. J Biotechnol 154:140–148

Vašicová P, Abrhámová Z, Nešvera J, Pátek M, Sahm H, Eikmanns B (1998) Integrating and autonomously replicating vectors for analysis of promoters in *Corynebacterium glutamicum*. Biotechnol Tech 12:743–746

Vašicová P, Pátek M, Nešvera J, Sahm H, Eikmanns B (1999) Analysis of the *Corynebacterium glutamicum dapA* promoter. J Bacteriol 181:6188–6191

von der Osten CH, Barbas CF 3rd, Wong CH, Sinskey AJ (1989) Molecular cloning, nucleotide sequence and fine-structural analysis of the *Corynebacterium glutamicum fda* gene: structural comparison of *C. glutamicum* fructose-1,6-biphosphate aldolase to class I and class II aldolases. Mol Microbiol 3:1625–1637

Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol 187:1591–1603

Youn JW, Jolkver E, Krämer R, Marin K, Wendisch VF (2008) Identification and characterization of the dicarboxylate uptake system DccT in *Corynebacterium glutamicum*. J Bacteriol 190:6458–6466

Zemanová M, Kadeřábková P, Pátek M, Knoppová M, Šilar R, Nešvera J (2008) Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*. FEMS Microbiol Lett 279: 195–201

Part II
Metabolic Engineering for Overproducing Chemicals and Materials

Chapter 11
Production of Fumaric Acid by Fermentation

Adrie J.J. Straathof and Walter M. van Gulik

Abstract Fermentative fumaric acid production from renewable resources may become competitive with petrochemical production. This will require very efficient processes. So far, using *Rhizopus* strains, the best fermentations reported have achieved a fumaric acid titer of 126 g/L with a productivity of 1.38 $g L^{-1} h^{-1}$ and a yield on glucose of 0.97 g/g. This requires pH control, aeration, and carbonate/CO_2 supply. Limitations of the used strains are their pH tolerance, morphology, accessibility for genetic engineering, and partly, versatility to alternative carbon sources. Understanding of the mechanism and energetics of fumaric acid export by *Rhizopus* strains will be a success factor for metabolic engineering of other hosts for fumaric acid production. So far, metabolic engineering has been described for *Escherichia coli* and *Saccharomyces cerevisiae.*

Keywords Citrate cycle • Neutralizing agents • Pellet formation • Pyruvate carboxylase • *Rhizopus* strains

Abbreviations

DSP	downstream processing
PYC	pyruvate carboxylase
RBC	rotary biofilm contactor
RoFUM	*Rhizopus oryzae* fumarase
RoMDH	*Rhizopus oryzae* malate dehydrogenase

A.J.J. Straathof (✉) • W.M. van Gulik
Department of Biotechnology, Delft University of Technology,
Julianalaan 67, 2628 BC Delft, The Netherlands
e-mail: a.j.j.straathof@tudelft.nl

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_11,

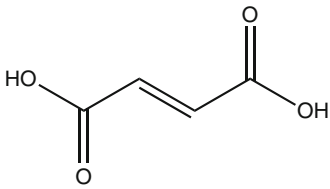


Fig. 11.1 Chemical structure of fumaric acid

11.1 Introduction

Production of chemicals such as alcohols or organic acids by fermentation is currently attracting a lot of attention. Production of organic acids, including fumaric acid, by filamentous fungi has been reviewed occasionally (Magnuson and Lasure 2004; Goldberg et al. 2006). Recently, Roa Engel et al. (2008) published a review focusing on development of fermentation based fumaric acid production processes. The current review revises this, and updates it with the new developments that have taken place. After briefly summarizing fumaric acid properties, applications, and production, fumaric acid production by fermentation will be addressed. Metabolic routes for fumaric acid production will be discussed as well as the factors determining the performance of the most prominent fumaric acid producing strains. Furthermore, different methods to optimize fermentation processes will be described and future prospects will be discussed.

11.2 Fumaric Acid and Fumarate: Properties and Applications

Fumaric acid is also known as (*E*)-2-butenedioic acid or *trans*-1,2-ethylenedicarboxylic acid (Fig. 11.1). It is a white solid with a low aqueous solubility (6–7 g/kg at 25°C; 50 g/kg at 80°C (Weiss and Downs 1923a; Lange and Sinks 1930)). As compared to other carboxylic acids, it is relatively acidic, having pK_a values of 3.10 and 4.64 (Das et al. 1980). The acidity depends on temperature and ionic strength.

The solubility of fumarate salts depends on the cation. Weiss and Downs (1923b) described cadmium, cobalt, copper, lead, manganese, nickel, silver and strontium salts with an aqueous solubility below 5 g/kg; calcium and zinc salts (hydrated) that were somewhat better soluble; and sodium salt with a high solubility of 228 g/kg. For calcium and sodium, hydrogenfumarate salts were described with a solubility of 211 g/kg (hydrated) and 69 g/kg, respectively. Fumarate and hydrogenfumarate salts are also known for ammonium and potassium.

Fumaric acid's carbon-carbon double bond and its two carboxylic acid groups make it suitable for many potential industrial applications (Table 11.1). It can act as starting material for polymerization and esterification reactions.

As monomer for production of polymers such as unsaturated polyester resins, maleic anhydride is currently preferred to fumaric acid because it is cheaper than fumaric acid. The latter is historically ~10% more expensive than maleic anhydride.

Table 11.1 Applications of fumaric acid (The Innovation Group, website 2006)

Application	Amount (tons/a)
Paper resins	31,500
Food and beverage	19,800
Unsaturated polyester resins	13,500
Alkyd resins	5,400
Plasticizers	4,500
Miscellaneous	15,300
Total	90,000

Nevertheless, fumaric acid could still be preferred, because of its non-toxic nature and because a greater hardness in the polymer structure can be achieved when fumaric acid is used.

Because of its acidity, fumaric acid has been used in food and beverage products since 1946. It is currently used in wheat and corn tortillas, sour dough and rye breads, refrigerated biscuit dough, fruit juice and nutraceutical drinks, gelatin desserts, gelling aids, pie fillings and wine (www.bartek.com). Fumaric acid improves quality and reduces costs of many food and beverage products.

Another application of fumaric acid is as supplement in animal feed. Studies indicate that a large reduction in the methane emissions of cattle can be achieved (up to 70%), if this cattle receives fumaric acid based additives as supplements in their diet (McGinn et al. 2004). Also, fumaric acid increases feed efficiency for pigs and poultry.

Fumarate salts can be enzymatically converted to L-aspartate and to L-malate by aspartase and fumarase, respectively. The industrial processes for these versatile products have been described by Liese et al. (2006).

A very different application of fumaric acids is in a medicine to treat psoriasis (Reich et al. 2009). Long-term treatment of patients with moderate and severe psoriasis with fumaric acid dimethyl ester showed a good and sustained clinical efficacy combined with a favorable safety profile. The dimethyl ester is also promising in treatment of multiple sclerosis (Killestein et al. 2011).

11.3 Fumaric Acid Occurrence and Production

Fumaric acid is a naturally occurring key intermediate in the citrate cycle and therefore present in a very wide range of organisms. Some microorganisms excrete fumaric acid as fermentation product. Fungi, in particular, are known for their organic acid producing capability and have been used in fermentation processes for fumaric acid production. This will be discussed later.

Fumaric acid production by fermentation was operated in the United States during the 1940s (Goldberg et al. 2006) but later this process was discontinued and replaced by chemical synthesis from maleic anhydride. Maleic anhydride was originally

produced by oxidation of benzene but in modern processes *n*-butane/*n*-butene mixtures are used (Felthouse et al. 2001). However, as petroleum prices are rising, maleic anhydride as a petroleum derivative increases in price as well. Therefore the interest has revived in fumaric acid production by fermentation of renewable resources. The fermentation process is also interesting because it involves carbon dioxide fixation, as will be discussed later.

11.4 Natural Fumaric Acid Producing Strains

Fumaric acid production by *Rhizopus nigricans* has been discovered more than a century ago by Ehrlich (1911). Foster and Waksman (1939) screened 41 strains from 8 different genera of Mucorales to identify high fumarate producing strains. The fumarate producing genera identified were *Rhizopus, Mucor, Cunninghamella*, and *Circinella* species. Among these strains, *Rhizopus* species (*nigricans, arrhizus, oryzae)* were the best producing ones, yielding fumaric acid under aerobic and anaerobic conditions (Foster and Waksman 1939; Rhodes et al. 1959; Kenealy et al. 1986; Cao et al. 1996; Carta et al. 1999). Abe et al. (2007) have indicated that fumaric acid producing *R. oryzae* strains should be renamed to *R. delemar*. This is not common practice, however, and in this review such strains are indicated by *R. oryzae* to avoid confusion.

Usually submerged batch fermentation is used for fumaric acid fermentation. Table 11.2 includes important results achieved with *Rhizopus* strains, and as is shown there *R. arrhizus* NRRL 1526 gave the highest volumetric productivity, product titer and product yield values (Ling and Ng 1989).

Other microorganisms, such as some *Aspergillus niger* and *A. flavus* strains, can excrete fumaric acid as a minor fermentation product (Goldberg et al. 2006), but such strains have so far not been used for fumaric acid production purposes.

11.5 Metabolic Pathways to Fumarate

In metabolic pathways such as the citrate cycle, fumarate is an intermediate between succinate and malate. In addition, fumarate plays a role in several biosynthetic and degradation pathways that will probably have no consequences for the fermentative production of fumaric acid.

Figure 11.2 shows the main carbon flows during fumarate production. These involve fumarate reactions in the mitochondria (through the citrate cycle) as well as in the cytosol. CO_2 fixation catalyzed by pyruvate carboxylase under aerobic conditions explained the high molar yields observed for fumarate production (Overman and Romano 1969). This CO_2 fixation leads to oxaloacetic acid formation (Osmani and Scrutton 1985), so that C_4 citrate cycle intermediates can be withdrawn for biosynthesis during the growth phase under aerobic conditions. When nitrogen becomes limiting and the growth phase stops, the metabolism of glucose and CO_2

Table 11.2 Fumaric acid production by different microorganisms

Strain	Fermentor	Substrate	Product titer (g/L)	Yield (g/g)	Vol. prod. (g L^{-1} h^{-1})	Time (h)	Final pH	References
R. nigricans 45	Shake flask	Glucose	18	0.46	0.15	120	6.5	Foster and Waksman (1938)
	Shake flask	Glucose	20	0.66	0.25	80	6.5	Romano et al. (1967)
R. arrhizus NRRL 2582	Stirred tank	Glucose	90	0.70	1.2	72	6.0	Rhodes et al. (1962)
	Stirred tank	Glucose	107	0.89	2.00	53	6.0	Ng et al. (1986)
	Stirred tank	Glucose	73	0.72	0.50	147	5.5	Gangl et al. (1990)
R. arrhizus NRRL 1526	Shake flask	Glucose	98	0.81	1.0	96	6.0	Kenealy et al. (1986)
	Stirred tank	Glucose	126	0.97	1.38	91	–	Ling and Ng (1989)
	Shake flask	Starch	44	0.59	0.31	140	5.9	Moresi et al. (1991)
	Fluidized bed	Molasses	17	0.36	0.36	48	6.0	Petruccioli et al. (1996)
	Stirred tank	Glucose	41	0.31	0.38	108	5.5	Riscaldati et al. (2000)
	Stirred tank	Glucose	75	0.58	1.15	65	6.8	Guettler et al. (2010)
R. oryzae ATCC 20344	RBC[a] plus adsorption	Glucose	92	0.85	4.2	24	4.5	Cao et al. (1996)
	Stirred tank	Glucose	65	0.65	0.90	72	5.0	Cao et al. (1996)
	RBC[a]	Glucose	75	0.75	3.8	24	5.0	Cao et al. (1997)
	10-L air lift	Glucose	38	0.75	0.81	46	5.0	Du et al. (1997)
	Stirred tank	Glucose	36	0.60	0.90	40	5.5	Zhou (1999)
	Bubble column	Glucose	37	0.53	1.0	36	5.0	Zhou et al. (2002)
	Shake flask	Corn straw	28	0.35	0.33	84	–	Xu et al. (2010)
E. coli JM125	Shake flask	Glucose	31	0.38	0.64	48	6.5	Jiang et al. (2010)

[a]Rotatory biofilm contactor

fixation might continue and lead to an accumulation of C_4 acids (Romano et al. 1967). Apparently, *Rhizopus* selectively exports fumarate in the presence of malate. The maximal theoretical yield in a non-growth situation is 2 moles of fumaric acid per mole of glucose consumed (1.29 g/g), upon fixation of 2 moles of CO_2 via reductive pyruvate carboxylation. However, the reductive pyruvate carboxylation does not produce ATP for maintenance or transport purposes. Then, ATP must be produced via the oxidative branch of the citrate cycle in the presence of oxygen (Rhodes et al. 1959; Kenealy et al. 1986), requiring aeration, or by fermentative routes such as ethanol formation. Either way will be at the expense of the fumaric acid yield.

Pyruvate carboxylase, the CO_2 carboxylation enzyme, is known to be localized exclusively in the cytosol, together with NAD-malate dehydrogenase and fumarase (which are also present in the cytosol and in the mitochondria), leading to fumaric acid synthesis in the cytosol (Osmani and Scrutton 1985). Peleg et al. (1989) indicated higher activities of these enzymes (especially the cytosolic isoenzymes) during fumaric acid production. Besides, it was also found that addition of cycloheximide virtually eliminated the cytosolic fumarase isoenzyme and therefore caused a large decrease in the amount of fumaric acid produced by *R. oryzae* (Peleg et al. 1989). Using mitochondrial inhibitors, Kenealy et al. (1986) found no direct involvement of these inhibitors of the citrate cycle on fumarate production, but carbon-labeling studies demonstrated the simultaneous utilization of both the citrate cycle and the reductive pyruvate carboxylation pathways under aerobic conditions (Fig. 11.2).

Although the glyoxylate bypass might be important for maximizing the fermentative production of fumaric acid, it cannot be a major pathway for fumaric acid production in studied *Rhizopus* strains (Rhodes et al. 1959; Romano et al. 1967). The main evidence for rejecting the glyoxylate bypass mechanism was that the key enzyme of the glyoxylate pathway, isocitrate-glyoxylate lyase, was repressed when high glucose concentrations were present such as used in the experiments used for fumaric acid production (Romano et al. 1967). However, Kenealy et al. (1986) do not rule out that a minor portion of fumarate is formed via glyoxylate.

Fumaric acid producing strains of *Rhizopus* usually also produce smaller amounts of other carboxylic acids such as malic, acetic, lactic, succinic and citric acid (Rhodes et al. 1959; Carta et al. 1999). Ethanol and glycerol can also be produced, but it was found that ethanol production can be reduced by a sufficient supply of oxygen to the culture (Cao et al. 1996).

11.6 Transport of Fumaric Acid Across Cell Membranes

The transport mechanism of fumaric acid in *Rhizopus* strains has not been studied yet. Increasing the number or activity of the dicarboxylic acid transporters could lower the intracellular fumarate concentration and could therefore have a positive effect on the production yield.

Transport of L-malic acid and other dicarboxylic acids, including fumaric acid, has been studied in yeasts that may be comparable to fumarate-producing fungi. For *S. cerevisiae* under anaerobic conditions, it was concluded that the uptake of fumaric

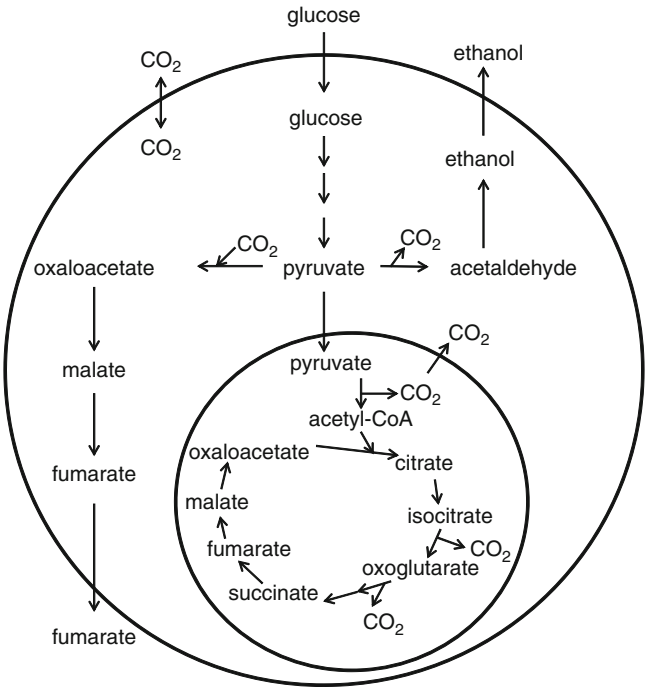


Fig. 11.2 Main carbon flows during fumaric acid production by *Rhizopus* species. The *inner circle* represents the mitochondrial membrane, with citrate cycle reactions occurring in the mitochondria, while the *outer circle* represents the cytoplasmic membranes

acid is not only happening via passive diffusion but also is facilitated by two or more membrane transporter proteins (Jamalzadeh et al. 2012). Studies with *Schizosaccharomyces pombe*, *Candida utilis*, *Candida sphaerica* and *Hansenula anomala* (Corte-Real et al. 1989; Corte-Real and Leao 1990; Saayman et al. 2000) showed that the initial uptake of malic acid was accompanied by disappearance of extracellular protons suggesting that the anionic form of the acid was transported by an accumulative dicarboxylate proton symporter. As fumaric acid seems to be a competitive inhibitor of L-malic acid uptake, it was suggested that fumaric acid uses the same import system. These studies also showed that undissociated dicarboxylic acid enters the cell slowly by simple diffusion.

11.7 Carbon Sources

For *Rhizopus* strains, glucose is not the only suitable monosaccharide for fumaric acid production. Xylose is converted as well, although at a lower rate (Rhodes et al. 1962; Kautola and Linko 1989). Using a newly isolated *Rhizopus* strain, Moon et al. (2004) observed rapid conversion of glucose, fructose and somewhat slower conversion of galactose into fumaric acid.

Amongst disaccharides, maltose was rapidly and lactose was slowly converted into fumaric acid, whereas sucrose yielded no fumaric acid (Moon et al. 2004).

Starch containing materials can also be used for the fumaric acid production, because *R. arrhizus* NRRL 2582 contains amylase that readily hydrolyzes starch (Rhodes et al. 1962). Moresi et al. 1991 used potato flour as feedstock for *R. arrhizus* and although fumaric acid was the main metabolic product, a volumetric productivity of only 0.35 $g\,L^{-1}\,h^{-1}$ was achieved. Recently, lignocellulose hydrolysate was used with *R. arrhizus* DSM 5772 (Xu et al. 2010; Julio et al. 2011).In all cases the performance of the fermentation was lower than with glucose, but might still be economically attractive because feedstock costs are important.

Foster and Waksman (1938) supplied ethanol as the only carbon source and found that 70% was converted to fumaric acid. Using glucose, transient accumulation of ethanol is often observed, and apparently this might later be converted into fumaric acid. Glycerol, another potential side-product of fumaric acid fermentation, can also serve as carbon source for formation of fumaric acid (Moon et al. 2004).

11.8 Nutrient Requirements

Nutritional requirements for fermentative fumaric acid production have been studied extensively (Foster and Waksman 1938; Rhodes et al. 1959, 1962; Zhou 1999). For example, Zhou (1999) found that trace metal concentrations of 500, 4, and 100 ppb for Mg^{2+}, Zn^{2+}, and Fe^{2+}, respectively, were optimal for the formation of small (1 mm) spherical pellets that produced high concentrations of fumaric acid (36 g/L). The most important finding, however, is that *Rhizopus* species should enter a phase of limited growth during the fermentation in order to achieve high fumaric acid yield. The most popular way to achieve this is by first performing cell cultivation without growth restriction, and then imposing nitrogen limitation. This limitation is often expressed by the ratio of glucose to nitrogen. For example, a yield of fumaric acid on glucose of 0.85 g/g was obtained using an initial C:N molar ratio of 200:1 for *R. arrhizus* 2582. When trying to produce ammonium fumarate, nitrogen limitation was not desired (Riscaldati et al. 2000). Then phosphorus limitation was successfully used instead.

CO_2 is used by *Rhizopus* for oxaloacetate formation from pyruvate by pyruvate carboxylase (Fig. 11.2). Carbonates such as $CaCO_3$ or Na_2CO_3, which are used in many cases as a neutralizing agent, seem to be also important as CO_2 source during the production phase of the fermentation. If no carbonate is used, gaseous CO_2 can be supplied together with air. Roa Engel et al. (2011) achieved better results with 10% CO_2 in air than with 0, 20, 40 or 70%. In case that neither CO_2 nor carbonate is added, the pyruvate carboxylation depends on CO_2 liberated in the citrate cycle. If all liberated CO_2would be completely used for pyruvate carboxylation, the maximum theoretical yield would be 1.5 mole of fumaric acid per mole of glucose (0.97 g/g). This theoretical value is close to the value of 1.32 mole of fumaric acid per mole of glucose reported by Cao et al. (1996), who optimized a fermentation process without CO_2 or carbonate feeding.

11.9 Neutralizing Agents

Continuous neutralization of the pH has been necessary to obtain optimal yields in fumaric acid production by fermentation. $CaCO_3$ suspensions have mostly been used as a neutralizing agent because this salt slowly dissolves during fumaric acid production. At the same time this compound is causing heavy foaming (Rhodes et al. 1962) and viscosity problems due to the low aqueous solubility of calcium fumarate (21 g/L at 30°C) (Gangl et al. 1990) that precipitates as a fermentation product. Furthermore the cells can interact with the precipitated product, as has been found for *R. oryzae*, resulting in a highly viscous suspension. This has a detrimental effect on the rate of oxygen transfer which can be achieved and hence the fermentation might fail due to oxygen limitation problems. Therefore, replacement of $CaCO_3$ by other neutralizing agents has been studied by different authors. However, the fumarate production rates are the highest when $CaCO_3$ is used as a neutralizing agent.

Because of the high solubility of sodium and magnesium fumarate, fermentative production of fumarate using Na_2CO_3 or $MgCO_3$ as neutralizing agent leads to simpler downstream processing than when $CaCO_3$ is used because there is no need of heating to recover the fermentation product, and cells might be reused (Gangl et al. 1990; Zhou et al. 2002; Guettler et al. 2010). A similar situation was obtained when *R. arrhizus* growth was limited by phosphorus so that $(NH_4)_2CO_3$ could be used as neutralizing agent (Riscaldati et al. 2000). Nevertheless, it has been argued that a fermentation process without the use of neutralizing agents and at the same time preventing product inhibition, will improve the economics of the general process (Gangl et al. 1990). However, when a high yield process is developed without carbonate as neutralizing agent, the required CO_2 must be supplied by other sources. Fumaric acid fermentation systems without the use of neutralizing agents will be discussed in a later section.

Inhibition by fumarate salts has been discussed to some extent. Rhodes et al. (1962) reported that production of soluble sodium or potassium fumarate was inhibited when the concentration of fumarate reached values of 34–40 g/L using *R. arrhizus* as a producer strain. However, Gangl et al. (1990) found that addition of sodium fumarate (71 g/L) was not inhibiting the same strain although the cells needed 35 h to adapt to the high sodium fumarate concentrations. On the other hand, undissociated fumaric acid does inhibit its own production much stronger than sodium fumarate does (Roa Engel et al. 2011). At very low pH, excreted fumaric acid will passively diffuse back through the plasma membrane of the fungus, thus decreasing its intracellular pH until the fermentation stops.

11.10 Morphology and Oxygen Transfer

One difficulty of fermenting *Rhizopus* species is the morphology of these fungi. They tend to grow on the walls and on the stirrer of the reactor and sometimes clumps are formed. This can lead to oxygen limitation. One way to solve this problem is to

stimulate formation of small spherical cell pellets (Zhou 1999). Roa Engel et al. (2011) showed from measurements with a microsensor that 0.25 mm inside a pellet the O_2 level was less than 5% of the O_2 concentration outside the pellet. Zhou et al. (2011) demonstrated that smaller pellets produce more fumaric acid, which could have been caused by the absence of oxygen limitation inside the pellets. By reducing clumps formation the viscosity of the broth is also reduced, even if $CaCO_3$ is present. Moreover, pellets can facilitate biomass retention. For *R. oryzae* a large number of factors have been optimized to favor pellet formation during cell cultivation (Zhou et al. 2000, 2011; Liu et al. 2008; Roa Engel et al. 2011). These include pH, temperature, nutrients, spore concentration and age, stirring frequency, and culture volume.

In another morphology improvement study, Cao et al. (1997) used a rotary biofilm contactor (RBC) as fermentor with self-immobilized *R. oryzae* to produce fumaric acid. During the fermentation, vertical discs with adhered cells were rotating, moving the cells from the gas phase of the fermentor to the liquid phase and back again. When the cells are exposed to the air, high oxygen transfer rates can be reached, while the cells can take up substrate and excrete fumaric acid when they are submerged. In this system additional agitation was not needed. The volumetric productivity was very high, as compared to the volumetric productivity of an equivalent stirred vessel fermentor setup (Table 11.2). However, upscaling the RBC fermentor might be difficult.

Furthermore, as the oxygen mass transfer resistance through the boundary layer on the liquid side of the gas–liquid interface can affect the interfacial oxygen transfer from the gas phase to the liquid phase, pressure pulsation was applied in a stirred tank fermentor using *R. oryzae* (Zhou 1999), to reduce this resistance. Yield on glucose and volumetric productivity of fumarate were 0.70 g/g and 0.99 $g L^{-1} h^{-1}$, respectively, which were higher than for traditional stirred tank fermentations. The same organism was studied in an airlift loop reactor as a fermentation system for fumaric acid production (Du et al. 1997). Here the airlift loop reactor with porous sparger produced favorable conditions for mass transfer and also higher yields and productivities were reached than in stirred tank fermentations.

11.11 Metabolic Engineering for Fumaric Acid Production

Genetic engineering has been applied to achieve fumaric acid production in non-fumaric acid producing organisms. The main motivations for this are that the natural producing *Rhizopus* species are difficult to cultivate due to their morphology (see Sect. 11.10, above), while they are less genetically accessible than generally used industrial micro-organisms like for example *Saccharomyces cerevisiae* and *Escherichia coli*. This hampers the optimization of fermentation processes for fumaric acid production with *Rhizopus*. So far only initial attempts have been undertaken to achieve fumaric acid production in other organisms. Pines et al. (1997)

overexpressed native cytosolic malate dehydrogenase (MDH2) in *S. cerevisiae*, resulting in a 6- to 16-fold increase in activity. This resulted in the excretion of malic (11.8 g/L), but also of fumaric (0.37 g/L) and citric acid (30.6 g/L). The authors stated that pyruvate carboxylase appeared to be a limiting factor, thus being a target for further metabolic engineering. Liu et al. (2011) introduced fumarase and dehydrogenase genes from *R. oryzae* NRRL 1526 in *S. cerevisiae* and obtained fumaric acid production, although the titer of 26 mg/L was quite low. Xu et al. (2012) also introduced malate dehydrogenase (RoMDH) and fumarase (RoFUM1) from *R. oryzae* into *S. cerevisiae*, but also overexpressed endogenous pyruvate carboxylase (PYC2). The resulting strain was able to accumulate 3.18 ± 015 g/L of fumaric acid in batch culture.

The potential of *S. cerevisiae* for overproduction of dicarboxylic acids has been demonstrated by Zelle et al. (2008), who constructed a strain that produced 59 g/L of malic acid by engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Also the overproduction of succinic acid by *S. cerevisiae* appeared so successful that the company DSM announced to open a commercial scale (10 kt/year) bio-based succinic acid plant in 2012 (DSM press release 9 May 2011).

Jiang et al. (2010) engineered *E. coli* to obtain fumaric acid. Pathways to acetyl-CoA, acetate, formate and succinate were blocked, whereas pyruvate carboxylation was overexpressed. The best example given with strain JM125 converted in 48 h 80 g/L glucose anaerobically into 30.7 g/L fumarate, using 15 g/L $MgCO_3$. These are promising data, although not yet competitive with the best *Rhizopus* fermentations. A disadvantage of using *E. coli* for fumaric acid production is that the cultivations have to be carried out at neutral pH as the organism is not tolerant to acidic conditions. This prohibits a low pH fermentation processes with this organism, which would be advantageous for downstream processing (see Sect. 11.12, below). From this point of view *S. cerevisiae* would be a better choice as host organism, as it can be cultivated under acidic conditions, i.e. pH 3 (Jamalzadeh et al. 2012).

11.12 Overall Processes for Fermentation and Recovery of Fumaric Acid

Downstream processing (DSP) has been studied less for fumaric acid than for related fermentation products such as succinic acid (Kurzrock and Weuster-Botz 2010), citric acid (Heinzle et al. 2007) and lactic acid (Joglekar et al. 2006). For all these carboxylic acids, the main DSP issue is to convert the carboxylate salt that is obtained during fermentation into the carboxylic acid. The default option leads to a stoichiometric amount of waste: A strong mineral acid is added to filtered fermentation broth, and a mineral waste salt is formed together with the produced carboxylic acid. Although it might seem that the same methods can be used for all carboxylic acids, the low aqueous solubility of fumaric acid leads to some different approaches. Upon the acidification, one might aim at precipitating fumaric acid, whereas for

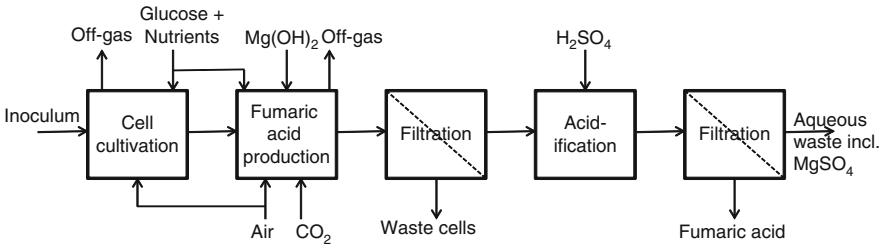


Fig. 11.3 Block scheme for fermentative fumaric acid production as derived from Guettler et al. (2010)

other carboxylic acids one may rather aim at precipitating the insoluble mineral waste salt. This has implications for the choice of the neutralizing base during fermentation (see earlier section). Figure 11.3 shows a block scheme of a process as described by Guettler et al. (2010). This is the basis of pilot scale production (3 m^3) of fumaric acid by the company MBI. A comparable process scheme, using Na_2CO_3 instead of $Mg(OH)_2$ as base, has been proposed by Gangl et al. (1990). The main disadvantage of such processes is the cost of consumption of stoichiometric amounts of mineral acid and base.

Fumaric acid can also be recovered without such acid and base consumption. Roa Engel (2010) performed a fermentation using *R. oryzae* ATCC 20344 at a final pH as low as 3.6 and obtained crystalline fumaric acid from filtered fermentation broth upon merely cooling it to 0–5°C. The mother liquor with uncrystallized fumaric acid and fumarate, side products, and unconverted nutrients was returned for continued fermentation. A continuous version of such a process would be feasible, but the efficiency of the system would be much better if the fermentation could be performed at pH 2–3, where much less (soluble) fumarate salt would be present, and the extent of recovery during the cooling crystallization would be much higher. Developing strains that would tolerate this pH without compromising the fermentation performance too much is an obvious target for future research.

Simultaneous fermentation and recovery is an attractive target. Removal of fumaric acid during its formation allows control of fermentation pH at desired values without adding mineral base. It may also minimize product inhibition (Cao et al. 1996; Zhou 1999). Cao et al. (1996) used a rotary biofilm contactor (RBC) setup as fermentor for *R. oryzae*. The produced fumaric acid was removed from the broth using an adsorption column in a recycle loop, reducing product inhibition and thus increasing the production rate and sustaining cell viability. Polyvinyl pyridine anion exchange resin in the hydroxide form was selected as adsorbent because it yielded the highest loading capacity for fumaric acid (0.31 g/g dry wt). The RBC, coupled with the adsorption column, increased the fumaric acid productivity significantly, to 4.25 $g\,L^{-1}\,h^{-1}$, because the total fermentation time was much less than in traditional stirred tank fermentations (see Table 11.2). This volumetric productivity is the highest reported in literature; also yield and product titer were high. In this integrated system,

constant removal of produced fumarate and liberation of OH^- from the adsorption column kept the fermentation pH at 4.5. The fumarate was desorbed from the adsorption resin as sodium salt by using 0.4 M NaOH. Additional processing will be necessary to convert the sodium fumarate into neutral fumaric acid.

A related product recovery process using anion exchange was developed by Zhou (1999). Fumarate was recovered from a fermentation broth without the use of neutralizing agents while keeping the pH at 5, by cycling the broth over a column of a resin (Amberlite IRA-900 with OH^- as counterion). Upon eluting the loaded column with ammonium hydroxide, an ammonium fumarate solution was obtained. This was passed through a Y-zeolite column, which retained the ammonium and liberated the fumaric acid. By using thermal regeneration of the zeolite, the ammonium hydroxide solution can be recovered and recycled for pH control at the fermentation (Zhou 1999). Although this integrated process did not surpass yield values obtained in a stirred tank fermentor when fumaric acid fermentation was performed using $CaCO_3$, a higher productivity value of 1.09 $g L^{-1} h^{-1}$ was reached (Zhou et al. 2000). Besides, waste salt production can be avoided using the Y-zeolite.

11.13 Future Prospects

Fumaric acid, together with the related succinic and malic acids, has been identified as one of the top ten building block chemicals that can be produced from sugars via biological or chemical conversion (Werpy and Petersen 2004). However, it disappeared from this list in a re-evaluation (Bozzell 2010). One reason has been that developments have been faster for some other building block chemicals. Fermentatively produced fumaric acid is probably not yet a cheaper alternative to petrochemically based maleic acid for use as unsaturated dibasic acid in polyester resins. So far, the organisms with the highest productivity and yield of fumaric acid are *Rhizopus* strains. To maximize the fermentation performance, the origin of the limitations in yield and productivity for this microorganism should be understood. Although metabolic engineering of *Rhizopus* strains will be explored, introducing an optimized pathway in a genetically more accessible and easier fermentable industrial host organism will probably become the main trend. Then, understanding of the mechanism and energetics of fumaric acid export by *Rhizopus* strains is an obvious requirement. If metabolic engineering could lead to efficient fumaric acid production by acid-resistant microorganisms, the costs of auxiliary chemicals can be minimized and the product recovery will be facilitated. *In-situ* product removal may reduce the required acid tolerance of such strains.

Acknowledgements This work was carried out within the BE-Basic R&D Program, which was granted a FES subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I).

References

Abe A, Oda Y, Asano K, Sone T (2007) *Rhizopus delemar* is the proper name for *Rhizopus oryzae* fumaric-malic acid producers. Mycologia 99:714–722

Bozzell JJ (2010) Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. Green Chem 12:539–554

Cao NJ, Du JX, Gong CS, Tsao GT (1996) Simultaneous production and recovery of fumaric acid from immobilized *Rhizopus oryzae* with a rotary biofilm contactor and an adsorption column. Appl Environ Microbiol 62:2926–2931

Cao NJ, Du JX, Chen CS, Gong CS, Tsao GT (1997) Production of fumaric acid by immobilized *Rhizopus* using rotary biofilm contactor. Appl Biochem Biotechnol 63–5:387–394

Carta FS, Soccol CR, Ramos LP, Fontana JD (1999) Production of fumaric acid by fermentation of enzymatic hydrolysates derived from cassava bagasse. Bioresour Technol 68:23–28

Corte-Real M, Leao C (1990) Transport of malic acid and other carboxylic acids in the yeast *Hansenula anomala*. Appl Environ Microbiol 56:1109–1113

Corte-Real M, Leao C, van Uden N (1989) Transport of L-malic acid and other dicarboxylic acids in the yeast *Candida sphaerica*. Appl Microbiol Biotechnol 31:551–555

Das RC, Dash UN, Panda KN (1980) Thermodynamics of dissociation of DL-malic, maleic and fumaric acids in water and water + dioxan mixtures. J Chem Soc Farad Trans I 76:2152–2158

Du JX, Cao NJ, Gong CS, Tsao GT, Yuan NJ (1997) Fumaric acid production in airlift loop reactor with porous sparger. Appl Biochem Biotechnol 63–65:541–556

Ehrlich P (1911) Über die Bildung von Fumarsäure durch Schimmelpilze. Ber Deu Chem Ges 44:3737–3742

Felthouse TR, Burnett JC, Horrell B, Mummey MJ, Kuo YJ (2001) Maleic anhydride, maleic acid, and fumaric acid. In: Kirk-Othmer encyclopedia of chemical technology, 4th edn. Wiley, vol 15, pp 1–58

Foster JW, Waksman SA (1938) The specific effect of zinc and other heavy metals on growth and fumaric acid production by *Rhizopus*. J Am Chem Soc 37:599–617

Foster JW, Waksman SA (1939) The production of fumaric acid by molds belonging to the genus *Rhizopus*. J Am Chem Soc 61:127–135

Gangl IC, Weigand WA, Keller FA (1990) Economic comparison of calcium fumarate and sodium fumarate production by *Rhizopus arrhizus*. Appl Biochem Biotechnol 24–25:663–677

Goldberg I, Rokem JS, Pines O (2006) Organic acids: old metabolites, new themes. J Chem Tech Biotechnol 81:1601–1611

Guettler MV, Hanchar RJ, Rumler DS, Kleff S (2010) Fermentation method using a magnesium compound containing oxygen. WO2010147920

Heinzle E, Biwer AP, Cooney CL (2007) Development of sustainable bioprocesses: modelling and assessment. Wiley, Chichester

Jamalzadeh E, Verheijen PJT, Heijnen JJ, van Gulik WM (2012) pH-dependent uptake of fumaric acid in *Saccharomyces cerevisiae* under anaerobic conditions. Appl Environ Microbiol 78: 705–716

Jiang M, Ma J, Wang Y, Yu L, Huang X (2010) Novel constructed high-yield fumaric acid gene engineering bacterium and method for producing fumaric acid thereby. CN101240259

Joglekar HG, Rahman I, Babu S, Kulkarni BD, Joshi A (2006) Comparative assessment of downstream processing options for lactic acid. Sep Purif Technol 52:1–17

Julio RL, Antonio DS, Diana MG, Aloia R, Juan CP (2011) Fermentative production of fumaric acid from Eucalyptus globulus wood hydrolyzates. J Chem Technol Biotechnol 87(7):1036–1040

Kautola H, Linko YY (1989) Fumaric acid production from xylose by immobilized *Rhizopus arrhizus* cells. Appl Microbiol Biotechnol 31:448–452

Kenealy W, Zaady E, Dupreez JC, Stieglitz B, Goldberg I (1986) Biochemical aspects of fumaric acid accumulation by *Rhizopus arrhizus*. Appl Environ Microbiol 52:128–133

Killestein J, Rudick RA, Polman CH (2011) Oral treatment for multiple sclerosis. Lancet Neurol 10:1026–1034

Kurzrock T, Weuster-Botz D (2010) Recovery of succinic acid from fermentation broth. Biotechnol Lett 32:331–339

Lange BA, Sinks MH (1930) The solubility, specific gravity and index of refraction of aqueous solutions of fumaric, maleic and *i*-malic acids. J Am Chem Soc 52:2602–2604

Liese A, Seelbach K, Wandrey C (2006) Industrial biotransformations, 2nd edn. Wiley-VCH, Weinheim

Ling LB, Ng TK (1989) Fermentation process for carboxylic acids. US4877731

Liu Y, Liao W, Chen SL (2008) Study of pellet formation of filamentous fungi *Rhizopus oryzae* using a multiple logistic regression model. Biotechnol Bioeng 99:117–128

Liu, L, Xu G, Duan N, Zuang H, Chen J (2011) *Saccharomyces cerevisiae* genetic engineering strain for producing fumaric acid as well as construction method and application thereof. CN102031227

Magnuson JK, Lasure LL (2004) Organic acid production by filamentous fungi. In: Tracz JS, Lange L (eds) Advances in fungal biotechnology for industry, agriculture and medicine. Kluwer Academic/Plenum Publishers, New York, pp 307–340

McGinn SM, Beauchemin KA, Coates T, Colombatto D (2004) Methane emissions from beef cattle: effects of monensin, sunflower oil, enzymes, yeast and fumaric acid. J Anim Sci 82:3346–3356

Moon SK, Wee YJ, Yun JS, Ryu HW (2004) Production of fumaric acid using rice bran and subsequent conversion to succinic acid through a two-step process. Appl Biochem Biotechnol 113–116:843–855

Moresi M, Parente E, Petruccioli M, Federici F (1991) Optimization of fumaric acid production from potato flour by *Rhizopus arrhizus*. Appl Microbiol Biotechnol 36:35–39

Ng TK, Hesser RJ, Stieglitz B, Griffiths BS, Ling LB (1986) Production of tetrahydrofuran/1,4 butanediol by a combined biological and chemical process. Biotechnol Bioeng Symp 17:355–363

Osmani SA, Scrutton MC (1985) The sub-cellular localization and regulatory properties of pyruvate carboxylase from *Rhizopus arrhizus*. Eur J Biochem 147:119–128

Overman SA, Romano AH (1969) Pyruvate carboxylase of *Rhizopus nigricans* and its role in fumaric acid production. Biochem Biophys Res Commun 37:457–463

Peleg Y, Battat E, Scrutton MC, Goldberg I (1989) Isoenzyme pattern and subcellular localization of enzymes involved in fumaric acid accumulation by *Rhizopus oryzae*. Appl Microbiol Biotechnol 32:334–339

Petruccioli M, Angiani E, Federici F (1996) Semi continuous fumaric acid production by *Rhizopus arrhizus* immobilized in polyurethane sponge. Process Biochem 31:463–469

Pines O, Shemesh S, Battat E, Goldberg I (1997) Overexpression of cytosolic malate dehydrogenase (MDH2) causes overproduction of specific organic acids in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 48:248–255

Reich K, Thaci D, Mrowietz U, Kamps A, Neureither M, Luger T (2009) Efficacy and safety of fumaric acid esters in the long-term treatment of psoriasis: a retrospective study (FUTURE). J Dtsch Dermatol Ges 7(7):603–611

Rhodes RA, Moyer AJ, Smith ML, Kelley SE (1959) Production of fumaric acid by *Rhizopus arrhizus*. Appl Microbiol 7:74–80

Rhodes RA, Lagoda AA, Jackson RW, Misenhei TJ, Smith ML, Anderson RF (1962) Production of fumaric acid in 20 liter fermentors. Appl Microbiol 10:9–15

Riscaldati E, Moresi M, Federici F, Petruccioli M (2000) Direct ammonium fumarate production by *Rhizopus arrhizus* under phosphorous limitation. Biotechnol Lett 22:1043–1047

Roa Engel CA (2010) Integration of fermentation and crystallisation to produce organic acids. PhD thesis. Delft University of Technology, Delft

Roa Engel CA, Straathof AJJ, Zijlmans TW, van Gulik WM, van der Wielen LAM (2008) Fumaric acid production by fermentation. Appl Microbiol Biotechnol 78:379–389

Roa Engel CA, van Gulik WM, Marang L, van der Wielen LAM, Straathof AJJ (2011) Development of a low pH fermentation strategy for fumaric acid production by *Rhizopus oryzae*. Enzyme Microb Technol 48:39–47

Romano AH, Bright MM, Scott WE (1967) Mechanism of fumaric acid accumulation in *Rhizopus nigricans*. J Biotechnol 93:600–604

Saayman M, van Vuuren HJJ, van Zyl WH, Viljoen-Bloom M (2000) Differential uptake of fumarate by *Candida utilis* and *Schizosaccharomyces pombe*. Appl Microbiol Biotechnol 54:792–798

Weiss JM, Downs CR (1923a) The physical properties of maleic, fumaric and malic acids. J Am Chem Soc 45:1003–1008

Weiss JM, Downs CR (1923b) The salts of maleic, fumaric and inactive malic acids. J Am Chem Soc 45:2341–2349

Werpy T, Petersen G (2004) Topten value added chemicals from biomass feedstocks. US Department of Energy, Washington, DC

Xu Q, Li S, Fu YQ, Tai C, Huang H (2010) Two–stage utilization of corn straw by *Rhizopus oryzae* for fumaric acid production. Bioresour Technol 101:6262–6264

Xu G, Liu L, Chen J (2012) Reconstruction of cytosolic fumaric acid biosynthetic pathways in *Saccharomyces cerevisiae*. Microb Cell Fact 24:11–24

Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JM, van Dijken JP, Pronk JT, van Maris AJ (2008) Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Appl Environ Microbiol 74:2766–2777

Zhou Y (1999) Fumaric acid fermentation by *Rhizopus oryzae* in submerged systems. PhD thesis, Purdue University, West Lafayette

Zhou Y, Du J, Tsao GT (2000) Mycelial pellet formation by *Rhizopus oryzae* ATCC 20344. Appl Biochem Biotechnol 84–86:779–789

Zhou Y, Du J, Tsao GT (2002) Comparison of fumaric acid production by *Rhizopus oryzae* using different neutralizing agents. Bioproc Biosyst Eng 25:179–181

Zhou ZX, Du GC, Hua ZZ, Zhou JW, Chen J (2011) Optimization of fumaric acid production by *Rhizopus delemar* based on the morphology formation. Bioresour Technol 102:9345–9349

Chapter 12
Metabolic Engineering of Microorganisms for Vitamin C Production

Jingwen Zhou, Guocheng Du, and Jian Chen

Abstract Vitamin C, an important organic acid, is widely used in the industries of pharmaceuticals, cosmetics, food, beverage and feed additives. Compared with the Reichstein method, biotechnological production of vitamin C is an attractive approach due to the low cost and high product quality. In this chapter, biosynthesis of vitamin C, including one-step fermentation processes and two-step fermentation processes are discussed and compared. Furthermore, the prospects of the biotechnological production of vitamin C are also presented.

Keywords Vitamin C • 2-keto-L-gulonic acid • Fermentative production • One-step fermentation

Abbreviations

AA-2G	2-O-alpha-D-glucopyranosyl-L-ascorbic acid
2,5-DKG	2,5-diketo-D-gluconate
2,5-DKGR	2,5-diketo-D-gluconic acid reductase
GUL oxidase	L-gulono-1,4-lactone oxidase
2-KLG	2-keto-L-gulonic acid
2-KLGR	2-keto-L-gulonic acid reductase
ORFs	open reading frames
PQQ	pyrro-quinoline quinone
SDH	L-sorbose dehydrogenase
SLDH	D-sorbitol dehydrogenase
SNDH	L-sorbosone dehydrogenase

J. Zhou.G. Du. J. Chen (✉)
School of Biotechnology, Jiangnan University, 1800 Lihu Road, Wuxi 214122, China
e-mail: jchen@jiangnan.edu.cn

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_12,

Table 12.1 Application of vitamin C and its derivatives

Fields	Usages
Pharmaceutical industry	For treatment of encephalitis convulsions, idiopathic thrombocytopenic purpura, atherosclerosis, viral myocarditis, cancer, secondary erythroderma, nitrite poisoning, bronchialasthma and acute viral hepatitis;
Food industry	Inhibit oxidation and browing; change the flavor of food; food color protection; inhibit corrosion of the tank wall; inhibit lipid oxidation.
Beverage industry	Antiseptic preservation function as an additive;
Cosmetic industry	For treatment of chloasma; free radical scavenging, anti-aging;
Feed industry	Improve animal anti-stress ability, enhance immune function, accelerate bone growth, improve the rate of reproduction;
Biochemical reagent	Used for the determination of phosphorus.

12.1 Introduction

Vitamin C, also named as L-ascorbic acid, is a kind of essential vitamin and antioxidant for human beings and some primates (Zhang et al. 2011). Vitamin C is widely used in pharmaceutical, foods, beverages, cosmetics and feeds industries (Table 12.1). Along with the expansion of vitamin C in different field, the market is still growing. The world market for vitamin C and its direct derivatives (such as calcium salt, potassium salt, sodium salt, and glycosylation derivatives) is beyond 0.5 billion US dollars.

The earliest commercial vitamin C production process began from the Reichstein route (Table 12.2), which is developed by Reichstein in 1934, Germany. Reichstein developed a strategy for conversion of glucose to 2-keto-L-gulonic acid (2-KLG) with five steps of chemical reactions and one step of bio-reaction (Reichstein and Grussner 1934). The 2-KLG was further converted into vitamin C with esterification (Fig. 12.1). The process was industrialized and dominated the vitamin C market for over 60 years. However, the Reichstein process has many disadvantages, such as high-energy consuming, high-amount of organic solvent requirement and serious environment pollution. Therefore, many researchers find out several different processes to improve the performance of classical Reichstein process. One of the most successful routes is the "two-step" fermentation process, which was applied in industrial scale for more than 40 years. Along with the development of industrial biotechnology, researchers further developed several other processes to improve the economic and environmental aspects of the two-step fermentation process.

According to the production process, vitamin C production by biotechnology can now be divided into two-step fermentation and one-step fermentation. The two-step fermentation including tandem fermentation process that uses glucose as substrate and the fermentation process that uses D-sorbitol as substrate (Fig. 12.2). For the

Table 12.2 Manufacturing methods of vitamin C

Routes	Fermentative products	Substrates	Microbial/chemical methods	Characteristics			References
				Concentration (g/L)	Intensity (g/L/h)	Yield (wt%)	
Reichstein routes	2-KLG	D-Glucose	5 steps of chemical reactions, 1 step of bioreaction	—	—	50	Bremus et al. (2006)
Biotechnology methods	2-KLG	D-gluconic acid	*G. oxydans* ATCC 9937 and *Corynebacterium* sp. ATCC 31090	9.43	0.13	38	Ji and Gao (1998)
		D-Sorbitol	*G. melanogenus* Z84	60	0.42	60	Sugisawa et al. (1990)
			G. oxydans NB6939/pSDH-tufB1	88	1.22	88	Saito et al. (1997)
		L-Sorbose	*G. melanogenus* U13	60	0.42	60	Sugisawa et al. (1990)
			G. oxydans IGO112 and *B. megaterium* IBM302	75.8	1.58	94.8	Xu et al. (2004)
			G. oxydans SCB329 and *B. thuringiensis* SCB933	130.92	2.85	90	Yin et al. (1997)
		L-Sorbone	*G. oxydans* U13(p7A6 4)	32.7	0.2725	83.4	Shinjoh et al. (1995)
	Ca-KLG	D-Glucose	*Erwinia* sp. SHS 2629001 and *Corynebacterium* sp. SHS 752001	106.3	1.16	84.6	Sonoyama et al. (1982)
	Vitamin C	D-Glucose	*Xanthomonas campestris* 2286	20.4	0.408	5.1	Rao and Sureshkumar (2000)
		D-Sorbitol	*K. vulgare* DSM 4025^{TP}	0.09	0.0038	0.11	Sugisawa et al. (2005)
		L-Sorbose	*K. vulgare* DSM 4025^{TP}	0.908	0.045	1.14	Sugisawa et al. (2005)
		L-Sorbone	*K. vulgare* DSM 4025^{TP}	1.37	0.34	27.4	Sugisawa et al. (2005)
		D-Galacturonic acid	*Candida norvegensis*	1.3	0.027	8.7	Petrescu et al. (1992)
		L-galactose	*Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*	0.1	6.7×10^{-4}	40	Sauer et al. (2004)

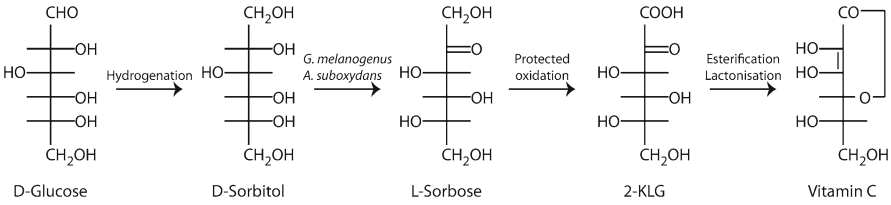


Fig. 12.1 Reichstein process for vitamin C production. The D-glucose was hydrogenated to form D-sorbitol. The D-sorbitol was converted into L-sorbose by acetic bacteria. The L-sorbose was further oxidized with protection to form 2-KLG. The 2-KLG was then esterified and lactonized to form vitamin C

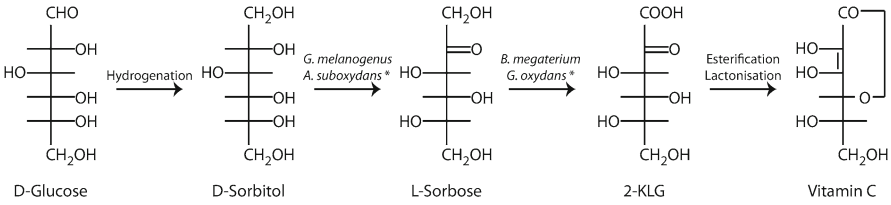


Fig. 12.2 Classical two-step fermentation process. The D-glucose was hydrogenated to form D-sorbitol. The D-sorbitol was converted into L-sorbose by acetic acid bacteria. The L-sorbose was further oxidized with a mixture culture system with *B. megaterium* and *K. vulgare* to form 2-KLG. The 2-KLG was then esterified and lactonized to form vitamin C. The only difference between the classical two-step process and the Reichstein process is the replacement of low efficient protective oxidation with a fermentation process. *The *G. oxydans* here was further identified to be *K. vulgare*

former, D-glucose is converted into 2,5-diketo-D-gluconate (2,5-DKG) by *Erwinia sp.*, which is subsequently converted into 2-KLG by *Corynebacterium*. The latter was studied the earliest and most intensively. In this process, L-sorbose is formed from D-sorbitol by *Gluconobacter oxydans*, and then by *Ketogulonicigenium vulgare* with *Bacillus megaterium* as the associated bacterium, L-sorbose is converted into 2-KLG (Zhang et al. 2011). *K. vulgare* is a 2-KLG producing bacterium, its growth is a lengthy, inefficient process with a low 2-KLG yield when cultured alone. *B. megaterium* is an associated bacterium which can promote *K. vulgare* growth and acid production but it does not produce 2-KLG (Liu et al. 2011a).

The optimization of the fermentation process has been continuing ever since, tremendous work has been done to improve the process stability and yield, such as the medium optimization, the relationship between two bacteria, the ecological regulation, etc. (Zhang et al. 2010). One-step fermentation is based on the bacteria of the two-step fermentation, it aims to achieve one-step fermentation from glucose or sorbitol to 2-KLG by using genetic engineering techniques to construct the genetically engineered bacteria.

12.2 Elucidation of the Mechanisms in Mutualism Between *Bacillus megaterium* and *Ketogulonicigenium vulgare*

12.2.1 Previous Research on the Classical Two-Step Fermentation Process

Because of co-culture of *K. vulgare* and *B. megaterium* during the classical two-step fermentation period of vitamin C, the optimal culture condition not only very suitable for growing condition of two bacteria but the excretion of acid by *K. vulgare* (Zhang et al. 2011). Zhou et al. discovered that carbon source, nitrogen source, dissolved oxygen, growth factors of seed culture medium and environmental factors etc., could regulate the growth and metabolism of strains effectively, and make the organisms grow under an optimal condition (Zhou et al. 2002). Therefore, there would be considerable impacts on the fermentation. The metabolism of mixed bacteria in vitamin C fermentation was more complex compared with that of a single strain.

The commonly used urea and corn steep liquor were nitrogen source in the industrial production of vitamin C by the classical two-step fermentation process. The metabolic properties were listed as follow: (1) There were two kinds of roles with urea addition, one was as a physiological alkaline substance to adjust pH, the other was providing the nitrogen sources for bacterial metabolism; (2) The protein content of the system was increasing with fermentation time; (3) The 17 kinds of amino acids in corn steep liquor were divided into three categories (Li et al. 1996). The large number of unclear natural components and the different batches of corn steep liquor can also cause the instability of 2-KLG production and fermentation cycle. The specific growth factors of cell growth and acid production were a prerequisite to control the stability of 2-KLG production.

Dissolved oxygen was also an important factor in the fermentation process of vitamin C. Giridhar and Srivastava (2000) found that in the batch and fed-batch fermentation processes of D-sorbitol to L-sorbose, adding 4% of the oxygen vector (N-hexadecane) could shorten the period for 2 and 5 h, and increase the production by 17 and 26%, respectively.

12.2.2 Elimination the Inhibition Effects of Substrate to Improve the Production and Yield of 2-KLG

The microbial transformation of D-sorbitol to L-sorbose is the most important industrial fermentation process of vitamin C in “Reichstein process” and two-step fermentation. Microorganisms are inhibited severely by high concentration of

D-sorbitol. The oxidation rate of D-sorbitol decrease drastically, which lead to the low concentration of L-sorbose in broth. Thereby, it will increase the downstream cost of the separation and purification. Control of initial concentration of sorbitol with 200 g/L, the production and yield of L-sorbose were 200 g/L and 14.2 g/L/h (Giridhar and Srivastava 2002). Besides this, with constant speed feeding of 600 g/L D-sorbitol, the concentration of L-sorbose was 1.6 times greater than batch fermentation; whereas 272.37 g/L L-sorbose achieved with linear feeding strategy (Giridhar and Srivastava 2001); exponential feeding of 700 g/L D-sorbitol after 10 h batch fermentation with 200 g/L D-sorbitol made 290 g/L L-sorbose (Srivastava and Lasrado 1998).

When L-sorbose was further converted to 2-KLG, huge amount of L-sorbose will make the bacteria metabolize abnormally, meanwhile the production and yield reduced. With the fermentation of 80 g/L sorbose for 35–40 h, the concentration of 2-KLG was 65–75 g/L. In the industrial process, constant feeding strategy was used widely. During the whole fermentation process, the curve of acids producing rate shaped a saddle while an ideal curve should be only a peak. Feeding sorbose according to the curve of fermentation rate would make the curve of acids producing fit the ideal model closer. Feeding L-sorbose at 10 and 20 h during fermentation, L-sorbose was 100 and 140 g/L, the concentration of 2-KLG reached 120–135 g/L, and the transformation efficiency was 90%.

12.2.3 Improvement of High Temperature Tolerance

Previous researches showed that *B. megaterium* could grow very well under 33 °C. However, *K. vulgare* could only grow under 30 °C. Therefore, the energy cost could be significantly improved by improving the higher temperature tolerance of *K. vulgare*. Because the mechanisms inside the tolerance of bacteria to high temperature are obscure, the rational metabolic engineering of *K. vulgare* to improve its tolerance to higher temperature is difficult. Current improvement of the high temperature tolerance of *K. vulgare* is mainly focused on the random mutagenesis (Yan et al. 2006) or *de novo* screening from nature (Moonmangmee et al. 2000). Yan et al. improved the performance of the mix-culture system consists of *K. vulgare* G0 and *B. megaterium* B0 under 33 °C by obtaining of mutant GI13 by ion implantation. The conversion ratio of the mutant GI13 at 33°C was improved to 94%. Besides, the L-sorbose dehydrogenase (SDH) activity of the mutant strain was further improved by 100% compared to the wild-type strain G0 (Yan et al. 2006).

G. oxydans, which is used for the conversion of the D-sorbitol to L-sorbose, could only grow under 32 °C (Yang and Lim 1997). Moonmangmee et al. (2000) screened a *G. oxydans* CHM54 from plants. The strain could grow under 37 °C well and has a conversion ratio of 80%.

12.2.4 Inhibition of the Degradation Pathway of 2-KLG

During the fermentation process of the mix-culture of *Pseudomonas putida* ATCC 21812 and *Gluconobacter melanogenus* IFO 3293 for the 2-KLG production, it was discovered that there was an enzyme, which needs NAD(P)H as cofactor, could further convert the 2-KLG to L-idonic acid. Furthermore, a similar enzyme was also found in *K. vulgare* WB0104 (Yang et al. 2006). Besides, the 2-ketoaldonate reductase A and B in *Erwinia herbicola* SCB125 could also degrade 2,5-DKG and 2-KLG. These downstream pathways could inhibit the further accumulation of 2-KLG. Therefore, knockout of these genes is essential for the further improvement of 2-KLG production. Chen et al. (2000) knocked out the *thrA* gene with a streptomycin resistance gene and successfully decreased the 2-KLG degradation.

12.2.5 Effects of the Relationship Between Two Bacteria in Mixed Culture on Vitamin C Production

The two-step fermentation process which uses D-sorbitol as the substrate is a classic example where a mixed-culture method produces a metabolic product. The second step fermentation is completed by mixed fermentation of 2-KLG-producing bacteria *K. vulgare* and associated bacteria *B. megaterium*. *K. vulgare* is difficult to culture alone and 2-KLG production capacity is very low simultaneously. *B. megaterium* does not produce 2-KLG, but it can promote the growth and acid production of *K. vulgare*. There are many kinds of associated bacteria, In addition to the commonly used *B. megaterium*, *Bacillus cereus*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus thuringiensis*, and some yeast strains also possess the role.

It is a collaborative symbiotic relationship between acid producing bacteria and associated bacteria (Zhou et al. 2002). The enzyme activity that synthesizes of 2-KLG is directly proportional to the amount of *K. vulgare* cells and has nothing to do with the amount of *B. megaterium* cells. Further studies have shown that the 30–50 KDa (Feng et al. 2000) and >100 KDa (Lv et al. 2001) biological active substances of *B. megaterium* released in the metabolic process, participating in the growth and acid production process of *K. vulgare*. The extracellular active substances are proteins.

Three bacteria were involved in the classical two-step fermentation process, for example *G. oxydans*, *B. megaterium* and *K. vulgare*. The denomination of the bacteria for *G. oxydans* and *K. vuglare* is highly speculative because of the lack of efficient strain identification methods. Furthermore, there are also some updates on the two strains. In early times, *G. oxydans* were named as *G. melanogenus*, *Acetobacter suboxydans*, or *A. melanogenus*. Because most of these stains could not

be sufficiently distinguished for further confirmation, all these strains that could efficiently catalyze D-sorbitol to L-sorbose are now designated as *G. oxydans*. The species classified as *K. vulgare* were finally defined as *K. vulgare*. The bacterium was previously known as *G. oxydans* or some other *Acetobacter* sp. strains. However, the strain that could only grow well with *B. megaterium* or other *Bacillus* sp. is now referred to as *K. vulgare*.

In the classical two-step fermentation process, the D-sorbitol was synthesized by hydrogenation from D-glucose by chemical reaction. The D-sorbitol was then converted into L-sorbose by the sorbitol dehydrogenase (encoded by *sldh*) in *G. oxydans*. The culture broth containing L-sorbose was then transferred to another bioreactor, added with other raw materials and sterilized for the second time. Then the L-sorbose was converted in to L-sorbosone by the L-sorbose dehydrogenase (encoded by *sdh*) in *K. vulgare*. The L-sorbosone was further converted into 2-KLG by L-sorbone dehydrogenase (encoded by *sndh*) in *K. vulgare*.

The fantastic phenomenon in the classical two-step fermentation process is the second step, which consists of the *B. megaterium* and *K. vulgare*. The *K. vulgare* itself could not grow well after supplement of different kinds of substrates. The addition of *B. megaterium*, *B. thuringiensis* or *B. cereus* could significantly enhanced both the cell growth and 2-KLG production by the *K. vulgare*. There are many reports aiming to demonstrate the mutualism between the two bacteria. In the past few decades, much effort has been devoted to elucidating the detailed mechanism by which *B. megaterium* enhances *K. vulgare* growth and 2-KLG production (Bremus et al. 2006; Shinjoh et al. 1995; Sugisawa et al. 2005; Tsukada and Perlman 1972). Studies have showed that both internal and external metabolites from *B. megaterium*, through biochemical and molecular methods and certain proteins or amino acids, play a role in this process (Zhang et al. 2011; Zhao et al. 2008). Further research has shown that *B. megaterium* secretes two kinds of proteins of about 30–50 kDa and >100 kDa that improve *K. vulgare* growth and increase 2-KLG productivity (Lu et al. 2003; Xu et al. 2004; Zhao et al. 2008). However, researchers have not specifically determined which proteins play a role in promoting *K. vulgare* growth and 2-KLG production. Based on these results, Zhang et al. (2010) used lysozyme to damage the *B. megaterium* cell wall structure to release the intracellular components, as a consequence, the growth rate of *K. vulgare*, its sorbose consumption rate, and 2-KLG productivity increased 27.4, 37.1, and 28.2%, respectively. Because most of the companion strains for *K. vulgare* are *Bacillus* sp., it was proposed that the spore formation of these *Bacillus* strains should have some impact on the mutualism process. Zhu et al. (2012) first revealed that not only the spore formation, but also the spore stability, played key role in the mutualism process.

Based on the development of next-generation high-throughput sequencing technologies, all the genomes of three strains involved in the classical two-step fermentation process were obtained in last several years. One of the *B. megaterium* was sequenced. The 4.14-Mb genome of *B. megaterium* WSH-002 contains four replicons, a circular chromosome (4.04 Mb) encoding 5,186 predicted open reading frames (ORFs), and three circular plasmids, named pBME_100 (0.074 Mb),

pBME_200 (9,699 bp), and pBME_300 (7,006 bp), with mean GC contents of 39.1, 36, 32.2, and 33.2%, respectively. There are 5,482 protein-encoding genes, 99 tRNAs, and 10 rRNA operons. Among them, 2,460 functional descriptions, 1,327 gene abbreviations, and 856 EC numbers were assigned to the WSH-002 genome by function annotation. Furthermore, 782 genes were assigned according to the Kyoto Encyclopedia of Genes and Genomes metabolic pathways (Liu et al. 2011c).

There are two *K. vulgare* strains independently. According to phylogenetics, both *K. vulgare* Y25 and WSH-001 has the same ancestor. The two strains were used by the two main vitamin C manufactures in China, North China Pharmaceutical Group Corporation and Jiangsu Jiangshan Pharmaceutical Company, respectively. However, the multiple-mutagenesis with different methods may lead to some significant differences in these two strains. The genome of *K. vulgare* Y25 consists of a circular chromosome and two plasmids. The chromosome is composed of 2,776,084 bp, with a GC content of 61.72%. One plasmid contains 268,675 bp, with a GC content of 61.35%, and the other contains 243,645 bp, with a GC content of 62.63%. There are a total of 3,290 putative ORFs (2,807 [chromosome], 256 [pYP1], and 227 [pYP2]) using Glimmer, giving a coding intensity of 91.05% (Xiong et al. 2011). The complete genome sequence of *K. vulgare* WSH-001 is composed of a circular, 2,766,400-bp chromosome and two circular plasmids named pKVU_100 (267,986 bp) and pKVU_200 (242,715 bp) with mean GC contents of 61.69, 61.33, and 62.58%, respectively. There are 2,604 protein-encoding genes, three rRNA operons, and 51 tRNA-encoding genes in the chromosome and 246 and 215 protein-encoding genes in plasmids pKVU_100 and pKVU_200, respectively. 2497 functional descriptions, 1,279 gene abbreviations, and 820 EC numbers were assigned in the WSH-001 genome by function annotation. Among them, the genes for the 2-KLG synthesis pathway from L-sorbose were annotated; four genes encoding L-sorbose/L-sorbosone dehydrogenase, responsible for converting L-sorbose to L-sorbosone, are highly homologous to *ssdA1* (AB092515), *ssdA2* (AB092516), *ssdA3* (AB092517), and *ssdB* (AB092518) of *K. vulgare* strain DSM 4025. The *sndh* gene that is responsible for the conversion of L-sorbosone to 2-KLG was located in plasmid pKVU_200 (Liu et al. 2011b). It seems that the genome of *K. vulgare* (~2.7 M) is much smaller than other common bacteria. Function annotation indicated that the *K. vulgare* lacks of most of the genes or gene clusters for biosynthesis of many kinds of amino acids, nucleotides and cofactors.

Based on the release of genome sequences for *K. vulgare* and *B. megaterium*, and the integration of other -omics data, the understanding of the mechanisms in the mutualism were further illuminated (Liu et al. 2011b, c; Xiong et al. 2011). Based on the global pathway analysis, Zhang et al. (2011) reported most of the essential amino acids for the dependent growth of *K. vulgare* and developed a definite culture medium for both *K. vulgare* growth and 2-KLG production (Zhang et al. 2011). Furthermore, Liu et al. (2011a) reconstructed the biosynthesis pathways for amino acids in *K. vulgare* and found out many gaps in the amino acid metabolic pathways,

and consequently developed a novel strategy with gelatin supplement, which contains the several kinds of key amino acids that could not be *de novo* biosynthesized in *K. vulgare*, and significantly enhanced the *K. vulgare* growth and 2-KLG productivity.

Zhou et al. (2011) found that the microorganisms interact by exchanging a number of metabolites with time-of-flight mass spectrometry. Both intracellular metabolism and cell-cell communication via metabolic cooperation were essential in determining the population dynamics of the ecosystem. The contents of amino acids and other nutritional compounds in *K. vulgare* were rather lower in comparison to those in *B. megaterium*, but the levels of these compounds in the medium surrounding *K. vulgare* were fairly high, even higher than in fresh medium. Erythrose, erythritol, guanine, and inositol accumulated around *B. megaterium* were consumed by *K. vulgare* upon its migration. The oxidization products of *K. vulgare*, including 2-KLG, were sharply increased. Upon co-culturing of *B. megaterium* and *K. vulgare*, 2,6-dipicolinic acid (the biomarker of sporulation of *B. megaterium*), was remarkably increased compared with those in the monocultures. Therefore, the interactions between *B. megaterium* and *K. vulgare* were a synergistic combination of mutualism and antagonism.

12.3 Metabolic Engineering Based on the Two-Step Vitamin C Fermentation Process

The classical two-step fermentation process is the most successful route for vitamin C production for its high yield of 2-KLG on D-sorbitol. Though there are two fermentation process, the yield of L-sorbose on D-sorbitol and the yield of 2-KLG on L-sorbose could achieve to more than 99.5 and 97%, respectively. Few of the industrial process could achieve this level. Therefore, the metabolic engineering on the classical two-step fermentation process is always undergoing.

12.3.1 Interruption of the Downstream Metabolism Pathways

It has already been reported that the 2-KLG synthesized by *K. vulgare* could be further degraded into iduronic acid by 2-KLG reductase in the bacteria. Therefore, the earliest metabolic engineering for the classical two-step fermentation process is to knockout the gene that encodes the 2-KLG reductase. Saito et al. (1997) obtained a *K. vulgare* strain NB6939 (it was named as *G. oxydans* in the original literature) with interrupted L-iduronic acid by mutagenesis using N-methyl-N'-nitro-N-nitrosoguanidine, the 2-KLG concentration of the resultant strain was improved to 31 g/L, which was 200% of the wild-type strain.

12.3.2 Enhancement of the Cofactor Metabolism Pathway

Conversion of D-glucose or D-sorbitol to 2-KLG or vitamin C involves a number of dehydrogenases that need different kinds of cofactors. These dehydrogenase include D-sorbitol dehydrogenase (SLDH), L-sorbose dehydrogenase (SDH), L-sorbosone dehydrogenase (SNDH), 2,5-diketo-D-gluconic acid reductase (2,5-DKGR), 2-keto-L-gulonic acid reductase (2-KLGR). The SDH needs FAD or Pyrroloquinoline Quinone (PQQ) as cofactor in different strains. SNDH from *G. melanogenus* UV10 use NAD(P) as cofactor. 2,5-DKGR in *Corynebacterium* ATCC 31090 (Anderson et al. 1985) and 2-KLGR in *K. vulgare* (former *G. oxydans*) (Jiang et al. 1997) use NAD(P)H as the cofactor.

Therefore, rational regulation of the existence, concentration or preference of cofactors could improve 2-KLG production. Banta et al. (2002) showed that there were two kinds of 2,5-DKGR in *Corynebacterium* sp., i.e., 2,5-DKGR A and 2,5-DKGA B. The 2,5-DKGR A could use both NADH and NADPH as cofactor. However, the dissociation constant (K_d) of the enzyme with NADH is 260 times of that with NADPH as cofactor. Therefore, the 2,5-DKG A tend to use NADPH as cofactor. However, the NADH concentration is three times higher than the NADPH and has higher stability. Sanli et al. decreased the K_d of 2,5-DKGR A with NADH to 1/3 compared to the wild-type enzyme using double mutagenesis of F22Y and A272G (Sanli et al. 2001). The tetra-mutagenesis of the 2,5-DKGR A (F22Y/K232G/R238H/A272G) decreased the K_d to only 1/4 of the wild-type enzyme (Banta et al. 2002). This will significantly decreased the cost for 2-KLG production by 2,5-DKGR A in vitro.

Primary analysis showed that the *K. vulgare* could not synthesize folate, which referred to a series of folic acid derivatives that is distinguished by the state of oxidation, one-carbon substitution of the pteridine ring, and the number of glutamate residues (Sybesma et al. 2003). Folate is a kind of important cofactor, which donates one-carbon compounds in many reactions, and is crucial in the *de novo* biosynthesis of amino acids, purines, and pyrimidines (Sybesma et al. 2003). Leduc et al. (2004) showed that the derivatives of folate could enhanced the cell growth and 2-KLG production of *K. vulgare* LMPP-20356, while the vitamins, glutathione, hemin, PQQ, pyrimidines could not achieve the same results. By introducing a folate biosynthesis genes cluster found in *Lactococcus lactis* MG1363, which encodes enzymes catalyzing reactions involved in the assembly of folate and its various derivatives, into *K. vulgare*, both the cell density and 2-KLG production by the classical two-step fermentation process were significantly enhanced (Cai et al. 2012).

12.4 One-Step Vitamin C Fermentation Process

Previously, researchers from both industry and academia have invariably noted the inherent disadvantages of the two-step fermentation vitamin C production process, such as long period fermentation, additional sterilizing, control of the mix-culture

system, which lead to increased consumption for both raw materials and energy. Compared to other similar large block fermentation products, the manufacturing cost for vitamin C is much higher than others according to its final concentration of more than 100 g/L. The higher cost is mostly caused by its two-step fermentation process.

Therefore, if the vitamin C or its direct precursor, 2-KLG, could be produced by one single step fermentation process, the final cost of the vitamin C will be significantly decreased. Current research on the one-step fermentation process is focused mainly on the following three aspects: (1) Classical two-step based one-step fermentation process for 2-KLG production; (2) Innovative two-step based one-step fermentation process 2-KLG; (3) Direct production of vitamin C from glucose.

12.4.1 Classical Two-Step Based One-Step Fermentation Process

The *K. vulgare* contains the genes and proteins for the *sdh* and *sndh*, which could convert the L-sorbose to 2-KLG. The cofactor requirements of the *sdh* and *sndh* is somehow confused. However, it could be confirmed that at least one of the dehydrogenase is PQQ dependent. The PQQ is a kind of cofactor that was discovered by Hauge as the third redox cofactor after NAD^+ and FAD^+ in bacteria (Hauge 1964). Anthony and Zatman (1967) found the unknown redox cofactor in alcohol dehydrogenase and named it methoxatin (Anthony and Zatman 1967). In 1979, Salisbury and colleagues as well as Duine and colleagues extracted this prosthetic group from methanol dehydrogenase of methylotrophs and identified its molecular structure (Salisbury et al. 1979; Westerling et al. 1979). The existence of PQQ in *Acetobacter* sp. was first reported by Adachi's group (Ameyama et al. 1981).

Because the *G. oxydans* has *sldh* gene and the complete PQQ *de novo* biosynthesis and regeneration genes, if the *sdh* and *sndh* from *K. vulgare* was introduced into the *G. oxydans*, the resultant *G. oxydans* strain should produce 2-KLG from D-sorbitol by one-step fermentation process. Early research had already found out that the *G. oxydans* T100 could produce 2-KLG from sorbitol directly. However, both the final 2-KLG concentration and yield of 2-KLG on sorbitol was relatively low; 7.0 g/L and 13.1%, respectively. Overexpression of the sorbitol dehydrogenase gene *sldh* and sorbitone dehydrogenase gene *sndh* from *G. oxydans* T100 into the strain *G. oxydans* G624 with expression vector pSDH155 could significantly improve the 2-KLG production (Fig. 12.3). The final concentration of 2-KLG and the yield of 2-KLG on sorbitol was improved to 16 g/L and 30%, respectively (Saito et al. 1997). Furthermore, by inhibition of the L-idose pathway in *G. oxydans* by chemical mutagenesis and the replacement of *tufB1* promoter from *E. coli*, the recombinant strain could finally produce 2-KLG with a concentration of 88 g/L and a yield of 2-KLG on sorbitol of 82.6% (Saito et al. 1998).

G. oxydans is a kind of Gram-positive bacteria and lacks of investigation in the genetic engineering for both multi-gene overexpression and gene knockout.

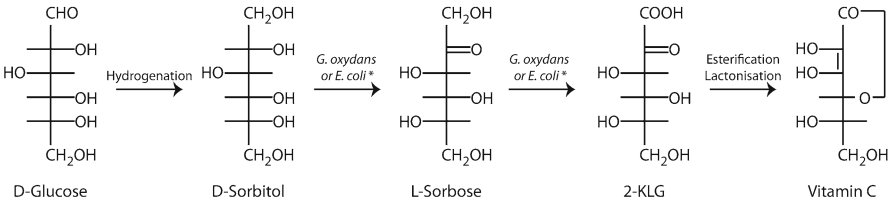


Fig. 12.3 Classical two-step based one-step fermentation process (*Recombinant strains with different dehydrogenase). The D-glucose was hydrogenated to form D-sorbitol. The D-sorbitol was directly converted into 2-KLG with a *G. oxydans* strain or a *E. coli* strain (or any other potential strains) with the three different kinds of dehydrogenases. The 2-KLG was then esterified and lactonized to form vitamin C

Another choice is to metabolic engineer other model microorganisms, such as *E. coli* or *S. cerevisae*. Because most of the dehydrogenase genes are located on the cell membrane, the difference between prokaryotes and eukaryotes makes *S. cerevisiae* a poor choice. Therefore, the aim would be to express all of the dehydrogenase and PQQ metabolism related pathways in *E. coli* for one-step biosynthesis of 2-KLG from D-sorbitol. Previous reports had already achieved the *de novo* biosynthesis of PQQ in *E. coli* by overexpression of the PQQ biosynthesis gene cluster *pqqABCDE* from *G. oxydans* (Yang et al. 2010). However, the regeneration system of the PQQ is still not well elucidated. Therefore, current works are mainly focused on the bacteria that has its own PQQ metabolism system, such as *G. oxydans* and *Paracoccus denitrificans* (Xia et al. 2003). To solve this problem, it is essential to elucidate the PQQ regeneration system on gene level.

However, after the report of the innovative two-step fermentation process, the research on the classical two-step based one-step fermentation process seems to be suspended. Few literatures about metabolic engineering of *G. oxydans* for one-step vitamin C production could be found after then.

12.4.2 Innovative Two-Step Based One-Step Fermentation Process

Though the classical two-step fermentation process has got highly promoted results, it still needs D-sorbitol as substrate, which need additional hydrogenation step by chemical process. Therefore, the direct fermentation of D-glucose to 2-KLG is still presumed. There have been some attempts to add an additional step that could convert the D-glucose to D-sorbitol, in order to supplement the classical two-step fermentation process. However, the lack of efficient enzyme makes this impossible.

Therefore, an innovative two-step fermentation process were discovered to resolve this problem. In the new innovative two-step fermentation process, the D-glucose was converted into 2,5-diketo-gluonic acid (2,5-DKG) by *Erwinia herbicola* ATCC 21988 or some similar strains with glucose dehydrogenase, gluconate dehydrogenase

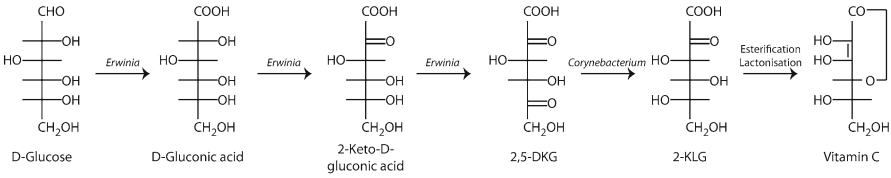


Fig. 12.4 Innovative two-step based one-step fermentation process. The D-glucose was converted into 2,5-DKG by *Erwinia* sp. The 2,5-DKG was converted into 2-KLG by *Corynebacterium* sp. The 2-KLG was then esterified and lactonized to form vitamin C. The innovative two-steps fermentation process need not chemical hydrogenation process

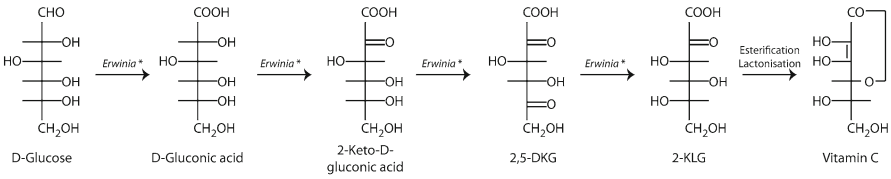


Fig. 12.5 Innovative two-step based one-step fermentation process. The D-glucose was directly converted into 2,5-DKG by a metabolic engineered *Erwinia* sp. strain containing a 2,5-DKGR from *Corynebacterium* sp. The 2-KLG was then esterified and lactonized to form vitamin C

and 2-keto-D-gluconate dehydrogenase. Then the 2,5-DKG could be transformed into 2-KLG by *Corynebacterium* ATCC 31090 or other similar strains with 2,5-DKG reductase. With this innovative two-step fermentation process, the D-glucose could be transformed into 2-KLG within two steps (Anderson et al. 1985) (Fig. 12.4). However, 2,5-DKG is highly unstable during the fermentation process. Additional sterilization could completely destroy this chemical.

Based on this innovative two-step fermentation process, Anderson et al. (1985) has expressed a 2,5-DKG reductase from *Corynebacterium* ATCC 31090 into the *Erwinia herbicola* ATCC 21988 with the expression vector ptrp1-35. The final recombinant *E. herbicola* strain could produce 1 g/L of 2-KLG from saturated D-glucose solution. By the protoplast fusion, Lin et al. (1999) fused an *Erwinia herbicola* and a *Corynebacterium* strain. The resultant strain could produce 2.07 g/L of 2-KLG (Fig. 12.5).

12.4.3 Direct Production of Vitamin C from D-Glucose

The vast majority of animals and plants are able to synthesize vitamin C through a sequence of four enzyme-driven steps, which convert D-glucose to vitamin C (Fig. 12.6). In eukaryotes, vitamin C biosynthesis occurs via two highly divergent biochemical pathways. In rats and other mammals, glucose is converted into L-gulono-1,4-lactone via a complex pathway involving uronic acid derivatives and

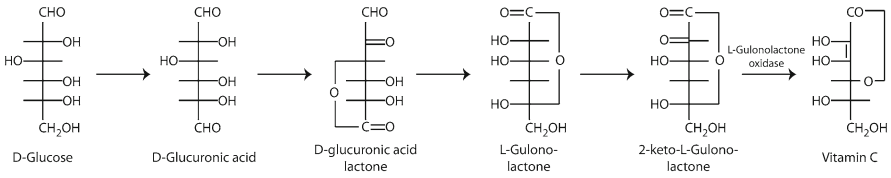


Fig. 12.6 Vitamin C production in higher eukaryotes

the 'inversion' of the original carbon skeleton (Isherwood and Cruickshank 1954; Touster 1962). The widespread occurrence of this pathway in the animal kingdom is deduced by the ubiquity of its terminal enzyme L-gulono-1,4-lactone oxidase (GUL oxidase). The gene coding for this enzyme is mutated in humans and some other animals, resulting in the loss of vitamin C synthesizing capacity.

Vitamin C biosynthesis generally exists in most of the plant cells (Zhang et al. 2007). It plays a crucial role in the detoxification of peroxide, ozone and free radicals and is essential for photosynthetic activity via the regeneration of membrane-soluble antioxidants (α-tocopherol), zeaxanthin and the pH-mediated modulation of PS II activity. Plant cells can accumulate large amount of vitamin C, particularly in green tissues and storage organs (Smirnoff 2000). Because plant cells have the whole pathway for the synthesis of vitamin C from D-galactose, attempts have been made for exploiting plant cell or microalgae cultures for the commercial production of vitamin C via fermentation.

Running et al. (2002) isolated *Chlorella pyrenoidosa* strains that were capable of producing over 70-fold more vitamin C than wild-type strains (0.64 mg/g dry cell weight for wild-type compared with 45 mg/g dry cell weight for the most-productive mutants) (Running et al. 2002). These improvements were achieved through repeated rounds of chemical mutagenesis and optimization of fermentation process conditions. Further claims were made regarding a method for the extracellular stabilization of vitamin C by reduction of medium pH (Running et al. 2002), resulting in the majority of vitamin C becoming harvestable in the fermentation medium. Yields of vitamin C microalgal cultures are not yet competitive with the traditional processes and it is likely that full economic exploitation of these systems for one-step fermentation strategies for the synthesis of vitamin C will require genetic engineering of the pathway. However, progress in the definition of the vitamin C biosynthetic pathway in plants has been singularly slow. Original models for the biosynthetic pathway in plants were based on the animal model (Isherwood and Cruickshank 1954). However, plant cells do not contain GUL oxidase and do not invert the glucose carbon skeleton during its conversion into vitamin C (Smirnoff 2000). Over the past years several pathways were proposed to accommodate these observations, but direct evidence of operation of any of these pathways has not been obtained. Only recently, with the identification of vitamin C-deficient *Arabidopsis* mutants (Conklin et al. 1996), the pathway in plants has been resolved (Wheeler et al. 1998). This made the direct fermentation production from glucose or even starch or cellulose achievable. However, because most of the genes from plants or animals had weaker

reactive dynamics characteristics, how to improve the productivity by involving these pathways from higher organisms is the key bottleneck for this route.

Furthermore, though the direct production of vitamin C from D-glucose seems to be an ideal process, vitamin C is renowned for its instability and susceptibility to oxidation. Therefore, a ventilation fermentation process may destroy most of the vitamin C. Some vitamin C derivatives, such as 2-O-alpha-D-glucopyranosyl-L-ascorbic acid (AA-2G), are relatively stable and perform almost identical physiological functions. In consequence, AA-2G is now widely used in luxury cosmetics and some of beverages. It can survive for a longer period than in subjects administered ascorbic acid 2-phosphate, which is a conventional vitamin C derivative. Therefore, the simultaneously glycosylation of vitamin C during the fermentation process may be a direct route for vitamin C production.

12.5 Perspectives

Based on the current reports, though there are many researches on the strain screening and mutagenesis, process optimization, relationships between *B. megaterium* and *K. vulgare*, metabolic engineering of *K. vulgare*, development of new two-step fermentation process to avoid mix-culture of *B. megaterium* and *K. vulgare*, development of one-step fermentation process based on both classical two-step fermentation process on new two-step fermentation process, the significant proceedings in the new processes are still sought for industrial applications for different reasons. It seems that evolution of the classical two-step fermentation process of vitamin C seems to have encountered many problems.

Based on current results, the further development of vitamin C production by microorganisms with metabolic engineering should be focused on the following area: (1) Demonstration of mutual relationships between the *B. megaterium* and *K. vulgare* and therefore enhancing cell growth and 2-KLG production by *K. vulgare*; (2) Elucidation of the genes involved in the biosynthesis and transportation of 2-KLG and its intermediate metabolites in both classical two-step and new two-step fermentation steps; (3) Investigation of the mechanisms of PQQ-dependent incompletion oxidation dehydrogenases and the regeneration of PQQ; (4) Discovery of efficient metabolic pathways for the direct fermentation of vitamin C and even its stable derivatives.

References

Ameyama M, Matsushita K, Ohno Y, Shinagawa E, Adachi O (1981) Existence of a novel prosthetic group, PQQ, in membrane-bound, electron transport chain-linked, primary dehydrogenases of oxidative bacteria. FEBS Lett 130:179–183

Anderson S, Marks CB, Lazarus R, Miller J, Stafford K, Seymour J, Light D, Rastetter W, Estell D (1985) Production of 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a genetically modified *Erwinia herbicola*. Science 230:144–149

Anthony C, Zatman LJ (1967) The microbial oxidation of methanol. The prosthetic group of the alcohol dehydrogenase of *Pseudomonas* sp. M27: a new oxidoreductase prosthetic group. Biochem J 104(3):960–969

Banta S, Swanson BA, Wu S, Jarnagin A, Anderson S (2002) Optimizing an artificial metabolic pathway: engineering the cofactor specificity of *Corynebacterium* 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis. Biochemistry 41:6226–6236

Bremus C, Herrmann U, Bringer-Meyer S, Sahm H (2006) The use of microorganisms in L-ascorbic acid production. J Biotechnol 124:196–205

Cai L, Yuan MQ, Li ZJ, Chen JC, Chen GQ (2012) Genetic engineering of *Ketogulonigenium vulgare* for enhanced production of 2-keto-L-gulonic acid. J Biotechnol 157:320–325

Chen F, Chen CS, Li Y, Yin GL (2000) Studies on gene knocking out of 2-keto aldose reductases from *Erwinia* sp. SCB125. Acta Microbiol Sin 40(5):475–481

Conklin PL, Williams EH, Last RL (1996) Environmental stress sensitivity of an ascorbic acid-deficient *Arabidopsis* mutant. Proc Natl Acad Sci U S A 93:9970–9974

Feng S, Zhang Z, Zhang CG, Zhang ZZ (2000) Effect of *Bacillus megaterium* on *Gluconobacter oxydans* in mixed culture. Chin J Appl Ecol 11:119–122

Giridhar R, Srivastava AK (2000) Fed-batch cultivation of *Acetobacter suboxydans* for the microbial oxidation of D-sorbitol to L-sorbose. Bioproc Eng 23:575–577

Giridhar R, Srivastava AK (2001) Computer coupled substrate feeding strategies for efficient conversion of D-sorbitol to L-sorbose in fed-batch culture. Proc Biochem 36:829–834

Giridhar RN, Srivastava AK (2002) Productivity improvement in L-sorbose biosynthesis by fedbatch cultivation of *Gluconobacter oxydans*. J Biosci Bioeng 94:34–38

Hauge JG (1964) Glucose dehydrogenase of bacterium anitratum: an enzyme with a novel prosthetic group. J Biol Chem 239:3630–3639

Isherwood FA, Cruickshank DH (1954) Chromatographic separation and analysis of mixtures of pyruvic, oxalacetic and α-ketoglutaric acids. Nature 173:121–122

Ji AG, Gao PJ (1998) Synthesis of 2-keto-L-gulonic acid from gluconic acid by co-immobilized *Gluconobacter oxydans* and *Corynebacterium* sp. Biotechnol Lett 20:939–942

Jiang YY, Guo ZY, Zhang CG (1997) Study on the purification of 2-keto-L-gulonate reductase and its physical, chemical and enzymic properties. Chin J Biotechnol 13:400–405

Leduc S, de Troostembergh JC, Lebeault JM (2004) Folate requirements of the 2-keto-L-gulonic acid-producing strain *Ketogulonigenium vulgare* LMP P-20356 in L-sorbose/CSL medium. Appl Microbiol Biotechnol 65:163–167

Li Q, Diao JY, Xiang BT, Cao ZA (1996) Studies on metabolism of nitrogen source in fermentation of 2-keto-L-gulonic acid. Acta Microbiol Sin 1:19–24

Lin HY, Chen CS, Yin GL (1999) Intergeneric cell fusion of *Corynebacterium* and *Erwinia*. Chin Microbiol 26:3–6

Liu LM, Chen KJ, Zhang J, Liu J, Chen J (2011a) Gelatin enhances 2-keto-L-gulonic acid production based on *Ketogulonigenium vulgare* genome annotation. J Biotechnol 156:182–187

Liu LM, Li Y, Zhang J, Zhou ZM, Liu J, Li XM, Zhou JW, Du GC, Wang L, Chen J (2011b) Complete genome sequence of the industrial strain *Ketogulonicigenium vulgare* WSH-001. J Bacteriol 193:6108–6109

Liu LM, Li Y, Zhang J, Zou W, Zhou ZM, Liu J, Li XM, Wang L, Chen J (2011c) Complete genome sequence of the industrial strain *Bacillus megaterium* WSH-002. J Bacteriol 193:6389–6390

Lu SJ, Jun W, Yao JM, Yu ZL (2003) Study on the effect of mutated *Bacillus megaterium* in two-stage fermentation of vitamin C. Plasma Sci Technol 5:2011–2016

Lv SX, Feng S, Zhang ZZ, Liu Y, Xie ZW, An HY (2001) The effect of *Bacillus megaterium* in vitamin C two-step fermentation. Acta Microbiol Sin 28:10–13

Moonmangmee D, Adachi O, Ano Y, Shinagawa E, Toyama H, Theeragool G, Lotong N, Matsushita K (2000) Isolation and characterization of thermotolerant Gluconobacter strains catalyzing oxidative fermentation at higher temperatures. Biosci Biotechnol Biochem 64:2306–2315

Petrescu S, Hulea SA, Stan R, Avram D, Herlea V (1992) A yeast strain that uses D-galacturonic acid as a substrate for L-ascorbic acid biosynthesis. Biotechnol Lett 14:1–6

Rao YM, Sureshkumar GK (2000) Direct biosynthesis of ascorbic acid from glucose by *Xanthomonas campestris* through induced free-radicals. Biotechnol Lett 22:407–411

Reichstein T, Grussner A (1934) A productive synthesis of L-arcorbic acid (C-vitamin). Helv Chim Acta 17:311–328

Running JA, Severson DK, Schneider KJ (2002) Extracellular production of L-ascorbic acid by *Chlorella protothecoides*, *Prototheca species*, and mutants of *P. moriformis* during aerobic culturing at low pH. J Ind Microbiol Biotechnol 29(2):93–98

Saito Y, Ishii Y, Hayashi H, Imao Y, Akashi T, Yoshikawa K, Noguchi Y, Soeda S, Yoshida M, Niwa M, Hosoda J, Shimomura K (1997) Cloning of genes coding for L-sorbose and L-sorbosone dehydrogenases from *Gluconobacter oxydans* and microbial production of 2-Keto-L-Gulonate, a precursor of L-ascorbic acid, in a recombinant *G-oxydans* strain. Appl Environ Microbiol 63:454–460

Saito Y, Ishii Y, Hayashi H, Yoshikawa K, Noguchi Y, Yoshida S, Soeda S, Yoshida M (1998) Direct fermentation of 2-keto-L-gulonic acid in recombinant *Gluconobacter oxydans*. Biotechnol Bioeng 58:309–315

Salisbury SA, Forrest HS, Cruse WB, Kennard O (1979) A novel coenzyme from bacterial primary alcohol dehydrogenases. Nature 280:843–844

Sanli G, Blaber SI, Blaber M (2001) Reduction of wobble-position GC bases in *Corynebacteria* genes and enhancement of PCR and heterologous expression. J Mol Microbiol Biotechnol 3:123–126

Sauer M, Branduardi P, Valli M, Porro D (2004) Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. Appl Environ Microbiol 70:6086–6091

Shinjoh M, Tomiyama N, Asakura A, Hoshino T (1995) Cloning and nucleotide sequencing of the membrane-bound L-sorbosone dehydrogenase gene of *Acetobacter liquefaciens* IFO 12258 and its expression in *Gluconobacter oxydans*. Appl Environ Microbiol 61:413–420

Smirnoff N (2000) Ascorbic acid: metabolism and functions of a multi-facetted molecule. Curr Opin Plant Biol 3:229–235

Sonoyama T, Tani H, Matsuda K, Kageyama B, Tanimoto M, Kobayashi K, Yagi S, Kyotani H, Mitsushima K (1982) Production of 2-keto-L-gulonic acid from D-glucose by two-stage fermentation. Appl Environ Microbiol 43:1064–1069

Srivastava AK, Lasrado PR (1998) Fed-batch sorbitol to sorbose fermentation by *A suboxydans*. Bioproc Eng 18(6):457–461

Sugisawa T, Hoshino T, Masuda S, Nomura S, Setoguchi Y, Tazoe M, Shinjoh M, Someha S, Fujiwara A (1990) Microbial production of 2-keto-L-gulonic acid from L-sorbose and D-sorbitol by *Gluconobacter melanogenus*. Agr Biol Chem 54:1201–1209

Sugisawa T, Miyazaki T, Hoshino T (2005) Microbial production of L-ascorbic acid from D-sorbitol, L-sorbose, L-gulose, and L-sorbosone by *Ketogulonicigenium vulgare* DSM 4025. Biosci Biotechnol Biochem 69:659–662

Sybesma W, Starrenburg M, Kleerebezem M, Mierau I, de Vos WM, Hugenholtz J (2003) Increased production of folate by metabolic engineering of *Lactococcus lactis*. Appl Environ Microbiol 69:3069–3076

Touster O (1962) Carbohydrate metabolism. Ann Rev Biochem 31:407–450

Tsukada Y, Perlman D (1972) The fermentation of L-sorbose by *Gluconobacter melanogenus*. I. General characteristics of the fermentation. Biotechnol Bioeng 14:799–810

Westerling J, Frank J, Duine JA (1979) The prosthetic group of methanol dehydrogenase from Hyphomicrobium X: electron spin resonance evidence for a quinone structure. Biochem Biophys Res Commun 87:719–724

Wheeler GL, Jones MA, Smirnoff N (1998) The biosynthetic pathway of vitamin C in higher plants. Nature 393:365–369

Xia ZX, Dai WW, He YN, White SA, Mathews FS, Davidson VL (2003) X-ray structure of methanol dehydrogenase from *Paracoccus denitrificans* and molecular modeling of its interactions with cytochrome c-551i. J Biol Inorg Chem 8:843–854

Xiong XH, Han S, Wang JH, Jiang ZH, Chen W, Jia N, Wei HL, Cheng H, Yang YX, Zhu B, You S, He JY, Hou W, Chen MX, Yu CJ, Jiao YH, Zhang WC (2011) Complete genome sequence of the bacterium *Ketogulonicigenium vulgare* Y25. J Bacteriol 193:315–316

Xu A, Yao J, Yu L, Lv S, Wang J, Yan B, Yu Z (2004) Mutation of *Gluconobacter oxydans* and *Bacillus megaterium* in a two-step process of L-ascorbic acid manufacture by ion beam. J Appl Microbiol 96:1317–1323

Yan B, Xu A, Zhang W, Zhou W, Wang J, Yao JM, Yu ZL (2006) Accumulation of 2-keto-L-gulonate at 33 degrees C by a thermotolerant *Gluconobacter oxydans* mutant obtained by ion beam implantation. Plasma Sci Technol 8:237–241

Yang FC, Lim YH (1997) Kinetic study of the bioconversion of D-sorbitol to L-sorbose by *Acetobacter pasteurianus*. Proc Biochem 32:233–236

Yang F, Jia Q, Xiong Z, Zhang X, Wu H, Zhao Y, Yang J, Zhu J, Dong J, Xue Y, Sun L, Shen Y, Jin Q (2006) Complete genome analysis of *Ketogulonigenium* sp. WB0104. Chin Sci Bull 51(8):941–945

Yang XP, Zhong GF, Lin JP, Mao DB, Wei DZ (2010) Pyrroloquinoline quinone biosynthesis in *Escherichia coli* through expression of the *Gluconobacter oxydans pqqABCDE* gene cluster. J Ind Microbiol Biotechnol 37:575–580

Yin GL, He JM, Ren SX, Song Q, Ye Q, Lin YH, Chen CS, Guo XY (1997) Production of vitamin C precursor-2-leto-L-gulonic acid from L-sorbose by a novel bacterial component system of SCB329-SCB933. Ind Microbiol 27:1–7

Zhang L, Wang Z, Xia Y, Kai G, Chen W, Tang K (2007) Metabolic engineering of plant L-ascorbic acid biosynthesis: recent trends and applications. Crit Rev Biotechnol 27:173–182

Zhang J, Liu J, Shi ZP, Liu LM, Chen J (2010) Manipulation of *B-megaterium* growth for efficient 2-KLG production by *K-vulgare*. Process Biochem 45:602–606

Zhang J, Zhou JW, Liu J, Chen KJ, Liu LM, Chen J (2011) Development of chemically defined media supporting high cell density growth of *Ketogulonicigenium vulgare* and *Bacillus megaterium*. Bioresour Technol 102:4807–4814

Zhao SG, Yao LM, Su CX, Wang T, Wang J, Tang ML, Yu ZL (2008) Purification and properties of a new L-sorbose dehydrogenase accelerative protein from *Bacillus megaterium* bred by ion-beam implantation. Plasma Sci Technol 10:398–402

Zhou B, Li Y, Liu YP, Zhang ZZ, Zhu KL, Liao DM, Gao YT (2002) Microbiological eco-regulation in Vc two-step fermentation. Chin J Appl Ecol 13:1452–1454

Zhou J, Ma Q, Yi H, Wang LL, Song H, Yuan YJ (2011) Metabolome profiling reveals metabolic cooperation between *Bacillus megaterium* and *Ketogulonicigenium vulgare* during induced swarm motility. Appl Environ Microbiol 77:7023–7030

Zhu YB, Liu J, Du GC, Zhou JW, Chen J (2012) Sporulation and spore stability of *Bacillus megaterium* enhance *Ketogulonigenium vulgare* propagation and 2-keto-L-gulonic acid biosynthesis. Bioresour Technol 107:399–404

Chapter 13
Molecular Mechanisms and Metabolic Engineering of Glutamate Overproduction in *Corynebacterium glutamicum*

Takashi Hirasawa, Jongpill Kim, Tomokazu Shirai, Chikara Furusawa, and Hiroshi Shimizu

Abstract Glutamate is a commercially important chemical. It is used as a flavor enhancer and is a major raw material for producing industrially useful chemicals. A coryneform bacterium, *Corynebacterium glutamicum*, was isolated in 1956 by Japanese researchers as a glutamate-overproducing bacterium and since then, remarkable progress in glutamate production has been made using this microorganism. Currently, the global market for glutamate is over 2.5 million tons per year. Glutamate overproduction by *C. glutamicum* is induced by specific treatments—biotin limitation, addition of fatty acid ester surfactants such as Tween 40, and addition of β-lactam antibiotics such as penicillin. Molecular biology and metabolic engineering studies on glutamate overproduction have revealed that metabolic flow is significantly altered by these treatments. These studies have also provided insight into the molecular mechanisms underlying these changes. In this chapter, we review our current understanding of the molecular mechanisms of glutamate overproduction in *C. glutamicum*, and we discuss the advances made by metabolic engineering of this microorganism.

T. Hirasawa • H. Shimizu (✉)
Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: shimizu@ist.osaka-u.ac.jp

J. Kim
Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

T. Shirai
RIKEN Biomass Engineering Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

C. Furusawa
Quantitative Biology Center, RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_13,

Keywords Anaplerotic pathways • *Corynebacterium glutamicum* • Glutamate • Metabolic engineering • NCgl1221 • OdhI • 2-Oxoglutarate dehydrogenase complex

Abbreviations

^{13}C-MFA	^{13}C metabolic flux analysis
CoA	coenzyme A
FBA	flux balance analysis
FHA	forkhead-associated
GC/MS	gas chromatography/mass spectrometry
GDH	glutamate dehydrogenase
ICDH	isocitrate dehydrogenase
K_m	Michaelis-Menten constant
MFA	metabolic flux analysis
MSG	mono sodium glutamate
NMR	nuclear magnetic resonance
ODHC	2-oxoglutarate dehydrogenase complex
PC	pyruvate carboxylase
PEPC	phosphoenolpyruvate carboxylase
TCA	tricarboxylic acid

13.1 Introduction

In 1908, Dr. Kikunae Ikeda identified monosodium glutamate (MSG) as the compound imparting the taste of Umami to numerous foods. Umami is distinct from other tastes such as bitter, sour, salty and sweet. Ikeda and his colleagues began to industrially produce MSG by subjecting wheat protein gluten to acid hydrolysis. Ajinomoto Co. Inc. was the first company to produce MSG on an industrial scale.

In 1956, Japanese researchers at Kyowa Hakko Kogyo Co. Ltd. (currently, Kyowa Hakko Bio Co. Ltd.) isolated a microorganism secreting high amounts of glutamate (Kinoshita et al. 1957; Udaka 1960). This microorganism was initially named *Micrococcus glutamicus* but was renamed *Corynebacterium glutamicum*. Using *C. glutamicum*, a fermentation process to produce glutamate directly from sugar and ammonia was developed. At present, *C. glutamicum* is widely used as a host species for producing glutamate and other amino acids such as lysine, arginine, threonine, and valine (Nakayama et al. 1961; Nakayama and Yoshida 1972; Sano and Shiio 1970; Shiio and Nakamori 1970; Tsuchida et al. 1975).

C. glutamicum is a facultatively anaerobic, rod-shaped, high G+C Gram-positive bacterium (Fig. 13.1). Recently, this microorganism has attracted attention as a host for producing useful compounds such as lactate and succinate as well as amino

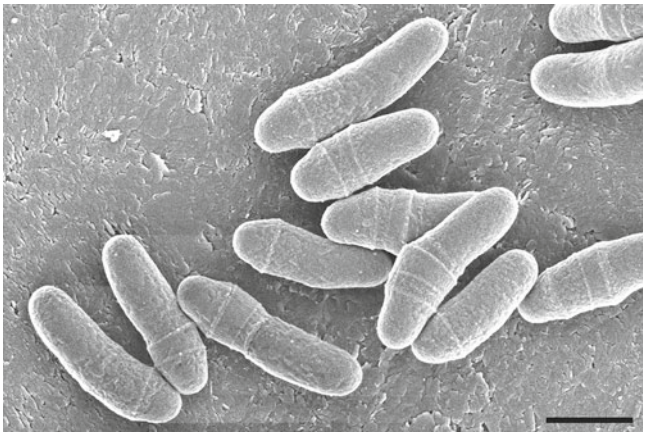


Fig. 13.1 Scanning electron micrograph of *C. glutamicum*. The *bar* represents 1 μm

acids (Okino et al. 2005; Wendisch et al. 2006). It has also been employed for secreted protein production (Kikuchi et al. 2003; Umakoshi et al. 2011). The complete genomic DNA sequence of *C. glutamicum* has been determined (Ikeda and Nakagawa 2003; Kalinowski et al. 2003; Yukawa et al. 2007). A related species *Corynebacterium efficiens* was also isolated as a glutamate producing bacterium under high temperature growth conditions (Fudou et al. 2002). The complete genome sequence of this species is also available (Nishio et al. 2003).

Glutamate has been used as a flavor enhancer and a raw material for producing useful chemicals for many years. The global market of glutamate is the largest among the amino acids (Shimizu and Hirasawa 2007) and is over 2.5 million tons per year (Becker and Wittmann 2012).

Following the discovery of *C. glutamicum*, numerous molecular biology and metabolic engineering studies for glutamate overproduction have been conducted. In this section, we review recent progress in the metabolic engineering of glutamate overproduction by *C. glutamicum* and discuss our current understanding of the underlying molecular mechanisms.

13.2 Glutamate Production by *C. glutamicum*

Glutamate is synthesized from 2-oxoglutarate in the tricarboxylic acid (TCA) cycle by glutamate dehydrogenase (GDH) reaction, which is the main pathway for glutamate biosynthesis when nitrogen supply is sufficient.

C. glutamicum requires biotin (vitamin B7) for growth. Wild-type *C. glutamicum* does not secrete glutamate when excess biotin is present in the culture medium. However, when biotin is limiting, *C. glutamicum* can produce large amounts of glutamate (Shiio et al. 1962). Even in the presence of excess biotin,

certain treatments can induce glutamate overproduction by *C. glutamicum*—addition of fatty acid ester surfactants such as polyoxyethylene sorbitan monopalmitate (Tween 40) and polyoxyethylene sorbitan monostearate (Tween 60) (Takinami et al. 1965), addition of β-lactam antibiotics such as penicillin (Nara et al. 1964), and addition of the antimycobacterial agent ethambutol (Radmacher et al. 2005). These treatments affect cell surface structures including the cytoplasmic membrane, cell wall peptidoglycan layer and outer mycolic acid-containing layer of *C. glutamicum*; biotin is a cofactor for the enzyme for fatty acid biosynthesis. Similarly, additions of fatty acid ester surfactants, β-lactam antibiotics and ethambutol affect fatty acid biosynthesis, peptidoglycan biosynthesis and formation of mycolic acid-containing layer, respectively. Therefore, it was thought that overproduced glutamate passively leaked out of *C. glutamicum* due to the increased cell surface permeability. However, this leak model of glutamate production cannot account for the high glutamate production observed (more than 60–80 g/L), based on the differences between intracellular and extracellular glutamate levels.

Shigu and Terui (1971) reported a change in the enzyme activity of the 2-oxoglutarate dehydrogenase complex (ODHC), which is located at the branch point between the TCA cycle and the glutamate biosynthesis pathway, during glutamate production by *C. glutamicum*. Kinoshita (1985) proposed that glutamate production by *C. glutamicum* is regulated by the ODHC activity. Later, Kawahara and colleagues found that ODHC activity decreases during glutamate overproduction induced by biotin limitation, Tween 40 addition and penicillin addition (Kawahara et al. 1997). Moreover, the importance of a balanced supply of acetyl-CoA and oxaloacetate has been emphasized (Hasegawa et al. 2008), particularly for oxaloacetate-supplying anaplerotic reactions, which are important for glutamate production (Sato et al. 2008; Shirai et al. 2007). Very recently, a novel protein, OdhI, was identified as a protein contributing to the decrease in ODHC activity by interacting with OdhA, a subunit of the ODHC. In addition, phosphorylation status of OdhI was found to influence its ability to decrease in ODHC activity and affect glutamate production (Kim et al. 2010, 2011; Niebisch et al. 2006; Schultz et al. 2007). Furthermore, a mechanosensitive channel, NCgl1221, was found to be important for glutamate overproduction by *C. glutamicum* (Nakamura et al. 2007).

13.3 Molecular Mechanisms of Glutamate Overproduction in *C. glutamicum*

Molecular biology studies of glutamate overproduction by *C. glutamicum* have been focused on cell surface structures and metabolic flux changes and have revealed some important factors for glutamate overproduction (Fig. 13.2). Particularly, studies on DtsR1, NCgl1221, and OdhI have contributed to our understanding of the mechanism of glutamate overproduction. In this section, the roles of these important proteins in glutamate overproduction are reviewed. Moreover, we discuss the

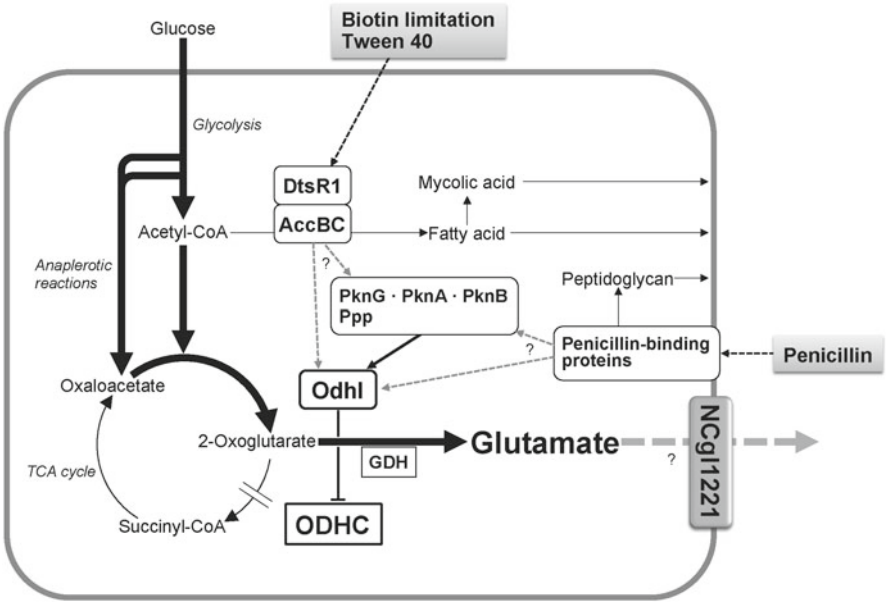


Fig. 13.2 Possible molecular mechanisms of glutamate overproduction by *C. glutamicum* involving DtsR1, OdhI and NCgl1221

relationship between cell surface structure and glutamate production, as well as genome-wide analyses of glutamate overproduction.

13.3.1 Role of DtsR1 in Glutamate Overproduction

As described above, glutamate overproduction in *C. glutamicum* is achieved by addition of Tween 40 to the culture medium. Kimura et al. (1996) analyzed the Tween 40-sensitive mutant of *C. glutamicum* and cloned a gene, *dtsR1*, which rescued the Tween 40-sensitivity. The *dtsR1* gene product shows high homology with the β subunit of propionyl-coenzyme A (CoA) carboxylase from other microorganisms, suggesting that it is a subunit of biotin-containing fatty acid biosynthesis enzyme complex, i.e. the acyl-CoA carboxylase complex (AccBC-DtsR1).

The *dtsR1*-disrupted strain shows auxotrophy for oleic acid and its ester (Tween 80). It can produce glutamate in the presence of excess biotin (Kimura et al. 1997) and exhibits decreased ODHC activity compared with the wild-type strain. Moreover, the intracellular level of DtsR1 is decreased by biotin limitation and Tween 40 addition, but is not changed by penicillin addition (Kimura et al. 1999). These results suggest that the target of biotin limitation and Tween 40 treatment is the acyl-CoA carboxylase complex, including DtsR1, and these treatments seem to inhibit fatty acid biosynthesis in *C. glutamicum*.

13.3.2 Role of the Mechanosensitive Channel Encoded by the NCgl1221 Gene in Glutamate Overproduction and Secretion

Many researchers have sought to understand the mechanisms of glutamate overproduction and its secretion. Until the early 1990s, it was thought that passive leakage of glutamate through the membrane and cell wall was the main mode of glutamate secretion induced by triggers mentioned above. In 2007, however, possible glutamate exporter protein encoded by NCgl1221 gene was identified.

It was reported that disruption of the *odhA* gene, encoding the catalytic subunit of ODHC, enhances glutamate production without the need for induction treatments (Asakura et al. 2007). However, some *odhA* disruptants do not produce glutamate, probably due to genetic instability of the disruptants. Nakamura et al. (2007) found that glutamate production in some *odhA* disruptants is not induced by *odhA* disruption, but is caused by other mutation in the gene NCgl1221. The NCgl1221 gene product shows high homology with mechanosensitive channels, such as the *Escherichia coli* YggB protein (Levina et al. 1999). Mechanosensitive channels sense changes in membrane tension and mediate adaptation to changes in osmotic pressure (Berrier et al. 1992). Electrophysiological analyses using NCgl1221-expressing bacteria revealed that NCgl1221 protein is, in fact, a mechanosensitive channel (Börngen et al. 2010; Hashimoto et al. 2010).

The specific mutations in NCgl1221, which probably involve in the conformational changes induced by various triggers, lead to glutamate overproduction without requiring *odhA* disruption. Moreover, disruption of NCgl1221 abolishes glutamate secretion due to the increase in intracellular glutamate level and NCgl1221 overexpression increases glutamate production with the need for triggers, i.e. biotin limitation, Tween 40 addition, or penicillin treatment. These results suggest that the activation of the NCgl1221 due to the changes in membrane tension induced by triggers is important for glutamate overproduction in *C. glutamicum*, and that NCgl1221 protein is a possible glutamate exporter.

13.3.3 OdhI and Decreased ODHC Activity During Glutamate Overproduction

As described in Sect. 13.2, the activity of the ODHC is decreased during glutamate overproduction induced by biotin limitation, Tween 40 addition, and penicillin treatment (Kawahara et al. 1997). However, the molecular mechanism responsible for the decrease in ODHC activity during glutamate production was unknown until 2006, when a crucial protein was identified.

Niebisch et al. (2006) identified the protein OdhI based on an analysis of a *C. glutamicum* mutant for *pknG*, encoding a serine/threonine protein kinase. OdhI is a 15-kDa protein and is a substrate for PknG serine/threonine kinase. The OdhI protein

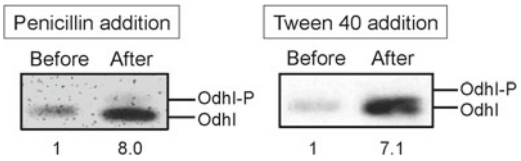


Fig. 13.3 Phosphorylation status of OdhI protein during penicillin- and Tween 40-induced glutamate overproduction in *C. glutamicum*. Phosphorylation status of OdhI protein before and after penicillin and Tween 40 addition was analyzed by western blotting with rabbit polyclonal OdhI antiserum. Values bellows the OdhI protein bands indicate the amount of unphosphorylated OdhI relative to that before penicillin and Tween 40 addition

has a forkhead-associated (FHA) domain, which binds to phosphothreonine epitopes on proteins and mediates interactions with other proteins. Two threonine residues Thr14 and Thr15, in OdhI are phosphorylated by PknG. Other protein kinases—PknA, PknB and PknL—can also phosphorylate the OdhI protein (Schultz et al. 2009). Phosphorylated OdhI is dephosphorylated by the Ppp phosphatase (Schultz et al. 2007, 2009). OdhI interacts with the OdhI protein, which is a catalytic subunit of ODHC, and unphosphorylated OdhI can inhibit ODHC activity by this direct interaction.

Barthe et al. (2009) solved the structure of the phosphorylated and unphosphorylated forms of OdhI. The phosphothreonine residue in OdhI interacts with its own FHA domain, and as a result, phosphorylated OdhI cannot interact with OdhA to inhibit ODHC activity. In contrast, the FHA domain in unphosphorylated OdhI can freely interact with OdhA, and thereby inhibit ODHC activity. Biochemical analyses revealed that the FHA domain of unphosphorylated OdhI interacts with the C-terminal dehydrogenase domain of OdhA (Krawczyk et al. 2010).

The deletion of *odhI* abolishes glutamate production in *C. glutamicum* triggered by biotin limitation, Tween 40 addition and penicillin addition (Schultz et al. 2007). Moreover, the phosphorylation status of OdhI was analyzed using western blotting. Boulahya et al. (2010) reported that OdhI is dephosphorylated under biotin limitation condition. Kim et al. (2011) analyzed Tween 40- and penicillin-induced glutamate overproduction in *C. glutamicum*, and suggested that the unphosphorylated OdhI increases substantially following both triggers (Fig. 13.3). These results suggest that OdhI has an important role in glutamate production by *C. glutamicum* induced by biotin limitation, Tween 40 addition and penicillin treatment, and that unphosphorylated form of OdhI is abundant during glutamate overproduction.

13.3.4 Relationship Between Cell Surface Structure and Glutamate Production by C. glutamicum

As described above, glutamate overproduction by *C. glutamicum* is induced by biotin limitation, fatty acid ester surfactant addition, and penicillin treatment. Biotin limitation and addition of fatty acid esters affect fatty acid biosynthesis in

C. glutamicum. Penicillin is a β-lactam antibiotic that binds to penicillin-binding proteins and inhibits the peptidoglycan biosynthesis. These treatments affect the cell surface integrity of *C. glutamicum*. Thus, it was thought that glutamate leaked passively through the membrane. Consequently, some relationship between cell surface state changes and glutamate overproduction were investigated.

In *C. glutamicum*, the cytoplasmic membrane is covered with a thick peptidoglycan layer, which is, in turn, surrounded by an arabinogalactan layer. The peptidoglycan of *C. glutamicum* consists of an L-alanyl-D-glutamyl-*meso*-diaminopimelyl-D-alanine peptide with β-1,4 linked *N*-acetylglucosamine and *N*-acetylmuramic acid containing side chains, as in *E. coli*. It should be noted that the *C. glutamicum* cell wall contains long chain fatty acids, called mycolic acids, similar to Mycobacteria. Mycolic acids are α-alkyl-β-hydroxylated fatty acids, i.e. R_1–CH(OH)–CH(R_2)–COOH (R_1 and R_2 represent alkyl chains). The alkyl chain lengths of mycolic acids in *C. glutamicum* (32–36) are shorter than that in Mycobacteria (50–56).

Hoischen and Kramer (1990) analyzed the relationship between membrane structure and glutamate production. The total amount of lipids including phospholipids is reduced, and the ratio of saturated to unsaturated fatty acids is changed under biotin limitation. Nampoothiri et al. (2002) examined the impact of overexpression or deletion of the genes for lipid or fatty acid biosyntheses genes on glutamate production by *C. glutamicum*, but no significant correlation between phospholipid composition and glutamate efflux was found.

Analyses of lysozyme-sensitive mutants of *C. glutamicum* have also been performed. Since *C. glutamicum* exhibits tolerance to the lytic enzyme lysozyme, due to the presence of mycolic acid-containing layer, lysozyme-sensitive mutants of *C. glutamicum* are expected to have defects in cell surface structure. Hirasawa et al. (2000) cloned the *ltsA* gene, which complements the temperature-sensitive growth and lysozyme-sensitivity of the *C. glutamicum* mutant strain KY9714. The *ltsA* gene product exhibits high homology with *purF*-type glutamine-dependent asparagine synthetases, which belongs to the glutamine-dependent amidotransferases, of various organisms. However, *ltsA* cannot complement asparagine auxotrophy of *E. coli asnA asnB* double mutant, suggesting that the LtsA protein is a novel glutamine-dependent amidotransferase involved in formation of cell surface structure, particularly the mycolic acid-containing layer. Moreover, the *ltsA* mutant strains can produce glutamate by elevating culture temperature, suggesting that the defects in cell surface structure by *ltsA* mutations can elicit glutamate overproduction (Hirasawa et al. 2000, 2001).

Hashimoto et al. (2006) examined the relationship between mycolic acid layer formation and glutamate overproduction in *C. glutamicum*. The cellular content of mycolic acid decreased following treatment with the various triggers, and the content of short chain mycolic acids increased under biotin limitation, indicating that formation of the mycolic acid layer is perturbed by treatments that induce glutamate overproduction. It is thought that the mycolic acid layer functions as a permeability barrier to glutamate secretion. Thus, changes in the mycolic acid layer during glutamate production might alter membrane tension, which is sensed by the mechanosensitive channel NCgl1221.

13.3.5 *Genome-Wide Analyses of Glutamate Overproduction Mechanism in C. glutamicum*

Genome-wide analyses including transcriptomic and proteomic analyses during glutamate overproduction have been conducted by a number of research groups. Kataoka et al. (2006) reported the transcriptomic analysis of *C. glutamicum* during glutamate overproduction by biotin limitation, Tween 40 addition, and penicillin treatment, using DNA microarray. The results of DNA microarray analysis indicated that the genes related to glycolysis, the pentose phosphate pathway, and the TCA cycle are downregulated. In particular, the expression of *odhA* and *sucB* encoding subunits of the ODHC is reduced. In contrast, the genes NCgl2944, NCgl2945, NCgl2946, and NCgl2975 are upregulated during glutamate overproduction, but the functions of proteins encoded by these genes are unknown.

Proteomic analysis during glutamate production by *C. glutamicum* was performed by Kim et al. (2010). They examined the effect of chloramphenicol, an antibiotic that inhibit *de novo* protein synthesis, on penicillin-induced glutamate overproduction by *C. glutamicum*, and found that protein synthesis within 4 h after penicillin addition is required for glutamate production. To identify the proteins whose synthesis was induced by penicillin addition, proteomic analysis (2-dimensional gel electrophoresis) was performed. The analysis indicated that penicillin increased the expression of 13 proteins including the OdhI protein. Moreover, *odhI* overexpression induced glutamate overproduction without requiring triggers. These results indicate that OdhI synthesis is required for penicillin-induced glutamate overproduction.

13.4 Metabolic Engineering of Glutamate Overproduction in *C. glutamicum*

Metabolic engineering can be defined as the directed improvement of product formation or cellular properties through the modification of intracellular biochemical reaction(s) in complex metabolic networks (Stephanopoulos et al. 1998). To improve production of target product in bioprocesses using cells, both genetic modification of metabolic pathways and process operation strategies are important. In metabolic engineering, a system for evaluating cellular metabolism after genetic manipulation is crucial, and is called metabolic flux analysis (MFA). Metabolic flux is defined as the rate of intracellular biochemical reaction per unit cell. MFA is conducted based on measurements of extracellular metabolic reaction rates or based on ^{13}C isotope labeling experiments and measurement of ^{13}C enrichment in metabolites using nuclear magnetic resonance (NMR) spectroscopy, gas chromatography/mass spectrometry (GC/MS), etc. Based on measurement of ^{13}C enrichment in metabolites (^{13}C-MFA), metabolic flux can be precisely determined. For example, the metabolic fluxes for forward and reverse reactions of reversible reactions can be

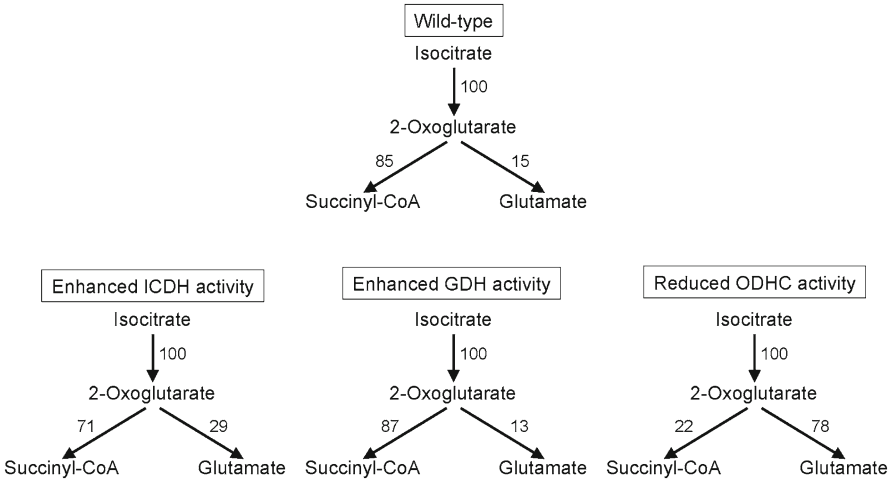


Fig. 13.4 Impact of change in activities of ICDH, GDH and ODHC on metabolic flux redistribution at the 2-oxoglutarate branch in *C. glutamicum*. Glucose was used as a carbon source in this experiment. Input flux to 2-oxoglutarate from isocitrate was normalized to 100, and redistributed fluxes to succinyl-CoA and glutamate from 2-oxoglutarate are shown

individually determined. In this section, we review *in vivo* and *in silico* MFA of glutamate production by *C. glutamicum*. We also discuss metabolic engineering of glutamate overproduction in *C. glutamicum*.

13.4.1 Metabolic Flux Analysis at the 2-Oxoglutarate Branch in the TCA Cycle During Glutamate Production

MFA has been performed on *C. glutamicum* during glutamate overproduction under biotin limitation. Here, MFA was performed based on the extracellular reaction rates including sugar consumption, glutamate production, and CO_2 evolution, and the flux redistribution around the 2-oxoglutarate branch in the TCA cycle during glutamate production was examined.

The impact of changes in enzyme activity at the 2-oxoglutarate branch on glutamate production by *C. glutamicum* was assessed based on MFA (Shimizu et al. 2003) (Fig. 13.4). Metabolic flux distributions in the *icd*- and *gdh*-overexpressing *C. glutamicum* recombinant strains under sufficient biotin supply were determined. The *icd* gene encodes isocitrate dehydrogenase (ICDH), which converts isocitrate to 2-oxoglutarate and *gdh* encodes GDH, which converts 2-oxoglutarate to glutamate (Börmann et al. 1992; Eikmanns et al. 1995). A 3.0-fold enhancement of ICDH activity was observed in the *icd*-overexpressing strain, and a 3.2-fold increase in GDH activity was achieved in the *gdh*-overexpressing strain. In addition, the impact of decreased ODHC activity under biotin limitation on metabolic flux redistribution at the 2-oxoglutarate branch

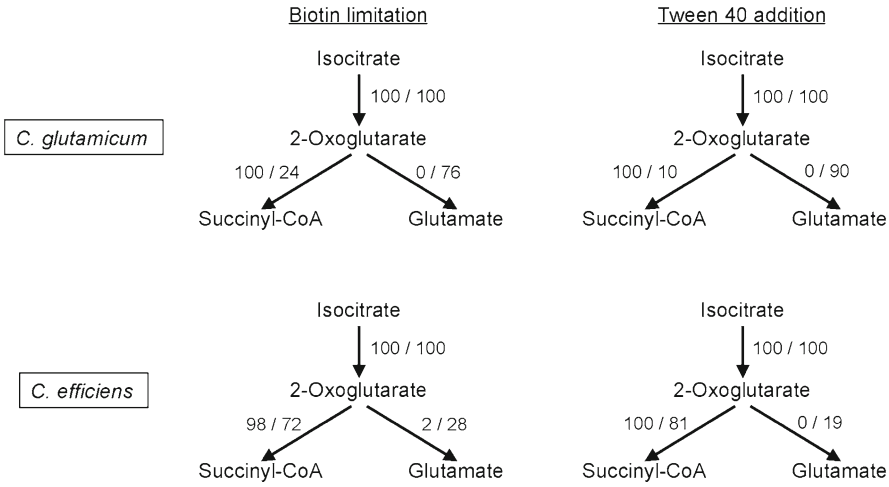


Fig. 13.5 Comparison of metabolic flux redistributions at the 2-oxoglutarate branch between *C. glutamicum* and *C. efficiens* in glutamate overproduction under biotin limitation and Tween 40 addition. Glucose was used as a carbon source in this experiment. Input flux to 2-oxoglutarate from isocitrate was normalized to 100, and redistributed fluxes to succinyl-CoA and glutamate from 2-oxoglutarate are shown

was also analyzed. When grown with sufficient biotin, the elevated ICDH and GDH activities had little impact on glutamate production by *C. glutamicum*—approximately 70% of carbon flux still flowed to the TCA cycle. In contrast, significant flux changes for GDH and ODHC were observed when the ODHC activity was decreased by biotin limitation—78% of carbon was utilized for glutamate formation. These results indicate that the ODHC activity has a great impact on glutamate production by *C. glutamicum*.

MFA of another coryneform bacterium, *C. efficiens*, was also performed (Shirai et al. 2005). *C. efficiens* was isolated by Ajinomoto Co., Ltd. and can grow at higher temperatures than *C. glutamicum*. Therefore, cooling during fermentation process is not necessary, and as a result, the costs of production can be decreased by using this bacterium. *C. efficiens* can produce glutamate following biotin limitation and Tween 40 addition. Thus, MFA of *C. efficiens* was performed and flux redistribution at the 2-oxoglutarate branch was compared with that of *C. glutamicum* during glutamate production (Fig. 13.5). Glutamate production in *C. glutamicum* was larger than that in *C. efficiens* under biotin limitation and Tween 40 addition. In both species, activities of ICDH and GDH were not affected by biotin limitation or Tween 40 addition, whereas ODHC activity was significantly decreased. Flux redistribution occurred after ODHC activity was reduced by biotin limitation and Tween 40 addition in both species. However, the carbon flow toward glutamate production depended on the magnitude of the decrease in ODHC activity—the carbon flow to glutamate production was larger in *C. glutamicum* than in *C. efficiens*, and was consistent with the difference in glutamate production between the two species. Moreover, the

Michaelis-Menten constants (K_m) for ICDH and ODHC to 2-oxoglutarate were lower than that of GDH, suggesting that the affinity of 2-oxoglutarate for GDH was lower than for ICDH and ODHC. Thus, the accumulation of 2-oxoglutarate mediated by decreased ODHC activity is important for glutamate overproduction in both species. In addition, the K_m of GDH was lower in *C. glutamicum* than in *C. efficiens*, and this phenomenon may underlie the difference in glutamate production levels between the two species.

13.4.2 Analyses of Importance of Anaplerotic Reactions During Glutamate Overproduction Based on ^{13}C Metabolic Flux Analysis

For glutamate production, a balanced supply of acetyl-CoA and oxaloacetate is important to proceed TCA cycle reactions toward 2-oxoglutarate, because it is a substrate for glutamate biosynthesis catalyzed by GDH. In particular, anaplerotic reactions, which supply oxaloacetate from the glycolytic intermediates, are crucial. *C. glutamicum* possesses two enzymes for anaplerotic reactions; phosphoenolpyruvate carboxylase (PEPC) encoded by *ppc* and pyruvate carboxylase (PC) encoded by *pyc* (Börmann et al. 1992; Eikmanns et al. 1995; O'Regan et al. 1989; Peters-Wendisch et al. 1998). PC requires biotin as a cofactor for its activity. MFA was conducted to understand the role of these anaplerotic reactions in glutamate production.

Sato et al. (2008) genetically and metabolically analyzed the roles of PEPC and PC on glutamate production by *C. glutamicum* under biotin limitation using the *ppc* and *pyc* disruptants. The *pyc* disruptant produced high amounts of glutamate under biotin limitation, but lactate was simultaneously produced. In contrast, the *ppc* disruptant could not produce glutamate under biotin limitation. MFA, based on the ^{13}C-labeling experiment and measurement of ^{13}C enrichment in glutamate using NMR spectroscopy, revealed that the flux for anaplerotic reactions in the *pyc* disruptant was lower than in the wild-type under biotin limitation, whereas the flux for lactate production was higher. Disruption of *ldh* encoding lactate dehydrogenase, which is involved in lactate production, enhanced glutamate production in the *pyc* disruptant of *C. glutamicum*. These results indicate that the PEPC reaction is important for glutamate production under biotin limitation, because PC is a biotin-containing enzyme that does not function under biotin limitation.

Shirai et al. developed an analysis system for precise ^{13}C-MFA for coryneform bacteria, and examined the role of anaplerotic reactions in Tween 40-induced glutamate overproduction by *C. glutamicum* (Shirai et al. 2006, 2007) (Fig. 13.6). Metabolic reaction model of *C. glutamicum* considering the reversibility of the reversible reaction and two anaplerotic reactions was constructed, and ^{13}C-MFA was performed based on time course data of ^{13}C enrichment of proteinogenic amino acids measured by GC/MS during cell growth and glutamate production phases. In this ^{13}C-MFA, metabolic flux distribution is tuned as the measured GC/MS data agreed with the GC/

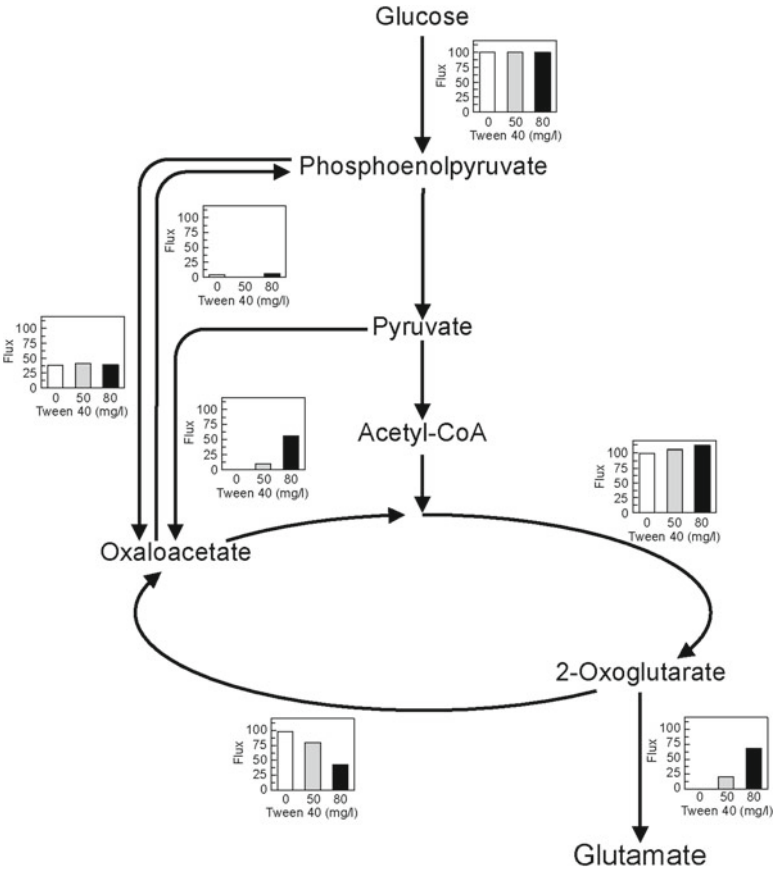


Fig. 13.6 ^{13}C-MFA of *C. glutamicum* during Tween 40-induced glutamate overproduction. Metabolic flux distribution of *C. glutamicum* focused on anaplerotic reactions and the 2-oxoglutarate branch in the absence and presence of Tween 40 is shown. The flux for glucose uptake is normalized to 100

MS data calculated from the tuned metabolic flux distribution. The results indicated that PEPC rather than PC is active during the growth phase, whereas PC is active during Tween 40-induced glutamate production phase. These ^{13}C-MFA results are consistent with reported enzyme activity measurement (Hasegawa et al. 2008).

13.4.3 In Silico Prediction of Metabolic Flux Profiles Using a Genome-Scale Metabolic Model of C. glutamicum

Recently, genome-scale cellular metabolic networks have been reconstructed and metabolic flux balance analysis (FBA) has been conducted using reconstructed

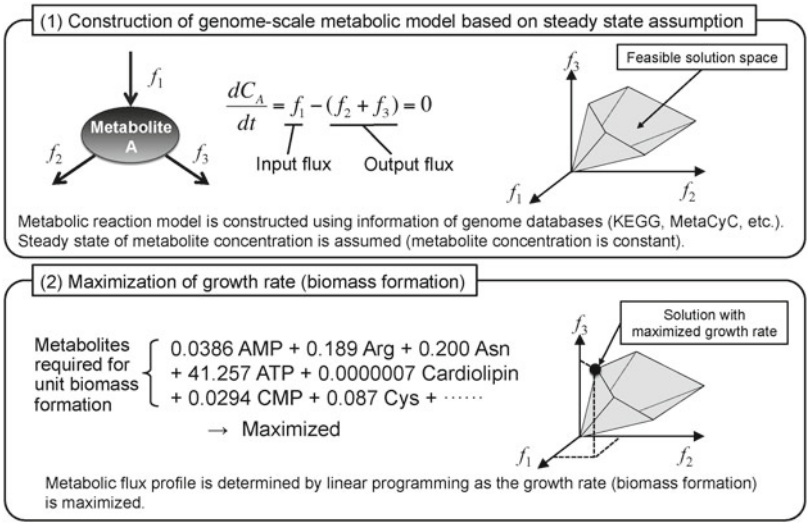


Fig. 13.7 Concept of prediction of metabolic flux profiles using genome-scale metabolic model based on flux balance analysis

genome-scale metabolic networks for bacteria, archaea and eukarya (Oberhardt et al. 2009). FBA is the analysis of metabolic flux profiles in which steady state of metabolic flux is assumed, and metabolic flux profiles are calculated by optimizing an objective function (Fig. 13.7). In FBA, the biomass production rate is generally used as the objective function to be optimized. Since it can be assumed that organisms maximize their growth rate by adaptation and evolution, prediction of metabolic profiles by FBA using genome-scale metabolic models is performed by maximizing the biomass production as an optimization process (Edwards and Palsson 2000; Edwards et al. 2001).

The genome-scale metabolic models of *C. glutamicum* were constructed and the metabolic flux profiles have been successfully predicted (Kjeldsen and Nielsen 2009; Shinfuku et al. 2009). Using genome-scale metabolic model of *C. glutamicum*, metabolic flux profiles under various oxygen supply conditions were simulated, and the predicted yield of CO_2 and organic acids agreed well with experimental data (Shinfuku et al. 2009).

Using genome-scale model of *C. glutamicum*, amino acid production can be also simulated as recently performed for lysine production (Shinfuku et al. 2009). In this simulation, lysine production was used as the objective function and was maximized with fixed glucose uptake and biomass production rates by Vallino and Stephanopoulos (1993). The simulated metabolic flux profile in central metabolism during maximized lysine production agreed with that determined experimentally. However, it is difficult to simulate metabolic flux during glutamate production in *C. glutamicum* using genome-scale metabolic models. As described above, various factors regulate glutamate production in *C. glutamicum*, including ODHC activity,

activities of anaplerotic reactions and conformational changes in mechanosensitive channel. Therefore, other modeling strategies that incorporate these regulatory mechanisms are necessary to simulate metabolic flux profiles during glutamate overproduction in *C. glutamicum*.

13.4.4 Enhancement of Glutamate Production Using Metabolic Engineering

Several studies have reported efforts to enhance glutamate production by *C. glutamicum*. An important target for improvement of glutamate production is the 2-oxoglutarate branch in the TCA cycle. As described above, deletion of the *odhA* gene encoding a catalytic subunit of the ODHC resulted in glutamate overproduction without the need for a trigger (i.e. biotin limitation, Tween 40 addition, or penicillin addition) (Asakura et al. 2007). Expression of *odhA* antisense RNA, which is expected to abolish translation of OdhA, enhanced Tween 40-induced glutamate production, whereas overexpression of *odhA* reduced Tween 40-induced glutamate overproduction (Kim et al. 2009). In contrast, overexpression of the *odhI* whose product inhibits ODHC activity resulted in continuous glutamate production without the need for a trigger (Kim et al. 2010) (Fig. 13.8).

Another target for enhancement of glutamate production is anaplerotic reactions. Peters-Wendisch et al. (2001) examined glutamate production by *pyc*-disrupted and *pyc*-overexpressing strains of *C. glutamicum* treated with Tween 60. The *pyc* gene encodes PC, and Tween 60 induces glutamate production similarly to Tween 40. Glutamate production in the *pyc*-overexpressing strains was higher than that in the wild-type strain. In addition, glutamate production could be decreased by disrupting *pyc*. Yao et al. (2009b) reported that *pyc* disruption enhanced glutamate production in the *dtsR1*-disruptant without requiring a trigger, but glutamate production in this strain could be reduced by disrupting *ppc*, which encodes PEPC.

13.4.5 Metabolic Engineering for Glutamate Production from Renewable Resources

Recently, a number of different groups have attempted to produce useful chemicals including amino acids from renewable resources, using microbial cells. Metabolic engineering has been applied to *C. glutamicum* to enhance its ability and to utilize biomass resources and waste byproducts obtained during industrial production processes for producing chemical compounds. In this section, we discuss metabolic engineering of glutamate production from renewable resources in *C. glutamicum*.

Utilization of pentose sugars such as xylose and arabinose, which are obtained by hydrolyzing lignocellulosic biomass, has been attempted in *C. glutamicum*.

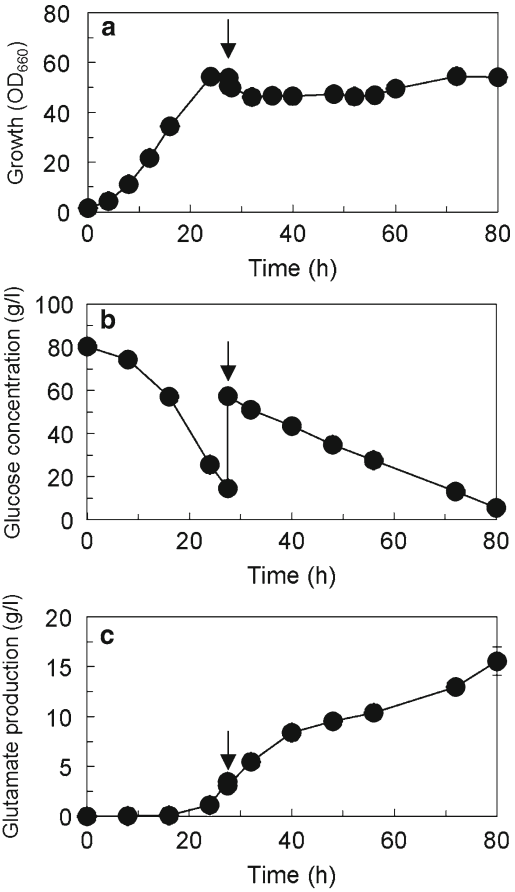


Fig. 13.8 Continuous glutamate production by the *odhI*-overexpressing *C. glutamicum* recombinant strain. The strain was cultivated aerobically and glucose was supplemented prior to glucose depletion (i.e. 27.5 h after starting cultivation, indicated by arrows). Time course of cell growth (**a**), glucose concentration in culture supernatant (**b**), and glutamate concentration in the culture supernatant (**c**) are shown

A *C. glutamicum* recombinant strain carrying *E. coli xylA* and/or *xylB* genes encoding xylose isomerase and xylulokinase, respectively, can grow on xylose as the sole carbon source (Kawaguchi et al. 2006). Moreover, Kawaguchi et al. (2008) indicated that a strain expressing the *E. coli araBAD* genes is able to grow on arabinose as the sole carbon source. Schneider et al. (2011) successfully achieved ethambutol-induced glutamate production from arabinose in a strain carrying this *E. coli araBAD* genes. Gopinath et al. (2011) constructed a strain expressing *E. coli xylA* and *araBAD* genes, which was able to grow on both xylose and arabinose, and this strain produced glutamate, which is induced by ethambutol addition, from a mixture of xylose, arabinose, rice straw and wheat bran hydrolysates.

Glycerol is known as a major byproduct of biodiesel production. Therefore, utilization of glycerol for useful chemical production is an important objective of metabolic engineering. However, *C. glutamicum* cannot grow well on glycerol as the carbon source. Rittmann et al. (2008) reported that the *C. glutamicum* recombinant strain harboring the glycerol utilization pathway from *E. coli* (i.e. expressing

the *E. coli glpP*, *glpK*, and *glpD* genes, which encode glycerol facilitator, glycerol kinase and glycerol-3-phosphate dehydrogenase, respectively) can grow on glycerol and produce glutamate from glycerol.

An alternative strategy to produce useful chemicals from biomass resources is to use the recombinant strains secreting carbohydrate-degrading enzymes. Tsuchidate et al. (2011) constructed a recombinant *C. glutamicum* strain secreting endoglucanases derived from various microorganisms, and succeeded in Tween 40-induced glutamate production from β-glucan by this recombinant. Yao et al. (2009a) constructed a recombinant *C. glutamicum* strain expressing α-amylase from *Streptococcus bovis* on the cell surface. This strain was able to produce glutamate from starch.

13.5 Conclusion

Although the molecular mechanisms of glutamate overproduction are increasingly well known and much progress has been made in the metabolic engineering of glutamate production in *C. glutamicum*, there are a number of unknown factors that regulate glutamate overproduction and secretion. For example, it is not known how *odhI* expression is induced by the various triggers—biotin limitation, Tween 40 addition, and penicillin treatment. Moreover, the molecular mechanisms of glutamate secretion are still unclear.

Glutamate production by *C. glutamicum* can be achieved without cell growth, i.e. cells can be switched completely from the growth phase to the glutamate production phase. If this switching mechanism can be applied more generally to other production process of useful chemicals using microorganisms, efficient and cost-effective production processes could be developed for numerous compounds. Genome DNA sequencing of the *C. glutamicum* strains used in industrial glutamate production is ongoing (Lv et al. 2011, 2012). Once complete, this sequence information will help uncover the molecular mechanisms of glutamate production mechanism in *C. glutamicum* and facilitate the metabolic engineering of this microorganism.

References

Asakura Y, Kimura E, Usuda Y, Kawahara Y, Matsui K, Osumi T, Nakamatsu T (2007) Altered metabolic flux due to deletion of *odhA* causes L-glutamate overproduction in *Corynebacterium glutamicum*. Appl Environ Microbiol 73:1308–1319

Barthe P, Roumestand C, Canova MJ, Kremer L, Hurard C, Molle V, Cohen-Gonsaud M (2009) Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism. Structure 17:568–578

Becker J, Wittmann C (2012) Systems and synthetic metabolic engineering for amino acid production – the heartbeat of industrial strain development. Curr Opin Biotechnol. doi:10.1016/j.copbio.2011.1012.1025

Berrier C, Coulombe A, Szabo I, Zoratti M, Ghazi A (1992) Gadolinium ion inhibits loss of metabolites induced by osmotic shock and large stretch-activated channels in bacteria. Eur J Biochem 206:559–565

Börmann ER, Eikmanns BJ, Sahm H (1992) Molecular analysis of the *Corynebacterium glutamicum gdh* gene encoding glutamate dehydrogenase. Mol Microbiol 6:317–326

Börngen K, Battle AR, Moker N, Morbach S, Marin K, Martinac B, Kramer R (2010) The properties and contribution of the *Corynebacterium glutamicum* MscS variant to fine-tuning of osmotic adaptation. Biochim Biophys Acta 1798:2141–2149

Boulahya KA, Guedon E, Delaunay S, Schultz C, Boudrant J, Bott M, Goergen JL (2010) OdhI dephosphorylation kinetics during different glutamate production processes involving *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 87:1867–1874

Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97:5528–5533

Edwards JS, Ibarra RU, Palsson BO (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130

Eikmanns BJ, Rittmann D, Sahm H (1995) Cloning, sequence analysis, expression, and inactivation of the *Corynebacterium glutamicum icd* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. J Bacteriol 177:774–782

Fudou R, Jojima Y, Seto A, Yamada K, Kimura E, Nakamatsu T, Hiraishi A, Yamanaka S (2002) *Corynebacterium efficiens* sp. nov., a glutamic-acid-producing species from soil and vegetables. Int J Syst Evol Microbiol 52:1127–1131

Gopinath V, Meiswinkel TM, Wendisch VF, Nampoothiri KM (2011) Amino acid production from rice straw and wheat bran hydrolysates by recombinant pentose-utilizing *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 92:985–996

Hasegawa T, Hashimoto K, Kawasaki H, Nakamatsu T (2008) Changes in enzyme activities at the pyruvate node in glutamate-overproducing *Corynebacterium glutamicum*. J Biosci Bioeng 105:12–19

Hashimoto K, Kawasaki H, Akazawa K, Nakamura J, Asakura Y, Kudo T, Sakuradani E, Shimizu S, Nakamatsu T (2006) Changes in composition and content of mycolic acids in glutamate-overproducing *Corynebacterium glutamicum*. Biosci Biotechnol Biochem 70:22–30

Hashimoto K, Nakamura K, Kuroda T, Yabe I, Nakamatsu T, Kawasaki H (2010) The protein encoded by NCgl1221 in *Corynebacterium glutamicum* functions as a mechanosensitive channel. Biosci Biotechnol Biochem 74:2546–2549

Hirasawa T, Wachi M, Nagai K (2000) A mutation in the *Corynebacterium glutamicum ltsA* gene causes susceptibility to lysozyme, temperature-sensitive growth, and L-glutamate production. J Bacteriol 182:2696–2701

Hirasawa T, Wachi M, Nagai K (2001) L-Glutamate production by lysozyme-sensitive *Corynebacterium glutamicum ltsA* mutant strains. BMC Biotechnol 1:9

Hoischen C, Kramer R (1990) Membrane alteration is necessary but not sufficient for effective glutamate secretion in *Corynebacterium glutamicum*. J Bacteriol 172:3409–3416

Ikeda M, Nakagawa S (2003) The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. Appl Microbiol Biotechnol 62:99–109

Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Kramer R, Linke B, McHardy AC, Meyer F, Mockel B, Pfefferle W, Puhler A, Rey DA, Ruckert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J Biotechnol 104:5–25

Kataoka M, Hashimoto KI, Yoshida M, Nakamatsu T, Horinouchi S, Kawasaki H (2006) Gene expression of *Corynebacterium glutamicum* in response to the conditions inducing glutamate overproduction. Lett Appl Microbiol 42:471–476

Kawaguchi H, Vertes AA, Okino S, Inui M, Yukawa H (2006) Engineering of a xylose metabolic pathway in *Corynebacterium glutamicum*. Appl Environ Microbiol 72:3418–3428

Kawaguchi H, Sasaki M, Vertes AA, Inui M, Yukawa H (2008) Engineering of an L-arabinose metabolic pathway in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 77:1053–1062

Kawahara Y, Takahashi-Fuke K, Shimizu E, Nakamatsu T, Nakamori S (1997) Relationship between the glutamate production and the activity of 2-oxoglutarate dehydrogenase in *Brevibacterium lactofermentum*. Biosci Biotechnol Biochem 61:1109–1112

Kikuchi Y, Date M, Yokoyama K, Umezawa Y, Matsui H (2003) Secretion of active-form *Streptoverticillium mobaraense* transglutaminase by *Corynebacterium glutamicum*: processing of the pro-transglutaminase by a cosecreted subtilisin-like protease from *Streptomyces albogriseolus*. Appl Environ Microbiol 69:358–366

Kim J, Hirasawa T, Sato Y, Nagahisa K, Furusawa C, Shimizu H (2009) Effect of *odhA* overexpression and *odhA* antisense RNA expression on Tween-40-triggered glutamate production by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 81:1097–1106

Kim J, Fukuda H, Hirasawa T, Nagahisa K, Nagai K, Wachi M, Shimizu H (2010) Requirement of de novo synthesis of the OdhI protein in penicillin-induced glutamate production by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 86:911–920

Kim J, Hirasawa T, Saito M, Furusawa C, Shimizu H (2011) Investigation of phosphorylation status of OdhI protein during penicillin- and Tween 40-triggered glutamate overproduction by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 91:143–151

Kimura E, Abe C, Kawahara Y, Nakamatsu T (1996) Molecular cloning of a novel gene, *dtsR*, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*. Biosci Biotechnol Biochem 60:1565–1570

Kimura E, Abe C, Kawahara Y, Nakamatsu T, Tokuda H (1997) A *dtsR* gene-disrupted mutant of *Brevibacterium lactofermentum* requires fatty acids for growth and efficiently produces l-glutamate in the presence of an excess of biotin. Biochem Biophys Res Commun 234:157–161

Kimura E, Yagoshi C, Kawahara Y, Ohsumi T, Nakamatsu T, Tokuda H (1999) Glutamate overproduction in *Corynebacterium glutamicum* triggered by a decrease in the level of a complex comprising DtsR and a biotin-containing subunit. Biosci Biotechnol Biochem 63:1274–1278

Kinoshita S (1985) Glutamic acid bacteria. In: Demain AL, Solomon NA (eds) Biology of industrial microorganisms. Benjamin Cummings, San Francisco, pp 115–145

Kinoshita S, Udaka S, Shimono M (1957) Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. J Gen Appl Microbiol 50:193–205

Kjeldsen KR, Nielsen J (2009) In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. Biotechnol Bioeng 102:583–597

Krawczyk S, Raasch K, Schultz C, Hoffelder M, Eggeling L, Bott M (2010) The FHA domain of OdhI interacts with the carboxyterminal 2-oxoglutarate dehydrogenase domain of OdhA in *Corynebacterium glutamicum*. FEBS Lett 584:1463–1468

Levina N, Totemeyer S, Stokes NR, Louis P, Jones MA, Booth IR (1999) Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. EMBO J 18:1730–1737

Lv Y, Wu Z, Han S, Lin Y, Zheng S (2011) Genome sequence of *Corynebacterium glutamicum* S9114, a strain for industrial production of glutamate. J Bacteriol 193:6096–6097

Lv Y, Liao J, Wu Z, Han S, Lin Y, Zheng S (2012) Genome sequence of *Corynebacterium glutamicum* ATCC 14067, which provides insight into amino acid biosynthesis in coryneform bacteria. J Bacteriol 194:742–743

Nakamura J, Hirano S, Ito H, Wachi M (2007) Mutations of the *Corynebacterium glutamicum* NCgl1221 gene, encoding a mechanosensitive channel homolog, induce L-glutamic acid production. Appl Environ Microbiol 73:4491–4498

Nakayama K, Yoshida H (1972) Fermentative production of L-arginine. Agric Biol Chem 36:1675–1684

Nakayama K, Kitada S, Kinoshita S (1961) Studies on lysine fermentation I. The control mechanism on lysine accumulation by homoserine and threonine. J Gen Appl Microbiol 7:145–154

Nampoothiri KM, Hoischen C, Bathe B, Mockel B, Pfefferle W, Krumbach K, Sahm H, Eggeling L (2002) Expression of genes of lipid synthesis and altered lipid composition modulates L-glutamate efflux of *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 58:89–96

Nara T, Samejima H, Kinoshita S (1964) Effect of penicillin on amino acid fermentation. Agric Biol Chem 28:120–124

Niebisch A, Kabus A, Schultz C, Weil B, Bott M (2006) Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. J Biol Chem 281:12300–12307

Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, Gojobori T (2003) Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. Genome Res 13:1572–1579

Oberhardt MA, Palsson BO, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320

Okino S, Inui M, Yukawa H (2005) Production of organic acids by *Corynebacterium glutamicum* under oxygen deprivation. Appl Microbiol Biotechnol 68:475–480

O'Regan M, Thierbach G, Bachmann B, Villeval D, Lepage P, Viret JF, Lemoine Y (1989) Cloning and nucleotide sequence of the phosphoenolpyruvate carboxylase-coding gene of *Corynebacterium glutamicum* ATCC13032. Gene 77:237–251

Peters-Wendisch PG, Kreutzer C, Kalinowski J, Patek M, Sahm H, Eikmanns BJ (1998) Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. Microbiology 144(Pt 4):915–927

Peters-Wendisch PG, Schiel B, Wendisch VF, Katsoulidis E, Mockel B, Sahm H, Eikmanns BJ (2001) Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. J Mol Microbiol Biotechnol 3:295–300

Radmacher E, Stansen KC, Besra GS, Alderwick LJ, Maughan WN, Hollweg G, Sahm H, Wendisch VF, Eggeling L (2005) Ethambutol, a cell wall inhibitor of *Mycobacterium tuberculosis*, elicits L-glutamate efflux of *Corynebacterium glutamicum*. Microbiology 151:1359–1368

Rittmann D, Lindner SN, Wendisch VF (2008) Engineering of a glycerol utilization pathway for amino acid production by *Corynebacterium glutamicum*. Appl Environ Microbiol 74:6216–6222

Sano K, Shiio I (1970) Microbial production of L-lysine III. Production by mutants resistant to S-(2-aminoethyl)-L-cysteine. J Gen Appl Microbiol 16(1):373–391

Sato H, Orishimo K, Shirai T, Hirasawa T, Nagahisa K, Shimizu H, Wachi M (2008) Distinct roles of two anaplerotic pathways in glutamate production induced by biotin limitation in *Corynebacterium glutamicum*. J Biosci Bioeng 106:51–58

Schneider J, Niermann K, Wendisch VF (2011) Production of the amino acids L-glutamate, L-lysine. L-ornithine and L-arginine from arabinose by recombinant Corynebacterium glutamicum. J Biotechnol 154:191–198

Schultz C, Niebisch A, Gebel L, Bott M (2007) Glutamate production by *Corynebacterium glutamicum*: dependence on the oxoglutarate dehydrogenase inhibitor protein OdhI and protein kinase PknG. Appl Microbiol Biotechnol 76:691–700

Schultz C, Niebisch A, Schwaiger A, Viets U, Metzger S, Bramkamp M, Bott M (2009) Genetic and biochemical analysis of the serine/threonine protein kinases PknA, PknB. PknG and PknL of Corynebacterium glutamicum: evidence for non-essentiality and for phosphorylation of OdhI and FtsZ by multiple kinases. Mol Microbiol 74:724–741

Shigu H, Terui G (1971) Studies on process of glutamic acid fermentation at the enzyme level. I. On the change of α-ketoglutaric acid dehydrogenase in the course of culture. J Ferm Technol 49:400–405

Shiio I, Nakamori S (1970) Microbial production of L-threonine. Part II. Production by α-amino-β-hydroxyvaleric acid resistant mutants of glutamate producing bacteria. Agric Biol Chem 34:448–456

Shiio I, Otsuka SI, Takahashi M (1962) Effect of biotin on the bacterial formation of glutamic acid. I. Glutamate formation and cellular premeability of amino acids. J Biochem 51:56–62

Shimizu H, Hirasawa T (2007) Production of glutamate and glutamate-related amino acids: molecular mechanism analysis and metabolic engineering. In: Wendisch VF (ed) Amino acid biosynthesis-pathways, regulation and metabolic engineering. Microbiology monograph, vol 5. Springer, Berlin, pp 1–38

Shimizu H, Tanaka H, Nakato A, Nagahisa K, Kimura E, Shioya S (2003) Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxoglutarate branch in glutamate production by *Corynebacterium glutamicum*. Bioprocess Biosyst Eng 25:291–298

Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H (2009) Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. Microb Cell Fact 8:43

Shirai T, Nakato A, Izutani N, Nagahisa K, Shioya S, Kimura E, Kawarabayasi Y, Yamagishi A, Gojobori T, Shimizu H (2005) Comparative study of flux redistribution of metabolic pathway in glutamate production by two coryneform bacteria. Metab Eng 7:59–69

Shirai T, Matsuzaki K, Kuzumoto M, Nagahisa K, Furusawa C, Shioya S, Shimizu H (2006) Precise metabolic flux analysis of coryneform bacteria by gas chromatography–mass spectrometry and verification by nuclear magnetic resonance. J Biosci Bioeng 102:413–424

Shirai T, Fujimura K, Furusawa C, Nagahisa K, Shioya S, Shimizu H (2007) Study on roles of anaplerotic pathways in glutamate overproduction of *Corynebacterium glutamicum* by metabolic flux analysis. Microb Cell Fact 6:19

Stephanopoulos G, Aristidou AA, Nielsen J (1998) Metabolic engineering: principles and methodologies. Academic, San Diego

Takinami K, Yoshii H, Tsuri H, Okada H (1965) Biochemical effects of fatty acid and its derivatives on L-glutamic acid fermentation. Part III. Biotin-Tween 60 relationship in the accumulation of L-glutamic acid and the growth of *Brevibacterium lactofermentum*. Agric Biol Chem 29:351–359

Tsuchida T, Yoshinaga F, Kubota K, Momose H (1975) Production of L-valine by 2-thiazolealanine resistant mutants derived from glutamic acid producing bacteria. Agric Biol Chem 39:1319–1322

Tsuchidate T, Tateno T, Okai N, Tanaka T, Ogino C, Kondo A (2011) Glutamate production from β-glucan using endoglucanase-secreting *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90:895–901

Udaka S (1960) Screening method for microorganisms accumulating metabolites and its use in the isolation of *Micrococcus glutamicus*. J Bacteriol 79:754–755

Umakoshi M, Hirasawa T, Furusawa C, Takenaka Y, Kikuchi Y, Shimizu H (2011) Improving protein secretion of a transglutaminase-secreting *Corynebacterium glutamicum* recombinant strain on the basis of ^{13}C metabolic flux analysis. J Biosci Bioeng 112:595–601

Vallino JJ, Stephanopoulos G (1993) Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol Bioeng 41:633–646

Wendisch VF, Bott M, Eikmanns BJ (2006) Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. Curr Opin Microbiol 9:268–274

Yao W, Chu C, Deng X, Zhang Y, Liu M, Zheng P, Sun Z (2009a) Display of α-amylase on the surface of *Corynebacterium glutamicum* cells by using NCgl1221 as the anchoring protein, and production of glutamate from starch. Arch Microbiol 191:751–759

Yao W, Deng X, Zhong H, Liu M, Zheng P, Sun Z, Zhang Y (2009b) Double deletion of *dtsR1* and *pyc* induce efficient L-glutamate overproduction in *Corynebacterium glutamicum*. J Ind Microbiol Biotechnol 36:911–921

Yukawa H, Omumasaba CA, Nonaka H, Kos P, Okai N, Suzuki N, Suda M, Tsuge Y, Watanabe J, Ikeda Y, Vertes AA, Inui M (2007) Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R. Microbiology 153:1042–1058

Chapter 14
Microbial Metabolic Engineering for L-Threonine Production

Xunyan Dong, Peter J. Quinn, and Xiaoyuan Wang

Abstract L-threonine, one of the three major amino acids produced throughout the world, has a wide application in industry, as an additive or as a precursor for the biosynthesis of other chemicals. It is predominantly produced through microbial fermentation the efficiency of which largely depends on the quality of strains. Metabolic engineering based on a cogent understanding of the metabolic pathways of L-threonine biosynthesis and regulation provides an effective alternative to the traditional breeding for strain development. Continuing efforts have been made in revealing the mechanisms and regulation of L-threonine producing strains, as well as in metabolic engineering of suitable organisms whereby genetically-defined, industrially competitive L-threonine producing strains have been successfully constructed. This review focuses on the global metabolic and regulatory networks responsible for L-threonine biosynthesis, the molecular mechanisms of regulation, and the strategies employed in strain engineering.

Keywords L-threonine biosynthesis • L-threonine production • Metabolic engineering • Regulation mechanism

X. Dong
Key Laboratory of Industrial Biotechnology of Ministry of Education,
School of Biotechnology, JiangnanUniversity, Wuxi 214122, China

P.J. Quinn
Department of Biochemistry, King's College London, 150 Stamford Street,
London SE1 9NH, United Kingdom

Wang (✉)
State Key Laboratory of Food Science and Technology, School of Biotechnology,
JiangnanUniversity, Wuxi 214122, China
e-mail: xiaoyuanwang@hotmail.com

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_14,

Abbreviations

SD sequence	Shine-Dalgarno sequence
ACT	Aspartate kinase, Chorismate mutase and TyrA

14.1 Introduction

L-threonine is currently one of the three major amino acids produced throughout the world with an annual market size of approximate 0.23 metric tons (Becker and Wittmann 2011). Among its wide industrial application, the most remarkable use of L-threonine is as a feed additive. Application of low protein level formula feeds supplemented with L-threonine improves the growth of livestock, relieves crude protein deficiency and lowers nitrogen emissions, thus contributing to the sustainable development of the society. Recent studies acknowledging L-threonine as the second and the third limiting amino acid in swine and poultry feeds respectively (Ajinomoto 2009) have stimulated the further expansion of the industry. Moreover, L-threonine can be used as precursor for the biosynthesis of L-isoleucine and L-homoalanine. This underpins recent developments in the efficient microbial production of L-threonine (Leuchtenberger et al. 2005).

The most efficient solution to improve productivity of the bioconversion and reduce costs is to develop highly productive strains. Thanks to traditional breeding methods, L-threonine producing strains of *Serratia marcescens* (Komatsubara et al. 1978), *Escherichia coli* (Furukawa et al. 1988) and *Corynebacterium glutamicum* (Morinage et al. 1987) have been developed. However, the traditional breeding procedures are time-consuming and labor-costly. Furthermore, strains surviving multiple rounds of random mutagenesis are genetically undefined and vulnerable to further changes, out of which at best either a marginal increase in yield or resistance to more stringent process requirements could be achieved. To overcome these difficulties in developing more efficient L-threonine producing strains, metabolic engineering through rational genetic manipulations offers a promising option for subsequent isolation of genetically defined hyper-producing strains. Selection criteria would include reduced by-product formation and expanded substrate spectra. Recruited strategies include increasing the biosynthetic metabolic flux of L-threonine, enhancing its excretion efficiency, and reducing unwanted carbon loss through the competing branches and its intracellular consumption. Most recently, systems metabolic engineering comprehensively combining all of these strategy elements has achieved success in restructuring a wild-type strain into a genetically definite, highly competitive L-threonine producer.

This chapter will focus on the global metabolic pathway of L-threonine together with regulation mechanisms and the progress in metabolic engineering of the dominating industrial bacteria, *E. coli* and *C. glutamicum*.

14.2 The Metabolic Pathway of L-Threonine Biosynthesis

L-threonine belongs to the aspartic family of amino acids. L-aspartate is synthesized from oxaloacetate, an intermediate of TCA cycle, by aspartate transaminase (Table 14.1) that is encoded by the *aspC* gene in *E. coli* (Fotheringham et al. 1986) and by *aspA* in *C. glutamicum* (Marienhagen et al. 2005). On substrates of carbohydrates,

Table 14.1 Enzymes and their coding genes related to the biosynthesis of L-threonine

Enzymes	Genes	Function
Phosphoenolpyruvate carboxylase	*ppc*	Carboxylate phosphoenolpyruvate to form oxaloacetate
Pyruvate carboxylase	*pyc*	Carboxylate pyruvate to form oxaloacetate
Aspartate transaminase	*aspC*; *aspA*	Transfer an amino group to oxaloacetate to form L-aspartate. The coding gene is *aspC* in *E. coli*, and is *aspA* in *C. glutamicum*
Aspartate kinase	*thrA*; *metL*; *lysC*	Phosphorylate aspartate to form aspartyl-P. There are three coding genes (*thrA*; *metL*; *lysC*) in *E. coli*, but only one (*lysC*) in *C. glutamicum*
Aspartyl semialdehyde dehydrogenase	*asd*	Deoxidize aspartyl-P to form aspartyl semialdehyde
Homoserine dehydrogenase	*thrA*; *metL*; *hom*	Remove the carboxyl group of aspartyl semialdehyde. There are two coding genes (*thrA*; *metL*) in *E. coli*, but only one (*hom*) in *C. glutamicum*
Homoserine kinase	*thrB*	Phosphorylate homoserine to form homoserine-P
Threonine synthase	*thrC*	Remove the phosphate group of homoserine-P to form threonine
Phosphoenolpyruvate carboxykinase	*pck*	Decarboxylate oxaloacetate to form phosphoenolpyruvate
Oxaloacetate decarboxylase	*odx*	Decarboxylate oxaloacetate to form pyruvate
Malic enzyme	*sfcA*; *maeB*; *malE*	Decarboxylate malate to form pyruvate. There are two coding genes (*sfcA*; *maeB*) in *E. coli*, but only one (*malE*) in *C. glutamicum*
Dihydrodipicolinate synthase	*dapA*	Consume L-aspartyl semialdehyde to produce L-lysine
Homoserine succinyltransferase	*metA*	Consume L-homoserine to produce L-methionine in *E. coli*
Homoserine acetyltransferase	*metX*	Consume L-homoserine to produce L-methionine in *E. coli*
Threonine dehydratase	*tdcB*; *ilvA*	Consume threonine to produce isoleucine
Threonine dehydrogenase	*tdh*	Consume threonine to produce glycine in *E. coli*
Threonine aldolase	*ltaE*	Consume threonine to produce glycine in *E. coli*
Serine hydroxymethyl transferase	*glyA*	Consume threonine to produce glycine in *C. glutamicum*
Permease	*rhtA*; *rhtB*; *rhtC*; *thrE*	Transport threonine from inside to the outside of cell. There are three coding genes (*rhtA*; *rhtB* and *rhtC*) in *E. coli*, but only one (*thrE*) in *C. glutamicum*

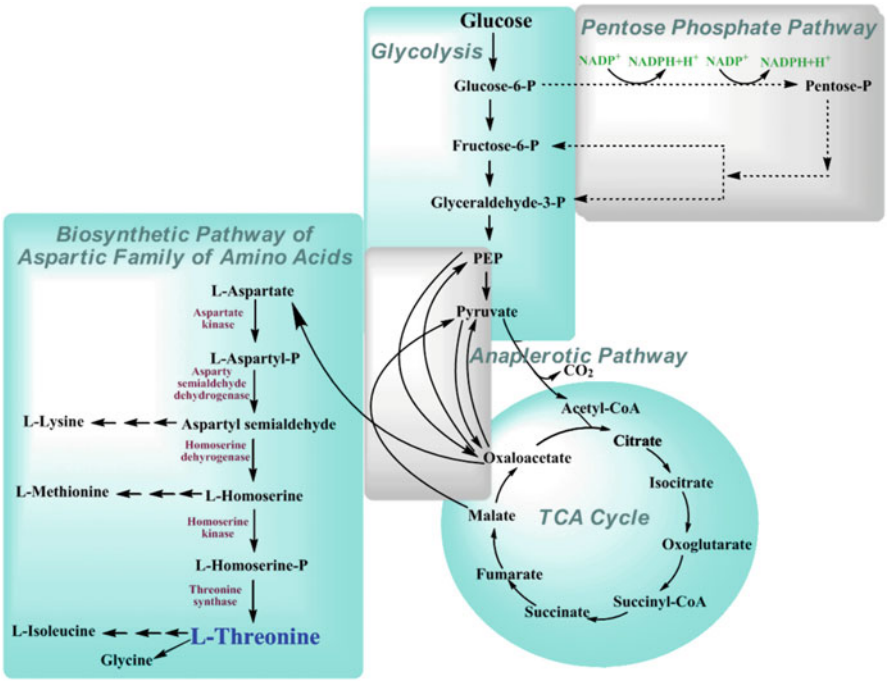


Fig. 14.1 The biosynthesis pathway of L-threonine. The pathway consists of centeral metabolic pathways and the threonine terminal pathways. The centeral metabolic pathways involve glycolysis, phosphate pentose pathway, TCA cycle and anaplerotic pathways. The threonine terminal pathway consists of five enzymetic steps. The first, third, and fourth reactions are catalyzed by the three key enzymes aspartate kinase, homoserine dehydrogenase, and homoserine kinase, respectively. There are four competing pathways that affect the biosynthesis of L-threonine, leading to formation of L-lysine, L-methionine, L-isoleucine, and glycine

L-threonine biosynthesis involves centeral metabolism including glycolysis, pentose phosphate pathway, TCA cycle and anaplerotic pathways between glycolysis and TCA cycle, prior to its terminal pathway (Fig. 14.1).

14.2.1 The Associated Anaplerotic Pathways

Anaplerotic reactions refer to C3-carboxylation and C4-decarboxylation around the phosphoenolpyruvate–pyruvate–oxaloacetate node, which interconnect the TCA cycle with glycolysis. These reactions result in direct oxaloacetate formation or depletion. Carboxylation of phosphoenolpyruvate catalyzed by phosphoenolpyruvate carboxylase and that of pyruvate by pyruvate carboxylase contribute to its formation. Accordingly, decarboxylation of oxaloacetate catalyzed by phosphoenolpyruvate carboxykinase and oxaloacetate decarboxylase form phosphoenolpyruvate and

pyruvate, respectively. The carbon interconversion between the TCA cycle and glycolysis can also be achieved by malic enzyme which decarboxylates malate to form pyruvate (Sauer and Eikmanns 2005). As malic enzyme depletes malate, the precursor of oxaloacetate in the TCA cycle, this reaction indirectly results in the depletion of oxaloacetate. *C. glutamicum* possesses all these enzymes, while *E. coli* has no oxaloacetate decarboxylase. In both organisms, phosphoenolpyruvate carboxylase is encoded by the *ppc* gene, phosphoenolpyruvate carboxykinase by *pck*, and pyruvate carboxylase by *pyc* (Sauer and Eikmanns 2005). Very recently, the oxaloacetate decarboxylase coding gene *odx* of *C. glutamicum* has been characterized (Klaffl and Eikmanns 2010). As for malic enzyme, unlike *C. glutamicum* which only has one encoded by *malE* (Gourdon et al. 2000), *E. coli* possesses two isoforms respectively encoded by *sfcA* and *maeB* (Bologna et al. 2007).

14.2.2 The Terminal Pathway of L-Threonine

Starting from the building block of L-aspartate, the biosynthesis of L-threonine comprises five successive reactions sequencially catalyzed by aspartate kinase, aspartyl semialdehyde dehydrogenase, homoserine dehydrogenase, homoserine kinase and threonine synthase.

Firstly, L-asparte is phosphorylated to form L-aspartyl-P by aspartate kinase, the first key enzyme in the L-threonine terminal pathway which serves to direct the carbon flux into the aspartate family of amino acids. *E. coli* possesses three aspartate kinase isoenzymes, aspartate kinase I, II and III (Chassagnole et al. 2001; Viola 2001). Aspartate kinase I and II, the most and the least abundant isoforms, respectively, exist as a catalytic domain in the bifunctional enzymes, aspartate kinase I-homoserine dehydrogenase I and aspartate kinase II-homoserine dehydrogenase II, correspondingly encoded by the *thrA* and *metL* genes (Katinka et al. 1980). Aspartate kinase III is a monofunctional enzyme encoded by the *lysC* gene (Shiio and Miyajima 1969). In *C. glutamicum*, the known aspartate kinase is encoded by the *lysC* gene. Although the enzymes involved in the terminal pathway of L-threonine are believed to have no isoenzyme components, deletion of *lysC* was detrimental but not lethal to a *C. glutamicum* strain grown on minimal medium (Jetten et al. 1995).

Secondly, L-aspartyl-P is deoxidized to form L-aspartyl semialdhyde by aspartyl semialdehyde dehydrogenase encoded by the *asd* gene in both *E. coli* and *C. glutamicum* (Boy and Patte 1972; Cremer et al. 1988).

Thirdly, L-homserine is synthesized from the deoxidization reaction of L-aspartyl semialdhyde by homoserine dehydrogenase, the second key enzyme of the pathway that controls carbon flux towards L-homoserine synthesis at the branchpoint of L-aspartyl semialdehyde. In *E. coli*, as mentioned above, two isoforms of homoserine dehydrogenase are known, both of which are present as the catalytic domain in the bifunctional aspartate kinaseI-homoserine dehydrogenase I and aspartate kinase II-homoserine dehydrogenase II. In *C. glutamicum*, homoserine dehydrogenase is encoded by *hom* (Follettie et al. 1988).

The conversion of L-homoserine to L-threonine is performed by homoserine kinase and threonine synthase. Homoserine kinase phosphorylates L-homoserine to form L-homoserine-P which is then dephosphorylated by threonine synthase to produce L-threonine. Homoserine kinase is the third key enzyme of the pathway; it controls carbon flux towards L-threonine synthesis at the branchpoint of L-homoserine. In both *E. coli* and *C. glutamicum*, homoserine kinase is encoded by the *thrB* gene and threonine synthase is by *thrC* (Follettie et al. 1988; Theze and Saint-Girons 1974).

14.2.3 The Efflux System of L-Threonine

After synthesis in the cell, L-threonrine could be excreted into the medium by both passive diffusion and carrier-mediated export, the latter accounting for over 90 % of the total efflux in *C. glutamicum* (Palmieri et al. 1996). Five L-threonine permeases have been characterized in *E. coli* (Eggeling and Sahm 2003). They confer tolerance to high concentration of L-threonine to the producing strains. However, only three of them, RhtA, RhtB and RhtC separately encoded by genes *rhtA*, *rhtB* and *rhtC*, show activity in exporting L-threonine out of cell (Kruse et al. 2002; Livshits et al. 2003). RhtA belonging to the drug/metabolite transporter super family can export both L-threonine and L-homoserine. RhtB and RhtC belonging to the RhtB translocator super family are specific exporters for L-threonine (Diesveld et al. 2009; Eggeling and Sahm 2001). The activity of RhtC is higher than that of RhtB. *C. glutamicum* is assumed to be deficient in the export system of L-threonine (Debabov 2003). The only identified transporter showing activity in exporting L-threonine in this organism is ThrE encoded by the *thrE* gene. Nevertheless, ThrE has a low affinity with L-threonine and exports both L-threonine and L-homoserine (Simic et al. 2001).

14.2.4 Carbon Flux Depleting Pathways

The loss of available carbon flux for L-threonine biosynthesis is mainly through the L-lysine and L-methionine competing branches existing in the terminal pathway and its intracellular depletion towards L-isoleucine and glycine. The L-lysine branch at the nexis of L-aspartyl semialdehyde is initiated by dihydrodipicolinate synthase encoded by the *dapA* gene in both organisms (Velasco et al. 2002). The L-methionine branch at L-homoserine is initiated by the *metA* gene encoding homoserine succinyltransferase in *E. coli*, while by *metX* encoding homoserine acetyltransferase in *C. glutamicum* (Rückert et al. 2003). It should be noted that *C. glutamicum* utilizes novel split pathways for both L-lysine and L-methionine biosynthesis (Lee and Hwang 2003; Schrumpf et al. 1991). The depletion of L-threonine towards L-isoleucine is initiated by threonine dehydratase. Both *E. coli* and *C. glutamicum* possess two isoforms of threonine dehydratase encoded by the *ilvA* and *tdcB* genes

respectively (Kalinowski et al. 2003; Mockel et al. 1992; Umbarger and Brown 1957). The expression of *tdcB* would take place only under anaerobic conditions in *E. coli* (Umbarger 1973). The depletion of L-threonine towards glycine in *E. coli* can be initiated by two enzymes, namely threonine dehydrogenase encoded by the *tdh* gene (Bell and Turner 1976) and threonine aldolase encoded by the *ltaE* gene (Liu et al. 1998). However, only when the intracellular glycine is lacking will the threonine aldolase function as a compensatory force. In *C. glutamicum*, a side activity of serine hydroxymethyl transferase encoded by the *glyA* gene fulfills the function of cleaving L-threonine directly into glycine and acetaldehyde (Simic et al. 2002). The main substrate of this enzyme is L-serine with which the cleavage activity is 24-fold higher than with L-threonine.

14.3 Regulation of L-Threonine Biosynthesis

The biosynthesis of L-threonine is subjected to strict regulation due to its physiological importance in the cellular metabolism. A computational analysis of the global metabolic network of *E. coli* by Almaas et al. (2004) indicated that the L-threonine biosynthesis along with its conversion into glycine was a component of the high-flux backbone of metabolism. Hartman (2007) reported that the biosynthesis and uptake of L-threonine was of great significance for *Saccharomyces cerevisiae* cell to maintain stability, as L-threonine could be converted into glycine and subsequently initiate *de-novo* purine synthesis. Curien et al. (2009) found that in *Arabidopsis* L-threonine played an integrative role in regulating the metabolic distribution of L-aspartate.

14.3.1 The Regulation of Repression

In *E. coli*, the regulation of repression occurs in the expression of all the genes in the L-threonine terminal pathway: *thrA*, *thrB* and *thrC* by L-threonine and L-isoleucine in a covalent manner (Theze and Saint-Girons 1974); *metL* by L-methionine; *lysC* by L-lysine; *asd* by L-lysine, L-threonine and L-methionine in a multivalent manner (Boy and Patte 1972).

The *E. coli thrA*, *thrB* and *thrC* are sequencially clustered in the *thr* operon along with a leader sequence *thrL* containing 178 base pairs preceding *thrA*. The simultaneous repression of the expression of *thrA*, *thrB* and *thrC* by L-threonine and L-isoleucine can be ascribed to the transcriptional attenuation of the operon through *thrL* (Theze and Saint-Girons 1974). An internal region of *thrL* potentially encodes a short peptide with 8 threonine codons and 4 isoleucine codons, 11 of which are tandemly arranged. The downstream region of the tandem codons in *thrL* tends to form a stem and loop structure, either a terminator or an antiterminator depending on the availability of L-threonine and L-isoleucine in the cell. When L-threonine and

L-isoleucine are both in excess, the translation of the short peptide proceeds, resulting in the formation of a terminator structure, so the transcription of the *thr* operon is pre-terminated. By contrast, when L-threonine and L-isoleucine are lacking, translation of the short peptide stalls, leading to the formation of an antiterminator structure, so the transcription of the *thr* operon can continue. Mutation of a G insertion at position -37 upstream of *thrA* could cause derepression, probably through destabilizing the terminator structure (Gardner 1979; Gardner and Reznikoff 1978).

The repression of *E. coli lysC* expression by L-lysine is exerted through translational attenuation. The adjoining upstream of *lysC* resided a leading sequence that can form a complex secondary structure of 6 helixes via internal base pairing during transcription. When L-lysine is in excess, one of the helixes formed covers the Shine-Dalgarno sequence (SD sequence), preventing the binding of the ribosome to mRNA, so the translation of *lysC* can not be initiated. By contrast, when L-lysine is deficient, a different secondary structure forms, releasing the SD sequence, thus the translation of *lysC* can proceed (Grundy et al. 2003).

In *C. glutamicum,* the expression of *hom* and *thrB* is subjected to the repression by L-methionine. The two genes are clustered in one operon in the direction of 5'-*hom-thrB*-3' with an internal 10 bp non-coding space. Also resided in this operon is a long reversibly repeated sequence upstream of *hom,* which is apt to form a stem structure with a Gibbs free energy of −16.2 kJ/mol (Mateos et al. 1994). The repression might be a transcriptional attenuation-like regulation exerted through the leading sequence. Such types of regulation may need the assistance of some regulatory proteins (Henkin and Yanofsky 2002; Mateos et al. 1994). However, the transcription mode of *hom-thrB* operon remains unclear (Diesveld et al. 2009; Mateos et al. 1994), and the evidence of transcriptional attenuation-like regulation needs further investigation.

14.3.2 The Regulation of Inhibition

E. coli aspartate kinase I is subjected to feed-back inhibition by L-threonine. The inhibition is partial, and the inhibitory mechanism is allosteric and competitive with L-aspartate. Homoserine dehydrogenase I is also partially inhibited by L-threonine, but the inhibitory mechanism is non-competitive (Chassagnole et al. 2001). The bifunctional aspartate kinase I-homoserine dehydrogenase I is a homotetramer and each peptide chain contains two catalytic domains, with the kinase site residing on the N-terminal region and the dehydrogenase site residing on the C-terminal region. Between the two catalytic domains located an interphase region responsible for the allosteric regulation by L-threonine (Fazel et al. 1983). Although it was reported previously that the inhibition of homoserine dehydrogenase I was mediated through the aspartate kinase I domain (Truffa-Bachi et al. 1974), the subsequent research by James and Viola (2002) also indicated that the inhibition might be exerted by the interphase region. The mechanism of inhibition remains ambiguous without a clear picture of the natural structure of aspartate kinase I-homoserine dehydrogenase I.

E. coli aspartate kinase III is subjected to feed-back inhibition by L-lysine. The inhibition is complete (Chassagnole et al. 2001), and the inhibitory mechanism is

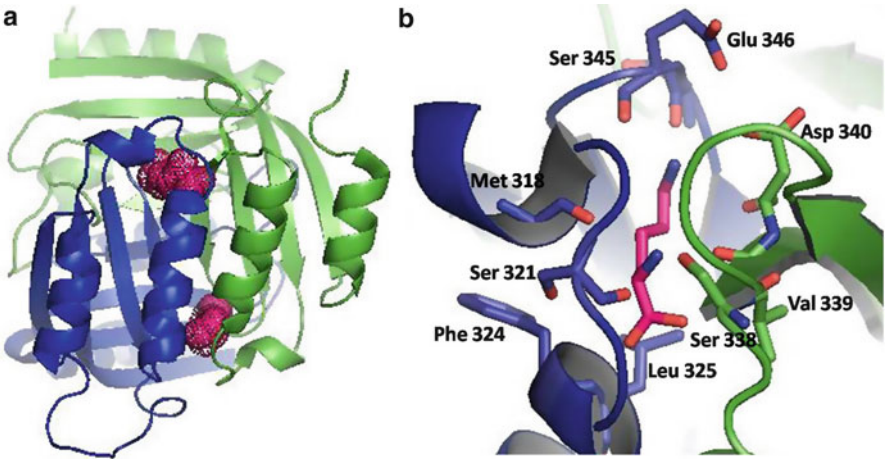


Fig. 14.2 L-lysine binding site in *E. coli* aspartate kinase III. (**a**) The dimer structure of the regulatory domains is shown. Regulatory domains from two different chains are shown in blue and green, respectively. The bound L-lysine molecules are shown in pink. (**b**) The bound L-lysine molecule and the amino acid residues involved in its binding are both shown in the manner of sticks. These models are built by using the PyMOL software, Protein Data Bank (accession number 2J0X) and the published information by Kotaka et al. (2006)

allosteric (Kotaka et al. 2006). The functional *E. coli* aspartate kinase III is a homodimer (Fig. 14.2a). The N-terminal region of each subunit functions as a catalytic domain, containing substrate binding sites for L-aspartate and ATP. The C-terminal region serves as a regulatory domain, containing two perpendicularly arranged ACT domains, ACT1 and ACT2. The ACT (Acronym for aspartate kinase, Chorismate mutase and TyrA) domain is a conserved structure responsible for the binding of small-molecule regulatory ligands found in functionally diverse proteins (Chipman and Shaanan 2001). The interface of the two ACT1 domains from different subunit shape binding sites for two L-lysine molecules (Fig. 14.2b). Mutations of T344M, S345L, T352I, all in the ACT1 domain, have been confirmed to be associated with partial L-lysine resistance of the enzyme. The ACT2 domains play the role of stabilizing the dimer structure and transmitting the L-lysine binding signal to the catalytic domain. The mechanism of inhibition is that the binding of L-lysine to the regulatory domains triggers tetramerization of two dimers and concomitant allosteric transition of the catalytic domains, resulting in blockage of the ATP binding site and eventually loss of activity. Furthermore, the allosteric transition disrupts one hydrogen bond at the L-aspartate binding site, which does not affect the substrate binding ability but may reduce the conversion rate (Kotaka et al. 2006).

The inhibition of *E. coli* homoserine kinase is complicated. It is inhibited by the substrates, L-homoserine at a concentration above 1 mM and ATP above 3 mM in a hypothetical "preferred order" manner, by L-threonine in a competitive manner, and by L-lysine in a non-competitive manner (Chassagnole et al. 2001).

In *C. glutamicum*, aspartate kinase is subjected to feed-back inhibition by L-lysine. The inhibitory mechanism is allosteric. *C. glutamicum* aspartate kinase is

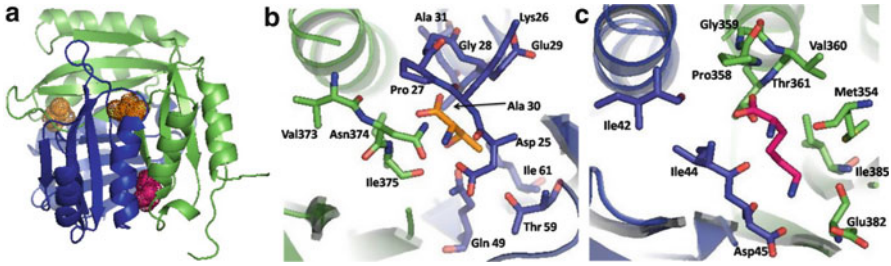


Fig. 14.3 L-threonine and L-lysine binding sites in *C. glutamicum* aspartate kinase. (**a**) The dimer structure of αβ-subunits is shown. The regulatory domain from α-subunit is shown in green and β-subunit is shown in blue. The bound L-threonine and L-lysine molecules are shown in orange and pink, respectively. (**b**) The bound L-threonine molecule and the amino acid residues directly involved in its binding are shown in the manner of sticks. (**c**) The bound L-lysine molecule and the amino acid residues directly involved in its binding are shown in the manner of sticks. These structural models are built by using the PyMOL software, Protein Data Bank (accession number 3AAW) and the published information by Yoshida et al. (2010)

a heterotetramer, comprised of two α subunits and two β subunits. As the two types of subunits are encoded by a single gene *lysC* with an in-frame sequence overlap (Kalinowski et al. 1990), the amino acid sequence of the β subunit is identical to 160 residues of the C-terminus of the α subunit. The N-terminal region of the α subunit functions as the catalytic domain, while its C-terminal region and the β subunit serve as the regulatory domain (Kato et al. 2004). Each regulatory domain contains two perpendicularly arranged ACT domains, ACT1 and ACT2 (Yoshida et al. 2007). Interaction of the β-subunit with the regulatory domain of the α-subunit causes the formation of four effector-binding sites, three of which are for the binding of two L-threonine and one L-lysine molecules (Fig. 14.3a). The L-threonine binding sites (Fig. 14.3b) are located in the interphase between ACT1 from the β-subunit and ACT2 from the α-subunit, and between ACT1 from the α-subunit and ACT2 from the β-subunit. The L-lysine binding site (Fig. 14.3c) is located in the interphase between ACT1 from the β-subunit and ACT2 from the α-subunit. The mechanism of inhibition may involve two steps. Firstly, the binding of L-threonine triggers dimerization of the β-subunit and the regulatory domain of the α-subunit. Then the binding of L-lysine triggers partial dissociation of the dimer and enables the interaction of the β-subunit with the catalytic domain of the α-subunit, resulting in an inactive conformation of the enzyme. Furthermore, L-lysine may also bind to the binding site for L-aspartate, which stabilizes the inactive conformation (Yoshida et al. 2010). Mutations in the regulatory domains of aspartate kinase were discovered in *C. glutamicum* strains with resistance to the L-lysine analog S-(2-aminoethyl)-l-cysteine (Yoshida et al. 2007).

C. glutamicum homoserine dehydrogenase is subjected to feed-back inhibition by L-threonine. The inhibition is almost complete when the concentration of L-threonine reaches 2 mM. The binding site for L-threonine is located in the C-terminus (Archer et al. 1991). Mutation of G378E conferred L-threonine resistance on homoserine

dehydrogenase (Reinscheid et al. 1991). The activity of the mutant was not affected by 25 mM L-threonine and was inhibited by 50 % when the concentration reached 100 mM. Furthermore, the residue 378 was confirmed to be essential for regulation as both the size and the charge of the amino acid at that position affected the enzyme's sensitivity to L-threonine. *C. glutamicum* homoserine kinase is inhibited by L-threonine with the concentration for half maximal inhibition of 25 mM (Colon et al. 1995). The inhibition is competitive and is relieved as the concentration of L-homoserine increases (Miyajima et al. 1968).

14.4 Metabolic Engineering for L-Threonine Production

The strategies recruited in the metabolic engineering for L-threonine production can be summarized as follows; (1) overexpressing the key enzymes coding genes of the biosynthesis pathway to condense carbon influx, (2) attenuating the competing branches to save more available precursors, (3) reducing intracellular depletion of L-threonine, (4) enhancing L-threonine secretion, (5) systems metabolic engineering. This section will focus on the effects of application of these strategies on L-threonine production in strains with almost defined genotypes.

14.4.1 Overexpressing the Genes of the Biosynthesis Pathway

Overexpressing the genes of the biosynthesis pathway, especially the deregulated ones encoding key enzymes, is usually the most productive strategy. Application of such strategy has achieved significant enhancements in L-threonine production. The priority of this strategy has been given to the genes of the L-threonine terminal pathway, and the outcome is better with mutated alleles coding deregulated key enzymes.

In *E. coli*, as *thrA*, *thrB*, and *thrC* are clustered in the *thr* operon, and aspartate kinase I encoded by *thrA* exists most abundantly among the three aspartate kinases, engineering of the L-threonine branch has been focused on this operon (Table 14.2). In most cases, the overexpressed *thr* operon contained a desensitized *thrA* gene. Introduction of a recombinant plasmid containing the *thr* operon from an L-threonine producer *E. coli* βIM4 into the very donor strain led to an increase in L-threonine production by threefold reaching 13.4 g/L (Miwa et al. 1983). Similarly, introduction of a recombinant plasmid containing the mutant $thrA_{442}BC$ operon into its donor strain *E. coli* MG442 also increased the L-threonine production from 8 to 18.4 g/L (Livshits et al. 2003). Recently, the wild-type *thr* operon and the mutant $thrA_{345}BC$ operon, both under their native promoters, were separately overexpressed on a high-copy number plasmid pMD19-T in the wild-type strain *E. coli* W3110. Overexpression of the *thr* operon and the $thrA_{345}BC$ operon both apparently increased the L-threonine production, from 0.036 to 2.590 and 9.223 g/L respectively (Zhang et al. 2009).

Table 14.2 Effects of metabolic engineering strategies on L-threonine production

Strategies	Strains	Increment[a] (fold)	Yield (g/L)	Reference
Overexpressing genes of L-threonine biosynthesis pathway	*E. coli*			
	βIM4(pBR322-*thrA*^{r})	3	13.4	Miwa et al. (1983)
	MG422(pAYC32-*thrA*^{r}*BC*)	1.3	18.4	Livshits et al. (2003)
	W3110(pMD19-*thrLABC*)	71	2.59	Zhang et al. (2009)
	W3110(pMD19-*thrLA*^{r}*BC*)	255	9.223	Zhang et al. (2009)
	C. glutamicum			
	DM368-3(pEK-*hom*^{r}-*thrB*)	1.13	1.7	Eikmanns et al. (1991)
	MH20-22B-(*hom*^{r}-*thrB*)^3	>77	7.7	Reinscheid et al. (1994)
	ATCC21799(pGC42)	>118	11.8	Colon et al. (1995)
Reducing the intracellular depletion of L-threonine	*C. glutamicum*			
	DM368-2*glyA'*	0.5	1.3	Simic et al. (2002)
	DM1800*ilvA*^{M} (pET-T18*hom*^{r}-*thrB-thrE*)	0.6	4	Diesveld et al. (2009)
Enhancing L-threonine efflux	*E. coli*			
	MG422(pTrc99A-*rhtB*)	1.4	–	Kruse et al. (2002)
	MG422(pTrc99A-*rhtC*)	2	–	Kruse et al. (2002)
	MG422(pTrc99A-$thrE_{Cg}$)	2.9	–	Kruse et al. (2002)
	MG422 *rhtA23*(pAYC32-*thrA*^{r}*BC*)	1	36.3	Livshits et al. (2003)
	C. glutamicum			
	MH20-22B-(*hom*^{r}-*thrB*)^3 (pEC-T18mob2-*thrE*)	0.05	8.1	Simic et al. (2002)
	DM368-2*glyA'*(pEC-T18mob2-*thrE*)	0.15	1.5	Simic et al. (2002)
	DM368-3(pEKEx2-$rhtC_{Ec}$)	3.11	3.7	Diesveld et al. (2009)
	DM1800(pET-T18*hom*^{r}-*thrB-thrE*) (pEKEx2-$rhtC_{Ec}$)	1.56	6.4	Diesveld et al. (2009)
Systems metabolic engineering	*E. coli*			
	TH28C (pBRThrABCR3)	–	82.4	Lee et al. (2007)
	MDS-205	–	40.1	Lee et al. (2009)

[a]Approximate value

In *C. glutamicum*, the strategy of co-expressing the three key enzymes coding genes, *lysC*, *hom* and *thrB*, has been confirmed feasible. Overexpression of *lysC* is the prerequisite for L-threonine production as aspartate kinase controls the total carbon-influx into the biosynthesis pathways of aspartic family of amino acids. Therefore *C. glutamicum* strains carrying a $lysC^r$ on the chromosome were mostly used as base strains for metabolic engineering (Colon et al. 1995; Eikmanns et al. 1991; Reinscheid et al. 1994). The mutation of S301Y in chromosomal *lysC* which conferred S-(2-aminoethyl)-l-cysteine-resistance on aspartate kinase additionally enhances the downstream *asd* expression, probably due to the accidental formation of a stronger internal promoter (Kalinowski et al. 1991). Although *C. glutamicum* ($lysC^r$) strains are L-lysine producers due to the absence of regulation for the L-lysine branch in this organism (Wittmann 2010), overexpression of the hom^r-*thrB* operon redirects the carbon flux from L-lysine branch into L-threonine branch, leading to the accumulation of both L-threonine and L-homoserine (Colon et al. 1995; Reinscheid et al. 1994). Introduction of a recombinant plasmid containing the hom^r-*thrB* operon into *C. glutamicum* DM368-3 (AEC^r, AHV^r) doubled L-threonine production from 0.8 to 1.7 g, meanwhile reduced L-lysine production from 1.3 to 0.018 g/L (Eikmanns et al. 1991). Integration of three additional copies of the hom^r-*thrB* operon into the chromosome in the L-lysine producing strain *C. glutamicum* MH20-22B resulted in the production of 7.7 g/L L-threonine and 2.5 g/L L-homoserine and a significant decrease in L-lysine production from 30.4 to 8.5 g/L (Reinscheid et al. 1994). Further increasing the expression of *thrB* to a higher level than that of *hom* efficiently circumvents L-homoserine accumulation (Colon et al. 1995). Introduction of a recombinant plasmid, on which hom^r was expressed constitutively under its native promoter and *thrB* was expressed inductively under the P_{tac} promoter, into *C. lactofermentum* ATCC21799 (AEC^r) led to an L-threonine production of 11.8 g/L with no L-homoserine accumulation, and a dramatic reduction in L-lysine production from 22.0 to 0.8 g/L (Colon et al. 1995).

14.4.2 Enhancing L-Threonine Efflux

When L-threonine accumulates to a certain level in the cell, the prompt secretion becomes limiting for production. Even for Gram-negative *E. coli*, whose cell envelope is not considered as a permeation barrier for amino acids, the intracellular L-threonine concentration exceeded the one detected in medium throughout the fermentation course, with a tenfold excess observed in early stages (Kruse et al. 2002). Gram-positive *C. glutamicum* possesses a characteristic cell wall structure, containing an outer layer of mycolic acids (Eggeling and Sahm 2001), the amino acid efflux in *C. glutamicum* is therefore significantly impaired. High intracellular concentrations of L-threonine down-regulate the biosynthesis enzymes in a feed-back manner, increase the precursor availability for the depletion pathway, and even inhibit cell growth. The strategy of overexpressing the specific permease-coding genes, $rhtA_{Ec}$, $rhtB_{Ec}$, $rhtC_{Ec}$ *and* $thrE_{Cg}$, has been confirmed effective in accelerating L-threonine secretion and thus contributing to an increased production.

In *E. coli*, individual overexpression of $rhtB_{Ec}$, $rhtC_{Ec}$ and $thrE_{Cg}$ on an episomal plasmid pTrc99A in strain MG422 increased the L-threonine production by 140, 200 and 290 %, respectively, as compared with the control strain MG422 (pTrc99A) (Kruse et al. 2002). In another case, strengthening the transcription of *rhtA* in strain MG422 (pAYC32-*thrA*^{r}*BC*), by introducing a point mutation (G→A) one base upstream the start codon of *rhtA* on the chromosome, increased L-threonine production from 18.4 to 36.3 g/L (Livshits et al. 2003).

As in the case of *C. glutamicum*, introduction of a recombinant plasmid containing $thrE_{Cg}$ into strain DM368-2*glyA'* led to an increase in L-threonine production from 1.3 to 1.5 g/L and a reduction in glycine accumulation. When the same recombinant plasmid was introduced into strain MH20-22B-(*hom*^{r}*-thrB*)[3], L-threonine production increased from 5.8 to 8.1 g/L with reduced accumulation of L-lysine, glycine and L-isoleucine (Simic et al. 2002). Overexpression of $rhtA_{Ec}$, $rhtC_{Ec}$ and $yeaS_{Ec}$ in different *C. glutamicum* strains also has a positive effect on L-threonine production. The best result was obtained when $rhtC_{Ec}$ gene was expressed on an episomal plasmid pEKEx2, which increased L-threonine production in strain DM368-3 (AEC^r, AHV^r) from 0.9 to 3.7 g/L and in DM1800 (pET-T18*hom*^{r}*-thrB-thrE*) from 4 to 6.4 g/L without L-homoserine accumulation (Diesveld et al. 2009).

14.4.3 Reducing the Intracellular Depletion of L-Threonine

The strategy of reducing L-threonine depletion towards L-isoleucine and glycine in the cell has been applied in both *E. coli* and *C. glutamicum*, and confirmed effective in increasing L-threonine production and reducing by-products formation. Reducing intracellular L-threonine conversion towards L-isoleucine in *E. coli* with genotype of *relA*$^+$ also results in activation of the *thr* operon via a "stringent response" mechanism in addition to a derepression effect (Debabov 2003). Lee et al. (2007) attenuated L-threonine depletion in *E. coli* by introducing a mutation of C290T into the chromosomal *ilvA* gene and deleting the chromosomal *tdh* gene. Simic et al. (2002) weakened the conversion of L-threonine to glycine in *C.glutamicum* DM368-2 (*lysC*^{r}, *hom*^{r}) by down-mutating the promoter of *glyA*. As a result, L-threonine production was increased to 1.3 g/L, and glycine accumulation was reduced from 0.5 to 0.3 g/L in minimal media. Similarly, Diesveld et al. (2009) cut down the conversion of L-threonine to L-isoleucine in *C.glutamicum* DM1800-T by down-mutating the promoter of *ilvA*, which increased L-threonine production from 2.5 to 4 g/L.

14.5 Systems Metabolic Engineering

Application of systems metabolic engineering gives rise to breakthroughs in strain construction. So far, a few *E. coli* L-threonine hyper-producers have been constructed by this means, and the total conversion rate of glucose to L-threonine and biomass approximates to the predicted theoretical values.

The most typical case of this strategy is the construction of *E. coli* TH28C (pBRThrABCR3) by Lee et al. (2007). A *lacI*$^-$ mutant strain of *E. coli* W3110 was used as the base strain so that promoters such as P_{tac} and P_{trc} could initiate transcription constitutively. First, the carbon influx for L-threonine biosynthesis was condensed by three steps. The feed-back inhibitions of aspartate kinase I and III were released through site-directed mutagenesis of C1034T in *thrA* and C1055T in *lysC* in chromosome; the feed-back repression to the chromosomal *thr* operon was released by substituting its native promoter with P_{tac}; and the deregulated *thr* operon was overexpressed on an episomal vector. Secondly, the carbon depleting pathways were attenuated. The production of L-lysine was blocked by deleting *lysA* encoding the last enzyme involved in L-lysine branch; the L-methionine branch was shut down by deleting *metX*; the depletion of L-threonine towards glycine was reduced by deleting *tdh*; and the depletion of L-threonine towards L-isoleucine was ablated by down-regulating threonine dehydratase activity through mutagenesis of C290T in *ilvA*. Thirdly, the availability of oxaloacetate was increased by enhancing the transcription of the *ppc* gene through substituting its native promoter with a stronger promoter P_{trc}, and enhancing the glyoxylate shunt through deleting the *iclR* gene encoding the repressor of isocitrate lyase and malate synthase. Fourthly, the efflux of L-threonine was enhanced by deleting the *tdcC* gene encoding an uptake-carrier and overexpressing *rhtA*, *rhtB* and *rhtC* on the same episomal vector as for the *thr* operon. Both the third and fourth steps were performed according to the transcriptome and *in silico* flux analysis. Finally, the production of acetate was attenuated by strengthening the transcription of the *acs* gene encoding acetyl-CoA synthetase through substituting its native promoter with the stronger promoter P_{trc}. The constructed strain TH28C (pBRThrABCR3) could produce 82.4 g/L L-threonine in 50 h' fed-batch fermentation. The L-threonine/glucose conversion rate was 39.3 %. No lactate was formed and the accumulation of acetate was 2.35 g/L.

In another successful case, Lee et al. (2009) constructed a plasmid-free L-threonine hyper-producer of *E. coli*, MDS-205, from a reduced-genome strain *E. coli* MDS42. In MDS-205, the *thr* operon together with its native promoter was substituted with the L-threonine resistant *thrA*^{r}*BC* operon under the stronger promoter P_{tac} on the chromosome, and the *lacI* gene encoding the P_{tac} repressor was deleted. The *tdh* gene was deleted to reduce the L-threonine depletion. The L-threonine uptake facilitator-coding genes, *tdcC* and *sstT*, were substituted with a mutant exporter gene *rhtA23*, which not only blocked the L-threonine uptake but also enhanced its export. The constructed strain showed robust growth and better performance in high cell-density fermentations. It produced 40.1 g/L L-threonine in 30 h' batch-fermentation with an L-threonine/glucose conversion rate of 39.3 % (Lee et al. 2009).

14.6 Conclusion

With the identification of more genes encoding the enzymes involved in L-threonine biosynthesis and the unveiling of the molecular mechanisms of regulation, metabolic engineering for L-threonine production has achieved preliminary success.

Systems metabolic engineering that globally modifies the metabolic network has shown its superiority in effectiveness over simply engineering one or two properties of the strain. Information collected from "Post genome study" on traditionally-bred L-threonine producers contributes to completing our knowledge of cellular metabolism and L-threonine biosynthesis (Kim et al. 2004; Lee et al. 2003, 2007). Knowledge gained from kinetic studies on biosynthesis enzymes reveals the essential biochemical characteristics of L-threonine biosynthesis (Chassagnole et al. 2001; Curien et al. 2009; Rodríguez-Prados et al. 2009). Evolution of methods and tools (Pátek and Nešvera 2011; Tan et al. 2012; Tyo et al. 2009; Xu et al. 2010) accelerates the engineering process. Fast development of systems biology and synthetic biotechnology enables genome-wide reprogramming of cells (Feist et al. 2007; Feist and Palsson 2008; Oberhardt et al. 2009; Shinfuku et al. 2009; Tyo et al. 2010). Thus it is promising to engineer microbes into more productive platforms for L-threonine production.

Acknowledgements This work was supported by the National key Basic Research Program of China (2012CB725202), the Basic Research Programs of Jiangsu Province (BK2009003), and the 111 Project (No. 111-2-06).

References

Ajinomoto (2009) http://www.ajinomoto.com/ir/pdf/Feed-useAA-Oct2009.pdf. Feed-use amino acids business, Accessed on Oct 2009

Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási AL (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature 427:839–843

Archer J, Solow-Cordero E, Sinskey AJ (1991) A C-terminal deletion in *Corynebacterium glutamicum* homoserine dehydrogenase abolishes allosteric inhibition by L-threonine. Gene 107:53–59

Becker J, Wittmann C (2011) Systems and synthetic metabolic engineering for amino acid production – the heartbeat of industrial strain development. Curr Opin Biotech 23:1–9

Bell SC, Turner JM (1976) Bacterial catabolism of threonine. Threonine degradation initiated by L-threonine-NAD + oxidoreductase. Biochem J 156(2):449–458

Bologna FP, Andreo CS, Drincovich MF (2007) *Escherichia coli* malic enzymes: two isoforms with substantial differences in kinetic properties, metabolic regulation, and structure. J Bacteriol 189:5937–5946

Boy E, Patte JC (1972) Multivalent repression of aspartic semialdehyde dehydrogenase in *Escherichia coli* K-12. J Bacteriol 112:84–92

Chassagnole C, Raïs B, Quentin E, Fell DA, Mazat JP (2001) An integrated study of threonine-pathway enzyme kinetics in *Escherichia coli*. Biochem J 356:415–423

Chipman DM, Shaanan B (2001) The ACT domain family. Curr Opin Struc Biol 11:694–700

Colon GE, Jetten MS, Nguyen TT, Gubler ME, Follettie MT, Sinskey AJ, Stephanopoulos G (1995) Effect of inducible *thrB* expression on amino acid production in *Corynebacterium lactofermentum* ATCC 21799. Appl Environ Microbiol 61:74–78

Cremer J, Treptow C, Eggeling L, Sahm H (1988) Regulation of enzymes of lysine biosynthesis in *Corynebacterium glutamicum*. J Gen Microbiol 134:3221–9

Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cárdenas ML, Dumas R (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. Mol Syst Biol 5:271–284

Debabov VG (2003) The threonine story. Adv Biochem Eng Biot 79:113–136
Diesveld R, Tietze N, Fürst O, Reth A, Bathe B, Sahm H, Eggeling L (2009) Activity of exporters of *Escherichia coli* in *Corynebacterium glutamicum*, and their use to increase L-threonine production. J Mol Microbiol Biotechnol 16:198–207
Eggeling L, Sahm H (2001) Review: the cell wall barrier of *Corynebacterium glutamicum* and amino acid efflux. J Biosci Bioeng 92:201–213
Eggeling L, Sahm H (2003) New ubiquitous translocators: amino acid export by *Corynebacterium glutamicum* and *Escherichia coli*. Arch Microbiol 180:155–160
Eikmanns BJ, Metzger M, Reinscheid D, Kircher M, Sahm H (1991) Amplification of three biosynthesis genes in *Corynebacterium glutamicum* and its influence on carbon flux in different strains. Appl Microbiol Biotechnol 34:617–622
Fazel A, Mueller K, Le Bras G, Garel JR, Veron M, Cohen GN (1983) A triglobular model for the polypeptide chain of aspartokinase I-homoserine dehydrogenase I of *Escherichia coli*. Biochemistry 22:158–165
Feist AM, Palsson B (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotechnol 26:659–667
Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson B (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121–138
Follettie MT, Shin HK, Sinskey AJ (1988) Organization and regulation of the *Corynebacterium glutamicum hom-thrB* and *thrC* loci. Mol Microbiol 2:53–62
Fotheringham IG, Dacey SA, Taylor PP, Smith TJ, Hunter MG, Finlay ME, Primrose SB, Parker DM, Edwards RM (1986) The cloning and sequence analysis of the *aspC* and *tyrB* genes from *Escherichia coli* K12. Comparison of the primary structures of the aspartate aminotransferase and aromatic aminotransferase of *E. coli* with those of the pig aspartate aminotransferase isoenzymes. Biochem J 234:593–604
Furukawa S, Ozaki A, Nakanishi T (1988) L-threonine production by L-aspartate- and L-homoserine-resistant mutant of *Escherichia coli*. Appl Microbiol Biotechnol 29:550–553
Gardner JF (1979) Regulation of the threonine operon: tandem threonine and isoleucine codons in the control region and translational control of transcription termination. Proc Natl Acad Sci USA 76(4):1706–1710
Gardner JF, Reznikoff WS (1978) Identification and restriction endonuclease mapping of the threonine operon regulatory region. J Mol Biol 126:241–258
Gourdon P, Baucher M-F, Lindley ND, Guyonvarch A (2000) Cloning of the malic enzyme gene from *Corynebacterium glutamicum* and role of the enzyme in lactate metabolism. Appl Environ Microbiol 66(7):2981–2987
Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA Websuite. Nucleic Acids Res 36:70–74
Grundy FJ, Lehman SC, Henkin TM (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. Proc Natl Acad Sci USA 100:12057–12062
Hartman JLIV (2007) Buffering of deoxyribonucleotide pool homeostasis by threonine metabolism. Proc Natl Acad Sci USA 104(10):11700–11705
Henkin TM, Yanofsky C (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays 24:700–707
James CL, Viola RE (2002) Production and characterization of bifunctional enzymes. Domain swapping to produce new bifunctional enzymes in the aspartate pathway. Biochemistry 41(11):3720–3725
Jetten MSM, Follettie MT, Sinskey AJ (1995) Effect of different levels of aspartokinase on the lysine production by *Corynebacterium lactofermentum*. Appl Microbiol Biotechnol 43:76–82
Kalinowski J, Bachmann B, Thierbach G, Pühler A (1990) Aspartokinase genes *lysC*α and *lysC*β overlap and are adjacent to the aspartate β-semialdehyde dehydrogenase gene asd in *Corynebacterium glutamicum*. Mol Gen Genet 224:317–324

Kalinowski J, Cremer J, Bachmann B, Eggeling L, Sahm H, Pühler A (1991) Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. Mol Microbiol 5:1197–1204

Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfefferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahm H, Wendisch VF, Wiegräbe I, Tauch A (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J Biotechnol 104:5–25

Katinka M, Cossart P, Sibilli L, Saint-Girons I, Chalvignac MA, Le Bras G, Cohen GN, Yaniv M (1980) Nucleotide sequence of the *thrA* gene of *Escherichia coli*. Proc Natl Acad Sci USA 77:5730–5733

Kato C, Kurihara T, Kobashi N, Yamane H, Nishiyama M (2004) Conversion of feedback regulation in aspartate kinase by domain exchange. Biochem Bioph Res Co 316:802–808

Kim YH, Park JS, Cho JY, Cho KM, Park YH, Lee J (2004) Proteomic response analysis of a threonine-overproducing mutant of *Escherichia coli*. Biochem J 381:823–829

Klaffl S, Eikmanns BJ (2010) Genetic and functional analysis of the soluble oxaloacetate decarboxylase from *Corynebacterium glutamicum*. J Bacteriol 192:2604–2612

Komatsubara S, Kisumi M, Murata K, Chibata I (1978) Threonine production by regulatory mutants of *Serratia marcescens*. Appl Environ Microbiol 35:834–840

Kotaka M, Ren J, Lockyer M, Hawkins AR, Stammers DK (2006) Structures of R- and T-state *Escherichia coli* aspartokinase III mechanisms of the allosteric transition and inhibition by lysine. J Biol Chem 281:31544–31552

Kruse D, Krämer R, Eggeling L, Rieping M, Pfefferle W, Tchieu JH, Chung YJ, Saier MH Jr, Burkovski A (2002) Influence of threonine exporters on threonine production in *Escherichia coli*. Appl Microbiol Biotechnol 59(2–3):205–210

Lee SH, Hwang BJ (2003) Methionine biosynthesis and its regulation in *Corynebacterium glutamicum*: parallel pathways of transsulfuration and direct sulfhydrylation. Appl Microbiol Biotechnol 62:459–467

Lee JH, Lee DE, Lee BU, Kim HS (2003) Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. J Bacteriol 185:5442–5451

Lee KH, Park JH, Kim TY, Kim YU, Lee SY (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149–156

Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC (2009) Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb Cell Fact 8:2–13

Leuchtenberger W, Huthmacher K, Drauz K (2005) Biotechnological production of amino acids and derivatives: current status and prospects. Appl Microbiol Biotechnol 69:1–8

Liu JQ, Dairi T, Itoh N, Kataoka M, Shimizu S, Yamada H (1998) Gene cloning, biochemical characterization and physiological characterization of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. Eur J Biochem 255:220–226

Livshits VA, Zakataeva NP, Aleshin VV, Vitushkina MV (2003) Identification and characterization of the new gene *rhtA* involved in threonine and homoserine efflux in *Escherichia coli*. Res Microbiol 154:123–135

Marienhagen J, Kennerknecht N, Sahm H, Eggeling L (2005) Functional analysis of all aminotransferase proteins inferred from the genome sequence of *Corynebacterium glutamicum*. J Bacteriol 187:7639–7646

Mateos LM, Pisabarro A, Pátek M, Malumbres M, Guerrero C, Eikmanns BJ, Sahm H, Martin JF (1994) Transcriptional analysis and regulatory signals of the *hom-thrB* cluster of *Brevibacterium lactofermentum*. J Bacteriol 176:7362–7371

Miwa K, Tsuchida T, Kurahashi O, Nakamori S, Sano K, Momose H (1983) Construction of L-threonine overproducing strains of *Escherichia coli* K-12 using recombinant DNA techniques. Agric Biol Chem 47:2329–2334

Miyajima R, Otsuka SI, Shiio I (1968) Regulation of aspartate family amino acid biosynthesis in *Brevibactenium flavum*; I. Inhibition by amino acids of the enzymes in threonine biosynthesis. J Biochem 63:139–148

Mockel B, Eggeling L, Sahm H (1992) Functional and structural analyses of threonine dehydratase from *Corynebacterium glutamicum*. J Bacteriol 174:8065–8072

Morinage Y, Takagi H, Ishida M, Miwa K, Sato T, Nakamori S, Sano K (1987) Threonine production by co-existence of cloned genes coding homoserine dehydrogenase and homoserine kinase in *Brevibacterium lactofermentum*. Agric Biol Chem 51:93–100

Oberhardt MA, Palsson B, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320–334

Palmieri L, Berns D, Krämer R, Eikmanns M (1996) Threonine diffusion and threonine transport in *Corynebacterium glutamicum* and their role in threonine production. Arch Microbiol 165:48–54

Pátek M, Nešvera J (2011) Sigma factors and promoters in *Corynebacterium glutamicum*. J Biotechnol 154:101–113

Reinscheid DJ, Eikmanns BJ, Sahm H (1991) Analysis of a *Corynebacterium glutamicum hom* gene coding for a feedback-resistant homoserine dehydrogenase. J Bacteriol 173:3228–3230

Reinscheid DJ, Kronemeyer W, Eggeling L, Eikmanns BJ, Sahm H (1994) Stable expression of *hom-1–thrB* in *Corynebacterium glutamicum* and its effects on the carbon flux to threonine and related amino acids. Appl Environ Microbiol 60:126–132

Rodríguez-Prados JC, de Atauri P, Maury J, Ortega F, Portais JC, Chassagnole C, Acerenza L, Lindley ND, Cascante M (2009) *In silico* strategy to rationally engineer metabolite production: a case study for threonine in *Escherichia coli*. Biotechnol Bioeng 103:609–620

Rückert C, Pühler A, Kalinowski J (2003) Genome-wide analysis of the L-methionine biosynthetic pathway in *Corynebacterium glutamicum* by targeted gene deletion and homologous complementation. J Biotech 104:213–228

Sauer U, Eikmanns BJ (2005) The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS Microbiol Rev 29:765–794

Schrumpf B, Schwarzer A, Kalinowski J, Puhler A, Eggeling L, Sahm H (1991) A functionally split pathway for lysine synthesis in *Corynebacterium glutamicium*. J Bacteriol 173:4510–4516

Shiio I, Miyajima R (1969) Concerted inhibition and its reversal by end products of aspartate kinase in *Brevibacterium flavum*. J Biochem 64:849–859

Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H (2009) Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. Microb Cell Fact 8:43–57

Simic P, Sahm H, Eggeling L (2001) L-threonine export: use of peptides to identify a new translocator from *Corynebacterium glutamicum*. J Bacteriol 183:5317–5324

Simic P, Willuhn J, Sahm H, Eggeling L (2002) Identification of *glyA* (Encoding serine hydroxymethyl transferase) and its use together with the exporter ThrE To increase L-threonine accumulation by *Corynebacterium glutamicum*. Appl Environ Microbiol 68:3321–3327

Tan YZ, Xu DQ, Li Y, Wang X (2012) Construction of a novel *sacB*-based system for marker-free gene deletion in *Corynebacterium glutamicum*. Plasmid 67:44–52

Theze J, Saint-Girons I (1974) Threonine locus of *Escherichia coli* K-12: genetic structure and evidence for an operon. J Bacteriol 118:990–998

Truffa-Bachi P, Veron M, Cohen GN (1974) Structure, function, and possible origin of a bifunctional allosteric enzyme, *Escherichia coli* aspartokinase I-homoserine dehydrogenase I. Crit Rev Biochem Mol Biol 2:379–415

Tyo KEJ, Ajikumar PK, Stephanopoulos G (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol 27:760–767

Tyo KEJ, Kocharin K, Nielsen J (2010) Toward design-based engineering of industrial microbes. Curr Opin Microbiol 13:255–262

Umbarger HE (1973) Threonine deaminase. Adv Enzymol 37:349–395

Umbarger HE, Brown B (1957) Threonine deamination in *Escherichia coli.* II. Evidence for two L-threonine deaminases. J Bacteriol 73:105–112

Velasco AM, Leguina JI, Lazcano A (2002) Molecular evolution of the lysine biosynthetic pathways. J Mol Evol 55:445–459

Viola RE (2001) The central enzymes of the aspartate family of amino acid biosynthesis. Accounts of Chemical Research 34:339–349

Wittmann C (2010) Analysis and engineering of metabolic pathway fluxes in *Corynebacterium glutamicum*. Adv Biochem Eng Biotechnol 120:21–49

Xu DQ, Tan YZ, Huan XJ, Hu XQ, Wang X (2010) Construction of a novel shuttle vector for use in *Brevibacterium flavum*., an industrial amino acid producer. J Microbiol Meth 80(1):86–92

Yoshida A, Tomita T, Kurihara T, Fushinobu S, Kuzuyama T, Nishiyama M (2007) Structural insight into concerted inhibition of α2β2-type aspartate kinase from *Corynebacterium glutamicum*. J Mol Biol 368:521–536

Yoshida A, Tomita T, Kuzuyama T, Nishiyama M (2010) Mechanism of concerted inhibition of α2β2-type hetero-oligomeric aspartate kinase from *Corynebacterium glutamicum*. J Biol Chem 285:27477–27486

Zhang X, Yan JA, Yu L, Zhang GQ, Zhang Y, Chen N, Wen T (2009) Construction of recombinant plasmids containing threonine operon and their effects on L-threonine accumulation. Acta Microbiol Sin 49:591–596

Chapter 15
The Production of Coenzyme Q10 in Microorganisms

Corinne P. Cluis, Dominic Pinel, and Vincent J. Martin

Abstract Coenzyme Q10 has emerged as a valuable molecule for pharmaceutical and cosmetic applications. Therefore, research into producing and optimizing coenzyme Q10 via microbial fermentation is ongoing. There are two major paths being explored for maximizing production of this molecule to commercially advantageous levels. The first entails using microbes that naturally produce coenzyme Q10 as fermentation biocatalysts and optimizing the fermentation parameters in order to reach industrial levels of production. However, the natural coenzyme Q10-producing microbes tend to be intractable for industrial fermentation settings. The second path to coenzyme Q10 production being explored is to engineer *Escherichia coli* with the ability to biosynthesize this molecule in order to take advantage of its more favourable fermentation characteristics and the well-understood array of genetic tools available for this bacteria. Although many studies have attempted to over-produce coenzyme Q10 in *E. coli* through genetic engineering, production titres still remain below those of the natural coenzyme Q10-producing microorganisms. Current research is providing the knowledge needed to alleviate the bottlenecks involved in producing coenzyme Q10 from an *E. coli* strain platform and the fermentation parameters that could dramatically increase production titres from natural microbial producers. Synthesizing the lessons learned from both approaches may be the key towards a more cost-effective coenzyme Q10 industry.

Keywords Coenzyme Q10 • *Escherichia coli* • Metabolic engineering • *Rhizobium radiobacter* • Ubiquinone

C.P. Cluis • D. Pinel • V.J. Martin (✉)
Department of Biology, Concordia University,
7141 Sherbrooke West, Montréal H4B 1R6, Québec, Canada
e-mail: vmartin@alcor.concordia.ca

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_15,

Abbreviations

8P-Ph	2-Octaprenylphenol
10P-Ph	2-Decaprenylphenol
CoQn	Coenzyme Qn
DCW	Dry cell weight
DMAPP	Dimethylallyl diphosphate
DNP	2,4-Dinitrophenol
DO	Dissolved oxygen
DPP	Decaprenyl diphosphate
DPS	Decaprenyl diphosphate synthase
DXP	1-Deoxy-D-xylulose 5-phosphate
E4P	Erythrose-4-phosphate
FPP	Farnesyl diphosphate
G3P	Glyceraldehyde-3-phosphate
IPP	Isopentenyl diphosphate
MEP	2-C-methyl-D-erythritol 4-phosphate
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
OPP	Octaprenyl diphosphate
PEP	Phosphoenolpyruvate
PHB	Para-hydroxybenzoic acid

15.1 Introduction

15.1.1 Physiological Functions and Applications of Coenzyme Q10 in Medicine

Coenzyme Qn, also known as ubiquinone, is a lipid-soluble molecule with high redox potential. It is formed of a benzoic group conjugated to a long isoprenoid chain with a varying number (*n*) of isoprene units (Fig. 15.1). Humans naturally produce coenzyme Q10 (CoQ10), which plays a major role in the transfer of electrons between respiratory complexes of the electron transport chain, located within the inner mitochondrial membrane (Kroger and Klingenberg 1973). CoQ10 also acts as an antioxidant, protecting membranes and proteins from oxidation by scavenging free radicals and by regenerating pools of tocopherols (Lass and Sohal 1998; Yoshida et al. 2006; Martin et al. 2007). There is evidence that CoQ10 is involved in the transcriptional regulation of genes, some of which play roles in inflammatory responses and in cholesterol metabolism (Schmelzer et al. 2007, 2010). CoQ10 supplements have been demonstrated to have positive effects on patients suffering from certain cardio-vascular conditions, such as conjunctive heart failure and acute myocardial infarction (Singh et al. 1998; Hodgson et al. 2002; Shults et al. 2002; Yang et al. 2010). Additionally, it has been shown that patients taking statin drugs

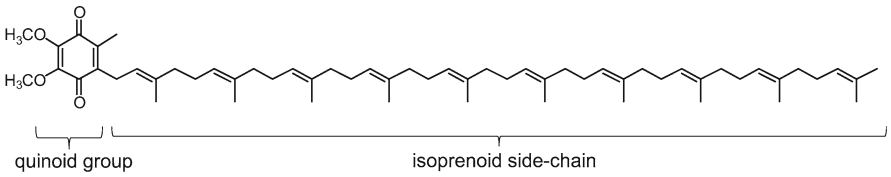


Fig. 15.1 Structure of coenzyme Q10

(HMG-CoA reductase inhibitors) to reduce cholesterol levels may suffer from diminished blood ubiquinone levels, which may be linked to muscle problems like myopathy and myositis that could be alleviated through CoQ10 supplementation (Schaars and Stalenhoef 2008). Furthermore, early research has suggested that CoQ10 supplementation may benefit patients suffering from male infertility, neurodegenerative disease and diabetes-associated nephropathy (Yang et al. 2010; Mancini and Balercia 2011; Sourris et al. 2012). Another major application of CoQ10 is as an ingredient of dermatological ointments, due to evidence that its topological application can reduce wrinkle formation (Inui et al. 2008).

This growing demand for CoQ10 from the pharmaceutical and cosmetic sectors has led to intensified efforts to decrease the cost of CoQ10 industrial production. Possible production methods for CoQ10 include: chemical synthesis (Negishi et al. 2002), semi-chemical synthesis using solanesol (Lipshutz et al. 2005), or by microbial fermentation using yeasts or bacteria. Microbial biosynthesis offers several advantages over chemical and semi-chemical synthesis, including specificity towards the all-trans biologically active isomer of CoQ10, and the reduced production of environmentally hazardous waste. Indeed, the viability of a microbial biosynthesis approach has been demonstrated by companies, such as Kaneka Corporation, that have already commercialized biologically-produced CoQ10 obtained from fermentation using proprietary yeast strains. In recent years, considerable research has emerged in attempts to isolate suitable microbial strains for the development of cost-effective bioprocesses for CoQ10 production.

15.1.2 Quinone Diversity Between and Within Species

The function and overall structure of ubiquinone is well conserved across species, but the length of its isoprenoid side-chain varies. For instance, humans mainly synthesize CoQ10 while rodents predominantly generate CoQ9; the budding yeast *Saccharomyces cerevisiae* makes CoQ6, but the closely related *Schizosaccharomyces pombe* makes CoQ10. Quinone diversity is also found among prokaryotes, where the ubiquinone chain length varies between eight-isoprene units (i.e. *Escherichia coli*, CoQ8), nine-isoprene units (i.e. *Acetobacter aceti*, CoQ9) or ten-units (i.e. *Paracoccus denitrificans*, CoQ10). In addition, prokaryotes may produce alternative quinone types bearing a naphthalene head group, depending on their species. These naphthoquinones include menaquinone and demethylmenaquione. Typically, strictly aerobic gram-negative species of bacteria contain solely ubiquinone, while gram-negative facultative

aerobic bacteria use both ubiquinone and menaquinone or demethylmenaquinone. Gram-positive bacteria, on the other hand, appear to largely use menaquinone or demethylmenaquinone (Collins and Jones 1981).

15.1.3 Physiological Functions of Ubiquinone in Prokaryotes

Similar to eukaryotes, prokaryotes use ubiquinone as an electron carrier during respiration. In *E. coli*, ubiquinone is used when O_2 or nitrate are the final electron acceptors. Menaquinone and demethylmenaquinone, which have lower midpoint potentials than ubiquinone, are used during anaerobic respiration when final electron acceptors, such as fumarate, have a low midpoint potential (Ingledew and Poole 1984). The redox state of ubiquinone also serves as a cue for cells to initiate aerobic or anaerobic cell functions. For example, in *E. coli*, the oxidized form of CoQ8 has been demonstrated to act as a direct negative signal for the Arc two-component signal transduction system (Georgellis et al. 2001). This regulatory switch consists of the membrane-associated sensor kinase ArcB, which undergoes autophosphorylation under anaerobic conditions, turning on its kinase activity. Activated ArcB then phosphorylates the transcription regulator ArcA, resulting in the transcription of genes involved in anaerobic metabolism and in the repression of genes taking part in aerobic metabolism. Therefore, by inhibiting the autophosphorylation of ArcB during aerobiosis, CoQ8 thus prevents the onset of anaerobic metabolism (Georgellis et al. 2001; Malpica et al. 2004). Also in *E. coli*, CoQ8 participates in disulfide bond formation by providing oxidizing power to the cytoplasmic membrane protein DsbB, which then oxidizes the periplasmic enzyme DsbA, responsible for introducing disulfide bonds to newly synthesized proteins (Bader et al. 2000). The many roles for CoQ8 that have been elucidated in *E. coli* demonstrate the importance quinones may have in many facets of cellular functioning.

15.1.4 Biosynthesis of Ubiquinone in Prokaryotes

Genetic and biochemical studies have led to almost complete elucidation of the CoQ8 biosynthesis pathway in *E. coli*. Homologues of genes involved in this pathway were identified in several other bacterial species, suggesting that ubiquinone synthesis is highly conserved among prokaryotes (Fig. 15.2).

Fig. 15.2 (continued) (26) 2-polyprenyl-6-methoxyphenol; (27) 2-polyprenyl-6-methoxy-1,4-benzoquinol; (28) 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol; (29) 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol; (30) ubiquinol; (31) ubiquinone (CoQn). Abbreviations: *ATP* adenosine triphosphate, *AMP* adenosine monophosphate, *NADPH/NADP*$^+$ nicotinamide adenine dinucleotide phosphate, *CTP* cytidine triphosphate, *CMP* cytidine monophosphate, $Fd_{red/ox}$ reduced/oxidized ferredoxin, *SAM S*-adenosylmethionine, *SAH S*-adenosylhomocysteine

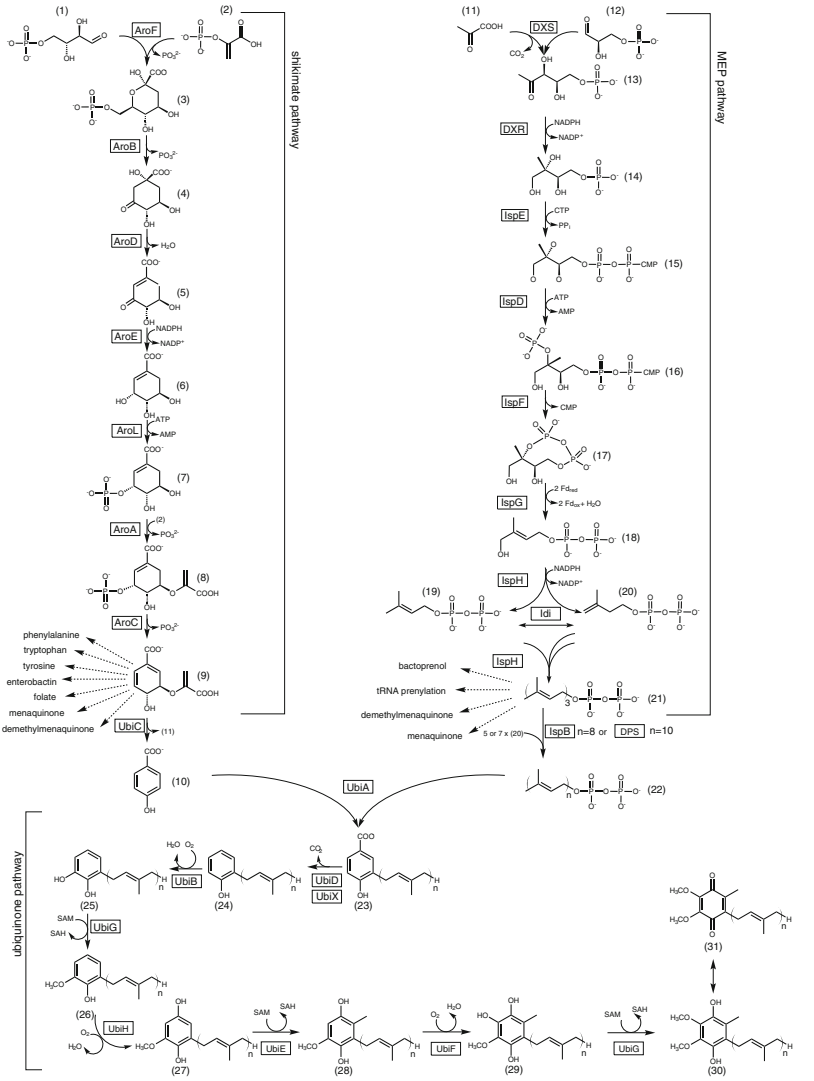


Fig. 15.2 Biosynthetic pathway of ubiquinone in prokaryotes. Enzyme names and intermediates are inferred from research studies in *E. coli*. The chain length of isoprenoids and ubiquinone pathway intermediates is kept unspecified (n) to generalize the pathway model to encompass all prokaryotes. (1) erythrose-4-phosphate; (2) phosphoenolpyruvate; (3) 3-deoxy-arabino-heptulosonate 7-phosphate; (4) 3-dehydroquinate; (5) 3-dehydroshikimate; (6) shikimate; (7) shikimate-3-phosphate; (8) 5-enolpyruvyl-shikimate-3-phosphate; (9) chorismate; (10) *para*-hydroxybenzoate; (11) pyruvate; (12) glyceraldehyde-3-phosphate; (13) 1-deoxy-D-xylulose-5-phosphate; (14) 2-C-methyl-D-erythrithol-5-phosphate; (15) 4-diphosphocytidyl-2C-methyl-D-erythritol; (16) 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate; (17) 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; (18) 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate; (19) demethylallyl diphosphate; (20) isopentenyl diphosphate; (21) farnesyl diphosphate; (22) polyprenyl diphosphate; (23) 3-polyprenyl-4-hydrobenzoate; (24) 2-polyprenylphenol; (25) 2-polyprenyl-6-hydroxyphenol;

The aromatic core of the CoQn molecule is derived from the shikimate pathway, which starts with the condensation of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP). The shikimate pathway leads to the production of the aromatic intermediate chorismate, a precursor for several essential aromatic molecules including aromatic amino acids and folate, which is converted to *para*-hydroxybenzoic acid (PHB) by the chorismate lyase UbiC. PHB, the first committed intermediate to ubiquinone biosynthesis, is then prenylated with an isoprenoid of varying length, depending on the species (Fig. 15.2).

In prokaryotes, long chain isoprenoids are derived from the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. The MEP pathway leads to the formation of two isoprenoid building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). DMAPP then serves as a primer for the elongation of a 15-carbon isoprenoid by the farnesyl diphosphate (FPP) synthase IspA, which successively condenses IPP molecules to form FPP. This 15-carbon isoprenoid is the substrate for a polyprenyl diphosphate synthase, which further elongates the isoprenoid chain that will eventually form the prenyl side-chain constituent for ubiquinone synthesis (Fig. 15.2). *E. coli* endogenously expresses an octaprenyl diphosphate (OPP) synthase, IspB, leading to the formation of OPP and finally to CoQ8. In this way, the product formation catalyzed by a given polyprenyl synthase determines the chain length of the ubiquinone produced in each organism. Indeed, the heterologous expression in a microbial host of a polyprenyl diphosphate synthase, catalyzing the formation of an isoprenoid of n subunits, is sufficient for the accumulation of CoQ species with the corresponding chain length, n. For instance, the expression in *E. coli* of *sdsA*, encoding the solanesyl diphosphate synthase from *Rhodobacter capsulatus*, results in the production of both CoQ8 (from the endogenous IspB) and CoQ9 (from the heterologous SdsA) (Okada et al. 1997).

The enzymatic steps downstream of PHB and isoprenoid formation, leading to the formation of ubiquinone, further described in this paragraph, will be referred to here as the ubiquinone pathway (Fig. 15.2). The prenylation of PHB is carried out by an integral membrane protein, UbiA, forming the quinoid intermediate 3-polyprenyl-4-hydroxybenzoate (Young et al. 1972). In *E. coli*, UbiA competes for OPP with MenA, which directs OPP towards the biosynthesis of menaquinone. The product of UbiA is then decarboxylated into the intermediate 2-polyprenylphenol. In *E. coli*, two decarboxylases, UbiD and UbiX, participate in this reaction (Cox et al. 1969; Leppik et al. 1976; Gulmezian et al. 2007). The next step of the ubiquinone pathway consists of the hydroxylation of 2-polyprenylphenol into 2-polyprenyl-6-hydroxyphenol. Genetic studies in *E. coli* have linked this reaction to the product of *ubiB* (Poon et al. 2000). However, UbiB is a kinase-like protein with no homology to known hydroxylases. The specific role of UbiB in this hydroxylation reaction therefore remains to be elucidated. 2-polyprenyl-6-hydroxyphenol is then methylated by UbiG, forming 2-polyprenyl-6-methoxyphenol, which is further hydroxylated by UbiH to form 2-polyprenyl-6-methoxy-1,4-benzoquinol (Young et al. 1973; Alexander and Young 1978; Hsu et al. 1996). A second methyltransferase, UbiE, transfers a methyl group from S-adenosylmethionine to 2-polyprenyl-6-methoxy-1,4-benzoquinol, resulting in 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol

(Young et al. 1971; Lee et al. 1997). The hydroxylase UbiF then adds a hydroxyl group, forming 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol (Young et al. 1971). Finally, this last intermediate is methylated in a second reaction carried out by UbiG, completing the formation of CoQn, in its reduced form (Stroobant et al. 1972; Leppik et al. 1976).

Overall, the biosynthesis of ubiquinone is complex, incorporating precursors generated from different branches of central metabolism, requiring a great number of enzymes working in synchrony and a myriad of genes whose timely expression is essential for the proper unfolding of the pathway. Full understanding of the different steps of this biosynthetic pathway, as well as of the regulatory mechanisms underlying its functionality, is the first integral step in rationally engineering an industrially-viable biocatalyst capable of high CoQ10 yields.

15.1.5 Natural Producers of CoQ10

CoQ10 is produced in naturally high quantities by several microorganisms including strains of gram-negative soil bacteria such as *Rhizobium radiobacter* (a.k.a. *Agrobacterium tumefaciens*), *P. denitrificans,* and *Protomonas extorquens*, as well some photoautotrophs including *Rhodobacter sphaeroides* (a.k.a. *Rhodopseudomonas sphaeroides*), and yeast species such as *Sporidiobolus johnsonii,* which accumulate intracellular CoQ10 to levels between 0.8 and 3.3 mg/g dry cell weight (DCW) (Urakami and Hori-Okubo 1988; Yoshida et al. 1998; Dixson et al. 2011).

Furthermore, the innate CoQ10 content of a number of naturally producing strains has been increased through chemical mutagenesis. In order to select high CoQ10-producing mutant strains, a number of indirect phenotypes were used as surrogate screens. For example, *R. radiobacter* mutants that could produce up to twofold higher CoQ10 content than the wild type parental strain were selected for their ability to grow on daunomycin, menadinone (quinone-like molecules generating hydroxyl radicals) and/or the methionine antagonist ethionine (Myers et al. 1987; Farr and Kogoma 1991; Yoshida et al. 1998). The mutations resulting in enhanced CoQ10 production in such mutants remain largely undetermined. However, in one instance, it was determined that an *R. radiobacter* mutant strain with a high CoQ10 content displayed up-regulated expression levels of a number of genes involved in the tricarboxylic acid cycle (Koo et al. 2010). In another example of screening for enhanced CoQ10 producers, *R. sphaeroides* mutants were chosen based on their colony colour as a surrogate selection for increased CoQ10 accumulation. In this case, a green colony colour was used to indicate reduced accumulation of cellular carotenoids, which are synthesized from the same isoprenoid precursors as CoQ10 and could indicate the allocation of more cellular isoprenoid resources towards CoQ10 production (Urakami and Hori-Okubo 1988; Sakato et al. 1992; Yoshida et al. 1998). It is also possible that CoQ10 synthesis was up-regulated in carotenoid-reduced mutant strains in order to balance the loss of one antioxidant (carotenoid), with another (CoQ10). As shown in Table 15.2, the selected mutants display 1.3–3-fold increases

in CoQ10 production compared to their parental strains. Overall, the above studies suggest that using natural CoQ10 producers as fermentation biocatalysts for industrial CoQ10 production may be plausible, especially if coupled with classical strain improvement techniques.

15.1.6 R. radiobacter as a Production Platform for CoQ10

Despite their high CoQ10 contents, most natural producers of CoQ10, wild type or mutant, are unsuitable for industrial applications due to their limited growth rate. For instance, even under optimized fermentor conditions, the cell mass obtained with a mutant strain of *R. sphaeroides* did not exceed 18.6 g DCW/liter (L) (Kien et al. 2010). One exception is *R. radiobacter*, which is gaining popularity as a host for the production of CoQ10. *R. radiobacter* mutants isolated for their improved CoQ10 content can be cultivated in fermentors at densities exceeding 65 g DCW/L (Ha et al. 2007a). Such strains of *R. radiobacter* are therefore good candidates to establish an industrial bioplatform for CoQ10 generation, as their high CoQ10 contents are matched by high fermentation biomass yields. There are two accounts where *R. radiobacter* was engineered for higher CoQ10 production by either over-expressing its 1-deoxy-D-xylulose 5-phosphate (DXP) synthase or its glyceraldehyde-3-phosphate (G3P) dehydrogenase (Lee et al. 2007; Koo et al. 2010). However, most improvements in CoQ10 titres were obtained by modifications in culture growth conditions and media composition (Ha et al. 2007a, b, 2009), which demonstrates the necessity of optimizing fermentation conditions for increasing CoQ10 production. Various optimization strategies involving media composition and growth conditions are presented in Sect. 15.3.

15.1.7 E. coli as a Production Platform for CoQ10

E. coli is commonly used as a model for the heterologous production of CoQ10 in microbes. It is an attractive host for development as an industrial production platform due to the following: (1) its robust fermentation characteristics, including fast growth rate (2) its amenability to genetic engineering and (3) the extensive knowledge that exists regarding its ubiquinone pathway and primary metabolism. As a result, rational genetic engineering strategies have been developed to allow production of CoQ10 in *E. coli*, and are presented here and in Sect. 15.2. These strategies have led to improvements in the yields of CoQ10 obtained from engineered *E. coli* strains. However, such yields still fall below those obtained from natural CoQ10 producers, such as *R. radiobacter*, indicating that the factors limiting CoQ10 accumulation in *E. coli* bear further investigation.

15.2 Engineering Strategies for High CoQ10 Production in Microbes

15.2.1 Heterologous Expression of a Decaprenyl Diphosphate Synthase

As mentioned above (Sect. 15.1.2), *E. coli* natively produces CoQ8 as its predominant ubiquinone molecule. However, it is possible for *E. coli* to produce CoQ10 by the expression of a heterologous decaprenyl diphosphate (DPP) synthase (DPS), which catalyzes the sequential condensation reaction between IPP and FPP to form decaprenyl disphosphate. For example, the *dps* genes from CoQ10-producing bacterial species, including *R. radiobacter*, *R. sphaeroides*, *Gluconobacter suboxydans* and *S. baekryungensis,* have been isolated and expressed in *E. coli* (Lee et al. 2004; Park et al. 2005; Zhang et al. 2007; Cluis et al. 2011). Additionally, a given heterologous DPS must be expressed in a Δ*ispB* background, where expression of *E. coli*'s endogenous OPP synthase has been abolished, in order to eliminate CoQ8 synthesis in favour of CoQ10 production. (Okada et al. 1998; Takahashi et al. 2003; Lee et al. 2004; Park et al. 2005; Kim et al. 2006; Zahiri et al. 2006a, b; Choi et al. 2009; Cluis et al. 2011; Huang et al. 2011). *E. coli* Δ*ispB* strains expressing a foreign DPS accumulate CoQ10, along with varying levels of CoQ8, CoQ9 and/or CoQ11, depending on the DPS expressed (Okada et al. 1998; Cluis et al. 2011). Most bacteria surveyed to date produce minor amounts of differing CoQ forms with shorter-length isoprenoid side-chains in addition to their main ubiquinone species, suggesting that DPS enzymes generally make other prenyl diphosphates in addition to DPP. However, it seems that this relaxed enzyme specificity with regards to product size is exacerbated when a DPS is heterologously expressed in *E. coli*. For instance, the expression of a DPS from *S. baekryungensis* in *E. coli* results in the production of CoQ10 and CoQ11, while *S. baekryungensis* produces CoQ10 with only minor traces of CoQ9, but no detectable CoQ11 (Cluis et al. 2011). The production of shorter or longer ubiquinone species in *E. coli* strains engineered for CoQ10 production poses a problem from an industrial point of view due to the reduced yields of CoQ10 that result when precursors are channelled towards non-desired products and the extra costs associated with purifying CoQ10 from additional quinone species.

15.2.2 Increasing Production of Precursors of CoQ10

As described in Sect. 15.1.2, the formation of CoQ10 in bacteria occurs through condensation between the aromatic molecule, PHB, derived from the shikimate pathway, and the long-chain isoprenoid, DPP, formed via the MEP pathway. A current strategy to improve the heterologous production of CoQ10 in *E. coli* is to over-express rate-limiting enzymes leading to the production of these precursor molecules (Zahiri et al. 2006b; Zhang et al. 2007; Cluis et al. 2011).

15.2.2.1 Increasing the Production of Isoprenoid Precursors

The first step of the MEP pathway is the condensation of pyruvate and G3P to yield DXP, catalyzed by the DXP synthase, DXS (Fig. 15.2). The heterologous production of the carotenoid lycopene in *E. coli*, which also requires isoprenoid precursors, has been improved by the over-expression of DXS (Harker and Bramley 1999; Kim and Keasling 2001). Using the same rationale, higher titers of CoQ10 have been achieved by over-expressing the *dxs* gene isolated from *Pseudomonas aeruginosa* in *E. coli* strains engineered for CoQ10 production (Kim et al. 2006; Choi et al. 2009) (Fig. 15.3). Using this strategy, up to 1.8-fold improvement in specific CoQ10 content was obtained in strains over-expressing both DPS and DXS compared to those expressing solely DPS (Table 15.1). An alternative to the over-expression of the endogeneous MEP pathway in *E. coli* is to express the heterologous mevalonate pathway, generally found in eukaryotes. This strategy has an added advantage over the endogenous pathway in that it bypasses native regulatory mechanisms that may make up-regulation for over-production more problematic (Martin et al. 2003) (Fig. 15.3). Synthetic mevalonate pathways are typically divided into two parts: an upper pathway that converts acetyl-coA to mevalonate and a lower pathway that converts mevalonate into DMAPP and IPP. Both parts of the pathway can be co-expressed, enabling the formation of isoprenoids from acetyl-coA. Alternatively, one can express the lower part of the pathway alone and feed the microbe exogenous mevalonate to enable the production of IPP and DMAPP. The expression of the lower part of the mevalonate pathway in *E. coli* strains expressing a recombinant *dps*, has led to twofold improvements in CoQ10 accumulation (Zahiri et al. 2006b; Cluis et al. 2011), while the expression of both the upper and lower sections of the pathway has thus far led to more modest improvements (1.5-fold) (Zahiri et al. 2006b) (Table 15.1).

15.2.2.2 Increasing the Production of Aromatic Precursors

As described in Sect. 15.1.4, the first committed aromatic precursor to ubiquinone, PHB, is formed through the shikimate pathway leading to chorismate (Fig. 15.2). As chorismate is a precursor molecule in several essential metabolic routes, the conversion of chorismate to PHB may limit the rate of production of ubiquinone in *E. coli*. It has been shown that the over-expression of the chorismate lysase UbiC in engineered *E. coli* increased the production of CoQ10 when the downstream PHB polyprenyltransferase, UbiA, was coexpressed (Zhu et al. 1995; Zhang et al. 2007) (Fig. 15.3). In *S. johnsonii*, the addition of PHB resulted in up to an 8-fold increase in CoQ10 content (Dixson et al. 2011). The exogenous addition of PHB can also improve the CoQ10 content of *E. coli* cells, but only when the supply of isoprenoid precursors is concomitantly increased (Zahiri et al. 2006b). As no transporter for this organic acid has been characterized to date in *E. coli*, it is unclear to what extent exogenous PHB can be actively brought into *E. coli* cells. In order to ensure the availability of increased levels of PHB *in situ*, an alternative strategy has proven

Table 15.1 Production of CoQ10 by engineered microorganisms

[a]	Organism	Overexpressed enzymes	Gene deletions	Metabolic pathways affected	Specific CoQ10 content (mg/g DCW)	Study
1A	*E. coli*	DPS_{Gsub}	N/A	Isoprenoid	0.29	Park et al. (2005)
1B	*E. coli*	DXS	N/A	Isoprenoid (MEP)	~0.80	Harker and Bramley (1999)
1C	*E. coli*	DPS_{Gsub}, DXS	N/A	Isoprenoid (MEP)	0.94	Kim et al. (2006)
1D	*E. coli*	DPS_{Atum}, DXS	*ispB*	Isoprenoid (MEP)	1.41	Choi et al. (2009)
1E	*E. coli*	DPS_{Atum}, PhbA, MvaK1, MvaD, MvaK2, MvaA, MvaS, idi	N/A	Isoprenoid (mevalonate)	2.428	Zahiri et al. (2006b)
1F	*E. coli*	DPS_{Gsub}, Pck	*ispB, pykA, pykF*	Isoprenoid, glycolysis	0.44	Huang et al. (2011)
1G	*E. coli*	DPS_{Gsub}, GapC	*ispB, pykA, pykF, gapA*	Isoprenoid, glycolysis	0.45	Huang et al. (2011)
1H	*E. coli*	DPS_{Gsub}, DXS, Idi, IspA, UbiC, UbiA, Pck, GapC	*ispB, pykA, pykF, gapA*	Isoprenoid, glycolysis, ubiquinone	3.24	Huang et al. (2011)
1I	*E. coli*	DPS_{Atum}, UbiA	N/A	Isoprenoid, ubiquinone	0.51	Zhang et al. (2007)
1J	*E. coli*	DPS_{Atum}, UbiA, UbiC	N/A	Isoprenoid, ubiquinone	0.77	Zhang et al. (2007)
1K	*E. coli*	DPS_{Atum}, UbiG	N/A	Isoprenoid, ubiniquinone	0.19	Zhang et al. (2007)
1L	*E. coli*	UbiC, UbiA	N/A	Ubiquinone	~0.45	Zhu et al. (1995)
1M	*E. coli*	IspB, UbiC, UbiA	N/A	Isoprenoid, ubiniquinone	~0.80	Zhu et al. (1995)
1N	*E. coli*	IspB, UbiC, UbiA, UbiB	N/A	Isoprenoid, ubiquinone	~0.85	Zhu et al. (1995)
1O	*E. coli*	IspB, UbiC, UbiA, UbiG, UbiH	N/A	Isoprenoid, ubiquinone	~0.95	Zhu et al. (1995)
1P	*E. coli*	DPS_{Sbae}, $UbiA_{NAPJ}$, TktA, $AroF^{FBR}$, AroB, AroL, AroA, AroC, UbiC, Erg12, Erg8, MVD1, IspA, Idi	*ispB*	Isoprenoid (mevalonate), shikimate, ubiquinone	~0.80	Cluis et al. (2011)
1Q	*R. radiobacter*	GapA	N/A	Glycolysis	5.27	Koo et al. (2010)
1R	*R. radiobacter*	DXS		Isoprenoid	8.3	Lee et al. (2007)

[a]Listed letters are referred to those shown in Fig. 15.3

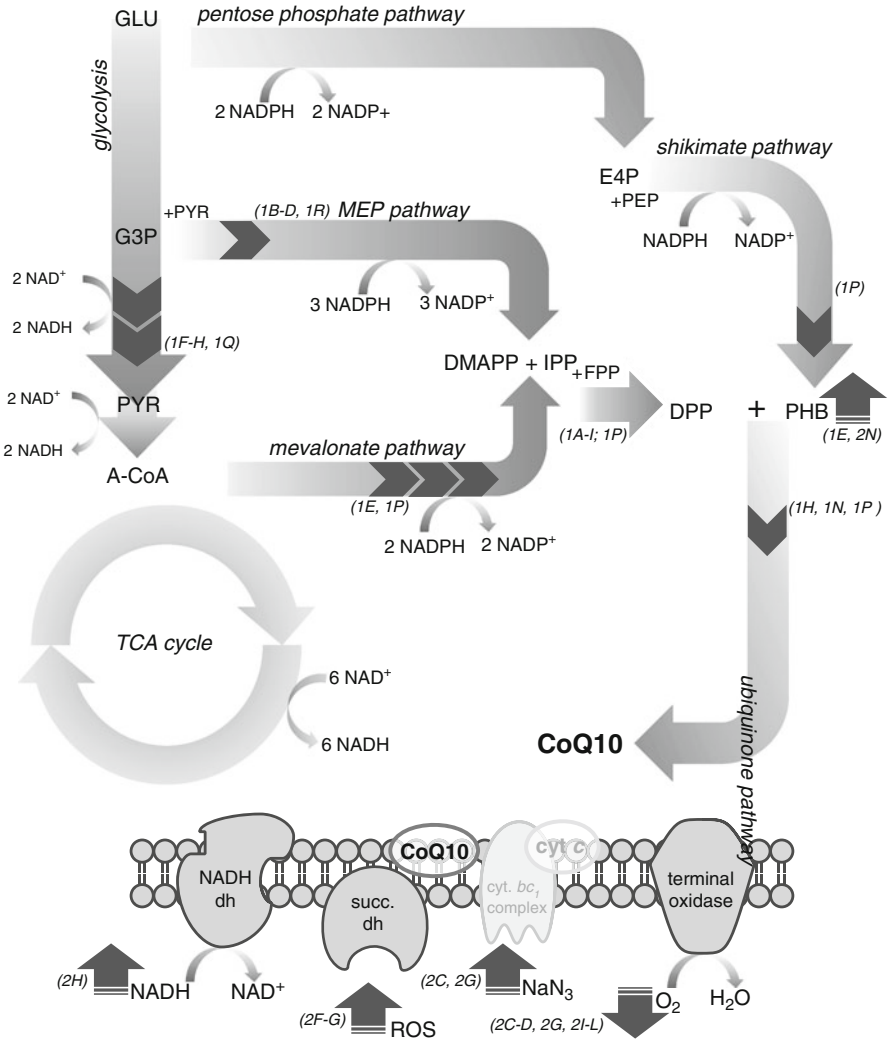


Fig. 15.3 Summary of strategies used to increase CoQ10 production in microbes. Numbers in parentheses refer to studies listed in Tables 15.1 and 15.2. Abbreviations: *GLU* glucose, *NADPH/NADP*$^+$ nicotinamide adenine dinucleotide phosphate, *NADH/NAD*$^+$ nicotinamide adenine dinucleotide, *G3P* glyceraldehydes-3-phosphate, *PYR* pyruvate, *A-CoA* acetyl coenzyme A, *TCA* tricarboxylic acid, *E4P* erythrose-4-phosphate, *PEP* phosphoenolpyruvate, *DMAPP* demethylallyl diphosphate, *IPP* isopentenyl diphosphate, *FPP* farnesyl diphosphate, *DPP* decaprenyl diphosphate, *PHB para*-hydroxybenzoate, *NADH dh* NADH dehydrogenase, *succ. dh* succinate dehydrogenase, *ROS* reactive oxygen species

successful, whereby rate-limiting enzymes of the shikimate pathway, along with UbiC, were over-expressed in a CoQ10-producing *E. coli* strain (Cluis et al. 2011) (Fig. 15.3). In this case, the expression of a first operon upregulated the enzymes that catalyze the conversion of E4P and PEP to shikimate, while enzymes encoded

by a second operon ensured higher flux from shikimate to PHB. This strategy resulted in a minor improvement in cellular CoQ10 content. However, when the over-expressed aromatic pathway was combined with an increased supply of isoprenoid precursors, a 2.6-fold improvement in CoQ10 production was observed (Cluis et al. 2011). Overall, these experiments suggest that the availability of PHB becomes rate-limiting to CoQ10 production mainly when the supply of isoprenoid increases, resulting in more DPP available for CoQ10 biosynthesis, or when the activity of UbiA goes up, leading to an increased flux of PHB to the ubiquinone pathway.

15.2.2.3 Over-Expression of the Ubiquinone Pathway

In *E. coli* it has been shown that the rate-limiting step to CoQ10 production occurs at the condensation reaction between the aromatic PHB and isoprenoid DPP, catalyzed by UbiA (Zhu et al. 1995; Zhang et al. 2007; Cluis et al. 2011). The over-expression of UbiA, along with DPS, resulted in as much as a 3.4-fold increase in CoQ10 content compared to strain expressing DPS alone (Zhang et al. 2007). Additionally, to overcome any potentially limiting steps stemming from the native regulatory mechanisms, the *ubiB*, *ubiG* and *ubiH* genes have been over-expressed in *E. coli*. The over-expression of these genes had only a limited effect on ubiquinone production (Zhu et al. 1995; Zhang et al. 2007) (Fig. 15.3). However, the *E. coli* background strains used in these experiments were subjected to endogenous regulation of aromatic and isoprenoid precursors. Therefore, it is possible that limited carbon flux to the ubiquinone pathway resulted in the modest increases in ubiquinone content reported in these studies. When precursors are produced in excess, CoQ10 biosynthesis may then become limited by the activity of one or more enzymes of the ubiquinone pathway.

15.2.3 Multi-Protein Enzyme Complexes in the Ubiquinone Pathway

The contiguous over-expression of the aromatic and isoprenoid pathways in CoQ10-producing *E. coli* strains leads to the accumulation of the ubiquinone pathway intermediate 2-decaprenylphenol (10P-Ph) (Cluis et al. 2011). Similarly, the accumulation of 2-octaprenylphenol (8P-Ph) has been shown in anaerobic *E. coli* cultures (Alexander and Young 1978), and also with *E. coli* strains bearing mutations in the ubiquinone pathway genes *ubiB*, *ubiH,* and *ubiG* (Young et al. 1973; Alexander and Young 1978; Hsu et al. 1996; Poon et al. 2000). When production of the aromatic and isoprenoid precursors to the ubiquinone pathway is up-regulated, flux through the pathway may increase and the activity of one of the above-mentioned Ubi enzymes may become rate-limiting and result in 10P-Ph accumulation. Furthermore, it has been shown that 10P-Ph accumulation stemming from a higher-flux-ubiquinone

pathway, due to increased isoprenoid and aromatic precursor availability, can be partly alleviated by UbiA over-expression and result in augmented CoQ10 accumulation (Cluis et al. 2011). UbiA, therefore, may be involved in modulating the activity of one or more downstream ubiquinone pathway enzymes. This theory is supported by studies of ubiquinone synthesis in *S. cerevisiae*, which have demonstrated strong evidence for the existence of a multi-protein complex involved in CoQ6 production (Gin et al. 2003; Gin and Clarke 2005; Johnson et al. 2005; Tran et al. 2006). It is theorized that Coq2p, the homolog of UbiA, does not physically interact with this multi-protein complex, but the Coq2p product, 3-hexaprenyl-4-hydroxybenzoate, is directly involved in the assembly and/or stability of the complex (Suzuki et al. 1994; Hsu et al. 2000; Hsieh et al. 2007). Studies with yeast also suggest that ubiquinone is synthesized in two phases: the first phase involves interactions between multiple proteins and CoQ6, leading to the formation of 2-hexaprenyl-6-methoxy-1,4-benzoquinol, while the second phase, ultimately leading to the conversion of hexaprenyl-6-methoxy-1,4-benzoquinol into CoQ6 (Fig. 15.2), is triggered by the transition to stationary phase and by linolenic acid-induced oxidative stress (Padilla et al. 2009). The activity of the monooxygenase responsible for the conversion of hexaprenyl-6-methoxy-1,4-benzoquinol to 2-hexaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol, Coq7p, is modulated by its phosphorylation state: phosphorylated Coq7p is deactivated and thus indirectly leads to a decrease in CoQ6 accumulation, while dephophorylated Coq7p results in increased CoQ6 synthesis (Martin-Montalvo et al. 2011). This suggests that in *S. cerevisiae*, the biosynthesis of ubiquinone is modulated post-translationally by protein kinases and phosphatases, thereby fine-tuning quinone production with the respiratory needs of the cell. Supporting a similar model in prokaryotes, *E. coli* studies using the translation inhibitor chloroamphenicol suggest that the increase in cellular ubiquinone pools that follow a switch from anaerobic to aerobic conditions can be attributed to a post-translational mechanism (Shestopalov et al. 1997). Although physical interactions between Ubi proteins are yet to be experimentally demonstrated in *E. coli,* early findings on the ubiquinone pathway suggest that there exists a soluble multi-protein complex that can convert 8P-Ph to CoQ8. Such a complex, which was isolated from the cytoplasmic membranes and was capable of converting 8P-Ph to CoQ8 *in vitro*, is hypothesized to remain in standby position during anaerobiosis and resume the conversion of 8P-Ph to ubiquinone upon transition to aerobic conditions (Knoell 1979, 1981). Engineering a CoQ10-over-producing *E. coli* strain may rely, therefore, on channeling 10P-Ph towards CoQ10 through increasing the expression of members of this putative protein complex, or on taking advantage of the post-translational mechanisms leading to its full activity.

15.2.4 Improving CoQ10 Production Using Central Metabolism

As an alternative or complementary strategy to increase carbon flux towards CoQ10 production in engineered *E. coli* strains, one may consider introducing alterations

into the central metabolic pathways. This strategy has been successfully used to increase the production of molecules of commercial interest in microbial hosts, notably to enhance lycopene production in recombinant *E. coli* strains (Farmer and Liao 2001; Martinez et al. 2008). As lycopene production also depends on the production of isoprenoid precursors, successful strategies to improve flux through the MEP pathway may be applicable to CoQ10 production in *E. coli.* The glycolysis intermediate G3P is, along with pyruvate, an initial building block of the MEP pathway, which in turn supplies the isoprenoid precursors necessary for the prenylation of ubiquinone (Fig. 15.2). By deleting *pykA* and *pykF*, coding for pyruvate kinases, and over-expressing the PEP carboxykinase Pck, carbon flux can be redirected towards G3P, which favours increased flux through the MEP pathway (Farmer and Liao 2001). Applying this approach to CoQ10 production, a 1.7-fold increase in CoQ10 content was obtained in an *E. coli* strain expressing a *G. suboxydans* DPS (Huang et al. 2011) (Fig. 15.3, Table 15.1).

Another attractive strategy to improve the production of a molecule of interest in a microbial host is to manipulate the levels of redox cofactors required for its synthesis. The production of one molecule of CoQ10 from glucose requires 21 molecules of nicotinamide adenine dinucleotide phosphate (NADPH) (20 molecules from the MEP pathway and 1 NADPH from the shikimate pathway). In order to increase the availability of NADPH for the MEP pathway, the NAD^+-dependent G3P dehydrogenase, GapA, of *E. coli* can be replaced with the $NADP^+$-dependent G3P dehydrogenase, GapC, from *Clostridium acetobutylicum* (Martinez et al. 2008; Huang et al. 2011). However, this strategy was shown to result in the down-regulation of the pentose-phosphate pathway, normally the major source of NAPDH in cells (Martinez et al. 2008). In the case of CoQ10 production, the down-regulation of the pentose-phosphate pathway may be detrimental as it is essential for the production of precursors to the shikimate pathway leading to PHB (Huang et al. 2011). Accordingly, when this approach was used to improve CoQ10 biosynthesis in engineered *E. coli* strains, only up to a 1.25-fold increase in CoQ10 content was observed (Table 15.1, Fig. 15.3). In a high CoQ10-producing mutant strain of *R. radiobacter*, a positive correlation was demonstrated between the nicotinamide adenine dinucleotide (NADH)/NAD^+ ratio and the specific CoQ10 content (Koo et al. 2010). Based on this finding, the native NAD^+-dependent G3P dehydrogenase, GapA, was over-expressed to increase the NADH/NAD^+ ratio and further improve CoQ10 production. This strategy resulted in a 25% increase in NADH/NAD^+ and in over twofold improvement in CoQ10 content (Koo et al. 2010) (Table 15.1, Fig. 15.3). A high NADH/NAD^+ ratio may increase the activity of the electron transport chain, which reoxidizes NADH through the activity of NADH-dehydrogenases. It is possible that the increased activity of these enzymes or the resulting changes in the oxidation state of the quinone pool triggers CoQ10 biosynthesis. The results of these two studies, which both take advantage of G3P dehydrogenases to alter the pools of different redox factors, appear somewhat conflicting (Koo et al. 2010; Huang et al. 2011). Nevertheless, these studies highlight the necessity to explore in more detail the relationship between central metabolism and CoQ10 production in microbes.

15.2.5 Improving CoQ10 Yields Through Growth Conditions

As discussed in Sect. 15.1.5, the CoQ10 productivity of natural producers, such as *R. radiobater* and *R. sphaeroides*, has been improved by random mutagenesis and selection. However, because of the limited genetic tools available for these organisms, the rational genetic strategies that can be applied to *E. coli* are largely unfeasible. As a result, most research has focused on increasing CoQ10 production by fine-tuning the growth conditions of these organisms, both in flasks and in fermentors.

Although there is still conflicting evidence on this matter (Wu et al. 2003), a number of reports have shown that CoQ10 production tends to increase under low aeration conditions, both in *R. radiobacter* and *R. sphaeroides* (Choi et al. 2005; Ha et al. 2007b; Yen and Chiu 2007; Yen and Shih 2009; Seo and Kim 2010) (Fig. 15.3). For instance, the CoQ10 content of a mutant strain of *R. radiobacter* increased fourfold by reducing dissolved oxygen concentrations from 20 to 5% (Choi et al. 2005) (Table 15.2). As part of the electron transport chain, it is possible that CoQ10 levels are modulated by variations in the electron flux or in the proton gradient brought about by the limited oxygen supply. To test these possibilities, the effect of the cytochrome *c* oxidase inhibitor, azide, and of 2,4-dinitrophenol (DNP), a proton ionophore that uncouples oxidative phosphorylation, on CoQ10 accumulation were assessed in *R. radiobacter* (Choi et al. 2005). The addition of DNP had no effect on CoQ10 accumulation, indicating that large-scale changes in the cellular proton gradient do not affect CoQ10 production (Choi et al. 2005). The addition of increasing concentrations of sodium azide, on the other hand, correlated to increasing CoQ10 content of bacterial cultures (Choi et al. 2005; Seo and Kim 2010). These results suggest that the restricted electron flux through the respiratory chain brought about by low aeration conditions, mimicked by the azide treatment, triggers the synthesis of CoQ10. Though the precise mechanism behind this phenomenon is unclear, it has been hypothesized that when the transfer of electrons through the electron transport chain to the final acceptor is slowed by a decrease in oxygen availability, electrons are trapped in upstream constituents of the respiratory chain, potentially leading to an accumulation of the reduced form of CoQ10, $CoQ10H_2$ (Bekker et al. 2007). Cells may thus be forced to synthesize more CoQ10 in order to fix this imbalance of $CoQ10/CoQ10H_2$ (Choi et al. 2005). Alternatively, partial blockage of the respiratory chain may lead to an accumulation of toxic free electrons in the inner membrane, hence triggering the synthesis of more CoQ10 to scavenge these electrons (Choi et al. 2005).

In addition to respiration, facultative photoautotrophs such as *R. sphaeroides* also use CoQ10 to transfer electrons from the reaction center to the cytochrome bc_1 complex, within the photosynthetic apparatus (Vermeglio and Joliot 1999). Growth of *R. sphaeroides* under anaerobic/light conditions increased the specific CoQ10 content over fourfold compared to the level attained under aerobic/dark growth conditions (Table 15.2). However, this growth strategy significantly decreased the growth rate of the already low biomass culture, therefore making it uncompetitive for industrial applications (Yen and Chiu 2007). Nevertheless, this finding suggests

Table 15.2 CoQ10 production and culture conditions of natural producer microbes

[a]	Organism	CoQ10 production (mg/L)	Specific CoQ10 content (mg/gDCW)	Optimized culture conditions	Study
2A	*R. radiobacter* ATTC4452	87.6	1.9	Not applicable	Yoshida et al. (1998)
2B	*R. radiobacter* WSH2601	32.1	1.91	Maximal DO of 40% (of air saturation)	Wu et al. (2003)
2C	*R. radiobacter* ATCC4452	Not reported	5.3	Decreased DO, azide (electron flux inhibitor) treatment	Choi et al. (2005)
2D	*R. radiobacter* KCCM 10413	458	8.54	Fed-batch with sucrose as carbon source, pH 7.0 and 0–10% DO	Ha et al. (2007b)
2E	*R. radiobacter* KCCM 10413	638	9.71	Decreased sucrose feeding to 10% w/v, led to less byproducts/osmotic stress	Ha et al. (2007a)
2F	*R. radiobacter* KCCM 10413	763.7	11.84	$CaCl_2$ supplementation	Ha et al. (2009)
2G	*R. radiobacter* T6102	14.6	1.95	Decreased DO; H_2O_2 and azide treatment	Seo and Kim (2010)
2H	*R. radiobacter* A603-35-gapA	Not reported	5.27	Increased NADH/NAD^+ ratio	Koo et al. (2010)
2I	*R. spheroides* KY 8598	770	Not reported	Controlled agitation speed and oxygen supply	Sakato et al. (1992)
2J	*R. sphaeroides* KY 4113	350	8.7	Decreased aeration	Yoshida et al. (1998)
2K	*R. sphaeroides* BCRC13100	10.5	8	0% DO and cultivation without light	Yen and Chiu (2007)
2L	*R. sphaeroides* BCRC13100	45.65	4.4	Use of airlift bioreactor for low aeration	Yen and Shih (2009)
2M	*R. sphaeroides* BCRC13100	83.8	4.5	Fed-batch system with molasses at 1% w/v	Yen et al. (2010)
2N	*S. johnsonii*	Not reported	10.5	Post log growth phase addition of PHB	Dixson et al. (2011)

[a]Listed letters are referred to those shown in Fig. 15.3

that the complete shutdown of respiration associated with the onset of photosynthetic reactions triggers the cells to accumulate more CoQ10 within their inner membrane. This observation, coupled with the CoQ10 accumulation that accompanies a slowed electron transport chain, yields strong evidence that hindering respiration is an effective strategy in increasing CoQ10 productivity.

Another successful strategy to improve CoQ10 yields in *R. radiobacter* was to increase the concentration of calcium in the growth media. It has been proposed that the increased Ca^{2+} concentration is a source of oxidative stress for the cell, triggering the onset of lipid peroxidation and the production of reactive oxygen species (Ha et al. 2009). Because CoQ10 has well-documented antioxidant properties, it may therefore be likely that increasing its production is the natural response to such oxidative stress. Supporting this hypothesis, an oxidative stress challenge, mediated by the addition of hydrogen peroxide to *R. radiobacter* cultures, also results in increases in CoQ10 content (Seo and Kim 2010). The selection of high CoQ10-producing mutants on free radical-generating growth inhibitors, described in Sect. 15.1.4, also supports this link between CoQ10 production and oxidative stress (Yoshida et al. 1998) (Fig. 15.3).

The effect of growth on select carbon and nitrogen sources on CoQ10 production has been explored in mutant strains of *R. radiobacter*. There is conflicting evidence of the impact that growth on different carbon sources has on accumulated CoQ10 content, with one study reporting sucrose, glucose and fructose as having a positive impact on CoQ10 content compared to xylose, lactose and galactose (Koo et al. 2010), while another report shows no differences in CoQ10 content for growth on these carbon sources (Ha et al. 2007b). Plant based nitrogen sources such as corn steep powder and soytone yielded the highest CoQ10 specific content compared to other nitrogen sources including yeast extract, meat extract and tryptone (Ha et al. 2007b; Koo et al. 2010). Notably, for both carbon and nitrogen sources, there was a positive correlation between the NADH/NAD^+ ratio and the resulting CoQ10 content (Koo et al. 2010). Overall, modifying growth conditions in order to modulate central metabolic processes appears to be an effective strategy in enhancing CoQ10 production in microbes (for further culture conditions that correspond to increased CoQ10 accumulation refer to Table 15.2).

15.3 Conclusions

Despite the careful rational engineering strategies and considerable research efforts put towards the development of recombinant *E. coli* strains capable of producing high levels of CoQ10, the specific CoQ10 contents obtained through these attempts still fall below those obtained with *R. radiobacter* and other natural producers of the molecule. Most strikingly, as demonstrated by the results summarized in Sect. 15.2.2, increasing supply of both the aromatic and isoprenoid precursors has had a relatively

minor effect on the synthesis of CoQ10, compared to theoretical yields. It follows that there may be unknown intrinsic factors specific to *E. coli* that prevent the accumulation of high amounts of ubiquinone. Such factors may include the ability to spatially accommodate excess ubiquinone in the inner membrane space, unique regulatory mechanisms to maintain ubiquinone concentrations within a precise range, and native properties of the respiratory chain. For instance, *E. coli*'s respiratory chain has the particularity of lacking cytochrome *c*, thus relying exclusively on ubiquinone and/or naphtoquinones to channel electrons between the different transport molecules of the electron transport chain. Transversely, naturally high producers of CoQ10 such as *R. radiobacter* and *R. sphaeroides* use cytochrome *c*, in addition to ubiquinone, and lack naphtoquinones. Such inherent differences may influence the cell's capacity to accumulate large amounts of CoQ10 in the membrane without inhibiting respiration. In addition, *E. coli* uses the redox state of its ubiquinone pool as a signal for the onset of aerobic/anaerobic metabolism (Georgellis et al. 2001). There may therefore be high pressure to maintain the ubiquinone pool within a given size range. In fact, promoter studies with different genes of the ubiquinone pathway have shown that the biosynthesis of ubiquinone is under the control of central transcriptional regulators including ArcA/B and FNR (Soballe and Poole 1997; Zhang and Javor 2003; Kwon et al. 2005). The accumulation of 10P-Ph, under conditions where the production of both PHB and isoprenoid precursors are deregulated, also points towards an intrinsic regulation system that prevents the accumulation of high amounts of CoQ10 in engineered *E. coli* (Cluis et al. 2011). Overall, a better understanding of the limiting factors in the accumulation of ubiquinone in *E. coli* will be essential for this organism to become a suitable platform for the industrial production of CoQ10.

The strains of *R. radiobacter* and *R. sphaeroides* that have been isolated for their high CoQ10 content may provide the key to unravelling the underlying mechanisms that favour the production of CoQ10 in microbes. Comparative genomic and metabolomic investigations of the ubiquinone pathway and of the respiratory chain in these species may elucidate the inherent abilities of these organisms to accumulate CoQ10 in their membrane. Lessons learned from these high producers may then be applied to inverse engineering industrially-suitable hosts such as *E. coli*. Alternatively, since some genetic tools and plasmids exist for *R. radiobacter*, targeted metabolic engineering strategies that have proven successful in *E. coli* might be applied to the high CoQ10-producing mutant strains of *R. radiobacter* to further enhance their inherent capacity to accumulate CoQ10. In addition, although studies in *R. radiobacter* have already explored a number of different conditions favouring the accumulation of CoQ10 (Ha et al. 2007b; Koo et al. 2010), there is a need to document more systematically how different carbon sources, nutrient ratios and electron acceptors influence CoQ10 production in the context of an over-producing strain. Such research may lead not only to further improvements in yields obtained in fermentors, but also open up new avenues for rationally engineering microbes to enhance CoQ10 accumulation.

References

Alexander K, Young IG (1978) Alternative hydroxylases for the aerobic and anaerobic biosynthesis of ubiquinone in *Escherichia coli*. Biochemistry 17:4750–4755

Bader MW, Xie T, Yu CA, Bardwell JC (2000) Disulfide bonds are generated by quinone reduction. J Biol Chem 275:26082–26088

Bekker M, Kramer G, Hartog AF, Wagner MJ, de Koster CG, Hellingwerf KJ, Teixeira de Mattos MJ (2007) Changes in the redox state and composition of the quinone pool of *Escherichia coli* during aerobic batch-culture growth. Microbiology 153:1974–1980

Choi G-S, Kim Y-S, Seo J-H, Ryu Y-W (2005) Restricted electron flux increases coenzyme Q10 production in *Agrobacterium tumefaciens* ATCC4452. Process Biochem 40:3225–3229

Choi JH, Ryu YW, Park YC, Seo JH (2009) Synergistic effects of chromosomal ispB deletion and dxs overexpression on coenzyme Q(10) production in recombinant *Escherichia coli* expressing *Agrobacterium tumefaciens* dps gene. J Biotechnol 144:64–69

Cluis CP, Ekins A, Narcross L, Jiang H, Gold ND, Burja AM, Martin VJ (2011) Identification of bottlenecks in *Escherichia coli* engineered for the production of CoQ(10). Metab Eng 13:733–744

Collins MD, Jones D (1981) Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. Microbiol Rev 45:316–354

Cox GB, Young IG, McCann LM, Gibson F (1969) Biosynthesis of ubiquinone in *Escherichia coli* K-12: location of genes affecting the metabolism of 3-octaprenyl-4-hydroxybenzoic acid and 2-octaprenylphenol. J Bacteriol 99:450–458

Dixson DD, Boddy CN, Doyle RP (2011) Reinvestigation of coenzyme Q10 isolation from *Sporidiobolus johnsonii*. Chem Biodivers 8:1033–1051

Farmer WR, Liao JC (2001) Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. Biotechnol Prog 17:57–61

Farr SB, Kogoma T (1991) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. Microbiol Rev 55:561–585

Georgellis D, Kwon O, Lin EC (2001) Quinones as the redox signal for the arc two-component system of bacteria. Science 292:2314–2316

Gin P, Clarke CF (2005) Genetic evidence for a multi-subunit complex in coenzyme Q biosynthesis in yeast and the role of the Coq1 hexaprenyl diphosphate synthase. J Biol Chem 280:2676–2681

Gin P, Hsu AY, Rothman SC, Jonassen T, Lee PT, Tzagoloff A, Clarke CF (2003) The *Saccharomyces cerevisiae* COQ6 gene encodes a mitochondrial flavin-dependent monooxygenase required for coenzyme Q biosynthesis. J Biol Chem 278:25308–25316

Gulmezian M, Hyman KR, Marbois BN, Clarke CF, Javor GT (2007) The role of UbiX in *Escherichia coli* coenzyme Q biosynthesis. Arch Biochem Biophys 467:144–153

Ha SJ, Kim SY, Seo JH, Moon HJ, Lee KM, Lee JK (2007a) Controlling the sucrose concentration increases Coenzyme Q10 production in fed-batch culture of *Agrobacterium tumefaciens*. Appl Microbiol Biotechnol 76:109–116

Ha SJ, Kim SY, Seo JH, Oh DK, Lee JK (2007b) Optimization of culture conditions and scale-up to pilot and plant scales for coenzyme Q(10) production by *Agrobacterium tumefaciens*. Appl Microbiol Biotechnol 74:974–980

Ha SJ, Kim SY, Seo JH, Jeya M, Zhang YW, Ramu T, Kim IW, Lee JK (2009) Ca2+ increases the specific coenzyme Q10 content in *Agrobacterium tumefaciens*. Bioprocess Biosyst Eng 32:697–700

Harker M, Bramley PM (1999) Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. FEBS Lett 448:115–119

Hodgson JM, Watts GF, Playford DA, Burke V, Croft KD (2002) Coenzyme Q10 improves blood pressure and glycaemic control: a controlled trial in subjects with type 2 diabetes. Eur J Clin Nutr 56:1137–1142

Hsieh EJ, Gin P, Gulmezian M, Tran UC, Saiki R, Marbois BN, Clarke CF (2007) *Saccharomyces cerevisiae* Coq9 polypeptide is a subunit of the mitochondrial coenzyme Q biosynthetic complex. Arch Biochem Biophys 463:19–26

Hsu AY, Poon WW, Shepherd JA, Myles DC, Clarke CF (1996) Complementation of coq3 mutant yeast by mitochondrial targeting of the *Escherichia coli* UbiG polypeptide: evidence that UbiG catalyzes both O-methylation steps in ubiquinone biosynthesis. Biochemistry 35:9797–9806

Hsu AY, Do TQ, Lee PT, Clarke CF (2000) Genetic evidence for a multi-subunit complex in the O-methyltransferase steps of coenzyme Q biosynthesis. Biochim Biophys Acta 1484:287–297

Huang M, Wang Y, Liu J, Mao Z (2011) Multiple strategies for metabolic engineering of *Escherichia coli* for efficient production of Coenzyme Q10. Chin J Chem Eng 19:316–326

Ingledew WJ, Poole RK (1984) The respiratory chains of *Escherichia coli*. Microbiol Rev 48:222–271

Inui M, Ooe M, Fujii K, Matsunaka H, Yoshida M, Ichihashi M (2008) Mechanisms of inhibitory effects of CoQ10 on UVB-induced wrinkle formation in vitro and in vivo. Biofactors 32:237–243

Johnson A, Gin P, Marbois BN, Hsieh EJ, Wu M, Barros MH, Clarke CF, Tzagoloff A (2005) COQ9, a new gene required for the biosynthesis of coenzyme Q in *Saccharomyces cerevisiae*. J Biol Chem 280:31397–31404

Kien NB, Kong IS, Lee MG, Kim JK (2010) Coenzyme Q10 production in a 150-l reactor by a mutant strain of *Rhodobacter sphaeroides*. J Ind Microbiol Biotechnol 37:521–529

Kim SW, Keasling JD (2001) Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. Biotechnol Bioeng 72:408–415

Kim SJ, Kim MD, Choi JH, Kim SY, Ryu YW, Seo JH (2006) Amplification of 1-deoxy-D-xyluose 5-phosphate (DXP) synthase level increases coenzyme Q10 production in recombinant *Escherichia coli*. Appl Microbiol Biotechnol 72:982–985

Knoell HE (1979) Isolation of a soluble enzyme complex comprising the ubiquinone-8 synthesis apparatus from the cytoplasmic membrane of *Escherichia coli*. Biochem Biophys Res Commun 91:919–925

Knoell HE (1981) Stand-by position of the dioxygen-dependent ubiquinone-8 synthesis apparatus in anaerobically grown *Escherichia coli* K-12. FEMS Microbiol Lett 10:59–62

Koo BS, Gong YJ, Kim SY, Kim CW, Lee HC (2010) Improvement of coenzyme Q(10) production by increasing the NADH/NAD(+) ratio in *Agrobacterium tumefaciens*. Biosci Biotechnol Biochem 74:895–898

Kroger A, Klingenberg M (1973) Further evidence for the pool function of ubiquinone as derived from the inhibition of the electron transport by antimycin. Eur J Biochem 39:313–323

Kwon O, Druce-Hoffman M, Meganathan R (2005) Regulation of the ubiquinone (coenzyme Q) biosynthetic genes ubiCA in *Escherichia coli*. Curr Microbiol 50:180–189

Lass A, Sohal RS (1998) Electron transport-linked ubiquinone-dependent recycling of alpha-tocopherol inhibits autooxidation of mitochondrial membranes. Arch Biochem Biophys 352:229–236

Lee PT, Hsu AY, Ha HT, Clarke CF (1997) A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the *Escherichia coli* ubiE gene. J Bacteriol 179:1748–1754

Lee JK, Her G, Kim SY, Seo JH (2004) Cloning and functional expression of the dps gene encoding decaprenyl diphosphate synthase from *Agrobacterium tumefaciens*. Biotechnol Prog 20:51–56

Lee JK, Oh DK, Kim SY (2007) Cloning and characterization of the dxs gene, encoding 1-deoxy-d-xylulose 5-phosphate synthase from *Agrobacterium tumefaciens*, and its overexpression in *Agrobacterium tumefaciens*. J Biotechnol 128:555–566

Leppik RA, Stroobant P, Shineberg B, Young IG, Gibson F (1976) Membrane-associated reactions in ubiquinone biosynthesis. 2-Octaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinone methyltransferase. Biochim Biophys Acta 428(1):146–156

Lipshutz BH, Lower A, Berl V, Schein K, Wetterich F (2005) An improved synthesis of the "miracle nutrient" coenzyme Q10. Org Lett 7:4095–4097

Malpica R, Franco B, Rodriguez C, Kwon O, Georgellis D (2004) Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. Proc Natl Acad Sci USA 101:13318–13323

Mancini A, Balercia G (2011) Coenzyme Q(10) in male infertility: physiopathology and therapy. Biofactors 37:374–380

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21:796–802

Martin SF, Buron I, Espinosa JC, Castilla J, Villalba JM, Torres JM (2007) Coenzyme Q and protein/lipid oxidation in a BSE-infected transgenic mouse model. Free Radic Biol Med 42:1723–1729

Martinez I, Zhu J, Lin H, Bennett GN, San KY (2008) Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from Clostridium acetobutylicum facilitates NADPH dependent pathways. Metab Eng 10:352–359

Martin-Montalvo A, Gonzalez-Mariscal I, Padilla S, Ballesteros M, Brautigan DL, Navas P, Santos-Ocana C (2011) Respiratory-induced coenzyme Q biosynthesis is regulated by a phosphorylation cycle of Cat5p/Coq7p. Biochem J 440:107–114

Myers CE, Muindi JR, Zweier J, Sinha BK (1987) 5-Iminodaunomycin. An anthracycline with unique properties. J Biol Chem 262:11571–11577

Negishi E, Liou SY, Xu C, Huo S (2002) A novel, highly selective, and general methodology for the synthesis of 1,5-diene-containing oligoisoprenoids of all possible geometrical combinations exemplified by an iterative and convergent synthesis of coenzyme Q(10). Org Lett 4:261–264

Okada K, Kamiya Y, Zhu X, Suzuki K, Tanaka K, Nakagawa T, Matsuda H, Kawamukai M (1997) Cloning of the sdsA gene encoding solanesyl diphosphate synthase from *Rhodobacter capsulatus* and its functional expression in *Escherichia coli* and *Saccharomyces cerevisiae*. J Bacteriol 179:5992–5998

Okada K, Kainou T, Tanaka K, Nakagawa T, Matsuda H, Kawamukai M (1998) Molecular cloning and mutational analysis of the ddsA gene encoding decaprenyl diphosphate synthase from *Gluconobacter suboxydans*. Eur J Biochem 255:52–59

Padilla S, Tran UC, Jimenez-Hidalgo M, Lopez-Martin JM, Martin-Montalvo A, Clarke CF, Navas P, Santos-Ocana C (2009) Hydroxylation of demethoxy-Q6 constitutes a control point in yeast coenzyme Q6 biosynthesis. Cell Mol Life Sci 66:173–186

Park YC, Kim SJ, Choi JH, Lee WH, Park KM, Kawamukai M, Ryu YW, Seo JH (2005) Batch and fed-batch production of coenzyme Q10 in recombinant *Escherichia coli* containing the decaprenyl diphosphate synthase gene from *Gluconobacter suboxydans*. Appl Microbiol Biotechnol 67:192–196

Poon WW, Davis DE, Ha HT, Jonassen T, Rather PN, Clarke CF (2000) Identification of *Escherichia coli* ubiB, a gene required for the first monooxygenase step in ubiquinone biosynthesis. J Bacteriol 182:5139–5146

Sakato K, Tanaka H, Shibata S, Kuratsu Y (1992) Agitation-aeration studies on coenzyme Q10 production using *Rhodopseudomonas spheroides*. Biotechnol Appl Biochem 16:19–22

Schaars CF, Stalenhoef AF (2008) Effects of ubiquinone (coenzyme Q10) on myopathy in statin users. Curr Opin Lipidol 19:553–557

Schmelzer C, Lindner I, Vock C, Fujii K, Doring F (2007) Functional connections and pathways of coenzyme Q10-inducible genes: an in-silico study. IUBMB Life 59:628–633

Schmelzer C, Okun JG, Haas D, Higuchi K, Sawashita J, Mori M, Doring F (2010) The reduced form of coenzyme Q10 mediates distinct effects on cholesterol metabolism at the transcriptional and metabolite level in SAMP1 mice. IUBMB Life 62:812–818

Seo MJ, Kim SO (2010) Effect of limited oxygen supply on coenzyme Q(10) production and its relation to limited electron transfer and oxidative stress in *Rhizobium radiobacter* T6102. J Microbiol Biotechnol 20:346–349

Shestopalov AI, Bogachev AV, Murtazina RA, Viryasov MB, Skulachev VP (1997) Aeration-dependent changes in composition of the quinone pool in *Escherichia coli*. Evidence of post-transcriptional regulation of the quinone biosynthesis. FEBS Lett 404:272–274

Shults CW, Oakes D, Kieburtz K, Beal MF, Haas R, Plumb S, Juncos JL, Nutt J, Shoulson I, Carter J, Kompoliti K, Perlmutter JS, Reich S, Stern M, Watts RL, Kurlan R, Molho E, Harrison M, Lew M (2002) Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Arch Neurol 59:1541–1550

Singh RB, Wander GS, Rastogi A, Shukla PK, Mittal A, Sharma JP, Mehrotra SK, Kapoor R, Chopra RK (1998) Randomized, double-blind placebo-controlled trial of coenzyme Q10 in patients with acute myocardial infarction. Cardiovasc Drugs Ther 12:347–353

Soballe B, Poole RK (1997) Aerobic and anaerobic regulation of the ubiCA operon, encoding enzymes for the first two committed steps of ubiquinone biosynthesis in *Escherichia coli*. FEBS Lett 414:373–376

Sourris KC, Harcourt BE, Tang PH, Morley AL, Huynh K, Penfold SA, Coughlan MT, Cooper ME, Nguyen TV, Ritchie RH, Forbes JM (2012) Ubiquinone (coenzyme Q10) prevents renal mitochondrial dysfunction in an experimental model of type 2 diabetes. Free Radic Biol Med 52:716–723

Stroobant P, Young IG, Gibson F (1972) Mutants of *Escherichia coli* K-12 blocked in the final reaction of ubiquinone biosynthesis: characterization and genetic analysis. J Bacteriol 109: 134–139

Suzuki K, Ueda M, Yuasa M, Nakagawa T, Kawamukai M, Matsuda H (1994) Evidence that *Escherichia coli* ubiA product is a functional homolog of yeast COQ2, and the regulation of ubiA gene expression. Biosci Biotechnol Biochem 58:1814–1819

Takahashi S, Nishino T, Koyama T (2003) Isolation and expression of Paracoccus denitrificans decaprenyl diphosphate synthase gene for production of ubiquinone-10 in *Escherichia coli*. Biochem Eng J 16:183–190

Tran UC, Marbois B, Gin P, Gulmezian M, Jonassen T, Clarke CF (2006) Complementation of *Saccharomyces cerevisiae* coq7 mutants by mitochondrial targeting of the *Escherichia coli* UbiF polypeptide: two functions of yeast Coq7 polypeptide in coenzyme Q biosynthesis. J Biol Chem 281:16401–16409

Urakami T, Hori-Okubo M (1988) Production of isoprenoid compounds in the facultative methylotroph *Protomonas extorquens*. J Ferment Technol 66:323–332

Vermeglio A, Joliot P (1999) The photosynthetic apparatus of *Rhodobacter sphaeroides*. Trends Microbiol 7:435–440

Wu Z, Du G, Chen J (2003) Effects of dissolved oxygen concentration and DO-stat feeding strategy on CoQ10 production with *Rhizobium radiobacter*. World J Microbiol Biotechnol 19:925–928

Yang X, Dai G, Li G, Yang ES (2010) Coenzyme Q10 reduces beta-amyloid plaque in an APP/PS1 transgenic mouse model of Alzheimer's disease. J Mol Neurosci 41:110–113

Yen HW, Chiu CH (2007) The influences of aerobic-dark and anaerobic-light cultivation on CoQ(10) production by *Rhodobacter sphaeroides* in the submerged fermenter. Enzyme Microb Technol 41:600–604

Yen HW, Shih TY (2009) Coenzyme Q10 production by *Rhodobacter sphaeroides* in stirred tank and in airlift bioreactor. Bioprocess Biosyst Eng 32:711–716

Yen HW, Feng CY, Kang JL (2010) Cultivation of *Rhodobacter sphaeroides* in the stirred bioreactor with different feeding strategies for CoQ(10) production. Appl Biochem Biotechnol 160:1441–1449

Yoshida H, Kotani Y, Ochiai K, Araki K (1998) Production of ubiquinone-10 using bacteria. J Gen Appl Microbiol 44:19–26

Yoshida Y, Hayakawa M, Habuchi Y, Niki E (2006) Evaluation of the dietary effects of coenzyme Q in vivo by the oxidative stress marker, hydroxyoctadecadienoic acid and its stereoisomer ratio. Biochim Biophys Acta 1760:1558–1568

Young IG, McCann LM, Stroobant P, Gibson F (1971) Characterization and genetic analysis of mutant strains of *Escherichia coli* K-12 accumulating the ubiquinone precursors 2-octaprenyl-6-methoxy-1,4-benzoquinone and 2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone. J Bacteriol 105:769–778

Young IG, Leppik RA, Hamilton JA, Gibson F (1972) Biochemical and genetic studies on ubiquinone biosynthesis in *Escherichia coli* K-12:4-hydroxybenzoate octaprenyltransferase. J Bacteriol 110:18–25

Young IG, Stroobant P, Macdonald CG, Gibson F (1973) Pathway for ubiquinone biosynthesis in *Escherichia coli* K-12: gene-enzyme relationships and intermediates. J Bacteriol 114:42–52

Zahiri HS, Noghabi KA, Shin YC (2006a) Biochemical characterization of the decaprenyl diphosphate synthase of *Rhodobacter sphaeroides* for coenzyme Q(10) production. Appl Microbiol Biotechnol 73(4):796–806

Zahiri HS, Yoon SH, Keasling JD, Lee SH, Won Kim S, Yoon SC, Shin YC (2006b) Coenzyme Q10 production in recombinant *Escherichia coli* strains engineered with a heterologous decaprenyl diphosphate synthase gene and foreign mevalonate pathway. Metab Eng 8(5): 406–416

Zhang H, Javor GT (2003) Regulation of the isofunctional genes ubiD and ubiX of the ubiquinone biosynthetic pathway of *Escherichia coli*. FEMS Microbiol Lett 223:67–72

Zhang D, Shrestha B, Li Z, Tan T (2007) Ubiquinone-10 production using *Agrobacterium tumefaciens* dps gene in *Escherichia coli* by coexpression system. Mol Biotechnol 35:1–14

Zhu X, Yuasa M, Okada K, Suzuki K, Nakagawa T, Kawamukai M, Matsuda H (1995) Production of ubiquinone in *Escherichia coli* by expression of various genes responsible for ubiquinone biosynthesis. J Biosci Bioeng 79:493–495

Chapter 16
Genetic Modification and Bioprocess Optimization for S-Adenosyl-L-methionine Biosynthesis

Xiaoqing Hu, Peter J. Quinn, Zhou Wang, Guoqiang Han, and Xiaoyuan Wang

Abstract *S*-Adenosyl-L-methionine is an important bioactive sulfur-containing amino acid. Large scale preparation of the amino acid is of great significance. *S*-Adenosyl-L-methionine can be synthesized from L-methionine and adenosine triphosphate in a reaction catalyzed by methionine adenosyltransferase. In order to enhance *S*-adenosyl-L-methionine biosynthesis by industrial microbial strains, various strategies have been employed to optimize the process. Genetic manipulation has largely focused on enhancement of expression and activity of methionine adenosyltransferase. This has included its overexpression in *Pichia pastoris*, *Saccharomyces cerevisiae* and *Escherichia coli*, molecular evolution, and fine-tuning of expression by promoter engineering. Furthermore, knocking in of *Vitreoscilla* hemoglobin and knocking out of cystathionine-β-synthase have also been effective strategies. Besides genetic modification, novel bioprocess strategies have also been conducted to improve *S*-adenosyl-L-methionine synthesis and inhibit its conversion. This has involved the optimization of feeding modes of methanol, glycerol and

X. Hu • X. Wang (✉)
State Key Laboratory of Food Science and Technology, Jiangnan University, 1800 Lihu Avenue, Wuxi 214122, China
e-mail: xiaoyuanwang@hotmail.com

P.J. Quinn
Department of Biochemistry, King's College London, 150 Stamford Street, London SE1 9NH, UK

Z. Wang • G. Han
Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, 1800 Lihu Avenue, Wuxi 214122, China

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_16,

L-methionine substrates. Taken together considerable improvements have been achieved in *S*-adenosyl-L-methionine accumulation at both flask and fermenter scales. This review provides a contemporary account of these developments and identifies potential methods for further improvements in the efficiency of *S*-adenosyl-L-methionine biosynthesis.

Keywords Bioprocess optimization • Genetic modification • *Pichia pastoris* • Promoter • *S*-adenosyl-L-methionine

Abbreviations

ATP	Adenosine triphosphate
CBS	Cystathionine-β-synthase
C-source	Carbon source
L-Met	L-methionine
MAT	Methionine adenosyltransferase
P_{AOX}	Promoter of alcohol oxidase 1 gene
P_{GAP}	Promoter of glyceraldehyde-3-phosphate dehydrogenase gene
SAH	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
VHb	*Vitreoscilla* hemoglobin

16.1 Introduction

16.1.1 The Structure, Function and Therapeutic Uses of S-Adenosyl-L-methionine

S-Adenosyl-L-methionine (SAM) was discovered in 1952 by Cantoni (1952), and its IUPAC name was designated as (2*S*)-2-Amino-4-[[(2*S*,3*S*,4*R*,5*R*)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolan-2-yl]methyl-methylsulfonio] butanoate. The molecular structure in Fig. 16.1 shows that SAM possesses a sulfonium ion with a high group transfer potential. Thus each of the attached carbons is activated toward nucleophilic attack. SAM is involved in three types of important biochemical reactions within living cells, including transmethylation, transsulfuration, and aminopropylation.

Transmethylation reactions are widely used in modification of a variety of biomolecules such as nucleic acids, proteins, phospholipids, and bioamines. In these reactions, SAM serves as the universal methyl donor (Mato et al. 1997). Methylation reactions yield the common product *S*-adenosyl-L-homocysteine (SAH), and most SAM-dependent methylations are strongly inhibited by accumulation of SAH (Mato

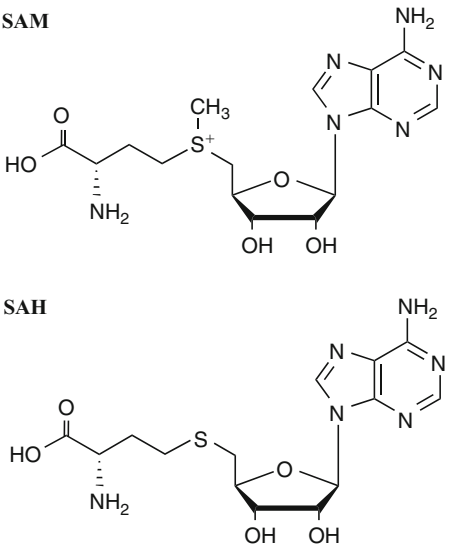


Fig. 16.1 Molecular structures of SAM and SAH. The existence of high energy sulfonium ion in SAM activates each attached carbon toward nucleophilic attack, and therefore, SAM mainly participates in three types of reactions (transmethylation, transsulfuration, and aminopropylation)

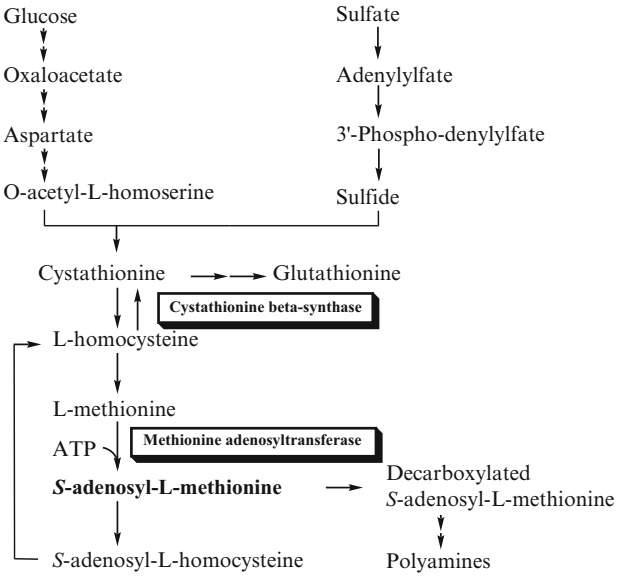


Fig. 16.2 The simplified pathway for SAM synthesis and conversion

et al. 1997). The ratio of SAM to SAH has been defined as the methylation potential, indexing the rate of flow of methyl groups within cells. SAH can be hydrolyzed to adenosine and homocysteine, and the latter can be further metabolized through two pathways: remethylation and transsulfuration. As shown in Fig. 16.2, in the

transsulfuration pathway, homocysteine and serine are condensed to form cystathionine, which is then hydrolyzed to release cysteine and α-ketobutyrate. The resultant cysteine can react with glutamate and glycine to synthesize glutathione, the major non-protein thiol source playing a pivotal role in the antioxidant status of cells (Stipanuk 1986). SAM also plays a key role in polyamine synthesis. SAM is decarboxylated and the aminopropyl group is transferred to putrescine or spermidine to form polyamines. SAM is not only the active form of L-methionine (L-Met), but also the methyl donor, precursor of aminopropyl groups and glutathione. These reactions are essential in almost all living organisms where they are involved in numerous metabolic pathways and processes. SAM has been developed for therapeutic applications such as treatment of liver disease (Lieber 1999), osteoarthritis (Barcelo et al. 1990), neurological and affective disorders (Torta et al. 1998). These applications rely on involvement of SAM in regulation of some biochemical processes (Friedel et al. 1989; Mato et al. 1997). In the United States and Canada, SAM is exploited as a nutritional supplement under the trade names SAM-e, SAME, or SAMe. SAM is also marketed in other countries such as Germany, Russia and Italy, under the Adomet, Admethionine, Gumbaral, Heptral and Samyr brand names as a prescription drug. It should be noted that the use of SAM for therapeutic applications and as a dietary supplement accelerated rapidly after the Dietary Supplement Health and Education Act (1994) came into force in the US.

16.1.2 SAM Preparation

The demand for SAM is increasing, which leads to the developing of new manufacture methods. To date SAM is prepared by chemical synthesis such as directed by United States Patent US 6,881,837 B2, enzymatic synthesis *in vitro* and biotransformation methods. Because of the potential reduction in cost biochemical methods have attracted greater interest recently. SAM can be synthesized by methionine adenosyltransferase (MAT, EC 2.5.1.6) from L-Met and adenosine triphosphate (ATP) in a reaction of the form:

$$L - Met + ATP \rightarrow SAM + PPi + Pi$$

A half century ago, MAT was purified from animal liver and employed to catalyze SAM synthesis *in vitro* (Catoni 1953). However, MAT cannot be extracted efficiently at low cost, thus there have been few advances using this method (Gross et al. 1983; Matos et al. 1987; Schlenk and Depalma 1957). Recently, Luo et al. (2008) constructed a recombinant *Pichia pastoris* KM71, which secreted MAT by exploiting an α-factor secretion signal. After cultivation at flask scale, MAT was harvested at a content about 200 mg/L and a specific activity of 23.84 U/mg.

Compared to the *in vitro* method, SAM synthesis *in vivo* employing microorganism is more competitive. Because the *in vivo* methods are more efficient, convenient

and robust they have attracted greater research interest worldwide. A large number of microbes with an ability to accumulate high concentrations of SAM was screened out and reconstructed a few decades ago when various expression systems and recombinant technologies became available. The resulting strains were cultured in media supplemented with a substrate of L-Met, and SAM was synthesized from L-Met taken up by the cells and ATP that was generated as a product of cellular metabolism. The main sources of ATP are HMP and TCA pathways in the presence of MAT expressed either natively or excessively. The high concentration of SAM after a long cultivation period was subsequently extracted from the cells and purified.

Many attempts have been made to synthesize and accumulate SAM using a culture medium enriched with L-Met with different microorganisms. Previous researches focused on isolation of various strains of known high producers, such as *Saccharomyces sake* (Shiozaki et al. 1984, 1986), *Kluyveromyces lactis* (Mincheva et al. 2002) and *Saccharomyces cerevisiae* (Lin et al. 2004; Shiomi et al. 1990). Recently, Shobayashi et al. (2006) developed a new efficient selection strategy for screening yeast with high SAM content. The abovementioned strains displayed a positive correlation between SAM accumulation and MAT, indicating SAM synthetase was a limiting factor in SAM formation. The employment of these microorganisms led to a significant increase in the harvest level and its percentage of cell weight. With the rapid development of various expression systems and recombinant technology, SAM formation can be further enhanced by overexpression of MAT in host strain such as *P. pastoris*, *S. cerevisiae* and *Escherichia coli* (Mato et al. 1995).

SAM accumulation can be enhanced in two ways: increased SAM synthesis, which can be achieved through increasing pool of the bottleneck factor, and the other is to reduce SAM conversion and turnover Ambits. Hence, approaches for improvement of SAM production fall into two categories: (1) Genetic modification, including enhancement of expression level of MAT, molecular evolution to elevate MAT specific activity, and knock-in of *vgb* to elevate ATP; (2) Fermentation optimization, including alternate feeding of methanol and glycerol during induction phase, unlimited supplement of glycerol by pulsed feeding mode, modification of L-Met feeding profiles. The reported strategies, as well as their mechanism and performance are summarized in Table 16.1.

16.2 Genetic Modifications

16.2.1 Modifying Expression Level and Specific Activity of MAT

16.2.1.1 Overexpression of MAT

SAM is produced in viable cell as an intermediate metabolite. Because it is readily converted to other active biomolecules, it does not accumulate and hence is in relatively low concentration. It is detected at a level of 0.04 g/L in wild

Table 16.1 The reported genetic and bioprocess optimization strategy for SAM production

Microorganism	Strategy and mechanism	SAM level (g/L)	Increasing range[a]	Reference
Enhancement of MAT[b] expression				
P_{AOX}-GS115[c]	Overexpressing MAT	1.58 (flask)	Not mentioned	Li et al. (2002)
P_{AOX}-GS115	Overexpressing MAT	0.74 (flask)	Almost 20-folds	He et al. (2006)
P_{GAP}-GS115	Overexpressing MAT	2.49 (flask)	57.6%	Yu et al. (2003)
S. cerevisiae H5M147	Overexpressing MAT	7.76 (flask); 9.64 (fermentor)	42.98%	Huang et al. (2011)
P_{GAP}-GS115/MAT[d]	Unlimitedly feeding glycerol to keep about 2% residual in broth, in order to increase glycerol consumption rate by enhancing oxygen transfer rate in broth	9.26 (fermentor)	77.39%[e]	Hu et al. (2008)
S. cerevisiae sake kyokai No. 6	Overexpressing MAT and *ERC1* that endowed strain tolerance to ethionine and capacity of accumulating more SAM	2 (flask)	Not mentioned	Lee et al. (2010)
Improvement of MAT specific activity				
P_{AOX}-GS115	Shuffling and screening of the over-expressed *mat*	6.14 (fermentor)	About 1.0-fold	Hu et al. (2009a)
Fine-tuning MAT expression				
P_{GAP}-GS115/DS56	Modification of P_{GAP} driving the over-expressed *mat*	11.04 (fermentor)	Not mentioned	Qin et al. (2011)
Increase of ATP generation				
P_{AOX}-GS115/MAT	Alternative feeding methanol and more utilizable glycerol	13.24 (fermentor)	34.30%	Hu et al. (2007)
Elevation of conversion ratio of L-Met to SAM				
P_{AOX}-GS115/DS56	Continuously feeding L-Met at 0.5 g/L/h, to ensure sufficient supply and no inhibition effect	8.46 (fermentor)	48.9%	Hu et al. (2009b)
Blocking the flux of SAM to cystathionine				
P_{AOX}-GS115/MAT	Knocking out of CBS	13.5 (fermentor)	About 2.8-fold	He et al. (2006)

[a]The value was not always that under optimized cultivation condition
[b]MAT was SAM synthetase II from *Saccharomyces cerevisiae* unless otherwise stated
[c]GS115 was *P. pastoris* GS115 strain
[d]P_{GAP}-GS115/MAT meant P_{GAP}-GS115 that transformed with MAT gene
[e]The value was achieved taking the best limited feeding scheme (strategy C) as reference

P. pastoris GS115, for example. The background activity of MAT is also correspondingly low. In order to achieve high SAM production the primary strategy is to enhance MAT expression. The activity MAT, such as *metK*-encoding enzyme in *E. coli*, is normally reduced when SAM accumulates in relatively high concentrations. It is therefore necessary to overcome the problem of product inhibition. This can be alleviated or even eliminated by addition of sodium *p*-toluenesulfonate, β-mercaptoethanol, acetonitrile, or urea. Furthermore two isozymes from *S. cerevisiae* differ in response to product inhibition. SAM synthase I is implicated in regulation mechanism of SAM metabolism and is markedly inhibited by product accumulation. SAM synthetase II encoded by *sam2* gene by contrast, is not inhibited by relatively high concentrations of SAM (Park et al. 1996). Therefore *sam2* was selected as the optimum MAT gene. Over-expression or modification of *sam2* is an obvious choice for optimizing the MAT gene. The following discussion is mainly related to *sam2*.

MAT Over-Expression in *P. pastoris*

During the past decade *P. pastoris* has been developed as a successful protein expression system. The yeast can grow on media employing methanol as the sole carbon source (C-source) because the peroxisomal enzyme alcohol oxidase (encoded by *AOX1* gene) can be over-expressed using methanol as inducer (Gleeson and Sudbery 1988). The induction of *AOX1* was attributed to the strong *AOX1* promoter (P_{AOX}) (Cregg et al. 1989), and P_{AOX} was successfully employed using several vectors to overproduce cytosolic and secreted proteins (Cregg et al. 1993; Scorer et al. 1994). The over-expressed protein retained within cell acts to catalyze the reaction in situ. In addition to P_{AOX}, the promoter of glyceraldehyde-3-phosphate dehydrogenase (P_{GAP}) has recently been investigated for the strong induction of expression of foreign proteins (Waterham et al. 1997).

P. pastoris has a unique advantage that it is capable of growing on minimal media to achieve a relatively high biomass (Cereghino and Cregg 2000). Furthermore, the excessive SAM can be sequestered in the vacuole of yeast cell, so yeast (such as *P. pastoris* and *S. cerevisiae*) can be regarded as an optimal host strain for SAM synthesis *in vivo* (Chan and Appling 2003; Shobayashi et al. 2007). Many reports have been published to document improvements of SAM production by *P. pastoris* (Li et al. 2002; Yu et al. 2003).

Li et al. (2002) transformed *sam2* into *P. pastoris* GS115 via pPIC3.5k, and G418 screening revealed that the resultant strain possessed two copies of expression cassette. Compared to the control strain GS115/pPIC3.5k, GS115/pPIC3.5k-*sam2* produced much more MAT, with 37-fold greater volumetric activity. These results showed that SAM level was elevated to 1.58 g/L. Likewise, Yu et al. (2003) also transformed pGAPz-based *sam2* into *P. Pastoris* GS115, and similarly, MAT expression was also found to increase in the resulting strain. Employing the P_{GAP}-driven GS115/pGAPz-*sam2*, SAM production was increased to 2.49 g/L in shake flasks, somewhat greater (58%) than GS115/pPIC3.5k-*sam2* described above.

MAT Over-Expression in *S. cerevisiae*

Improvements in SAM yields by over-expression of MAT were also reported in *S. cerevisiae*. By means of spaceflight culture, Huang et al. (2011) obtained a strain *S. cerevisiae* H5M147, with 87% greater SAM concentrations compared to the H5 strain of *S. cerevisiae*. Subsequently, P_{GAP} was cloned from *S. cerevisiae* H5, and inverted into pGAPZ A vector to create novel pGAPZ S, which was then used to transformed *mat2*. These resulted in a 43% increase in SAM yield compared to the parent H5M147 strain.

S. cerevisiae sake K6 was the first industrial strain isolated for SAM production. By UV mutagenesis, a mutant auxotroph strain K6-1 was isolated, and it displayed comparable growth rates and SAM yield to those of K6. Subsequent over-expression of *sam2* in K6-1 evidently resulted in improved SAM productivity. Expression of *ERC1*, which endowed the strain with tolerance to ethionine and was responsible for MAT activity, led to accumulation of considerably more SAM (Shiomi et al. 1991). Recently, *SAM2* and *ERC1* were co-expressed in strain ES1, leading to yield of 2 g/L SAM (Lee et al. 2010).

16.2.1.2 Molecular Evolution of MAT

In most cases, a positive correlation has been observed between SAM synthesis and MAT activity. In order to elevate the total MAT activity, it is a routine strategy to enhance its expression; however, elevation of the specific activity by molecular evolution is a more promising approach. To obtain modified MAT with elevated catalytic activity, three MAT genes from *S. cerevisiae*, *E. coli* D H5α, and *Streptomyces spectabilis* (GenBank Accession Nos. AE000377, M23368, and AF117274) that shared 52–62% nucleic acid identity were recombined by DNA shuffling. The shuffled genes were then transformed into *P. pastoris* GS115, and the strains with enhanced MAT activity and SAM production were screened out. In the two optimum mutants, twofold and 65% higher MAT activities were achieved in comparison to the wild type MAT, respectively. This resulted in SAM yields doubling or increasing by 65%, respectively (Hu et al. 2009a).

Five representative strains were selected, and the shuffled integrated MAT genes were cloned and sequenced. From the DNA sequences, the deduced amino acid sequences of the shuffled MAT (pDS16, pDS56, pDS4, pDS34, and pDS69) were analyzed. The pDS56 displayed the highest activity and consequently, SAM titer with the optimum strain GS115/DS56 reached to 6.14 g/L. The pDS56 displaying the greatest activity had four site mutations (K18R, L31P, I65V, and D341G), and furthermore, a frame shift mutation (395th-404th) was found in its DNA sequence, which resulted in a shortened MAT with 404 amino acid residues. However, pDS56 activity was greater than wild-type MAT from *S. cerevisiae* (pSA) and MAT from *S. spectabilis* (pST). Swiss-model server was employed to construct tertiary structure, and the K18R substitution is the only mutation in the conservative sites in pDS56 were most likely to have played a key role, even with

a truncated protein (Hu et al. 2009a). This work provided new insights into the relationship between structure and activity of MAT, and would benefit from further rational evolution.

16.2.1.3 Fine-Tuning of MAT Expression

At present the most common promoters in use are P_{AOX} and P_{GAP} the action of which have been demonstrated by the over-expression pDS56 under control of either promoter (Hu et al. 2009a). However, the action of P_{AOX} and P_{GAP} may not be strong enough for MAT expression required to maximize SAM synthesis. Nevertheless, excessive over-expression of foreign genes has a down side which can lead to metabolic burden, due to competition for carbon and nitrogen sources and energy. This may be the main reason why cell growth is inhibited which, in turn, leads to reduced SAM production (Hu et al. 2009a). To solve this problem, it is necessary to tune the expression strength.

It has been reported that modification of promoters with various strengths can be exploited to precisely control gene expression (Alper et al. 2005). However, this approach has been confined mainly to bacteria (Andersen et al. 2001; Bakke et al. 2009; De Mey et al. 2007). Reports of promoter engineering in yeast are few (Alper et al. 2005; Nevoigt et al. 2006; 2007), and the only promoter library has been constructed for P_{AOX} in *P. pastoris* (Hartner et al. 2008). By means of deletion and duplication of the putative transcription factor binding sites, a library of 46 mutant P_{AOX} was constructed (Hartner et al. 2008), which displayed a relative activity of between 6 and 160% of the control.

To investigate the effect of MAT activity on SAM synthesis, and then to optimize MAT expression for SAM synthesis, Qin et al. (2011) firstly constructed a library of P_{GAP} with the aid of error-prone PCR. The P_{GAP} mutants were then characterized and seven P_{GAP} mutants were selected to drive MAT expression. From this an optimal P_{GAP} mutant for MAT and SAM production was identified.

The gene of yeast-enhanced green fluorescent protein was employed to estimate the strength of P_{GAP} variant (Hu et al. 2009a). Eight mutants (G1, G2, G3, G4, G5, GAP, G6, and G7) with different P_{GAP} (including the strongest and extremely weak promoter) were selected from the promoter library to express MAT gene *ds56*, encoding the optimum MAT created by DNA shuffling. When a single copy of the expression cassette was inserted into the chromosome, the specific MAT activities of the eight variants spanned a 57-fold range. Promoter performance was also evaluated at the transcriptional level by quantitative reverse transcription-PCR. Results showed that there was a positive relationship between promoter strength and MAT activity. Furthermore, when MAT activity was less than 1.05 U/mg of protein in seven strains (from weak to strong: G7, G6, GAP, G5, G4, G3 and G2), an almost linear relationship between MAT activity and SAM yield was displayed, indicating enzyme was the rate-limiting factor in the current strain and cultivation conditions. By contrast, when MAT activity exceeded 1.33 U/mg in mutant G1, SAM concentration reached a plateau, suggesting under this strongest promoter, MAT expression

was no longer the critical factor limiting SAM synthesis. Besides, compared to G2, G1 showed a decreased biomass, which was most likely related to metabolic burden. As a consequence, the greatest SAM yield of 11.04 g/L was achieved in G2, 4.5% higher than G1. Thus, the mutant promoter in strain G2, was the strongest and the optimum for SAM accumulation (Qin et al. 2011).

16.2.2 Knocking in of Vitreoscilla *Hemoglobin (VHb) to Improve ATP Generation*

From the SAM synthesis reaction, it can be seen that MAT activity is only one limiting factor in SAM accumulation (Shiozaki et al. 1986), and sufficient supply of the two substrates for the reaction, including ATP mentioned in this subsection and L-Met in Sect. 16.3.3, can also become a bottleneck factor in certain instances. As Shiozaki et al. (1989) reported competition for ATP between energy-consuming processes and SAM synthesis leads to reduced synthesis of SAM. *P. pastoris* can grow to a cell density greater than 130 g/L, but under these conditions the oxygen requirement is extremely critical for two reasons: (1) Oxygen demand is high because *P. pastoris* prefers respiratory rather than production of SAM and, furthermore, methanol metabolism with P_{AOX}-*Pichia* has a greater demand for oxygen; (2) Oxygen solubility and transfer efficiency in the fermentation broth is poor, especially in late fermentation, since the high biomass and antifoam agents will reduce mass transfer efficiencies of oxygen and reduce its solubility.

VHb had been reported by many researchers to enhance oxygen utilization (Dikshit and Webster 1988). Knock in of *Vitreoscilla* VHb has proved to be an efficient strategy to improve oxygen uptake, as well as to promote growth and stimulate expression of proteins of interest (Bhave and Chattoo 2003; Minas et al. 1998). Chen et al. (2007) co-expressed VHb gene (*vgb*) and *mat* in *P. pastoris* GS115, both under the control of P_{AOX}. CO-difference spectrum confirmed the functional expression of VHb. When the methanol inducer was maintained at between 0.8 and 2.4% in the broth, VHb improved respiratory capability, cell growth rate and SAM production. Improvement in SAM formation by VHb was most likely to be due to stimulation of ATP synthesis.

16.2.3 Knocking Out of Cystathionine-β-Synthase (CBS) to Depress SAM Degradation

From metabolic fluxes illustrated in Fig. 16.2, it can be seen that SAM accumulation can be maximized not only by stimulation of synthesis, but also by inhibition of SAM degradation. Located at the critical branch point of SAM metabolism,

homocysteine has three alternative fates: to form methionine, or to revert back to SAH, or to form cystathionine in the presence of the enzyme cystathionine-β-synthase. CBS plays a key role in drawing homocysteine away from the SAM cycle, thus blocking the removal pathway by disrupting CBS is likely to enhance SAM accumulation. He et al. (2006) reported the disruption of CBS in SAM producer. *P. pastoris* Gsam that was transformed with MAT gene to produce 20-fold more SAM than GS115 was used as the starter strain. As expected, SAM content was doubled, and about 56-fold higher SAM than the initial strain GS115 was achieved. Knocking in of *mat* and knocking out of *cbs* were proved to have a synergistic effect on SAM accumulation as seen in shake flask (3.6 g/L) and in a 5 L fermenter (13.5 g/L).

16.3 Bioprocess Optimizations

The industrial production of SAM has been achieved in bioreactors with working volumes usually greater than 100 L. In order to enhance SAM synthesis while depress its conversion, strain modification is not sufficient and optimization of the process is also required.

16.3.1 Stimulation of MAT Expression

Driven by P_{GAP}, MAT was constitutively expressed throughout the fermentation cycle. MAT expression strength was not only controlled by the driving strength of P_{GAP}, but also influenced by the metabolic rate of C-source glycerol, which was further affected by the rate of oxygen uptake. Therefore, Hu et al. (2008) investigated different glycerol supplement regimes on MAT activity and SAM production. Five glycerol feeding strategies included: Policies A-C employed limited feeding mode correlated with DO at 50, 25 and 0%, respectively; Strategies D and E employed unlimited feeding mode (Glycerol residual level was maintained between 0 and 2.0% in fermentation D, while a comparable stable glycerol level at 2% during fed-batch cultivation in fermentation E). When strategies A-E were shifted in alphabetical order, a gradual increase in MAT activity was observed, with the greatest value of 9.05 $U\,g^{-1}$ for fermentation E. The reason was that the specific rate of consumption of glycerol was enhanced because of an elevation of specific oxygen uptake capacity. The improvement resulted from the stimulation of the oxygen transfer rate in the culture broth, since oxygen saturation content and volumetric oxygen transfer coefficient of the fermentation broth were improved with 2% glycerol residue. Strategy E, of course, resulted in the maximal levels of MAT, ATP and biomass. As a result, the best harvest level of SAM was obtained 9.26 $g\,L^{-1}$ (Hu et al. 2008).

16.3.2 Improvement of ATP Generation

The rate of ATP synthesis is closely related to the type of carbon source. As a strategy to boost SAM synthesis, Hu et al. (2007) reported a novel supplement strategy for inducible *Pichia*. The classical induction period was divided into several 10-h subintervals, and methanol (initial 6 h) and glycerol (subsequent 4 h) were fed alternately during every subinterval. Compared to the typical feeding mode, the novel scheme elevated growth and MAT volumetric activity, while decreased the specific MAT activity. The rate of SAM synthesis within *Pichia* cells, however, was still improved, mainly due to the enhancement of ATP supply. The switch from methanol to glycerol was invariably associated with increased ATP synthesis. SAM synthesis was also favored by an extension of the production period such that SAM content was reputed to be elevated to 13.24 g/L at 5-L bioreactor scale.

16.3.3 Elevation of L-Met Conversion Ratio to SAM

L-Met added as substrate for SAM synthesis can be consumed by microorganism in protein synthesis, desulfurization etc. The availability of L-Met in broth influenced its flow distribution, and affected the conversion ratio of L-Met to SAM, which bears directly on product cost of SAM. The supplement mode of L-Met has been investigated by a number of researchers. Hu et al. (2009b) compared the performance of batch addition mode (L-Met was added in three stages) and continuous feeding mode (0.2 g/L/h), and revealed that the latter strategy was of greater benefit for SAM formation (8.46 g/L) and efficient conversion of L-Met to SAM (41.7%). For the continuous feeding mode, different L-Met feeding rates at 0.1, 0.2, and 0.5 g/L/h were investigated. A low feeding rate at 0.1 g/L/h resulted in limiting L-Met supply, while a high rate at 0.5 g/L/h led to a reduction in activities of enzymes involved in tricarboxylic acid cycle and glycolytic pathway, which reduces SAM by reducing ATP synthesis.

16.4 Conclusion

SAM synthesis from L-Met and ATP that is catalyzed by MAT is a dynamic process. To elevate SAM productivity different strategies had been adopted at gene and bioprocess levels. The accumulation of SAM is influenced not only by its rate of synthesis but also by SAM acting as a substrate for further reactions. Two approaches to optimize SAM production have therefore been adopted: enhancing SAM synthesis, and inhibiting SAM conversion. To stimulate SAM formation, the strategies at the genetic level include modification of MAT expression and activity and knocking in of *vgb* to increase ATP. The first of these consists of over-expression of MAT

employing P_{AOX} and P_{GAP}, evolution of MAT by DNA shuffling, and fine-tuning of MAT expression level by promoter engineering. The strategies for a bioprocess scale include alternative feeding of methanol and glycerol during the induction period for P_{AOX}-*Pichia*, unlimited glycerol supplement with 2% residue in broth, and continuous L-Met feeding mode. To depress SAM conversion to other intermediate, knocking out of *cbs* has proven to be an effective procedure.

Acknowledgements Funding was provided by grants from National key Basic Research Program of China (2012CB725202), the Public Topic of Key Laboratory of Industrial Biotechnology, Ministry of Education (KLIB-KF201003) and State Key Laboratory of Food Science and Technology, Jiangnan University (SKLF-TS-201124).

References

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci USA 102:12678–12683

Andersen HW, Pedersen MB, Hammer K, Jensen PR (2001) Lactate dehydrogenase has no control on lactate production but has a strong negative control on formate production in *Lactococcus lactis*. Eur J Biochem 268:6379–6389

Bakke I, Berg L, Aune TE, Brautaset T, Sletta H, Tondervik A, Valla S (2009) Random mutagenesis of the PM promoter as a powerful strategy for improvement of recombinant-gene expression. Appl Environ Microbiol 75:2002–2011

Barcelo HA, Wiemeyer JC, Sagasta CL, Macias M, Barreira JC (1990) Experimental osteoarthritis and its course when treated with S-adenosyl-L-methionine. Rev Clin Esp 187:74–78

Bhave SL, Chattoo BB (2003) Expression of *vitreoscilla* hemoglobin improves growth and levels of extracellular enzyme in *Yarrowia lipolytica*. Biotechnol Bioeng 84:658–666

Catoni GL (1952) The nature of the active methyl donor formed enzymatically from L-methionine and adenosinetriphosphate. J Am Chem Soc 74:2942–2943

Catoni GL (1953) S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. J Biol Chem 204:403–416

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methalotropic yeast *Pichia pastoris*. FEMS Microbiol Rev 24(1):45–66

Chan SY, Appling DR (2003) Regulation of S-adenosylmethionine levels in *Saccharomyces cerevisiae*. J Biol Chem 278:43051–43059

Chen H, Chu J, Zhang S, Zhuang Y, Qian J, Wang Y, Hu X (2007) Intracellular expression of *Vitreoscilla* hemoglobin improves S-adenosylmethionine production in a recombinant *Pichia pastoris*. Appl Microbiol Biotechnol 74:1205–1212

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol Cell Biol 9:1316–1323

Cregg JM, Vedvick TS, Raschke WC (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. Biotechnol (NY) 11:905–910

De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. BMC Biotechnol 7:34. doi:10.1186/1472-6750-7-34

Dikshit KL, Webster DA (1988) Cloning, characterization and expression of the bacterial globin gene from *Vitreoscilla* in *Escherichia coli*. Gene 70:377–386

Friedel HA, Goa KL, Benfield P (1989) S-adenosyl-L-methionine. A review of its pharmacological properties and therapeutic potential in liver dysfunction and affective disorders in relation to its physiological role in cell metabolism. Drugs 38:389–416

Gleeson MA, Sudbery PE (1988) The methylotrophic yeasts. Yeast 4:1–15
Gross A, Geresh S, Whitesides GM (1983) Enzymatic synthesis of S-adenosyl-L-methionine from L-methionine and ATP. Appl Biochem Biotechnol 8:415–422
Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res 36:e76
He J, Deng J, Zheng Y, Gu J (2006) A synergistic effect on the production of S-adenosyl-L-methionine in *Pichia pastoris* by knocking in of S-adenosyl-L-methionine synthase and knocking out of cystathionine-beta synthase. J Biotechnol 126:519–527
Hu X, Chu J, Zhang S, Zhuang Y, Wang Y, Zhu S, Zhu Z, Yuan Z (2007) A novel feeding strategy during the production phase for enhancing the enzymatic synthesis of S-adenosyl-L-methionine by methylotrophic *Pichia pastoris*. Enzyme Microb Technol 40:669–674
Hu X, Chu J, Zhang Z, Zhang S, Zhuang Y, Wang Y, Guo M, Chen H, Yuan Z (2008) Effects of different glycerol feeding strategies on S-adenosyl-L-methionine biosynthesis by P_{GAP}-driven *Pichia pastoris* overexpressing methionine adenosyltransferase. J Biotechnol 137:44–49
Hu H, Qian J, Chu J, Wang Y, Zhuang Y, Zhang S (2009a) DNA shuffling of methionine adenosyltransferase gene leads to improved S-adenosyl-L-methionine production in *Pichia pastoris*. J Biotechnol 141:97–103
Hu H, Qian J, Chu J, Wang Y, Zhuang Y, Zhang S (2009b) Optimization of L-methionine feeding strategy for improving S-adenosyl-L-methionine production by methionine adenosyltransferase overexpressed *Pichia pastoris*. Appl Microbiol Biotechnol 83:1105–1114
Huang Y, Gou X, Hu H, Xu Q, Lu Y, Cheng J (2011) Enhanced S-adenosyl-L-methionine production in *Saccharomyces cerevisiae* by spaceflight culture, overexpressing methionine adenosyltransferase and optimizing cultivation. J Appl Microbiol 112:683–694
Lee SW, Park BS, Choi ES, Oh MK (2010) Overexpression of ethionine resistance gene for maximized production of S-adenosylmethionine in *Saccharomyces cerevisiae sake* kyokai No 6. Korean J Chem Eng 27(2):587–589
Li DY, Yu J, Tian L, Ji XS, Yuan ZY (2002) Production of SAM by recombinant *Pichia pastoris*. Chin J Biotechnol 3:295–299
Lieber CS (1999) Role of S-adenosyl-L-methionine in the treatment of liver diseases. J Hepatol 30:1155–1159
Lin JP, Tian J, You JF, Jin ZH, Xu ZN, Cen PL (2004) An effective strategy for the co-production of S-adenosyl-L-methionine and glutathione by fed-batch fermentation. Biochem Eng J 21:19–25
Luo Y, Yuan Z, Luo G, Zhao F (2008) Expression of secreted His-tagged S-adenosylmethionine synthetase in the methylotrophic yeast *Pichia pastoris* and its characterization, one-step purification, and immobilization. Biotechnol Prog 24:214–220
Mato JM, Pajares MA, Mingorance J, and Avarez L (1995) Production of S-adenosylmethionine (SAM) by fermentation of transformed bacteria. European Patent, vol 0647712A1, edited by E. P. Office.
Mato JM, Alvarez L, Ortiz P, Pajares MA (1997) S-adenosylmethionine synthesis: molecular mechanisms and clinical implications. Pharmacol Ther 73:265–280
Matos JR, Raushel FM, Wong CH (1987) S-Adenosylmethionine: study on chemical and enzymatic synthesis. Biotechnol Appl Biochem 9:39–52
Minas W, Brunker P, Kallio PT, Bailey JE (1998) Improved erythromycin production in a genetically engineered industrial strain of *Saccharopolyspora erythraea*. Biotechnol Prog 14:561–566
Mincheva K, Kamburova V, Balutzov V (2002) Production of S-adenosyl-L-methionine by a mutant strain of *Kluyveromyces lactis*. Biotechnol Lett 24:985–988
Nevoigt E, Kohnke J, Fischer CR, Alper H, Stahl U, Stephanopoulos G (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl Environ Microbiol 72:5266–5273
Nevoigt E, Fischer C, Mucha O, Matthaus F, Stahl U, Stephanopoulos G (2007) Engineering promoter regulation. Biotechnol Bioeng 96:550–558

Park J, Tai JZ, Roessner CA, Scott AI (1996) Enzymic synthesis of s-adenosyl-L-methionine on the preparative scale. Bioorg Med Chem 4:2179–2185

Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol 77:3600–3608

Schlenk F, Depalma RE (1957) The preparation of S-adenosylmethionine. J Biol Chem 229:1051–1057

Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K (1994) Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. Biotechnol (N Y) 12:181–184

Shiomi N, Fukuda H, Fukuda Y, Murata K, Kimura A (1990) Production of S-adenosyl-L-methionine by *Saccharomyces cerevisiae* cells carrying a gene for ethionine resistance. Biotechnol Bioeng 35:1120–1124

Shiomi N, Fukuda H, Fukuda Y, Murata K, Kimura A (1991) Nucleotide sequence and characterization of a gene conferring resistance to ethionine in yeast *Saccharomyces cerevisiae*. J Ferment Bioeng 71:211–215

Shiozaki S, Shimizu S, Yamada H (1984) Unusual intracellular accumulation of S-adenosyl-L-methionine by microorganisms. Agric Biol Chem 48:2293–2300

Shiozaki S, Shimizu S, Yamada H (1986) Production of S-adenosyl-L-methionine by *Saccharomyces sake*. J Biotechnol 4:345–354

Shiozaki S, Shimizu S, Yamada H (1989) S-Adenosyl-L-methionine production by *Saccharomyces sake*: optimization of the culture conditions for the production of cells with a high S-adenosyl-L-methionine content. Agric Biol Chem 53:3269–3274

Shobayashi M, Mukai N, Iwashita K, Hiraga Y, Iefuji H (2006) A new method for isolation of S-adenosylmethionine (SAM)-accumulating yeast. Appl Microbiol Biotechnol 69:704–710

Shobayashi M, Fujii T, Iefuji H (2007) Effects of accumulated S-adenosylmethionine on growth of yeast cells. Biosci Biotechnol Biochem 71:1595–1597

Stipanuk MH (1986) Metabolism of sulfur-containing amino acids. Annu Rev Nutr 6:179–209

Torta R, Cicolin A, Keller R (1998) Transmethylation and affective disorders. Arch Gerontol Geriatr Suppl 6:499–506

Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186:37–44

Yu ZL, Wu XJ, Li DY, Yang S, Zhou Z, Cai J, Yuan ZY (2003) Enhancement of the production of SAM by overexpression of SAM synthetase in *Pichia pastoris*. Acta Biochem Biophys Sin 35:127–132

Chapter 17
Manipulation of *Ralstonia eutropha* Carbon Storage Pathways to Produce Useful Bio-Based Products

Christopher J. Brigham, Natalia Zhila, Ekaterina Shishatskaya, Tatiana G. Volova, and Anthony J. Sinskey

Abstract *Ralstonia eutropha* is a Gram-negative betaproteobacterium found natively in soils that can utilize a wide array of carbon sources for growth, and can store carbon intracellularly in the form of polyhydroxyalkanoate. Many aspects of *R. eutropha* make it a good candidate for use in biotechnological production of polyhydroxyalkanoate and other bio-based, value added compounds. Manipulation of the organism's carbon flux is a cornerstone to success in developing it as a biotechnologically relevant organism. Here, we examine the methods of controlling and adapting the flow of carbon in *R. eutropha* metabolism and the wide range of compounds that can be synthesized as a result. The presence of many different carbon utilization pathways and the custom genetic toolkit for manipulation of those pathways gives *R. eutropha* a versatility that allows it to be a biotechnologically important organism.

C.J. Brigham
Department of Biology, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA

N. Zhila. E. Shishatskaya • T.G. Volova
Institute of Biophysics, Siberian Federal University, Krasnoyarsk 660036, Russian Federation

A.J. Sinskey (✉)
Department of Biology, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA

Engineering Systems Division, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA

Division of Health Sciences Technology, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: asinskey@mit.edu

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_17,

Keywords Biotechnology • Carbon flux • Metabolic engineering • Polyhydro xyalkanoate • *Ralstonia eutropha*

Abbreviations

A	Gas coefficient
C/N	Carbon/nitrogen ratio
DAPI	4′,6-Diamidino-2-Phenylindole
3HB	3-Hydroxybutyrate
4HB	4-Hydroxybutyrate
3HHx	3-Hydroxyhexanoate
3H4MV	3-Hydroxy-4-methyl valerate
3HV	3-Hydroxyvalerate
IR	Infrared
MCL	Medium chain length
MMBPP	Malaysia/MIT Biotechnology Partnership Programme
MTT	3-(4,5-Dimethylthiazol-2-Yl)-2,5-Diphenyltetrazolium Bromide
PHA	Polyhydroxyalkanoate
PhaB	3-Acetoacetyl-CoA reductase
P(HB-*co*-HV)	Poly(hydroxybutyrate-*co*-hydroxyvalerate)
P(HB-co-HHx)	Poly(hydroxybutyrate-*co*-hydroxyhexanoate)
PhaC	PHA synthase
PhaP	Phasin protein
PhaZ	PHA depolymerase
PHB	Polyhydroxybutyrate
PLA	Polylactic acid
POME	Palm oil mill effluent
q	Specific rate of gas substrate consumption, in kg/(kg·h)
RAD16-I::E	Self-assembling peptide with a sequence of arginine alanine and aspartate repeats ending in alanine and glutamate
SB RAS	Siberian Branch, Russian Academy of Sciences
SCL	Short chain length
TCA cycle	Tricarboxylic acid cycle
X_{CO}	Volumetric concentration of carbon monoxide in a gas mixture, in %

17.1 Introduction

Ralstonia species can accumulate large amounts of polyhydroxyalkanoate (PHA) under unbalanced growth conditions, where protein synthesis reactions are limited and carbon and energy are present in abundance. *Ralstonia eutropha* (also known as

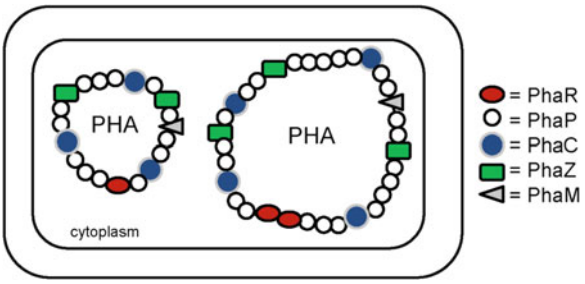


Fig. 17.1 Schematic drawing of a *R. eutropha* cell containing PHB granules. Granules are surrounded by many different proteins, all with different functions in PHB homeostasis

Cupriavidus necator, formerly known as *Alcaligenes eutrophus*) has become the paradigm for study of PHA biosynthesis, largely because of its genetic tractability (Brigham et al. 2010; Budde et al. 2010; York et al. 2003) and that large amounts of intracellular PHA (≥80% of cell dry weight) are typically accumulated under nitrogen or phosphate deficiency (Anderson et al. 1990; Anderson and Dawes 1990). PHA is typically sequestered intracellularly in the form of inclusion bodies, termed granules. These granules are complex nutrient sequestration and mobilization organelles (Jendrossek 2009) that control the flow of carbon and reducing potential in the cell. The anatomy of a PHA granule in *R. eutropha* is shown in Fig. 17.1. Despite the large body of literature on the biology of the PHA granule, in *R. eutropha* there is still much to be learned about the roles of several granule-associated proteins (e.g. phasins PhaP2-4, and PHB depolymerases PhaZ2, PhaZ3, PhaZ5). PHA is a family of polymers with over 150 types known to exist (Valentin and Steinbüchel 1994), all differentiated by the monomers incorporated into them. Wild-type *R. eutropha* produces PHA containing short chain length (SCL) monomers like PHB (containing only the 3-hydroxybutyrate, or 3HB, monomer, Fig. 17.2) and P(HB-*co*-HV) (containing 3HB and 3-hydroxyvalerate, or 3HV, monomer, Fig. 17.2). Due to the ease of genetic manipulation of *R. eutropha* and a maturing set of genetic tools, strains can be constructed that will produce medium chain length (MCL) PHA and polymer containing monomers of mixed chain length (SCL and MCL) (Budde et al. 2011; Loo et al. 2005; Mifune et al. 2008; Riedel et al. 2012a).

R. eutropha is an industrially relevant organism, and some of its advantageous traits are discussed in Table 17.1. Due to its autotrophic growth capabilities, *R. eutropha* was first considered for CO_2 mitigation in space vessels for astronauts (Schlegel and Lafferty 1971). However, it is generally used for high productivity biosynthesis of PHA biopolymer. PHA has become attractive to the "green" materials industry due to the fact that it is bio-based, biodegradable, biocompatible, and its material properties are adjustable in similar profile to those of several petroleum-based plastics. Thus, PHA is considered a more environmentally friendly alternative to chemically synthesized plastics, and many applications (industrial, household, medical, etc.) are being sought for this class of polymer.

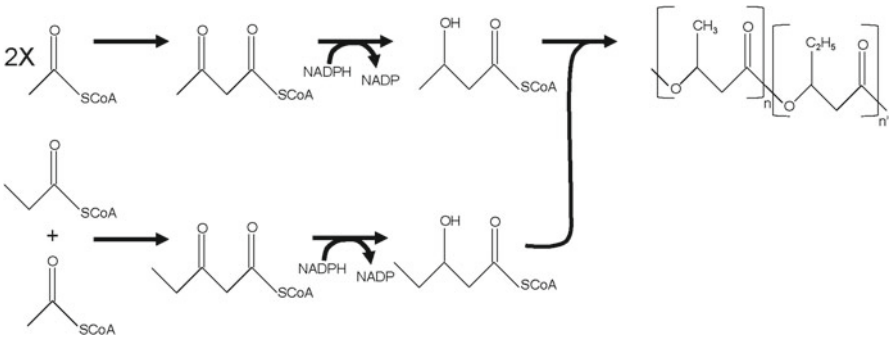


Fig. 17.2 Wild-type *R. eutropha* is capable of producing short chain length (SCL) polyhydroxyalkanoates, like PHB or P(HB-*co*-HV). For PHB production (top reactions), two molecules of acetyl-CoA are ligated together to produce acetoacetyl-CoA, which is then reduced to form 3HB-CoA, the precursor for 3HB monomers. For 3HV monomer production (bottom reactions), acetyl-CoA and propionyl-CoA are ligated together and then reduced to produce 3HV-CoA

Table 17.1 Characteristics of *Ralstonia eutropha* strains that make them suitable for use in industrial production of bio-based products

Characteristic/trait	Relevance in biomaterials and bioproducts production	References
Genetically manipulable	Can construct *R. eutropha* strains that produce different types of bio-based compounds, including many types of PHA	Budde et al. (2011), Loo et al. (2005), Aboulmagd et al. (2001)
Carbon source utilization range	Can produce value added products using plant oils or other inexpensive carbon sources, like agricultural and food processing waste streams	Brigham et al. (2010), Bruland et al. (2009), Volova et al. (2001), Yang et al. (2010)
Robust carbon storage pathway	Can produce intracellular biopolymers with a high productivity and purity	Kahar et al. (2004), Reinecke and Steinbüchel (2009), Riedel et al. (2012a)
Adjustable polymer material properties	Can produce variations of polymer makeup with ranges of medium and longer length monomers through fermentation process controls	Kahar et al. (2004), Reinecke and Steinbüchel (2009), Riedel et al. (2012a)
Autotrophic growth	Can utilize CO_2 for production of biopolymers and other products	Ishizaki et al. (2001), Volova et al. (2001)
Non-pathogenic/bio-compatible	Can be used to produced medical compounds and biopolymers for medical materials and devices	Shishatskaya and Volova (2004), Shishatskaya et al. (2002a)
Resistant to some toxic compounds	Is carbon monoxide resistant (see below) and can produce biopolymers from toxic mixtures, like syngas; can also potentially produce biopolymers from phenol	Savalieva (1979), Volova et al. (1988a)

Table 17.2 Some of the *Ralstonia eutropha* carbon flux pathways and their relevance in biotechnological production of materials and chemicals

Pathway	Relevance in biomaterials and bioproducts synthesis	References
Tricarboxylic acid (TCA) cycle	PHA biosynthesis; utilization of most carbon sources	Yu and Si (2004)
Calvin-Benson-Bassham cycle	Production of biomaterials and chemicals from CO_2	Bowien and Kusian (2002)
Entner-Doudoroff pathway	Utilization of sugars for growth and product formation	Lee et al. (2003)
Fatty acid β-oxidation	Production of biomaterials and chemicals from triacylglycerols and fatty acids	Brigham et al. (2010)
Glyoxylate cycle	Utilization of acetate/acetyl-CoA (incl. β-oxidation byproducts)	Wang et al. (2003), Yu and Si (2004)
Branched chain amino acid biosynthesis	Production of branched carbon chain products (e.g. isobutanol)	Li et al. (2012), Brigham et al. (2012), Lu et al. (2012)

Some of the important *R. eutropha* carbon flux pathways are listed in Table 17.2. These pathways come into play during production of biotechnologically relevant materials and/or chemicals. Alteration of carbon flux in *R. eutropha* does not occur solely due to the presence of heterologous genes and enzymes or the absence of chromosomally encoded pathway genes. A simple change in the sole carbon source of the bacteria will change the types of intermediates through which carbon flows. For example, PHA production using fructose or CO_2 as the sole carbon sources will result in production fructose 6-phosphate and fructose 1,6-bisphosphate intermediates through the pentose phosphate pathway and the Calvin-Benson-Bassham cycle, respectively, while PHA production using triacylglycerols will result in acetyl-CoA intermediates. Thus, we focus not only on genetic manipulations of *R. eutropha* that can alter the types and quantities of carbon-based intermediates in the cells, but also useful compound production and polymer makeup through different carbon sources that take advantage of some of the pathways listed in Table 17.2.

17.2 PHA Biosynthesis by *R. eutropha*

PHA can be produced by *R. eutropha* from various substrates, among them individual compounds (sugars, alcohols, organic acids, triacylglycerols); animal and plant oils (corn oil, lard, tallow, palm oil, palm kernel oil); and waste products of the alcohol, sugar, and hydrolysis industries and olive and palm oil production; etc. (Cromwick et al. 1996; Loo et al. 2005; Tanaka et al. 1995; Yang et al. 2010). As mentioned above, *R. eutropha* can grow and produce PHA autotrophically, using CO_2, H_2, and O_2 as the main growth substrates. A genome-scale metabolic network model has been constructed for *R. eutropha* (strain H16) PHB biosynthesis (Park et al. 2011), confirming the importance of culture pH and C/N ratio for optimal polymer production.

In this chapter, we discuss the many methods used to produce value added products from *R. eutropha* cultures, in particular discussions on feedstocks used and products made from these feedstocks. Manipulation of carbon flow in the organism is a hallmark of *R. eutropha* physiology studies. Two main methods are used for this carbon flux manipulation; genetic alteration of the genes, enzymes, and pathways present in the bacterium, or by changing the type and timing of feedstocks used to synthesize the product. From the summaries below, the versatility and industrial importance of *R. eutropha* and its metabolism is evident.

17.3 PHA Biosynthesis by *R. eutropha* from Syngas

One of the main principles of industrial biotechnology, feedstock availability, implies both the stable presence of economic material and the feasibility of rapid substitution of one feedstock with another, with minimal impact on the process technology or impairment of quality of the resultant product. Hydrogen-oxidizing bacteria like *R. eutropha* are grown on gases (carbon dioxide, oxygen, and hydrogen) that constitute the carbon and energy sources for the cell. A recent metabolic network reconstruction involving autotrophic growth of *R. eutropha* revealed that the growth rate of *R. eutropha* is sensitive to the CO_2/O_2 ratio in an autotrophic gas mixture (Park et al. 2011). Typically, hydrogen is present in vast excess, due to its poor solubility in aqueous solutions (Morinaga et al. 1978).

In an autotrophic gas substrate mixture, hydrogen is the component gas that is most difficult to obtain, and it is essential to find ways to lower its cost in order to develop hydrogen biosynthesis for practical purposes. Toward this end, various coals and coal byproducts have gained attention as substrates for biotechnological production (Fakoussa and Hofrichter 1999). Large reserves of coals and their relatively low cost make them very promising materials for the future large-scale industrial production of microbial bioplastics. It has been shown recently that the bacterium *Pseudomonas oleovorans* and *Rhodobacter ruber* can synthesise PHA polymers in growth media containing coal liquefaction products, i.e. mixtures of humic acids (Fuchtenbusch and Steinbüchel 1999). Using syngas as a growth substrate for hydrogen-oxidizing bacteria like *R. eutropha* is challenging due to the presence of a potentially inhibitory compound, carbon monoxide (CO). Among hydrogen bacteria there are unique organisms that are resistant to CO, including *R. eutropha* strain Z1 (Savalieva 1979; Volova et al. 1987) and its faster-growing variant, *R. eutropha* strain B5786 (Stasishina and Volova 1996).

Studies of *R. eutropha* growth and physico-chemical activities were performed on gaseous substrates containing carbon monoxide in order to determine if these strains could produce large quantities of PHA using gaseous substrates like syngas. While *R. eutropha* strains were shown to be able to grow in the presence of CO (Volova et al. 1988a, b), increases in CO concentrations in the growth media resulted in a decrease in the specific growth rate of the cells, increases in the activities of

hydrogenase enzymes and cytochrome concentrations, and enlargement of the cell membrane system (Volova et al. 1980, 1993).

It has been demonstrated that CO does not exert any significant effect on physiological-biochemical parameters of the nitrogen-limited culture of *R. eutropha* B5786 and to prove that CO-treated culture can produce high yields of polyhydroxyalkanoates (see above). The fact that CO produces no adverse effect on the PHA biosynthesis machinery in *R. eutropha* cell cultures offers opportunities for producing these polymers from gaseous products derived from coal via thermal treatment, such as coal gasification products. However, syngas used for this purpose must conform to certain requirements. First of all is that the H_2/CO ratio in the gaseous substrate must amount to at least 3:1. Second, the syngas must be free of carcinogenic tarry substances and any other biologically hazardous compounds. And finally, the process of brown coal gasification must be economically acceptable.

Syngas can be produced either involving the use of pure oxygen or through a multistage gasification process. In the latter case, heat is supplied into the gasifier with some heat carrier (gas, ash, ceramic balls, etc.) or through the reactor wall. However, processes of the latter kind are of rather low intensity. Moreover, in low-temperature processes of brown coal gasification, tar and tar-like substances make up a large percentage of the product, and it is technically difficult to utilize them for growth substrates. For the industrial gasification processes (Lurgi, Winkler, Koppers-Totzek processes) the CO:H_2 ratio is (48–58):(25–35). This is significantly different from the ratio required for PHA biosynthesis. Researchers proposed several methods for modifying the process of producing syngas and gas/vapor fuel mixture (Kuznetsov et al. 1995; Kuznetsov and Shchipko 1996). They attempted to optimize the process of gasification of the Kansk-Achinsk brown coal to produce syngas that could be used for synthesis of polyhydroxyalkanoates. This process consists of two main phases: oxidizing pyrolysis of coal and gasification of the resulting coal char by water vapor (Shchipko et al. 2003).

As the gas mixture for PHA biosynthesis must contain higher hydrogen concentrations, experiments were performed in which syngas with water vapor was converted into CO_2 and H_2, performed with excess water vapor (30%) at a temperature of 300–350°C. Another method examined was to feed excess vapor to the gasification step. Temperature and vapor feeding rates were regulated in such a way as to ensure that in the gas flux exiting the gasifier the ratio CO:H_2O was equal to 1:3 to 1:5. This method yielded the syngas of the necessary composition without involving the additional stage of catalytic conversion of CO, but the heat efficiency of this gasification process was significantly lower, due to large heat loss with excess vapor. Carbon dioxide was extracted using potassium hydroxide and was then used to for bacterial growth. Studies in which the hydrogenous products of brown coal gasification were used as substrate for growing cells of *R.eutropha* B5786 proved that they could be used in biotechnological processes (e.g. PHA production). PHA specimens synthesized in the presence of CO had a high molecular mass, independent of CO concentration. The temperature parameters of the PHA and its crystallinity degree were also similar to those of the polymers synthesized from electrolytic hydrogen. This new technology of coal gasification does not require the use of

supplemental oxygen, which can be very costly. Thus, the resulting syngas is purer than any other syngas derived from brown coal by gasification techniques. These advantages are attained through uniting of two processes: oxidizing pyrolysis of pulverized coal in the fluidized bed of open-hearth or boiler slag and gasification of the resulting coal char using water vapor, performed in one technological cycle (Shchipko et al. 2003).

As established by the works discussed above, *R. eutropha* was able to grow and produce 70–75% of PHA when grown on a syngas substrate, with stoichiometry of gas consumption, polymer yield, and total productivity unaffected by the presence of CO (Volova and Voinov 2003, 2004). However, CO concentration did affect the rate of consumption of gas components by the culture. This value was determined from the following equation:

$$q = A \exp\left(0.024 \cdot X_{co}\right)$$

where q is specific rate of gas substrate consumption, kg/(kg h); X_{CO} is volumetric concentration of carbon monoxide in the gas mixture, %; A is coefficient equal to 0.036 for hydrogen, 0.11 for carbon dioxide, and 0.17 for oxygen. Rates of gas consumption increased with increasing CO concentrations in the growth medium. The rate of gaseous substrate consumption by *R. eutropha* was only affected by CO during growth, and not during PHA production (Volova and Voinov 2003).

Given the findings discussed in the aforementioned works, CO should be rather regarded as an extraneous inhibitor. As has been previously shown (Volova et al. 1985), its oxidation rate by this culture is low (30–50 μM CO/min·g protein). Thus, as oxygen, carbon monoxide, and particularly hydrogen were consumed, CO concentration of the culture increased, while the partial pressure of the other gases decreased. This decreased the solubility of gas components in the liquid. As a result, in the media containing large amounts of CO, in the first stage, as the density of bacterial suspension increased, the hydrogen concentration in the culture decreased and even reached critical values. Thus, CO should be flushed from the culture and replaced with fresh syngas substrate. In order to minimize CO concentrations and optimize growth substrate gas concentrations in the culture, different modes of gas supply to the culture have been tested in experiments: (1) a single injection of gas into the culture; (2) recirculation of the gas mixture and ejection of part of the gas; and (3) complete utilization of the gas substrate and degassing of the culture medium to remove unutilized gases (Volova and Voinov 2003). The third option involving complete utilization of gas and degassing to remove extraneous CO yielded the best results. Under these gas exchange conditions, the 70-h fermentation cycle yielded 22 g biomass/L, with PHA concentration 75%, with hydrogen and oxygen utilization reaching 90%. Thus, with the same final PHA concentration and biomass production attained for the same culture period, variations in gas supply conditions significantly increase the completeness of gas substrate utilization, which minimizes costs.

These findings made it possible, for the first time in practical biotechnology, to synthesize high yields of polyhydroxyalkanoates from products derived from brown coals via modified gasification procedure. PHA biotechnology on syngas has been protected by RF Patent No. 2207375.

17.4 Synthesis of PHA Copolymers and Their Characterization

PHB was the first characterized of all the polyhydroxyalkanoates, but, in general, it does not have good thermoplastic properties. PHB is a very crystalline material that although strong, is also brittle. Mechanical and thermal properties of PHA can be improved as 3HB monomers are co-polymerized with longer chain length monomers, like 3HV or 3HHx. These copolymers typically exhibit better tensile strength, improved elongation to break, larger thermal processing window, and many other characteristics that are more favorable for a wide variety of applications (Noda et al. 2005).

Wild-type *R. eutropha* is capable of producing P(HB-*co*-HV) in the presence of 3HV precursor molecules, like propionate (Du et al. 2001) or valerate (Shang et al. 2004). Both propionate and valerate are toxic to *R. eutropha* at high concentrations (>0.4%, w/v), resulting in inhibition of growth and PHA production. For optimal P(HB-*co*-HV) biosynthesis by *R. eutropha*, feeding strategies must be developed for P(HB-*co*-HV) production (Du et al. 2001; Shang et al. 2004; Yu et al. 2005). Growth and PHA production on a mixture of organic acids derived from agricultural waste streams have been examined, resulting in production of a value added product from waste streams using wild type *R. eutropha* (Hassan et al. 2002; Yee et al. 2003). Following these works, a mixture analysis was developed for adjusting the ratios of the organic acids acetate, propionate, and butyrate to produce polymers that are "tailor made" with different 3HB and 3HV monomer contents (Yang et al. 2010), thus resulting in polymers with different properties (Fig. 17.3). P(HB-*co*-HV) containing a high molar % 3HV monomer has also been achieved with the aid of strain engineering. An *R. eutropha* strain expressing *phaC* and *zwf* (encoding a glucose 6P dehydrogenase) was shown to produce P(HB-*co*-58 mol%HV) when fed only 0.5% valerate as a 3HV precursor (Choi et al. 2003).

P(HB-co-HV) has been produced in fed-batch cultures of *R. eutropha* with productivities of up to 1.6 g PHA/L/h by employing various feeding strategies. Thermal and mechanical properties of resulting copolymers were tested, and it was determined that glass transition temperatures of P(HB-*co*-HV) copolymers were lower than PHB homopolymer. The effects of polymer aging on its mechanical properties were also tested and elongation to break values were shown to decrease as the polymer aged from 0 to 28 days (Madden and Anderson 1998). This is presumably because of crystallization occurring in the polymer over time. Aging effects resulting in brittleness of copolymer were witnessed in early works using commercial samples of P(HB-*co*-30%HV) (Scandola et al. 1989). Similar results have been seen in PHB (de Koning et al. 1992; Scandola et al. 1989), and also presumably from crystal formation in the polymer.

Even though copolymerization of 3HB and 3HV monomers offers improved properties of the resulting polymer, the incorporation of MCL monomers along with 3HB in PHA results in more dramatic improvement of the polymer properties (Budde et al. 2011; Noda et al. 2005). In the class of copolyesters of SCL and MCL PHA, the best studied member is poly(hydroxybutyrate-*co*-hydroxyhexanoate) (P(HB-co-HHx)). This P(HB-co-HHx) has a lower melting temperature, broader

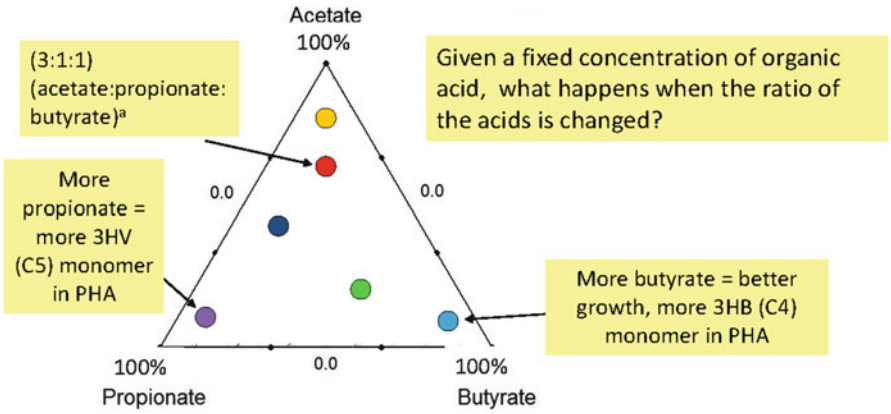


Fig. 17.3 Schematic of organic acid feedstock mixture analysis for P(HB-*co*-HV) production by *R. eutropha* cultures. Altering components in the ratio of an acetate, propionate, and butyrate mixture results in altered growth kinetics or PHA composition. Software used for initial analyses (Yang et al. 2010) can be found at the website http://www.statsoft.nl and http://www.minitab.com. [a]The original acetate, propionate, and butyrate ratio was published in the reference (Yee et al. 2003)

thermal processing window, lower Young's modulus, and a longer elongation to break than either PHB or P(HB-*co*-HV) (Noda et al. 2005), indicating that it is tougher and more flexible than either PHB or P(HB-*co*-HV). As with P(HB-*co*-HV), the thermal and mechanical properties of P(HB-*co*-HHx) depend on the amount of other (non-3HB) monomers present in the polymer (Brigham et al. 2011).

Several groups have produced P(HB-*co*-HHx) from engineered *R. eutropha* strains. Many of these engineered *R. eutropha* strains use the broad substrate specificity PHA synthase gene (*phaC*) from *Aeromonas caviae* (Kahar et al. 2004; Loo et al. 2005; Mifune et al. 2008). These papers discussed no other genetic modifications to *R. eutropha* beyond addition of the *A. caviae phaC* gene and/or inactivation of *R. eutropha*'s native *phaC* gene. With this modification, the maximum amount of 3HHx monomer present in the resulting copolymer was determined to be 5 mol% (Kahar et al. 2004; Loo et al. 2005). Other synthase genes that have been used for production of P(HB-co-HHx) by *R. eutropha* include *phaC* from *Pseudomonas fluorescens*, which also resulted in the biosynthesis of P(HB-*co*-5 mol%HHx) (Noda et al. 2005). P(HB-co-HHx) was also produced by engineered *R. eutropha* grown on fructose. In this engineered strain, PHA synthase and an *(R)*-specific enoyl-coA hydratase gene (*phaJ*) from *A. caviae* were used, along with a crotonyl-CoA reductase gene from *Streptomyces cinnamonensis* (Fig. 17.4a). In this work, the 3HHx-CoA precursor was synthesized *de novo* by the engineered *R. eutropha* strain and incorporated into PHA. The resulting HHx content in polymer was demonstrated to be ~2 mol% (Fukui et al. 2002). An earlier work had demonstrated de novo biosynthesis of MCL PHA precursors in *R. eutropha* strains expressing a PHA synthase gene and a 3-hydroxyacyl-ACP transferase gene (*phaG*) from *Pseudomonas* sp. (Matsumoto et al. 2001). Recently, *R. eutropha* strains were engineered to

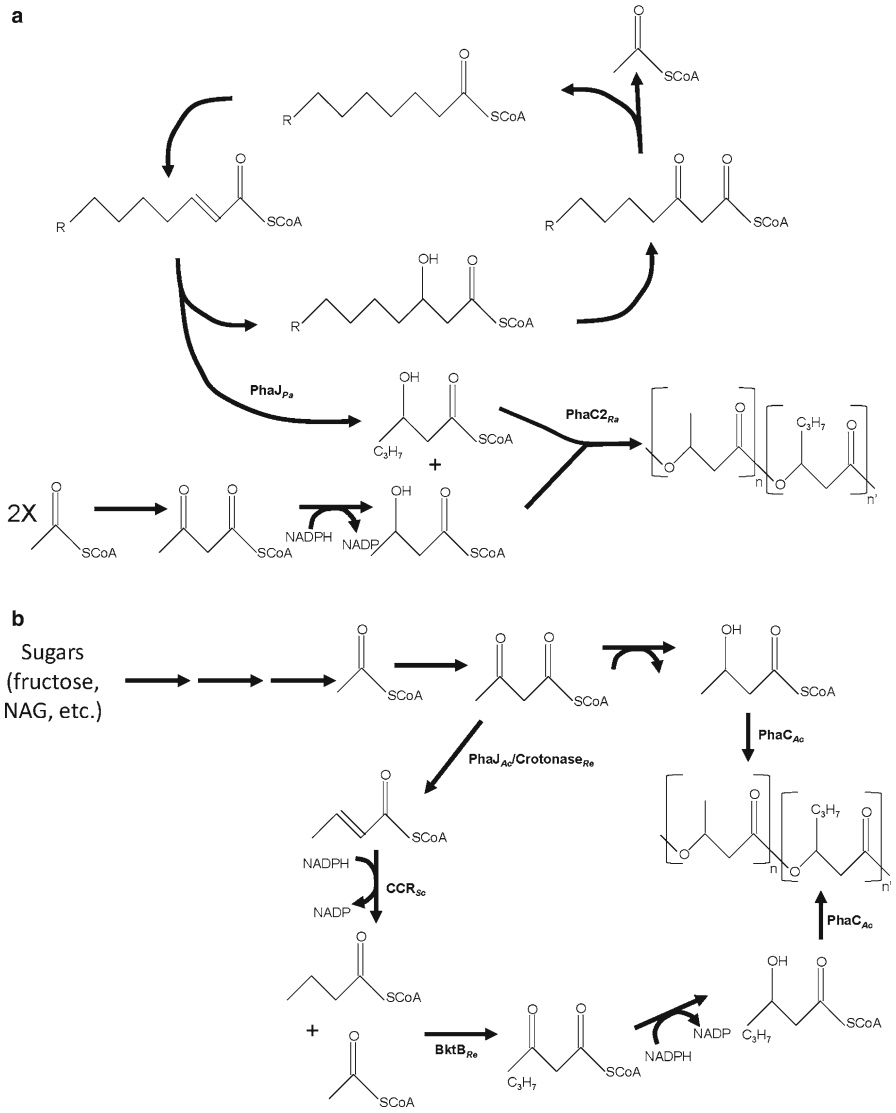


Fig. 17.4 (**a**) Schematic of P(HB-*co*-HHx) biosynthesis of *R. eutropha* strains capable of producing *de novo* HHx monomers. (**b**) Schematic of P(HB-*co*-HHx) biosynthesis of *R. eutropha* strains that produce high levels of HHx monomers from fatty acid beta-oxidation intermediates. *R. eutropha* strain in (**a**) was engineered by Budde et al. (2011), and *R. eutropha* strain in (**b**) was engineered by Fukui et al. (2002)

produce copolymer with >15 mol% HHx incorporated, when these strains were grown on plant oils (Budde et al. 2011). The schematic of the engineered P(HB-*co*-HHx) pathway is shown in Fig. 17.4b. In fed batch fermentations with palm oil as the sole carbon source using these strains, >140 g/L dried biomass was produced

containing >70% PHA. This polymer contained 17mol% 3HHx monomer. Productivity of fed batch fermentations was greater than 1 g PHA/L/h (Riedel et al. 2012), suggesting that this P(HB-*co*-HHx) production using engineered *R. eutropha* as the biocatalyst is a scalable process.

The copolymer poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) (P(3HB-*co*-4HB)) has also been produced by *R. eutropha* and has potential medical applications. As with other copolymers, thermal properties of P(3HB-co-4HB) improve with increasing amounts of 4HB monomer (Ishida et al. 2001; Kasuya et al. 1996). There are many instances in the literature where P(3HB-*co*-4HB) is produced (Cavalheiro et al. 2012; Ishida et al. 2001; Kim et al. 2005; Volova et al. 2011), and in most cases, a 4HB precursor molecule (γ-butyrolactone, 4-hydroxybutyrate, etc.) is typically added to the culture. Depending on the amount of precursor feeding, the 4HB fraction of PHA produced by *R. eutropha* can vary from 0 to 100 mol% (Cavalheiro et al. 2012; Ishida et al. 2001; Kim et al. 2005; Volova et al. 2011).

Other types of PHA polymers have been produced by *R. eutropha* in the laboratory. The 3-hydroxy-4-methyl valerate (3H4MV) monomer was incorporated into PHA from *R. eutropha* expressing a modified *phaC* gene from *Pseudomonas* sp. 61–3 (Tanadchangsaeng et al. 2009). Further studies indicated that leucine can be fed to cells as a precursor for 3H4MV, and the mol% of 3H4MV could be enhanced in PHA biosynthesis when leucine analog resistant mutant strains were used (Saika et al. 2011). The high 3H4MV polymers exhibited lower melting temperatures than either PHB or P(HB-*co*-HV), but along with this, lower molecular weight.

17.5 Pilot Scale Production of PHA

The palm oil industry is one of the biggest industries in Malaysia. With millions of tons of natural products (palm oil and palm oil products) and even larger waste streams (palm oil mill effluent, or POME) available (Hassan et al. 2002), there is strong incentive to produce scalable processes for PHA production using these carbon-rich streams as feedstocks. The Malaysia and MIT Biotechnology Partnership Programme (MMBPP) has undertaken this task with great success. The culmination of the group's studies of *R. eutropha* production strain construction (Budde et al. 2011), fermentation process control (Riedel et al. 2012a), and polymer recovery (Riedel et al. 2012b) was the design and construction of a pilot scale biopolymer production facility in 2011 by SIRIM Berhad, outside of Kuala Lumpur, Malaysia. The pilot plant has a 2,000 L capacity and the ability to produce various PHA polymers from palm oil products and treated POME. With bioplastics production using robust strains and plentiful carbon feedstocks, it is believed that the cost of production can be lowered significantly, down to almost USD$2 per kg (Ismail 2011).

Testing of these PHA production technologies in the Pilot Production Facility is a way to determine more accurately physicochemical parameters of fermentation, calculate the material and energy balance, work out technological regulations, and prepare sufficient amounts of the product for investigations. In a similar effort

to study scaleup of PHA production processes on various feedstocks, the first Russian pilot facility for researching PHA production was constructed and launched in January 2005 at the Krasnoyarsk Research Center of the Siberian Branch of the Russian Academy of Sciences. That facility was the outcome of cooperation between researchers of the Institute of Biophysics SB RAS and specialists of the Biotechnology Department at Biokhimmash company of Moscow, within the framework of the International Science & Technology Center project (Volova et al. 2006b).

The facility is based on the previously developed technology and can produce sufficient quantities of polymers to satisfy the requirements of all Russian research institutes and to perfect the technologies of manufacturing special polymer items. Furthermore, it can serve as a tool to optimize polyhydroxyalkanoate production technology: using new feedstocks, reducing the cost of the polymer, broadening the range of the synthesized polyhydroxyalkanoates, and obtaining the data for the further scale-up of the process to the commercial level. The pilot lots of variously structured polymers were used to develop the techniques and procedures of processing the polymers into various items. PHAs in different phase states proved to be usable for the production of films, 3D membranes, sutures, microparticles, and composites. PHAs and PHA-based items were examined in biomedical studies, exhibiting good functional properties, and proved to be biomedical grade materials (Volova et al. 2006a).

17.6 Applications of PHA Polymers

As mentioned above, PHA polymers of all types offer a biocompatible and biodegradable alternative to petroleum-based plastics in many applications. Household applications include linings of milk cartons and diapers, while agricultural applications include biodegradable mulch films and coatings for fertilizer pellets (Philip et al. 2007). Metabolix, Inc., from Cambridge, MA, USA produces a wide range of biodegradable PHA-based products under the name Mirel™. Such products include packaging materials, gift cards, erosion control devices, and compost bags (www.metabolix.com). Medical applications of PHA polymers are currently being tested, as discussed in detail below. PHAs biocompatibility means that it can be used in contact with animal tissues, blood, etc. Tepha, Inc., of Cambridge, MA, USA manufactures several medical materials and devices from PHA, such as surgical meshes and films. These medical products have been demonstrated to have favorable mechanical properties for use in surgical procedures and have become widely adopted, particularly in Europe (www.tepha.com).

Recent advances in transplant surgery and introduction of new materials in medicine have made biological safety of medical items, such as primarily implants, a critical issue. The materials for fabricating temporary implants (sutures, artificial pericardia, stents, etc.) must be not only biocompatible, but also prone to biodegradation, forming products that are non-toxic to the organism. Finding materials possessing these balanced properties is a great challenge. Implants made from biodegradable materials

are designed to repair tissues or organs, promote the growth of the surrounding tissues, and to be controllably degraded once they have been replaced by native biological structures.

It is generally recognized that biomedical-grade materials, intended to contact living organisms, must possess a complex set of necessary biological and physical-mechanical properties. They must be biocompatible at the level of cultured cells and tissues of the macroorganism and non-toxic both before and after degradation. These materials must also fulfill a supportive function for cells, favor their proliferation and differentiation, allow for availability of growth substrates, and allow for release of metabolites. The material must possess proper mechanical strength and flexibility, have tissue growth favorable surface characteristics, be tolerant of conventional sterilization methods and the effects of aggressive biological media, while also being easy to process using conventional production methods. Comprehensive preclinical investigations of polyhydroxyalkanoates, including in vitro assay systems and short- and long-term exposure experiments on laboratory animals, have shown that PHAs of different compositions, in this case produced by and purified from *R. eutropha* strains, can be regarded as medical grade materials. Polymers and polymeric items, such as film-based tissue engineering scaffolds and suture fibers, were evaluated in conventional and sanitary chemical, toxicological, and biomedical tests. The results of the tests demonstrated that the two main PHA types studied, PHB and P(HB-*co*-HV) both of *R. eutropha* origin, are highly biocompatible at the level of the cell, tissue, and macroorganism, and can be used in contact with blood (Sevastianov et al. 2003; Shishatskaya et al. 2002b, 2004, 2008; Volova et al. 2003).

To choose proper sterilization techniques, PHAs were tested for their heat resistance and radiation tolerance. PHA films, membranes, and sutures subjected to sterilization by conventional techniques did not exhibit any changes in their structure or function, suggesting a conclusion that PHAs can be sterilized using various conventional techniques: dry heat, autoclaving, disinfectant solutions, and γ-irradiation.

A very important issue is mechanism and kinetics of degradation of resorbable materials in biological media. Investigations of PHA degradation were conducted in biological media in vitro (in stabilized blood and serum) and in vivo with implants (films, sutures, and microparticles) immediately contacting with animal tissues and with implants enclosed in diffusion chambers. This in vivo process did not allow fibrous capsules to be formed and provided conditions for investigating biocompatibility and biostability of the material in terms of true cellular reaction. The PHAs used in this study were found to be degraded in biological media at low rates (over months and years) via humoral and cellular pathways, involving macrophages and foreign body giant cells with a high activity of acid phosphatase. It was also found that PHA biodegradation rate depends on the chemical structure of the polymer, the form of the item and implantation site (Shishatskaya et al. 2002a, b, 2005).

As previously mentioned, an important property of PHAs is their manufacturing processability to form special items from different phase states, using conventional techniques, without adding processing aids or plasticizers (Amass et al. 1998). Having investigated the dissolution and melting behavior of PHAs and physicochemical

properties of polymer solutions, gels, and melts, a series of various 2D and 3D matrices were prepared for cell cultures: flexible films and porous membranes, monofilament fibers between 0.15 and 0.40 mm in diameter, ultrathin fibers between 1 and 5 μm in diameter, microparticles, sponges, solid and porous 3D matrices, as well as composites with hydroxyapatite and collagen.

PHA solutions of various densities were used to prepare transparent flexible films. The surface properties of PHB and P(HB-*co*-HV) film scaffolds were similar to each other and to those of synthetic polyesters (polyethylene terephthalate, poly (methyl methacrylate), polyvinyl chloride, and polyethylene) (Shishatskaya 2007). The scaffold's surface properties are important for cell attachment and proliferation. To enhance cell adhesion to the surface, improve the gas-dynamic properties of scaffolds, and increase their permeability for substrates and cell metabolites, the scaffolds can be treated by physical factors or by chemical reagents. Biocompatibility of PHA scaffolds has been enhanced by immobilizing collagen film matrices on the scaffold surface and coating with chitosan and chitosan/polysaccharides (Hu et al. 2003).

One of the approaches to modification of PHA surfaces is to treat them with gas plasma. A more recent way to modify PHA properties is to use laser-cutting technique. The advantage of this technique is that the surface can be modified without destroying the material and releasing toxic byproducts. Laser processing has been shown to enhance the adhesive properties of the PHA matrix surface. PHA film scaffolds were laser processed and, since membranes are transparent in the near and far IR spectral regions, they can be processed using lasers that generate radiation of a far IR wavelength. Using a CO_2-laser with power ranging from 3.0 to 30.0 W, a series of films was prepared, with surface properties modified from pronounced roughness to perforations. The microstructure and surface properties of the film scaffolds were examined and it was observed that water contact angles were reduced to 50% and scaffold hydrophilic properties were improved without destroying the structure of the polymer. The adhesion of fibroblasts and osteoblasts to such scaffolds was 18% higher than to unprocessed ones (Shishatskaya 2007).

Cell culture scaffolds require mechanical strength and high bulk porosity. PHAs have the former, but one way to increase the porosity of a matrix is by fabrication using two-component mixtures. One method for fabrication of porous polymeric scaffolds is via a solution of two-component mixtures, containing a water-soluble component and a water-insoluble compound, and then to leach out one of the components from the scaffold in the solution. Leaching techniques usually involve the use of water-soluble sodium chloride or sucrose crystals. By varying the soluble compound concentrations, one can control the total porosity and the pore sizes in the scaffolds. Porous scaffolds can also be prepared using three-component mixtures containing polymer solvents and nonsolvents. These three-component systems can be used to prepare high-porosity polymer scaffolds with pore sizes <5–10 μm (Yang et al. 2001).

PHA membranes of different compositions have been constructed and characterized, including PHB, P(3HB-*co*-4HB), P(HB-*co*-HV), and P(HB-*co*-HHx), all of *R. eutropha* origin. Examination of microstructure of the surface of the membranes

revealed that membranes prepared from P(HB-*co*-HHx) had the roughest surface, and those prepared from P(HB-*co*-HV) had the smoothest surface. Copolymer membranes had smaller water contact angles and higher hydrophilicity than membranes prepared from highly-crystalline PHB. Testing biocompatibility of these scaffolds, mouse fibroblast NIH 3T3 cells were cultivated on PHA membranes, and results Romanovsky-type staining, DAPI staining, and the MTT assay showed that membranes prepared from PHAs of different chemical compositions did not exhibit cytotoxicity to cells cultured on them and proved to be highly biocompatible. Cell attachment and proliferation on PHA membranes were similar to those on polystyrene and better than those on membranes prepared from polylactic acid (PLA). These observations suggest that PHAs are biocompatible can be used as matrices on which to grow cell cultures. Among promising approaches to preparing ultrafine fibers, membranes, and micro- and nano-particles from PHA as models of cell scaffolds are nanotechnological methods such as microencapsulation and electrostatic spinning. Electrostatic spinning technique has been used to prepare ultrafine PHA fibers. These fibers, with diameters from 1 to 5 μm were used as scaffolds for growing mouse fibroblasts (NIH 3 T3). The results showed that a significant number of cells adhered to the fibers and proliferative activity was robust. Microencapsulation techniques, which have been developed in recent years (Sayin et al. 2006), are used to prepare micro- and nano-size PHA particles with large surface area for use as controlled drug delivery systems and cell culture scaffolds. Biocompatibility of the microparticles was demonstrated in cell cultures and in experiments with animals that received them via injection (Shishatskaya et al. 2009).

Development of cellular and tissue engineering offers new opportunities to reconstructive orthopedics. One of the approaches to improving mechanical properties of hydroxyapatite (reducing rigidity, increasing elasticity) is preparation of hydroxyapatite/synthetic polymer hybrids. A novel approach involves fabrication of hybrid materials based on hydroxyapatite and biodegradable polymers. As in vivo bioresorption rates of PHA are several times lower than those of other known biodegradable biomaterials (polylactide, polyglycolide), these polyesters can be used for prolonged regeneration of large bone defects. Hybrid matrices of PHB and hydroxyapatite were seeded with osteoblastic cells. Biocompatibility and functional properties of these matrices were confirmed both in vitro and in vivo, suggesting that PHB/HA hybrid 3D matrices have good osteogenic potential, facilitate bone formation, and can be used in further studies as bioactive constructs for bone defect repair (Shishatskaya 2006).

PHAs have also been studied as matrices for holding and delivering drugs or in controlled drug delivery systems. The most promising drug delivery systems seem to be polymer microparticles of diameters from 0.5 to 5.0 μm, which can function in vivo for up to 12 weeks and can be injected intramuscularly, intraperitoneally, and intravenously (Shishatskaya 2007; Shishatskaya et al. 2008). An experimental prolonged-action system in the form of microparticles demonstrated that a stable concentration of drug, delivered from a PHA matrix, in the blood and peritoneal fluid could be maintained for 10 days. It was also found that the experimental form of

rubomycin hydrochloride encapsulated into PHA microspheres and intraperitoneally injected to laboratory mice that had also been inoculated with a 100% lethal dose of Ehrlich ascites carcinoma cells significantly inhibited proliferation activity of carcinoma cells and enhanced the survival of tumor-bearing mice to 40% (Shishatskaya et al. 2008).

The use of PHA coatings to enhance biocompatibility of vascular stents was shown to be an effective modification. Self-expanding nitinol mesh vascular stents were coated with PHA and PHA loaded with an antiproliferative drug. Stents were placed in the femoral artery of dogs and examined for 120 days. The analysis of the state of the vessels and morphometric examination showed the effectiveness of coating vascular stents with PHA, especially PHA loaded with an antiproliferative drug. Reactive changes in the vessel wall were less pronounced and no complications occurred that are usually caused by implantation of uncoated metal stents (Protopopov et al. 2005, 2008).

Also, fully resorbable PHA (mesh, spiral, and tubular) stents were constructed and tested as implants for endobiliary surgery in laboratory animals (Markelova et al. 2008). The stents were implanted to the supraduodenal part of the bile duct of laboratory animals, fixed with PHA sutures, and monitored for 120 days. In the study, bilirubin levels of animals remained unimpaired, no unfavorable tissue responses to the implant were registered, the bile duct lumens at the implantation sites remained unchanged and retained their pre-surgery sizes, and no deformities or inflammations were observed at the site of implantation. Histological examination of the liver and the biliary tract did not reveal any unfavorable effects of stenting. These results suggest biocompatibility for the designed stents and provided a basis for local clinical trials.

PHA polymers recovered from cultures of *R. eutropha* have tremendous potential for development of medical technologies, including surgical reconstruction, orthopedic and trauma surgery, cardiovascular and abdominal surgery, and pharmacology. The matrices discussed in this section are biocompatible and show promise in many types of medical applications. PHA can be the next generation of medical material, with studies of polymer recovered *R. eutropha* at the forefront of the promising, burgeoning medical PHA field.

17.7 Manipulation of Carbon Flux in *R. eutropha* for Biosynthesis of Other Useful Compounds

As discussed in previous sections, *R. eutropha* is most widely studied for its ability to produce polyhydroxyalkanoates. However, given the ease of genetic manipulation and ability to utilize a wide range of carbon substrates, *R. eutropha* is an industrially relevant organism that can be designed to produce several important and value added natural products. Cyanophycin, a biodegradable alternative to polyacrylate (Aboulmagd et al. 2001; Schwamborn 1998), is typically synthesized by cyanobacteria albeit with low yields. Engineered *R. eutropha* strains have been designed to produce

intracellular cyanophycin, and can accumulate up to 32% of their cell dry weight with the compound (Voss and Steinbüchel 2006). Furthermore, high cell density cultures have been achieved using a modified cyanophycin gene expression system in recombinant *R. eutropha* (Lin et al. 2012). Self-assembling peptides can form nanofilament-like structures and have also been considered for tissue engineering. Reed, et al. have engineered *R. eutropha* strains to produce the RAD16-I::E peptide, which were purified from cells by fusion to cellulose binding modules (Reed et al. 2006).

With increased interest in microbial biofuels production, researchers are looking to *R. eutropha* as a recombinant fuels and chemicals producing organism. Utilizing the organism's branched chain amino acid biosynthetic pathway, a heterologously expressed decarboxylase gene (de la Plaza et al. 2004), and native alcohol dehydrogenase enzymes (Jendrossek et al. 1990), *de novo* production of isobutanol has been achieved, with the long term goal of producing isobutanol biofuel autotrophically (Brigham et al. 2012; Li et al. 2012; Lu et al. 2012). Longer chain fuel molecules are being produced by groups who have altered the flow of carbon in *R. eutropha* through to fatty acid biosynthesis to produce fatty acids and fatty acid methyl esters for use as biofuels (www.opxbio.com; arpa-e.energy.gov/ProgramsProjects/Electrofuels.aspx).

A microbial-based process for vanillin production has been sought in the flavoring and food industries. Vanillin has been biosynthesized by recombinant *Escherichia coli* (Yamada et al. 2008), *Rhodococcus* (Plaggenborg et al. 2006), and *R. eutropha* (Overhage et al. 2002). In the latter work, ferulic acid, a precursor to vanillin biotransformation, was converted by engineered *R. eutropha* strains from eugenol as the precursor. The engineered strain expressed eugenol catabolic genes from *Pseudomonas* sp. HR199, and achieved a productivity of ferulic acid conversion of 2.9 mmol/L/h (98 mol% yield). The 3-hydroxybutyrate monomer (3HB) has been shown to exhibit antimicrobial, antiviral, and insecticidal activities, suggesting that 3HB has wide industrial and medical applications (Tokiwa and Ugwu 2007). *R. eutropha* was subjected to UV mutagenesis to select for mutant strains that secrete 3HB. Strains were discovered that secrete 3HB, and the mutants were reported to contain a disruption of *phaB* (Ugwu et al. 2008). However, since at least two different *phaB* mutations were shown to knock down PHB production significantly in *R. eutropha* (Budde et al. 2010), it is unclear if the authors of this study mapped the actual mutations in their 3HB secreting strains.

Lastly, 2-methylcitric acid is a compound considered to be of importance for pharmaceutical applications, and is an intermediate of propionate metabolism in *R. eutropha* (Bramer and Steinbüchel 2001). Maximal production of 2-methylcitric acid was predicted to occur when the *prpB*, *prpD*, and *acnM* genes, whose products are all involved in the consumption of 2-methylcitrate, were deleted in *R. eutropha* (Park et al. 2011). In gluconate grown cultures of *R. eutropha* strains harboring a deletion of *acnM* and an additional *prpC* gene, a maximum production of 60 mM 2-methylcitric acid was achieved (Ewering et al. 2006), validating the genome-scale reconstruction predictions.

17.8 Outlook

Like many soil microorganisms, *R. eutropha* is able to utilize a versatile array of carbon sources for growth and biopolymer production. *R. eutropha* is genetically manipulable, increasing its value as an industrial microorganism for production of biopolymer and other engineered bioproducts. Understanding the flow of carbon in this organism is crucial to increasing its effectiveness in biotechnological endeavors, including high productivities of PHA synthesis. Studies have shown that *R. eutropha* is capable of using carbonic waste streams for production of biopolymer (Hassan et al. 2002; Yang et al. 2010; Yee et al. 2003). Scaling up fermentation efforts using waste streams and biosynthesis of polymer or other value added product is crucial for lowering costs associated with bio-based production. More effort must be made to understand the basic mechanisms of carbon flow in the organism. Study of the basic science of the metabolism of *R. eutropha* is a must to understanding how to unlock its full biotechnological potential. Microarray and proteomic studies have been performed using *R. eutropha* cultures (Brigham et al. 2010; Peplinski et al. 2010; Schwartz et al. 2009), but an in-depth look into the metabolomics and enzymology of the organism must not be far behind.

Acknowledgements We thank John W. Quimby for helpful suggestions and critical review of the manuscript. This manuscript was made possible, in part, by the work of the MIT/Malaysia Biotechnology Partnership Programme (MMBPP). The authors would like to thank the collaborators of this project for their support. This manuscript was also made possible, in part, by Project of SB RAS No. 96 and Project "Biotechnologies of novel biomaterials" (Agreement No. 11.G34.31.0013) in accordance with Resolution No. 220 of the Government of the Russian Federation of April 9, 2010, "On measures designed to attract leading scientists to the Russian institutions of higher learning." The authors declare that they have no conflict of interest.

References

Aboulmagd E, Voss I, Oppermann-Sanio FB, Steinbüchel A (2001) Heterologous expression of cyanophycin synthetase and cyanophycin synthesis in the industrial relevant bacteria *Corynebacterium glutamicum* and *Ralstonia eutropha* and in *Pseudomonas putida*. Biomacromolecules 2:1338–1342

Amass W, Amass A, Tighe BA (1998) Review of biodegradable polymers: uses, current developments in the synthesis and characterization of biodegradable polyesters, blends of biodegradable polymers and recent advances in biodegradation studies. Polym Int 47:89–144

Anderson AJ, Dawes EA (1990) Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev 54:450–472

Anderson AJ, Haywood GW, Dawes EA (1990) Biosynthesis and composition of bacterial poly(hydroxyalkanoates). Int J Biol Macromol 12:102–105

Bowien B, Kusian B (2002) Genetics and control of CO(2) assimilation in the chemoautotroph *Ralstonia eutropha*. Arch Microbiol 178:85–93

Bramer CO, Steinbüchel A (2001) The methylcitric acid pathway in *Ralstonia eutropha*: new genes identified involved in propionate metabolism. Microbiol 147:2203–2214

Brigham CJ, Budde CF, Holder JW, Zeng Q, Mahan AE, Rha C, Sinskey AJ (2010) Elucidation of beta-oxidation pathways in *Ralstonia eutropha* H16 by examination of global gene expression. J Bacteriol 192:5454–5464

Brigham CJ, Kurosawa K, Rha CK, Sinskey AJ (2011) Bacterial carbon storage to value added products. J Microbial Biochem Technol 83:S3–002

Brigham CJ, Gai CS Lu J, Speth DR, Worden RM, Sinskey AJ (2012) Engineering *Ralstonia eutropha* for production of isobutanol from CO_2, H_2, and O_2. In: Lee JW (ed) Advanced biofuels and bioproducts. Springer, New York, DOI: 10.1007/978-1-4614-3348-4_39

Bruland N, Voss I, Bramer C, Steinbüchel A (2009) Unravelling the C(3)/C(4) carbon metabolism in *Ralstonia eutropha* H16. J Appl Microbiol 109:79–90

Budde CF, Mahan AE, Lu J, Rha C, Sinskey AJ (2010) Roles of multiple acetoacetyl coenzyme A reductases in polyhydroxybutyrate biosynthesis in *Ralstonia eutropha* H16. J Bacteriol 192:5319–5328

Budde CF, Riedel SL, Willis LB, Rha C, Sinskey AJ (2011) Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from plant oil by engineered *Ralstonia eutropha* strains. Appl Environ Microbiol 77:2847–2854

Cavalheiro JM, Raposo RS, de Almeida MC, Teresa Cesario M, Sevrin C, da Grandfils C, Fonseca MM (2012) Effect of cultivation parameters on the production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) and poly(3-hydroxybutyrate-4-hydroxybutyrate-3-hydroxyvalerate) by *Cupriavidus necator* using waste glycerol. Bioresour Technol 111:391–397

Choi JC, Shin HD, Lee YH (2003) Modulation of 3-hydroxyvalerate molar fraction in poly(3-hydroxybutyrate-3-hydroxyvalerate) using *Ralstonia eutropha* transformant co-amplifying *phbC* and NADPH generation-related *zwf* genes. Enz Microbial Technol 32:178–185

Cromwick AM, Foglia T, Lenz RV (1996) The microbial production of poly(hydroxyalkanoates) from tallow. Appl Microbiol Biotechnol 46:464–469

de Koning GJM, Lemstra PJ, Hill DJT, Carswell TG, O'Donnell JH (1992) Ageing phenomena in bacterial poly[(R)-3-hydroxybutyrate]: 1. A study on the mobility in poly[(R)-3-hydroxybutyrate] powders by monitoring the radical decay with temperature after γ-radiolysis at 77 K. Polymer 33:3295–3297

de la Plaza M, Palencia PF, Peláez C, Requena T (2004) Biochemical and molecular characterization of alpha-ketoisovalerate decarboxylase, an enzyme involved in the formation of aldehydes from amino acids by *Lactococcus lactis*. FEMS Microbiol Lett 238:367–374

Du GC, Chen J, Yu J, Lun S (2001) Feeding strategy of propionic acid for production of poly (3-hydroxybutyrate-*co*-3-hydroxyvalerate) with *Ralstonia eutropha*. Biochem Eng J 8:103–110

Ewering C, Heuser F, Benolken JK, Bramer CO, Steinbüchel A (2006) Metabolic engineering of strains of *Ralstonia eutropha* and *Pseudomonas putida* for biotechnological production of 2-methylcitric acid. Metab Eng 8:587–602

Fakoussa RM, Hofrichter M (1999) Biotechnology and microbiology of coal degradation. Appl Microbiol Biotechnol 52:25–40

Fuchtenbusch B, Steinbüchel A (1999) Biosynthesis of polyhydroxyalkanoates from low-rank coal liquefaction products by *Pseudomonas oleovorans* and *Rhodococcus ruber*. Appl Microbiol Biotechnol 52:91–95

Fukui T, Abe H, Doi Y (2002) Engineering of *Ralstonia eutropha* for production of poly (3-hydroxybutyrate-co-3-hydroxyhexanoate) from fructose and solid-state properties of the copolymer. Biomacromolecules 3:618–624

Hassan MA, Nawata O, Shirai Y, Rahman NAA, Yee PL, Ariff AB, Ismail M, Karim A (2002) A proposal for zero emission from palm oil industry incorporating the production of polyhydroxyalkanoates from palm oil mill effluent. J Chem Eng Jpn 35:9–14

Hu SG, Jou CH, Yang MC (2003) Protein adsorption, fibroblast activity and antibacterial properties of poly(3-hydroxybutyric acid-co-3-hydroxyvaleric acid) grafted with chitosan and chitooligosaccharide after immobilized with hyaluronic acid. Biomater 24:2685–2693

Ishida K, Wang Y, Inoue Y (2001) Comonomer unit composition and thermal properties of poly (3-hydroxybutyrate-co-4-hydroxybutyrate)s biosynthesized by *Ralstonia eutropha*. Biomacromol 2:1285–1293

Ishizaki A, Tanaka K, Taga N (2001) Microbial production of poly-D-3-hydroxybutyrate from CO_2. Appl Microbiol Biotechnol 57:6–12

Ismail NR (2011) Bioplastic beauty. In: SIRIMLink. SIRIM Berhad, Shah Alam, pp 18–21
Jendrossek D (2009) Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). J Bacteriol 191:3195–3202
Jendrossek D, Kruger N, Steinbüchel A (1990) Characterization of alcohol dehydrogenase genes of derepressible wild-type *Alcaligenes eutrophus* H16 and constitutive mutants. J Bacteriol 172:4844–4851
Kahar P, Tsuge T, Taguchi K, Doi Y (2004) High yield production of polyhydroxyalkanoates from soybean oil by *Ralstonia eutropha* and its recombinant strain. Polym Degrad Stabil 83:79–86
Kasuya K, Inoue Y, Doi Y (1996) Adsorption kinetics of bacterial PHB depolymerase on the surface of polyhydroxyalkanoate films. Int J Biol Macromol 19:35–40
Kim JS, Lee BH, Kim BS (2005) Production of poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) by *Ralstonia eutropha*. Biochem Eng J 23:169–174
Kuznetsov BN, Shchipko ML (1996) Air–steam gasification of brown coal in fluidized and circulated layers of catalytically active slag. Chem Int Sus Dev 4:447–453
Kuznetsov BN, Shchipko ML, Golovin YG (1995) Some features of lignite gasification in fluidized bed of disposable catalyst. In: Pajares JA (ed) 8th International conference on coal science. Elsevier, Oviedo, pp 719–722
Lee JN, Shin HD, Lee YH (2003) Metabolic engineering of pentose phosphate pathway in *Ralstonia eutropha* for enhanced biosynthesis of poly-beta-hydroxybutyrate. Biotechnol Prog 19:1444–1449
Li H, Opgenorth PH, Wernick DG, Rogers S, Wu TY, Higashide W, Malati P, Huo YX, Cho KM, Liao JC (2012) Integrated electromicrobial conversion of CO_2 to higher alcohols. Science 335:1596
Lin K, Elbahloul Y, Steinbüchel A (2012) Physiological conditions conducive to high cell density and high cyanophycin content in *Ralstonia eutropha* strain H16 possessing a KDPG aldolase gene-dependent addiction system. Appl Microbiol Biotechnol 93:1885–1894
Loo CY, Lee WH, Tsuge T, Doi Y, Sudesh K (2005) Biosynthesis and characterization of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from palm oil products in a *Wautersia eutropha* mutant. Biotechnol Lett 27:1405–1410
Lu J, Brigham CJ, Gai CS, Sinskey AJ (2012) Studies on the production of branched-chain alcohols in engineered *Ralstonia eutropha*. Appl Microbiol Biotechnol. Epub before print. DOI: 10.1007/s00253-012-4320-9
Madden LA, Anderson AJ (1998) Synthesis and characterization of poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) polymer mixtures produced in high-density fed-batch cultures of *Ralstonia eutropha* (*Alcaligenes eutrophus*). Macromolecules 31:5660–5667
Markelova NM, Shishatskaya EI, Vinnic YS, Cherdansev DV, Beletskiy II, Kyznecov MN, Zykova LD (2008) *In vivo* justification of using endobiliary stents made of bioresorbable polyhydroxyalkanoates. Macromol Symposia 269:82–91
Matsumoto K, Nakae S, Taguchi K, Matsusaki H, Seki M, Doi Y (2001) Biosynthesis of poly (3-hydroxybutyrate-co-3-hydroxyalkanoates) copolymer from sugars by recombinant *Ralstonia eutropha* harboring the $phaC1_{Ps}$ and the $phaG_{Ps}$ genes of *Pseudomonas* sp. 61–3. Biomacromolecules 2:934–939
Mifune J, Nakamura S, Fukui T (2008) Targeted engineering of *Cupriavidus necator* chromosome for biosynthesis of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from vegetable oil. Can J Chem 86:621–627
Morinaga Y, Yamanaka S, Ishizaki A, Hirose Y (1978) Growth characteristics and cell composition of *Alcaligenes eutrophus* in chemostat culture. Agric Biol Chem 42:439–444
Noda I, Green PR, Satkowski MM, Schechtman LA (2005) Preparation and properties of a novel class of polyhydroxyalkanoate copolymers. Biomacromolecules 6:580–586
Overhage J, Steinbüchel A, Priefert H (2002) Biotransformation of eugenol to ferulic acid by a recombinant strain of *Ralstonia eutropha* H16. Appl Environ Microbiol 68:4315–4321
Park JM, Kim TY, Lee SY (2011) Genome-scale reconstruction and in silico analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production. BMC Syst Biol 5:101

Peplinski K, Ehrenreich A, Doring C, Bomeke M, Reinecke F, Hutmacher C, Steinbüchel A (2010) Genome-wide transcriptome analyses of the 'Knallgas' bacterium *Ralstonia eutropha* H16 with regard to polyhydroxyalkanoate metabolism. Microbiol 156:2136–2152

Philip S, Keshavarz T, Roy I (2007) Polyhydroxyalkanoates: biodegradable polymers with a range of applications. J Chem Technol Biotechnol 82:233–247

Plaggenborg R, Overhage J, Loos A, Archer JA, Lessard P, Sinskey AJ, Steinbüchel A, Priefert H (2006) Potential of *Rhodococcus* strains for biotechnological vanillin production from ferulic acid and eugenol. Appl Microbiol Biotechnol 72:745–755

Protopopov AV, Kochkina TA, Konstantinov EP, Shishatskaia EI, Efremov SN, Volova TG, Gitelson II (2005) Investigation of application of PHA coating to enhance biocompatibility of vascular stents. Dokl Biol Sci 401:85–87

Protopopov AV, Konstantinov EI, Shishatskaya EI, Efremov SN, Volova TG, Gitelzon II (2008) Use of bioresorbable polyesters for increasing of endobiliary stent biocompatibility. Technol Zhivl Sys 5:25–34

Reed DC, Barnard GC, Anderson EB, Klein LT, Gerngross TU (2006) Production and purification of self-assembling peptides in *Ralstonia eutropha*. Protein Expr Purif 46:179–188

Reinecke F, Steinbüchel A (2009) *Ralstonia eutropha* strain H16 as a model organism for PHA metabolism and for biotechnological production of technically interesting polymers. J Mol Microbiol Biotechnol 16:91–108

Riedel SL, Bader J, Brigham CJ, Budde CF, Yusof ZA, Rha C, Sinskey AJ (2012a) Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) by *Ralstonia eutropha* in high cell density palm oil fermentations. Biotechnol Bioeng 109:74–83

Riedel SL, Brigham CJ, Budde CF, Bader J, Rha C, Stahl U, Sinskey AJ (2012b) Recovery of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from *Ralstonia eutropha* cultures with non-halogenated solvents. Biotechnol Bioeng. Manuscript accepted for publication

Saika A, Watanabe Y, Sudesh K, Abe H, Tsuge T (2011) Enhanced incorporation of 3-hydroxy-4-methylvalerate unit into biosynthetic polyhydroxyalkanoate using leucine as a precursor. AMB Express 1:6

Savalieva ND (1979) Behavior of hydrogen bacteria towards carbon monoxide. Microbiol (Rus) 48:360–362

Sayin B, Calia S, Marangoz AS (2006) Clinical implantation of vancomycin microspheres in blend with human/rabbit bone grafts to infected bone defects. Microencapsul 5:553–556

Scandola M, Ceccorulli G, Pizzoli M (1989) The physical aging of poly(D-[beta]-hydroxybutyrate). Macromol Chem Rapid Commun 10:47–50

Schlegel H, Lafferty R (1971) Novel energy and carbon sources. Adv Biochem Eng 1:143–168

Schwamborn M (1998) Chemical synthesis of polyaspartates: a biodegradable alternative to currently used polycarboxylate homo- and copolymers. Polym Degrad Stabil 59:39–45

Schwartz E, Voigt B, Zuelhke D, Pohlmann A, Lenz O, Albrecht D, Schwarze A, Kohlmann Y, Krause C, Hecker M, Friedrich B (2009) A proteomic view of the facultatively chemolithoautotrophic lifestyle of *Ralstonia eutropha* H16. Proteomics 9:5132–5142

Sevastianov VI, Perova NV, Shishatskaya EI, Kalacheva GS, Volova TG (2003) Production of purified polyhydroxyalkanoates (PHAs) for applications in contact with blood. J Biomater Sci Polym Ed 14:1029–1042

Shang L, Yim SC, Park HG, Chang HN (2004) Sequential feeding of glucose and valerate in a fed-batch culture of *Ralstonia eutropha* for production of poly(hydroxybutyrate-co-hydroxyvalerate) with high 3-hydroxyvalerate fraction. Biotechnol Prog 20:140–144

Shchipko ML, Kuznetsov BN, Volova TG, Rudkovskii AV (2003) Development of the process of brown coal gasification for the production of syngas for polyhydroxyalkanoate biosynthesis. Chem Solid Fuel 5:58–67

Shishatskaya EI (2006) Biocompatible and functional properties of hybrid composite polyhydroxyalkanoate/hydroxyapatite. Bull Trans Artif Org 3:34–38

Shishatskaya EI (2007) Cellular matrices from bioresorbable polyhydroxyalkanoates. Cell Trans Tissue Eng 2:68–76

Shishatskaya EI, Volova TG (2004) A comparative investigation of biodegradable polyhydroxyalkanoate films as matrices for in vitro cell cultures. J Mater Sci Mater Med 15:915–923
Shishatskaya EI, Volova TG, Efremov SN, Puzyr' AP, Mogil'naya OA (2002a) Tissue response to biodegradable suture threads made of polyhydroxyalkanoates. Biomed Eng 36:210–217
Shishatskaya EI, Volova TG, Gitelson II (2002b) *In vivo* toxicological evaluation of polyhydroxyalkanoates. Dokl Biol Sci 383:109–111
Shishatskaya EI, Volova TG, Puzyr AP, Mogilnaya OA, Efremov SN (2004) Tissue response to the implantation of biodegradable polyhydroxyalkanoate sutures. J Mater Sci Mater Med 15:719–728
Shishatskaya EI, Volova TG, Zhemchugova AV (2005) A study of biodegradable polyhydroxyalkanoates as anti-tumor drug carrier. Antibiotik Khim 2:3–14
Shishatskaya EI, Voinova ON, Goreva AV, Mogilnaya OA, Volova TG (2008) Biocompatibility of polyhydroxybutyrate microspheres: in vitro and in vivo evaluation. J Mater Sci Mater Med 19:2493–2502
Shishatskaya EI, Goreva AV, Voinova ON, Kalacheva GS, Volova TG (2009) Distribution and resorption of polymeric microparticles in visceral organs of laboratory animals after intravenous injection. J Siberian Fed Univ Biol 4:453–465
Stasishina GN, Volova TG (1996) The strain of the bacterium *Alcaligenes eutrophus* – a producer of protein biomass. *BI* 1
Tanadchangsaeng N, Kitagawa A, Yamamoto T, Abe H, Tsuge T (2009) Identification, biosynthesis, and characterization of polyhydroxyalkanoate copolymer consisting of 3-hydroxybutyrate and 3-hydroxy-4-methylvalerate. Biomacromol 10:2866–2874
Tanaka K, Ishizaki A, Kanamaru T, Kawano T (1995) Production of poly(D-3-hydroxybutyrate) from CO(2), H(2), and O(2) by high cell density autotrophic cultivation of *Alcaligenes eutrophus*. Biotechnol Bioeng 45:268–275
Tokiwa Y, Ugwu CU (2007) Biotechnological production of (R)-3-hydroxybutyric acid monomer. J Biotechnol 132:264–272
Ugwu CU, Tokiwa Y, Aoyagi H, Uchiyama H, Tanaka H (2008) UV mutagenesis of *Cupriavidus necator* for extracellular production of (R)-3-hydroxybutyric acid. J Appl Microbiol 105:236–242
Valentin HE, Steinbüchel A (1994) Application of enzymatically synthesized short-chain-length hydroxy fatty acid coenzyme A thioesters for assay of polyhydroxyalkanoic acid biosynthesis. Appl Microbiol Biotechnol 40:699–709
Volova TG, Voinov NA (2003) Kinetic parameters of a culture of the hydrogen-oxidizing *Ralstonia eutropha*, grown under the regimen of biosynthesis of polyhydroxybutyrate. Prikl Biokhim Mikrobiol 39:189–193
Volova TG, Voinov NA (2004) Study of *Ralstonia eutropha* culture producing polyhydroxyalkanoates on products of coal processing. Prikl Biokhim Mikrobiol 40:296–300
Volova TG, Fyodorova YV, Trubachev IN (1980) Protein production from hydrogen. Nauka, Novosibirsk
Volova TG, Terskov IA, Sidko FY (1985) Microbial synthesis from hydrogen. Nauka, Novosibirsk, p 148
Volova TG, Kalacheva GS, Trubachev IN, Filippova VK (1987) The effect of physicochemical factors on growth and biochemical composition of carboxydobacteria. Microbiol (Rus) 57:973–978
Volova TG, Guseinov OA, Kalacheva GS, Medvedeva SE, Puzyr AP (1988a) The effect of carbon monoxide on metabolism and structure of carboxydobacteria. Microbiol (Rus) 57:793–797
Volova TG, Kalacheva GS, Trubachev IN, Filippova VK (1988b) A study of physiological and biochemical parameters of carboxydobacteria limited by deficiency of mineral elements. Microbiol (Rus) 57:61–64
Volova TG, Guseinov OA, Kalacheva GS, Medvedeva SE, Puzyr AP (1993) Effect of carbon monoxide on metabolism and ultrastructure of carboxydobacteria. World Appl Microbiol Biotechnol 9:160–163
Volova TG, Kalacheva GS, Altukhova OV (2001) Autotrophic synthesis of polyalkanoates by *Alcaligenes eutrophus* in the presence of carbon monoxide. Mikrobiologia 70:745–752

Volova TG, Shishatskaya EI, Sevastinov VI, Perova N (2003) Results of biomedical investigations of PHB and PHB/PHV fibers. Biochem Eng J 16:125–133

Volova TG, Sevastinov VI, Shishatskaya EI (2006a) Polyhydroxyalkanoates – biodegradable polymers for medicine. Platina, Krasnoyarsk, p 287

Volova TG, Voinov NA, Muratov VS, Bubnov NV, Gurulev KV, Kalacheva GS, Gorbunova NV, Plotnikov VF, Zhila NO, Shishatskaya EI, Belyaeva OG (2006b) Pilot production of biodegradable polymers. Biotekhnologiya 6:28–34

Volova TG, Zhila NO, Kalacheva GS, Sokolenko VA, Sinskey AJ (2011) Synthesis of 3-hydroxybutyrate-CO-4-hydroxybutyrate copolymers by hydrogen-oxidizing bacteria. Prikl Biokhim Mikrobiol 47:544–550

Voss I, Steinbüchel A (2006) Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16. Metab Eng 8:66–78

Wang ZX, Brämer CO, Steinbüchel A (2003) The glyoxylate bypass of *Ralstonia eutropha*. FEMS Microbiol Lett 228:63–71

Yamada M, Okada Y, Yoshida T, Nagasawa T (2008) Vanillin production using *Escherichia coli* cells over-expressing isoeugenol monooxygenase of *Pseudomonas putida*. Biotechnol Lett 30:665–670

Yang YY, Chung TS, Ng NP (2001) Morphology, drug istribution, and in vitro release profiles of biodegradable polymeric microspheres containing protein fabricated by double-emulsion solvent extraction/evaporation method. Biomater 22:231–241

Yang YH, Brigham CJ, Budde CF, Boccazzi P, Willis LB, Hassan MA, Yusof ZA, Rha C, Sinskey AJ (2010) Optimization of growth media components for polyhydroxyalkanoate (PHA) production from organic acids by *Ralstonia eutropha*. Appl Microbiol Biotechnol 87:2037–2045

Yee PL, Hassan MA, Shirai Y, Wakisaka M, Karim MIA (2003) Continuous production of organic acids from palm oil mill effluent with sludge recycle by the freezing-thawing method. J Chem Eng Jpn 36:707–710

York GM, Lupberger J, Tian J, Lawrence AG, Stubbe J, Sinskey AJ (2003) *Ralstonia eutropha* H16 encodes two and possibly three intracellular Poly[D-(−)-3-hydroxybutyrate] depolymerase genes. J Bacteriol 185:3788–3794

Yu J, Si Y (2004) Metabolic carbon fluxes and biosynthesis of polyhydroxyalkanoates in *Ralstonia eutropha* on short chain fatty acids. Biotechnol Prog 20:1015–1024

Yu ST, Lin CC, Too JR (2005) PHBV production by *Ralstonia eutropha* in a continuous stirred tank reactor. Process Biochem 40:2729–2734

Chapter 18
Metabolic Engineering of Inducer Formation for Cellulase and Hemicellulase Gene Expression in *Trichoderma reesei*

Bernhard Seiboth, Silvia Herold, and Christian P. Kubicek

Abstract The filamentous fungus *T. reesei* is today a paradigm for the commercial scale production of different plant cell wall degrading enzymes mainly cellulases and hemicellulases. Its enzymes have a long history of safe use in industry and well established applications are found within the pulp, paper, food, feed or textile processing industries. However, when these enzymes are to be used for the saccharification of cellulosic plant biomass to simple sugars which can be further converted to biofuels or other biorefinery products, and thus compete with chemicals produced from fossil sources, additional efforts are needed to reduce costs and maximize yield and efficiency of the produced enzyme mixtures. One approach to this end is the use of genetic engineering to manipulate the biochemical and regulatory pathways that operate during enzyme production and control enzyme yield. This review aims at a description of the state of art in this area.

Keywords Biorefineries • Cellulase production • Gene regulation • Lactose metabolism • *Trichoderma*

Abbreviations

CAZyme	Carbohydrate active enzymes
CCR	Carbon catabolite repression

B. Seiboth • S. Herold • C.P. Kubicek (✉)
Research Division Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorferstraße 1a, 166-5, A-1060 Vienna, Austria
e-mail: ckubicek@mail.tuwien.ac.at

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_18,

18.1 Introduction

Fungi are for a long time exploited for the industrial production of a diverse range of metabolites and enzymes. The decomposition of cellulosic plant biomass by fungi is only one example which demonstrates the importance of this group of organisms for our life on earth by the cycling of carbon and it shows how such essential processes can be exploited for industrial applications to produce biofuels and other biorefinery products from renewable carbon sources. Biorefining the sustainable processing of biomass into a spectrum of bio-based products and bioenergy may well play a major role in producing chemicals and materials which are today still produced from oil. Active research efforts in different areas will be necessary to make these processes cost-effective. One of the key issues for biotechnologists is the identification of novel enzymes, improvement of physicochemical properties and the efficient production of these enzymes in appropriate host organisms. A key to the latter is to understand under which conditions the host produces these enzymes to design strains with improved enzyme production properties. Therefore fundamental research is necessary to identify regulatory and metabolic networks which control the expression of the extracellular enzymes with the ultimate goal to improve the enzyme yield.

Fungi secrete a wide variety of extracellular enzymes to degrade different plant polymers including cellulose, hemicellulose, pectin and lignin which can be used by the fungus as carbon and energy source. Most of these enzymes active on the polysaccharides are known today under the name CAZymes (carbohydrate active enzymes) and the CAZyme database is dedicated to collect genomic, structural and biochemical information (Cantarel et al. 2009). Beside their role as an attractive resource for new enzymes, different fungi such as *Aspergillus* spp. or *Trichoderma reesei* are also used routinely as an industrial production host for these enzymes since they are well adapted to fermenter cultivations and secrete vast amounts of enzymes which are not reached by other microorganisms.

The filamentous fungus *T. reesei* is today a paradigm for the commercial scale production of different plant cell wall degrading enzymes mainly cellulases and hemicellulases. Maybe the most important and therefore best studied cellulase today is CEL7A/CBH1 of *T. reesei* for which also the first 3D model became available (Divne et al. 1994). The fungus and its enzymes have a long history of safe use in industry and well established applications are found within the pulp, paper, food, feed or textile processing industries. Today its enzymes are also employed for the saccharification of cellulosic plant biomass to simple sugars which can be further converted to biofuels or other biorefinery products (Bouws et al. 2008; Harman and Kubicek 1998; Kumar et al. 2008). Although *T. reesei* is already the main industrial source for such enzymes (Merino and Cherry 2007) and despite the fact that industrial strains produce more than 100 g/L of cellulases (Cherry and Fidantsef 2003), additional efforts are needed to reduce costs and maximize yield and efficiency of the produced enzyme mixtures (Carroll and Somerville 2009; Wilson 2009). Beside this prominent role in industrial applications, *T. reesei* serves today as a model organism for the regulation and biochemistry of (hemi) cellulose degradation

(Aro et al. 2005; Kubicek et al. 2009; Stricker et al. 2008a). Surprisingly, the genome sequence of *T. reesei* QM6a revealed that the fungus possess only a rather small set of cellulases and hemicellulases in comparison to other plant biomass degrading fungi. This underscores our still poor understanding of the process of cellulose degradation (Martinez et al. 2008).

18.2 Initiating the Degradation of Cellulose

One of the major obstacles in the degradation of plant cell walls is that their components are long polymers and often insoluble which prevents a direct uptake into the cell. Therefore *T. reesei* and other celluloytic fungi have developed an arsenal of extracellular enzymes – including cellulases and hemicellulases – which specifically degrade the different polysaccharides. But the production of extracellular enzymes is a highly energy consuming process and includes beside their transcription and translation – processes necessary for all proteins - also complex posttranslational modification steps such as *N* or *O*-linked glycosylation which are accomplished during the passage of these enzymes through the secretory pathway. These processing steps are completed by the export of the active enzymes from the fungal cell into the environment to cleave the plant biomass polymers. Therefore the expression of genes encoding such CAZymes is a tightly regulated process to minimize the energy loss for the fungus. To survive in such an ecosystem *T. reesei* has developed sophisticated mechanisms to guarantee an economic production of these enzymes. The degradation of different carbon sources that are present as complex mixtures follows an energy driven hierarchy which implies the presence of an efficient sensing and signaling system to detect the type (or composition) of the plant biomass. The regulation of enzyme expression is thus controlled by two major control circuits, i.e. specific induction and general carbon catabolite repression (CCR). Since most of the investigations on CCR have been performed with D-glucose, this type of repression is often termed glucose repression. However, this would be an oversimplification since also other carbon sources are able to provoke CCR, and in fact the rate of uptake is probably more critical for it than the sugar itself (Portnoy et al. 2011). As a result, most of the CAZyme genes necessary for the fragmentation of the plant cell wall polymers are repressed by the presence of increased concentrations of fast metabolizable carbon sources that can arise from the degradation of these polymers including D-glucose, D-galactose, D-xylose or others. This mechanism ensures that the extracellular rate of hydrolysis of the polymers does not occur in access to the uptake and intracellular metabolism of the break-down products, and ensures that other microorganisms present in this habitat do not benefit from the degradation efforts of *T. reesei* by competing for the different break-down products.

Most of the enzymes for plant cell wall degradation including cellulases and hemicellulases are usually only formed adaptively in the presence of their substrates. Therefore effective sensing and signaling networks must be present to communicate the presence of the substrate to the fungus. As a consequence, induction of the different

degradative enzymes is initiated. Since this process is essential for the survival of the fungus in its natural habitat, it is likely that it has developed several strategies that ensure the sensing of the lignocellulosic substrates in its environment.

One of these strategies is the formation of minute amounts of CAZymes constitutively. These enzymes are able to initiate degradation of cellulose, thereby releasing small amounts of oligosaccharides which are then taken up by the fungus to induce further cellulase biosynthesis. Evidence for this theory comes from the finding that addition of anti-cellulase antibodies block induction of cellulase formation by cellulose (El-Gogary et al. 1989) and that low levels of the major cellobiohydrolases *cel7A* and *cel7B* are present under non-inducing conditions (Carle-Urioste et al. 1997). A microarray analysis identified further candidate genes including *cel5b* that displayed a similar low basal expression level in the absence of cellulose (Foreman et al. 2003). This cellulase contains a putative binding site for the attachment of a glycosylphosphatidylinositol anchor which would allow a location of the enzyme close to the cell membrane and make it therefore an interesting candidate for inducer formation.

A second strategy to initiate cellulose degradation is connected to the production of conidia, the asexual form of spores in fungi. On the surface of these conidia an array of different plant polymer degrading enzymes is found including cellulases and hemicellulases (Kubicek 1987; Messner et al. 1991). Their removal by non-ionic detergents impairs germination of the conidia on cellulose (Kubicek et al. 1988) and it was also been shown that a *T. reesei* strain lacking the two main cellobiohydrolases is unable to initiate germination in the presence of cellulose (Seiboth et al. 1997). Metz et al. (2011) showed that numerous CAZyme genes are transcribed during the early phase of conidiation in the absence of any exogenous inducer and that their formation depends on the major cellulase and hemicellulase transcriptional regulator XYR1 (see below). Absence of XYR1 impairs germination of the spores on cellulose. The strong presence of different CAZymes might be a peculiarity of *T. reesei* since during conidiation of other fungi, i.e. *Neurospora crassa* or *Aspergillus fumigatus* no upregulation of cellulases or hemicellulase genes was detected. This strong presence of the different CAZymes on the conidia provides a selective advantage for *T. reesei* in the colonization of its lignocellulose rich habitat.

A third possibility to sense cellulose is suggested by the finding that cellulase transcripts are present in *T. reesei* mycelia 20–30 h after complete consumption of the carbon source (Ilmen et al. 1997). This phenomenon is not simply due to a relief from carbon catabolite repression and is not found after transfer to media lacking any carbon source. A conclusive interpretation for this observation is still missing, but it is possible that during this time carbohydrates (e.g. β-glucans) are released from the fungal cell wall which may induce cellulase formation. Alternatively, it is possible that this is due to the initiation of chlamydospore formation: *T. reesei* does this after exhaustion of a fast metabolizable carbon source, and it is thus possible that the detected cellulase transcripts are formed during chlamydospore development analogously to their appearance during conidiation (Metz et al. 2011).

Although different break-down products of the polymers might be produced by the action of other microorganisms or by abiotic factors in *T. reesei*'s natural environment, these three scenarios underscore how effectively *T. reesei* is adapted to sense cellulose and related polymers in its environment.

As a consequence of all these models for initiation of cellulose degradation different cellulose break-down products are produced which can serve as a specific inducer to boost the expression of further cellulases. The products of the synergistic action of the different cellulases are glucooligosaccharides, cellobiose and (in the case of high β-glucosidase activities) D-glucose. It would be logic to assume that one of these or a further metabolite thereof is the actual inducer of cellulases. D-Glucose can be ruled out for a number of reasons. Beside that it causes strong carbon catabolite repression, it is a carbohydrate that can be also derived from other polysaccharides (such as starch) as well and is therefore not a specific signal for the presence of cellulose. This is in accordance with findings that even a low feeding of D-glucose does not induce cellulose induction (Karaffa et al. 2006). On the other hand, growth in the presence of cellobiose has indeed been shown to induce cellulase expression in fungi including *T. reesei* but did not so in other fungi (Aro et al. 2005). A reason for the absence of induction in some cases could be the activity of extracellular β-glucosidase, which cleaves: cellobiose to D-glucose, which inhibits cellobiose transport (Kubicek et al. 1993) and also represses cellulase expression via CCR. Inhibition of transport may be the more relevant mechanism because it has been shown that altering the relative ratio of the uptake and hydrolysis rate (e.g. by the addition of β-glucosidase inhibitors, lowering of the pH below 3 which reduces β-glucosidase activity, or the use of β-glucosidase knock-out strains) increases cellulase induction by cellobiose (Fowler and Brown 1992; Fritscher et al. 1990; Sternberg and Mandels 1979). Since growth of the fungus on cellulose releases cellobiose in concentrations that favor the uptake rather than hydrolysis (the K_m of β-glucosidase for cellobiose is 20–40-fold higher than that of the β-glucoside permease; Kubicek et al. 1993), cellobiose could well be the in vivo inducer.

The most well studied inducer of cellulase formation in *T. reesei* is, however, sophorose (a glucosyl-ß-1,2-D-glucoside). It is the strongest known soluble inducer of cellulases for *T. reesei*, and its presence in the culture medium during growth on cellulose has been demonstrated (Mandels et al. 1962) as well as its formation by β-glucosidase via transglycosylation (Vaheri et al. 1979). Absence of the major extracellular β-glucosidase CEL3A resulted in a delay in induction of the other cellulase genes by cellulose, but not by sophorose, and a *cel3a*-multicopy strain formed higher amounts of cellulases than the parent strain under non-saturating concentrations of sophorose (Kubicek et al. 2009). However, also other β-glucosidases including the intracellular CEL1A are able to produce sophorose and cellobiose from glucose, and could therefore be involved in inducer formation. A potentially important property of sophorose is that it is much less efficiently hydrolysed by β-glucosidases, while still efficiently transported by the *T. reesei* cellobiose permease. Thus sophorose could be a gratuity inducer that mimicks the role of cellobiose in nature.

18.3 Engineering the Regulatory Systems of Cellulase and Hemicellulase Expression

18.3.1 Specific Transcriptional Activation and Repression

One key to improve the production of specific enzymes or a group of enzymes is to modulate the regulation of these enzymes on the level of transcription by overexpressing transcriptional activators or deactivating transcriptional repressors. A key to this is to understand the different regulatory networks which signal the presence, composition and abundance of possible substrate to the fungal nucleus to initiate or terminate the expression of the extracellular enzymes for polymer degradation, the sugar transporters and the intracellular metabolic enzymes for carbon assimilation or energy production.

The obligatory presence of an inducer for high cellulase gene expression implies tight regulation of their promoters. Most of the cellulase genes are also regulated in a coordinated way, although the relative ratio of their expression may differ in higher production mutants (Foreman et al. 2003). At least three transcriptional activators including XYR1, ACE2 and the HAP2/3/5 complex, as well as the two repressors CRE1 and ACE1 are involved in cellulase regulation.

XYR1 is clearly the major player in n the transcriptional regulation of cellulolytic and xylanolytic enzymes. Absence of this zinc binuclear cluster transcriptional activator virtually eliminates their expression by all known inducers (Stricker et al. 2006). XYR1 regulates also some α-L-arabinofuranosidases (Akel et al. 2009) and the catabolism of D-xylose, L-arabinose, lactose via induction of the reductase XYL1 and catabolism via XYL1 and the β-galactosidase BGA1 (Seiboth et al. 2007; Stricker et al. 2008b). Regulation is accomplished by binding to its target promoters by a $GGC(T/A)_4$ consensus (Furukawa et al. 2009). Although necessary for both cellulase and xylanase induction it is not clear how XYR1 is able to selectively induce cellulases or xylanases. As a transcriptional activator XYR1 must receive respective signals from the different inducers but the exact mode of activation is not known today. Noguchi et al. (2011) showed that in the absence of D-xylose the *Aspergillus niger* XYR1 homologue XlnR was present as a mixture of different phosphorylated forms while addition of D-xylose caused additional phosphorylation of the protein. It is however interesting that in the hypercellulolytic strain CL 847 an upregulation of *xyr1* transcription during growth on lactose was observed compared to other improved cellulase producers (Portnoy et al. 2011). Therefore the genetic background in which *xyr1* is expressed might play a role for the improvement of (hemi)cellulase production.

Beside XYR1 the zinc binuclear cluster type transcriptional activator ACE2 regulates hemi(cellulase) expression. Its deletion lowers the induction kinetics for different cellulase transcripts and reduces the cellulase activity to 30–70% when grown on cellulose. However, it does not affect cellulase induction by sophorose (Aro et al. 2001). ACE2 binds in vitro to the 5′-GGCTAATAA-3′ site present in the *cel7A* promoter. Therefore, both XYR1 and ACE2 are able to bind the same motif. Using the xylanase *xyn2* as a model system, Stricker et al. (2008b) suggested that ACE2 acts in

a dual role by (i) antagonizing early induction and (ii) enhancing of a continuous extension of *xyn2* expression. ACE2 *xyn2* promoter binding is dependent on its phosphorylation and dimerization. It is possible that ACE2 has evolved as a fine tuning component that binds to XYR1 and enables its response to the different inducing signals, as its gene knock out did not significantly affect cellulase gene transcription.

Induction of the *T. reesei* cellulase gene *cel6A* by sophorose is partially dependent on a CCAAT box located adjacently to its XYR1/ACE2 binding site. This motif is bound by the ubiquitously occurring trimeric HAP2/3/5 complex (Zeilinger et al. 2001) which contains a histone fold motif, suggesting an involvement in chromatin mediated regulation, eventually by association with acetyltransferases. No efforts to modify their expression were reported.

In contrast to the regulation of (hemi)cellulases in other fungi additional specific transcriptional repressors for cellulase or hemicellulase genes have been described for *T. reesei*, i.e. ACE1 and XPP1. ACE1 a Cys_2His_2-type zinc finger repressor and binds to the core sequence 5′-AGGCA-3′ (Saloheimo et al. 2000). Deletion of *ace1* increased the expression of all the main cellulase and xylanase genes in sophorose and cellulose induced cultures (Aro et al. 2003). A possible explanation for its action could be that the main regulator *xyr1* was found to be upregulated on D-xylose in a *ace1* deleted strain (Mach-Aigner et al. 2008). A strain deleted in both *ace1* and *ace2* expressed cellulases and xylanases similar to the *ace1* deleted train (Aro et al. 2003). The importance of ACE1 in cellulase production is also emphasized by the finding that the hyperproducing mutant *T. reesei* CL847 exhibits beside the above mentioned *xyr1* upregulation a strongly reduced *ace1* expression (Portnoy et al. 2011). A hint to a possible function of ACE1 comes from work with its orthologue in *Aspergillus nidulans* (Chilton et al. 2008). Its orthologue StzA was described here as a stress response factor and evidence of competition, or interaction, between StzA and AreA binding sites in promoters of *stzA*, as well as genes involved in the metabolism of amino acids was provided. Potential StzA binding sites were identified in *cpcA* (cross pathway control regulator of amino acid biosynthesis) and the presence of potential CpcA binding sites (5′-TGAC-3′) in the *stzA/ace1* promoters suggests an intriguing link between intracellular amino acid availability and cellulase gene expression. Support for a cross-talk between amino acid availability and cellulase formation comes from (Gremel et al. 2008) which reported an enhancement of cellulase gene expression by the addition of methionine in *T. reesei*.

Analysis of the *xyn2* promoter suggested the binding of another putative repressor protein XPP1 (Mach-Aigner et al. 2010) to an AGAA-box under glucose repressing conditions. A *xpp1* gene knock out is still required to see the potential of this putative regulator in xylanase gene transcription.

18.3.2 Wide Domain Regulators

Different wide domain regulators govern the expression of the different cellulases and hemicellulases. Their corresponding genes are usually not expressed during growth on glucose even in the presence of an inducer. This has been shown to be due

to both inducer exclusion (the inhibition of inducer uptake by D-glucose) and glucose repression, the latter generally called carbon catabolite repression (CCR). Consequently, one of the earliest attempts to improve cellulase production was the removal of CCR. Classical mutagenesis combined with selection for 2-desoxyglucose resistance under inducing conditions has led to increased cellulase producers such as *T. reesei* RUT C30, RL-P37 and CL847, thus supporting the importance of CCR in cellulase formation.

On the transcriptional level the key players for CCR is the Cys_2His_2 type transcription factor CRE1 (Ilmen et al. 1996). Numerous cellulase, hemicellulase and pectinase genes have been shown to be regulated by CRE1 in *T. reesei* (Aro et al. 2005). It is therefore not surprising that the analysis of RUT-C30 including its derivative CL847 identified a loss of a 2478-base pair fragment, which starts downstream of the region encoding the CRE1 zinc finger (Ilmen et al. 1996; Seidl et al. 2008).

In general, mutations of the *cre1* gene or of its binding motif 5′-SYGGRG lead to a (partial) derepression of enzyme gene expression on glucose as demonstrated for *cbh1* and *xyn1*. (Aro et al. 2005; Kubicek et al. 2009). Functional CRE1 binding sites have been shown to consist of two closely spaced 5′-SYGGRG motifs, and it has been suggested that direct CRE1 repression would occur only through such double binding sites. Phosphorylation of a serine residue in a conserved short stretch of *T. reesei* CRE1 has been demonstrated to regulate its binding to its target sequence (Cziferszky et al. 2002). Relieve from carbon catabolite repression alone is not sufficient for high cellulase production and cultivation on D-glucose or other carbon sources results only in low cellulase levels, indicating that cellulase hyperproduction is still inducer dependent. About sixfold-higher cellulase and 2.5-fold-higher xylanase levels were found during bioreactor cultivations (Nakari-Setälä et al. 2009). A further gene *cre2* is involved in cellulase gene repression and *cre2* deletion strains are consequently derepressed (Denton and Kelly 2011). CRE2 is orthologous to the *A. nidulans* CreB and involved in deubiquitinylation. These cysteine proteases target proteins for degradation by the proteasome. CreB forms a complex with the WD40-repeat protein CreC to protect CreB against proteolysis in the absence of carbon catabolite repression.

One of the highlights from the analysis of the genome sequence of the *T. reesei* genome (Martinez et al. 2008) was that – in contrast to all other ascomycetes whose genome had been sequenced – its cellulase, hemicellulase and other CAZyme genes were to be located in several discrete cluster. The reason for this and its potential advantage remained unclear. However, it was conspicuous that a number of genes for secondary metabolite production such as non-ribosomal polypeptide synthases and polyketide synthases were found in the same clusters. Such secondary metabolite cluster genes often occur near the telomere end of the chromosomes and have been demonstrated to be regulated at the level of chromatin remodelling by a the putative protein methyltransferase LaeA in *A. nidulans* (Reyes-Dominguez et al. 2010). Because of the co-clustering of CAZymes with secondary metabolite synthesis genes in the *T. reesei* genome, we hypothesized that cellulase expression may be regulated by a *T. reesei* LaeA orthologue. Production of cellulases and other CAZymes was indeed reduced almost to zero in the absence of LAE1, and dramatically

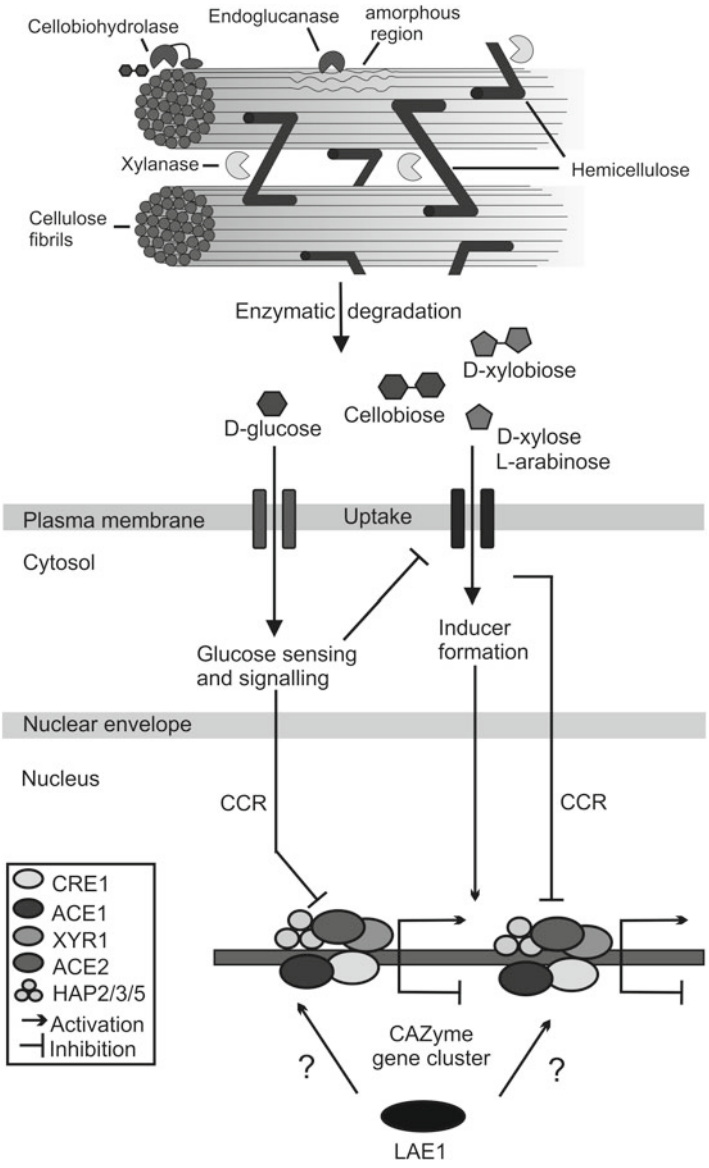


Fig. 18.1 Summary of steps involved in the regulation of cellulase and hemicellulase gene expression in *T. reesei*

increased by its overexpression showing that LAE1 is essential for cellulase and hemicellulase formation in *T. reesei* (Seiboth et al. 2012). The role of LAE1 in cellulase regulation is, however, less clear and deserves further efforts.

A summary of the events involved in the regulation of cellulase and hemicellulase induction by cellulose if given in Fig. 18.1.

18.4 Engineering the Signal Pathways for Cellulase and Hemicellulase Gene Expression

Although the regulatory network that governs (hemi)cellulase expression on the transcriptional level is already understood in some detail, the order of events by which different inducers and repressors signal their presence to modulate the different transcriptional regulators (e.g. by phosphorylation) is hardly known. The exogenous addition of dibutyryl-cyclic AMP or of inhibitors of cAMP-phosphodiesterase, to enhance intracellular cAMP levels, lead to secretion of increased endoglucanase activities in *T. reesei* (Sestak and Farkas 1993). cAMP, a classical messenger, is formed by the action of adenylate cyclase which in turn is activated by the G-alpha subunit of the G-protein complex associated with the cytoplasmic C-terminus of transmembrane receptors that receive the external signal to be transmitted. In *T. reesei*, the G-alpha proteins GNA1 and GNA3 seem to be only partially involved in cellulase gene expression: while loss of function strains and strains bearing constitutively activated *gna1* and *gna3* alleles showed a decreased and increased *cel7A* and *cel6A* transcription under some conditions, respectively, cellulase gene expression was still dependent on the presence of an inducer, thus implying that the signaling molecule that binds to the receptor and activates GNA1 and GNA3 cannot be the cellulase inducer (Schmoll et al. 2009; Seibel et al. 2009). The main target of cAMP in this signaling cascade is cAMP dependent protein kinase A (PKA), which – after binding of cAMP to the regulatory subunit that inhibits the activity of PKA – dissociates from it and phosphorylates its target proteins. Schuster et al. (2012) have shown that the adenylate cyclase of *T. reesei* is required for cellulase induction, because disruption of the gene encoding the catalytic subunit strongly impairs cellulase gene transcription. Interestingly, however, protein kinase A (PKA1) appears to influence cellulase gene transcription negatively because a knock-out in the *pka1* gene actually increases *cel7A* and *cel6A* expression. Thus the positive effect of cyclic AMP must act via another downstream player, which remains to be identified, and PKA1 may be involved via yet another mechanism.

18.4.1 Engineering the Lactose Catabolic Pathway

Induction of cellulases is – beside their natural substrate cellulose and its derivatives – also accomplished by a number of other inducing carbon sources, among which lactose – a heterodisaccharide (1,4-*O*-β-D-galactopyranosyl-D-glucose), a by-product of the cheese and whey processing industries – is also the most economical one. It induces the whole set of plant cell wall degrading enzymes (Ivanova, Seiboth, Kubicek unpublished results). Its inducing effect is surprising since lactose occurs in nature only in the milk of mammals and is therefore not present in the natural habitat of *T. reesei*. Thus it may rather mimick plant oligosaccharides with β-galactoside configuration that become available during the attack of pre-digested plant biomass.

The obvious advantage of lactose over cellulose as carbon sources for fermentations is that it is soluble, an attribute which facilitates the downstream processing of recombinant proteins, and it is therefore often used as inducing carbon source in combination with the strong cellulase promoters (Kubicek and Harman 1998; Penttilä et al. 2004). Although cellulase yields produced on lactose are usually reported to be lower compared to cellulose (Andreotti et al. 1980), it should be noted that the highest amount of cellulases produced by the publicly available *T. reesei* strain RUT C30 (Durand et al. 1988) was achieved on lactose as a carbon source.

18.4.1.1 The Initial Steps in Lactose Catabolism

Analysis of lactose metabolism in other fungi including the yeast *Kluyveromyces lactis* or *A. nidulans* identified two key players: a lactose permease responsible for its uptake and an intracellular β-galactosidase (GH family 2) for subsequent hydrolysis to D-glucose and D-galactose. In *K. lactis* the *LAC12* permease and *LAC4* β-galactosidase genes share an intergenic promoter region, and a similar situation is found also in other fungi (Fekete et al. 2012) but not in *T. reesei*. Orthologues of these genes are however absent from *T. reesei* and it was so far assumed that lactose is hydrolysed extracellularly by the GH family 35 β-galactosidase BGA1 (EC 3.2.1.23) and other cell-wall bound β-galactosidase activities (Seiboth et al. 2005). BGA1 has high galactosyltransferase activity and its galacto-β-D-galactanase activity points to its natural role, the hydrolysis of terminal non-reducing β-D-galactose residues in plant cell wall polymers. Galactitol, formed by the reductive D-galactose catabolic pathway, has been claimed to be the inducer of BGA1 expression during growth on lactose (Fekete et al. 2007). However, a genome-wide transcriptional analysis identified recently a novel lactose permease, whose knock out completely abolishes growth on lactose (Ivanova, Seiboth, Kubicek unpublished results). The further fate of the so taken up lactose is unclear, since a candidate for an intracellular β-galactosidase activity is still missing.

18.4.1.2 Intracellular D-Galactose Utilization

As a result of the extracellular lactose hydrolysis, D-galactose and D-glucose are taken up and are further assimilated. We will focus our attention on D-galactose catabolism as D-glucose does not induce cellulases even at low growth rates whereas D-galactose is able to induce cellulases (see below). A key for the understanding of the further catabolism is found in the stereospecificity of the D-galactopyranose released from lactose. D-galactose arising by BGA1 hydrolysis is the found in its β-anomeric form. The β-anomeric D-galactose is usually found in an equilibrium with its α-anomer but the chemical mutarotation necessary for this process is a slow, pH dependent process (Pettersson and Pettersson 2001). Therefore, many organisms enhance this conversion enzymatically by an aldose 1-epimerase (mutarotase). This enzymatic mutarotation is, however, necessary for a fast D-galactose utilization

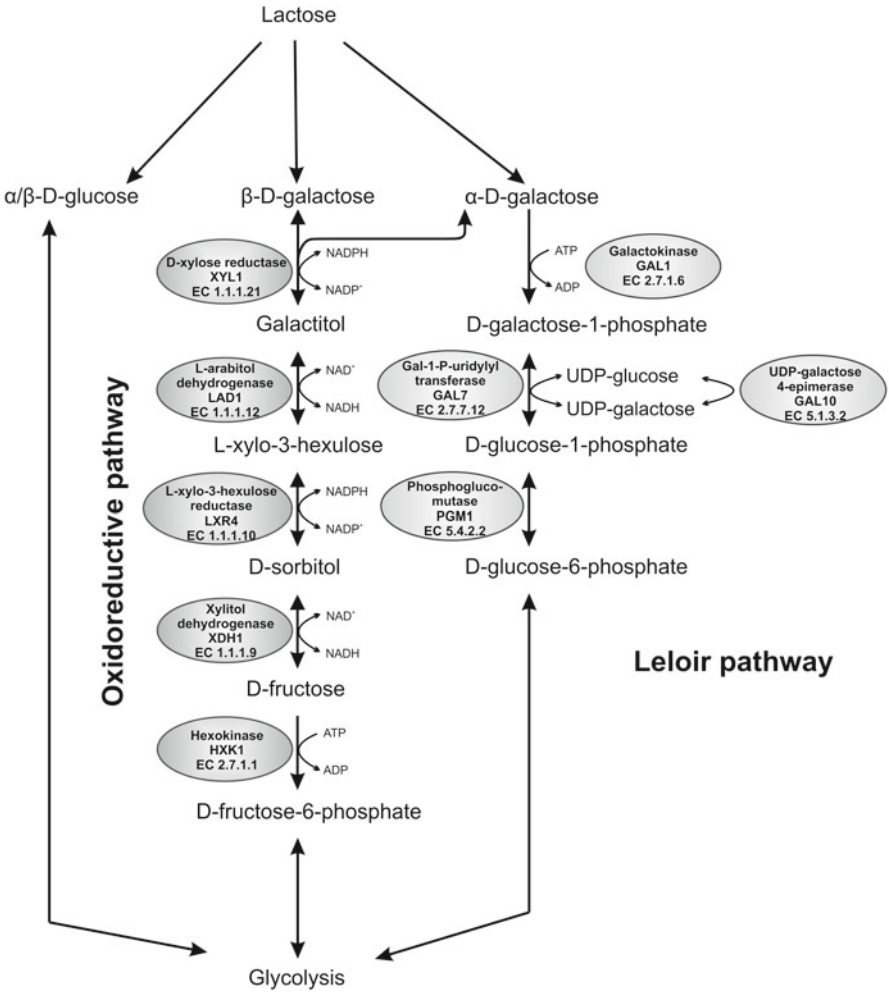


Fig. 18.2 Metabolic pathways involved in the degradation of lactose by *T. reesei*

via the canonical Leloir pathway since galactokinase which initiates D-galactose catabolism by C1 phosphorylation is specific for the α-anomeric form (Frey 1996; Holden et al. 2003). Therefore, it is not surprising that many fungi have developed a further possibility to degrade D-galactose independent of its anomeric form. In the case of *T. reesei* or *Aspergillus spp.* this is accomplished by the reductive pathway for D-galactose utilization.

In the Leloir pathway (Fig. 18.2) the sequence of galactokinase (GAL1; EC 2.7.1.6), D-galactose-1-phosphate uridylyl transferase (GAL7; EC 2.7.7.12), UDP-galactose 4-epimerase (UDP-glucose 4-epimerase GAL10; EC 5.1.3.2) and phosphoglucomutase (EC 2.7.5.1) convert α-D-galactose to the glycolytic pathway

intermediate D-glucose-6-phosphate (Holden et al. 2003). The aldose 1-epimerase (D-galactose mutarotase, EC 5.1.3.3) which provides the galactokinase with the α-D-galactose anomer is especially important for lactose/β-D-galactose catabolism and its absence resulted in a decreased growth rate on lactose in *E. coli* (Bouffard et al. 1994). In the yeasts *S. cerevisiae* and *K. lactis* UDP-galactose 4-epimerase and aldose 1-epimerase are combined to a single bifunctional protein Gal10p whereas the GAL10 of *T. reesei* and other fungal orthologues lack this C-terminal aldose 1-epimerase domain. The genome of *T. reesei* contains three putative aldose 1-epimerase encoding genes (*aep1-3*), however, none of these genes was expressed during growth on lactose and a mutarotase activity was absent during growth on lactose (Fekete et al. 2008). Expression of the C-terminal aldose 1-epimerase part of the *S. cerevisiae* Gal10 in *T. reesei* increased growth on lactose, which implies that the enzymatically catalyzed β-D-galactose mutarotation is either absent or inefficient and that the operation of the Leloir pathway for D-galactose during growth on lactose depends mainly on chemical mutarotation.

Other important difference to the model systems of *S. cerevisiae* and *K. lactis* are found in the physiological role of the Leloir pathway and its regulation. The *S. cerevisiae GAL* regulon is a eukaryotic model systems for transcriptional regulation in which induction of the *GAL* genes is controlled by the interplay of the transcriptional activator, Gal4p, a repressor, Gal80p, and an inducer, Gal3p (Bhat and Murthy 2001; Sellick et al. 2008). Gal3p is a paralogue of the galactokinase Gal1p, but misses its galactokinase activity. Regulation of transcriptional induction is similar in *K. lactis* (Rubio-Texeira 2005) with a number of differences including the combination of the inducer (Gal3p) and galactokinase (Gal1p) function in a single Gal1p protein. Similarly, *T. reesei* and other fungi miss a Gal3p protein. Expression of the *gal* genes is found in *T. reesei* under non-inducing and D-glucose repressing conditions, and *gal1* and *gal7* can be further induced (about two- to threefold) by D-galactose and L-arabinose (Seiboth et al. 2002a, b, 2004). Regulation of the other Leloir pathway genes is, however, completely independent of the presence of the galactokinase protein or its activity (Hartl et al. 2007) and an orthologue of the transcriptional activator Gal4p is absent. Another difference to the two model yeasts is that *T. reesei* strains deleted in the galactokinase encoding gene *gal1* were still capable of growing on D-galactose and this growth was not due to a residual galactokinase activity (Seiboth et al. 2004). A first hint towards this second pathway was obtained by the transient accumulation of galactitol during growth of a *T. reesei gal1* knockout strain on lactose which (Seiboth et al. 2004). This second D-galactose pathway is initiated by the D-xylose reductase XYL1 (EC 1.1.1.21) (Seiboth et al. 2007) which catalyzes the NADPH dependent reduction of D-galactose (beside D-xylose and L-arabinose) to galactitol. The second step in this pathway, i.e. the NAD^+ dependent oxidation of galactitol is catalysed by the *T. reesei* L-arabinitol 4-dehydrogenase LAD1. Δ*lad1* strains are unable to grow on galactitol and a simultaneous block of the Leloir pathway at the galactokinase step (Δ*lad1*Δ*gal1*) led to the inability of this strain to grow on D-galactose at all. LAD1 was actually the first identified enzyme of this second D-galactose pathway (Seiboth et al. 2004). Characterization of the recombinant produced LAD1 led to the unexpected finding that galactitol is oxidized

to L-xylo-3-hexulose (Pail et al. 2004). The enzyme acting on this uncommon sugar has recently been identified in *T. reesei* as the NADPH-dependent L-xylo-3-hexulose reductase LXR4 (Metz, B. and Seiboth, B., unpublished data). L-xylo-3-hexulose is converted to sorbitol and finally oxidized by another NAD-dependent polyol dehydrogenase, XDH1 (termed xylitol dehydrogenase) to D-fructose. The latter is phosphorylated by hexokinase to fructose-6-phosphate, an intermediate of glycolysis. The importance of this oxidoreductive pathway has been demonstrated in the catabolism of lactose in *T. reesei* (Seiboth et al. 2007). Despite that ß-D-galactose generated from lactose must be catabolised via the reductive galactose catabolizing pathway, disruption of the *gal1* gene still results in a decrease in growth rate on lactose (Seiboth et al. 2004) and overexpression of an aldose-1 epimerase improved the growth rate on lactose (Fekete et al. 2008). Mutants in the aldose reductase XYL1 display a strongly reduced growth on lactose. This pathway has also been termed a "stowaway shunt" because the oxidoreductive enzymes involved are the same that catalyze the catabolism of L-arabinose in *T. reesei*. The physiological relevance of this pathway for D-galactose utilization seems to differ between fungi. In *A. nidulans* the pathway can fully compensate for the loss of the Leloir pathway (Fekete et al. 2004; Roberts 1970) while in *T. reesei* inactivation of the Leloir pathway leads already to strains which are significantly impaired in their growth on D-galactose. One reason for this might be found in the fact that the enzymes of this pathway are adapted for their role in pentose catabolic pathways and their expression is in general lower on D-galactose than on the pentose sugars. For *A. niger* it was recently shown that at least some of the steps are catalyzed by more specific enzymes which are not involved in the degradation of L-arabinose (Koivistoinen et al. 2012; Mojzita et al. 2012). The major importance of this pathway is, however, found in its role in β-D-galactose and lactose catabolism by the formation of the inducer galactitol for β-galactosidase BGA1 (Fekete et al. 2007) and its importance for cellulase induction.

18.4.1.3 Metabolic Engineering of Cellulase Gene Expression on Lactose

Cellulase expression during growth on lactose is a growth rate dependent process with highest cellulase production during low growth rates (Karaffa et al. 2006; Pakula et al. 2005). The rate of the extracellular lactose hydrolysis is therefore critical for cellulase gene expression. A lowering of its rate by deletion of the *bga1* gene in *T. reesei* encoding the major extracellular ß-galactosidase BGA1 affected mainly growth on lactose but not cellulase expression. However, the constitutive overexpression of *bga1* led to higher growth rates on lactose and significantly inhibited cellulase expression during growth on lactose (Seiboth et al. 2005). The extracellular lactose hydrolysis leads to an interesting question: are its monomers also able to induce cellulases? During normal growth on these carbon sources neither D-glucose nor D-galactose nor mixtures of them are able to induce cellulase transcription, even in a carbon catabolite derepressed *cre1*-negative background (Seiboth et al. 2004). However, growth on D-glucose and D-galactose is much fast than on lactose

as carbon source. By using defined growth rates in carbon limited chemostat cultivations some cellulase induction by D-galactose and a mixture of D-galactose and D-glucose occurred at low growth rates. No induction was apparent under the same conditions with D-glucose as limiting carbon source. Nevertheless, cellulase expression under these conditions was significantly lower then found during growth on lactose at the same low growth rate, thus implying that lactose is still a superior inducer that a slow metabolism of D-galactose alone is not sufficient to account for the efficient cellulase induction. A equimolar mixture of D-galactose and galactitol, also led to cellulase induction at higher growth rates but still did not reach the levels obtained for lactose (Karaffa et al. 2006).

In the absence of a pure β-anomer the experiments described above were performed with D-galactose which is a mixture of the α- and β-anomer. Since enzymatically catalysed mutarotation of β-D-galactose is either absent or inefficient in *T. reesei* we investigated if the availability of β-D-galactose arising from lactose may thus be a relevant parameter in the induction of cellulase gene expression. Strains which expressed the C-terminal aldose 1-epimerase domain of the *S. cerevisiae* Gal10p showed a moderate increased growth rate on lactose but also led to a significant down-regulation of cellulase gene transcription. While this points towards a prominent role of the β-D-galactose anomer in cellulase induction by lactose, the actual mechanism must be more complex: despite of that D-galactose generated from lactose will be mainly catabolized via the second D-galactose pathway, deletion of the *gal1* gene results also in a strong decrease of cellulase formation on lactose (Seiboth et al. 2004). In *S. cerevisiae* the inactive galactokinase Gal3p acts as a coactivator for the induction of the *GAL* genes (Bhat and Murthy 2001; Sellick et al. 2008) and is needed for the fast induction of the *GAL* genes. In these Δ*gal1* strains, cellulase induction can be restored by retransformation with the structurally unrelated galactokinase gene from *E. coli* but cannot be restored by the introduction of an enzymatically inactive *T. reesei* GAL1 (Hartl et al. 2007). Therefore, galactokinase activity itself is important for cellulase induction by lactose. Inactivation of the subsequent step in the Leloir pathway, in which the *T. reesei* GAL7 converts D-galactose-1-phosphate into UDP-galactose, has no negative effect on cellulase induction except that the cellulase transcripts have a longer half-life (Seiboth et al. 2002a). So in contrast to the regulation of the *gal* genes (which showed that galactokinase is dispensable for *gal7* induction), transcription of the cellulases genes by lactose requires galactokinase activity. Similar to the situation in the Leloir pathway deletion of later steps did not affect cellulase expression, but affected growth on lactose. A *xyl1/gal1*-double knock-out did not display an additive action but exhibited an expression level similar to that of the *gal1* knock out, which would indicate that a blockage in either of the two pathways acts on the same target (Hartl et al. 2007). At the moment different models exist to explain the effects of the engineering of the galactose pathway. One explanation would be that an inducer is formed during lactose catabolism and that this oligosaccharide is composed of metabolites of both from the Leloir and the second pathway. A metabolomic analysis of intracellular oligosaccharides formed in *T. reesei* showed that several galacto-oligosaccharides were detected, whose intracellular concentrations changed in a consistent way with

cellulase formation (our unpublished results). Another possibility is that the interruption of one of the two pathways interferes with induction because of accumulation of a metabolite that would disturb the metabolic balance. Further investigations in this area are therefore clearly needed to clarify this point.

18.4.2 Engineering the Pentose Catabolic Pathways for D-Xylose and L-Arabinose

In contrast to cellulose, the hemicelluloses are composed of monomers that serve as inducing carbon sources for different hemicellulases. The fact that some of them (D-xylose, D-mannose and D-galactose) can exert carbon catabolite repression at high concentrations (Aro et al. 2005; Mach-Aigner et al. 2010) does not interfere with this assumption because they may naturally not accumulate in the medium at concentrations high enough to cause this effect. Also disaccharides (xylobiose, sophorose, glucose-β-1,2-D-xylose, lactose) have been reported to induce xylanases and other hemicellulases. The effect of xylobiose may be due to the formation of D-xylose at a slow rate by the action of β-xylosidase. Likewise, L-arabinose and L-arabitol have been shown to induce expression of the genes encoding enzymes involved in the degradation of arabinans (Aro et al. 2005).

18.4.2.1 The Interconnected Pathway for D-Xylose and L-Arabinose Utilization

The fungal oxidoreductive catabolism of these two pentoses is unique for fungi and produces ultimately D-xylulose 5-phosphate, which is an intermediate of the canonical pentose phosphate pathway (Fig. 18.3). In contrast, prokaryotes use an isomerase step to convert D-xylose to D-xylulose, and L-arabinose to L-ribulose. D-xylulose 5-phosphate is then either formed by the action of xylulokinase (in the case of D-xylose) or a sequence of L-ribulokinase and L-ribulose-5-phosphate 4-epimerase (in the case of L-arabinose) (Mishra and Singh 1993).

This fungal oxidoreductive pathway for pentose catabolism consists of a series of reactions in which an NADPH dependent reduction is followed by an NAD^+ dependent oxidation. Thus, although the overall redox balance of these two pathways is neutral, yet it results in an imbalance of the NAD+/NADH and NADP+/NADPH ratios. The pathways for D-xylose and L-arabinose catabolism are interconnected and share two enzymes, i.e. xylitol dehydrogenase and the xylulokinase. The enzyme(s) that catalyze(s) the first NADPH-dependent reduction of L-arabinose and D-xylose differs depending on the fungus: gene deletion analysis of *T. reesei xyl1* encoding the D-xylose reductase XYL1 shows that this enzyme is responsible for both D-xylose and L-arabinose catabolism because its absence results in significantly impaired growth and reductase activity on both pentoses (Akel et al. 2009; Seiboth et al. 2007). In *A. niger*, an additional reductase LarA was identified

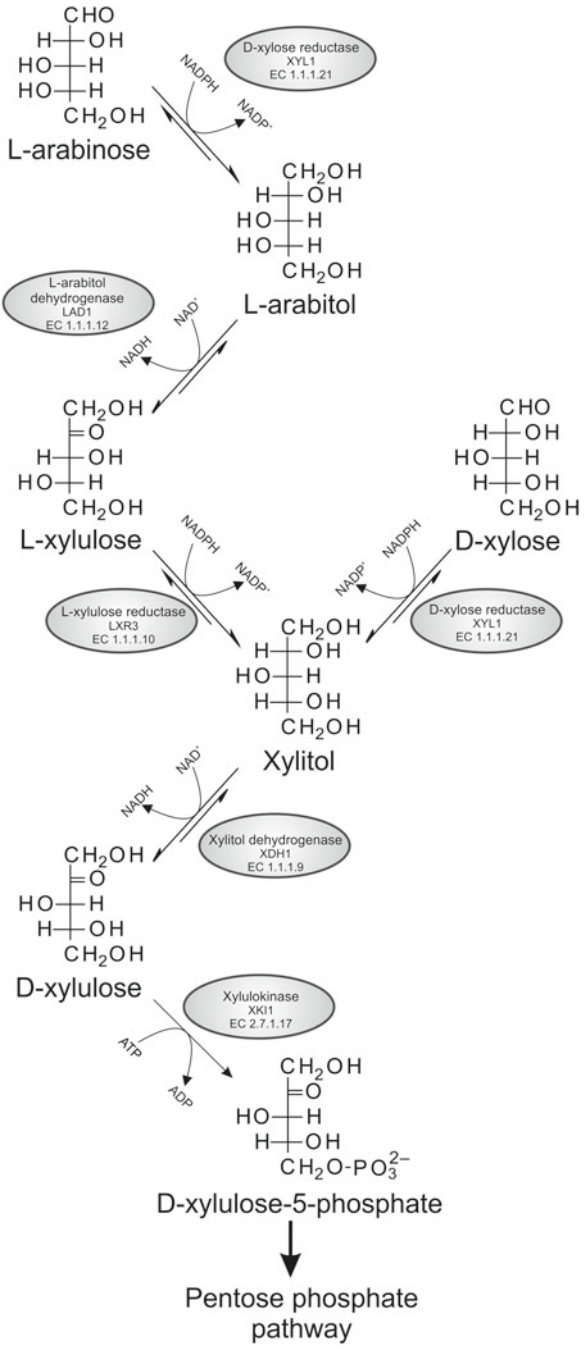


Fig. 18.3 The oxidoreductive pathway of D-xylose and L-arabinose catabolism in fungi

with a preferred specificity for L-arabinose and whose absence significantly reduced growth on L-arabinose but not D-xylose (Mojzita et al. 2010). The xylitol dehydrogenase XYL1 is the second enzyme of D-xylose pathway and the first enzyme of the common part of the pathway with L-arabinose. Genetic evidence showed that xylitol dehydrogenase is not essential for D-xylose and xylitol catabolism and that another polyol dehydrogenase, L-arabinitol 4-dehydrogenase LAD1, can partially compensate for its loss (Seiboth et al. 2003). However deletion of both *lad1* and *xdh1* in *T. reesei* fully impairs growth on D-xylose and xylitol. LAD1 is also the second enzyme in L-arabinose catabolism following XYL1 and is essential for L-arabinose and L-arabitol catabolism (Pail et al. 2004). Identification of L-xylulose reductase (LXR), the enzyme that catalyzes the third step, proved to be a difficult task due to presence of many potential candidates with L-xylulose reductase activity in the fungal genomes. The first gene that encoded a fungal enzyme with L-xylulose reductase activity (LXR1) was identified in *T. reesei* (Richard et al. 2002) but subsequently demonstrated to actually be a D-mannitol dehydrogenase that is involved in developmental processes (Metz et al. 2009). Only recently, an NADPH-dependent L-xylulose reductase (LXR3) could be identified in *T. reesei*, whose deletion results in strongly reduced growth on L-arabinose and L-arabinitol (B. Metz, C.P. Kubicek and B. Seiboth, unpublished data). Phosphorylation of D-xylulose at C-5 is catalyzed by xylulokinase, an enzyme present in *T. reesei*. Its functional characterization in *A. niger* showed that it is essential for L-arabinose and D-xylose catabolism.

18.4.2.2 Engineering the Pentose Catabolic Pathway for Hemicellulase Overexpression

T. reesei contains three GH11 (XYN1, XYN2 and XYN5), one GH10 (XYN3) and at least one GH30 xylanase (XYN4). All but XYN3, which has previously been described to be only expressed upon in a mutant strain (*T. reesei* PC-3-7) on cellulose, L-sorbose and sophorose, are coordinately induced by D-xylose and L-arabinose. Mach-Aigner et al. (2011) recently reported that D-xylose does not induce *xyn1* and *xyn2* gene expression in a delta-*xyl1* (xylose reductase) mutant and therefore concluded that a metabolite of xylose catabolism must be the inducer of xylanase gene expression. Unfortunately, they assayed xylanase transcription essentially after 2 h only, which – as we have shown here – is a too early time point where the real effect of D-xylose is not yet visible. Based on their findings that L-arabinitol also induces xylanase gene expression, they concluded that D-xylose may have to be converted to L-arabinitol to act as an inducer. However, we show here that addition of D-xylose also resulted in increased xylanase gene expression in a delta-*lxr4* (L-xylulose reductase) mutant, which renders this hypothesis impossible. We have recently investigated the induction of expression of *xyn1, xyn2, xyn4* and *xyn5* by D-xylose and L-arabinose in mutants blocked in individual steps of the oxidoreductive pentose catabolic pathway (Herold et al. unpublished data). Based on the use of the delta-*xlr3* strain, which allows to distinguish between the contribution between the D-xylose and L-arabinose catabolic pathway, as it blocks the interconversion of the two

sugars, we found that D-xylose and L-arabinose induce xylanases via different transmitters: D-xylose probably by itself or its transport through the plasma membrane, whereas L-arabinose likely induces via L-arabinitol. This finding of two independent inducers of xylanase gene expression in *T. reesei* is also supported by the synergistic effect of induction by D-xylose and L-arabinose. By this, the fungus may benefit from the fact that there are two major types of xylans: one, which occurs in the cell wall of cereals, which often contain large quantities of L-arabinose and are consequently termed arabinoxylans; and another one, hardwood xylan, which contains large amounts of D-glucuronic acid linked to the backbone (glucuronoxylans; Scheller and Ulvskov 2010). Mutants blocked in the steps after D-xylose and L-arabinitol have been shown to lead to a massive increase in expression of *xyn1, xyn2, xyn4* and *xyn5* (Herold et al. unpublished data).

T. reesei also contains four genes for the degradation of arabinan, i.e. the α-L-arabinofuranosidase encoding genes *abf1, abf2*, and *abf3* and also *bxl1*, which encodes a ß-xylosidase with a separate α-L-arabinofuranosidase domain and activity, but lacks endo-arabinanases. All of these genes is induced by L-arabinose and to a lesser extent by L-arabinitol and absent on D-glucose. Using the above mentioned pathway mutants, Akel et al. (2009) found that the induction by L-arabinitol is strongly enhanced in a Δ*lad1* strain lacking L-arabinitol dehydrogenase activity and severely impaired in an aldose reductase (Δ*xyl1)* strain, suggesting a cross talk between L-arabinitol and the aldose reductase XYL1.

18.4.3 Engineering the Pentose Phosphate Pathway

The pentose phosphate pathway (PPP) is the major source of NADPH, needed for the biosynthesis of many biomolecules, and provides intermediates for the synthesis of aromatic amino acids and histidine. Consequently, the PPP plays an important role in the production of proteins and may become limiting in an organism under conditions of massive protein overproduction. To this end, Limon et al. (2011) eliminated phosphoglucose isomerase activity in *T. reesei* RUT C30 by *pgi1* knock out in order to redirect the carbon flux through the PPP. The corresponding mutants grew in media with D-fructose, D-galactose, D-xylose, glycerol or lactose but did not in media with D-xylose and D-glucose. Interestingly, they secreted enhanced cellulase activity on glucose, whereas no increase was found on lactose. The authors speculated therefore that the observed increase would not be due to an increased PPP (in fact, they found that the expression of the gene encoding glucose-6-phosphate dehydrogenase, the first enzyme of the PPP, was not increased in the Δ*pgi1* strain) but rather related to interaction with CRE1-dependent carbon catabolite repression.

The way in which fungi sense the presence of D-glucose and signal carbon catabolite repression to CRE1 is only poorly understood. In *A. nidulans* and *T. reesei* a single glucokinase and a single hexokinase are present and only in the absence of both of them carbon catabolite derepression is observed (Flipphi et al. 2003). Strauss et al. (1999) have shown data that suggested that *A. nidulans* requires a

certain intracellular concentration of D-glucose-6-phosphate resulting from the action of both kinases to stimulate CreA-mediated carbon catabolite repression.

18.5 Conclusion

Research with *T. reesei* has strongly benefited from the availability of its genome sequence. Moreover, all the tools needed for systematic and high throughput metabolic pathway manipulation are now available, and can be used to produce regulatory or pathway mutants within a short time. Systems biological approaches (including genomic, transcriptomic, proteomic and metabolomic tools) have already been started towards the identification of alterations that had been introduced into the genomes of high(er) producer strains during classical mutagenesis (Kubicek 2012). In view of the information that is already available on the biochemical pathways and their regulation that lead to the induction of cellulases and hemicellulases, it can be anticipated that systems biological investigation of these engineered strains will be the next step towards a global understanding of the physiology required to form biomass degrading enzymes in *T. reesei*.

References

Akel E, Metz B, Seiboth B, Kubicek CP (2009) Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*). Eukaryot Cell 8:1837–1844

Andreotti RE, Medeiros JE, Roche C, Mandels M (1980) Effects of strain and substrate on production of cellulases by *Trichoderma reesei* mutants. In: Ghose TK (ed) Bioconversion and bioengineering symposium. BERC, IIT, New Delhi, pp 353–371

Aro N, Saloheimo A, Ilmen M, Penttilä M (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. J Biol Chem 276:24309–24314

Aro N, Ilmen M, Saloheimo A, Penttilä M (2003) ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. Appl Environ Microbiol 69:56–65

Aro N, Pakula T, Penttilä M (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. FEMS Microbiol Rev 29:719–739

Bhat PJ, Murthy TV (2001) Transcriptional control of the *GAL/MEL* regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. Mol Microbiol 40:1059–1066

Bouffard GG, Rudd KE, Adhya SL (1994) Dependence of lactose metabolism upon mutarotase encoded in the *gal* operon in *Escherichia coli*. J Mol Biol 244:269–278

Bouws H, Wattenberg A, Zorn H (2008) Fungal secretomes–nature's toolbox for white biotechnology. Appl Microbiol Biotechnol 80:381–388

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The carbohydrate-active enZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238

Carle-Urioste JC, Escobar-Vera J, El-Gogary S, Henrique-Silva F, Torigoi E, Crivellaro O, Herrera-Estrella A, El-Dorry H (1997) Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. J Biol Chem 272:10169–10174

Carroll A, Somerville C (2009) Cellulosic biofuels. Annu Rev Plant Biol 60:165–182

Cherry JR, Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. Curr Opin Biotechnol 14:438–443

Chilton IJ, Delaney CE, Barham-Morris J, Fincham DA, Hooley P, Whitehead MP (2008) The *Aspergillus nidulans* stress response transcription factor StzA is ascomycete-specific and shows species-specific polymorphisms in the C-terminal region. Mycol Res 112:1435–1446

Cziferszky A, Mach RL, Kubicek CP (2002) Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). J Biol Chem 277:14688–14694

Denton JA, Kelly JM (2011) Disruption of *Trichoderma reesei cre2*, encoding an ubiquitin C-terminal hydrolase, results in increased cellulase activity. BMC Biotechnol 11:103

Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT, Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 265:524–528

Durand H, Clanet H, Tiraby G (1988) Genetic improvement of *Trichoderma reesei* for large scale cellulase production. Enzyme Microbiol Technol 10:341–346

El-Gogary S, Leite A, Crivellaro O, Eveleigh DE, el-Dorry H (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. Proc Natl Acad Sci USA 86:6138–6141

Fekete E, Karaffa L, Sandor E, Banyai I, Seiboth B, Gyemant G, Sepsi A, Szentirmai A, Kubicek CP (2004) The alternative D-galactose degrading pathway of *Aspergillus nidulans* proceeds via L-sorbose. Arch Microbiol 181:35–44

Fekete E, Karaffa L, Kubicek CP, Szentirmai A, Seiboth B (2007) Induction of extracellular β-galactosidase (Bga1) formation by D-galactose in *Hypocrea jecorina* is mediated by galactitol. Microbiology 153:507–512

Fekete E, Seiboth B, Kubicek CP, Szentirmai A, Karaffa L (2008) Lack of aldose 1-epimerase in *Hypocrea jecorina* (anamorph *Trichoderma reesei*): a key to cellulase gene expression on lactose. Proc Natl Acad Sci USA 105:7141–7146

Fekete E, Karaffa L, Seiboth B, Fekete E, Kubicek CP, Flipphi M (2012) Identification of a permease gene involved in lactose utilisation in *Aspergillus nidulans*. Fungal Genet Biol 49(6):415–425

Flipphi M, van de Vondervoort PJ, Ruijter GJ, Visser J, Arst HN Jr, Felenbok B (2003) Onset of carbon catabolite repression in *Aspergillus nidulans*. Parallel involvement of hexokinase and glucokinase in sugar signaling. J Biol Chem 278:11849–11857

Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, Goedegebuur F, Houfek TD, England GJ, Kelley AS, Meerman HJ, Mitchell T, Mitchinson C, Olivares HA, Teunissen PJ, Yao J, Ward M (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J Biol Chem 278:31988–31997

Fowler T, Brown RD Jr (1992) The *bgl1* gene encoding extracellular β-glucosidase from *Trichoderma reesei* is required for rapid induction of the cellulase complex. Mol Microbiol 6:3225–3235

Frey PA (1996) The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. FASEB J 10:461–470

Fritscher C, Messner R, Kubicek CP (1990) Cellobiose metabolism and cellobiohydrolase I biosynthesis by *Trichoderma reesei*. Exp Mycol 14:451–461

Furukawa T, Shida Y, Kitagami N, Mori K, Kato M, Kobayashi T, Okada H, Ogasawara W, Morikawa Y (2009) Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. Fungal Genet Biol 46:564–574

Gremel G, Dorrer M, Schmoll M (2008) Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*). BMC Microbiol 8:174

Harman GE, Kubicek CP (1998) *Trichoderma* and *Gliocladium*, enzymes, biological control and commercial applications, vol 2. Taylor & Francis Ltd., London

Hartl L, Kubicek CP, Seiboth B (2007) Induction of the *gal* pathway and cellulase genes involves no transcriptional inducer function of the galactokinase in *Hypocrea jecorina*. J Biol Chem 282:18654–18659

Holden HM, Rayment I, Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. J Biol Chem 278:43885–43888

Ilmen M, Thrane C, Penttilä M (1996) The glucose repressor gene *cre1* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. Mol Gen Genet 251:451–460

Ilmen M, Saloheimo A, Onnela ML, Penttilä ME (1997) Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. Appl Environ Microbiol 63:1298–1306

Karaffa L, Fekete E, Gamauf C, Szentirmai A, Kubicek CP, Seiboth B (2006) D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. Microbiology 152:1507–1514

Koivistoinen OM, Richard P, Penttila M, Ruohonen L, Mojzita D (2012) Sorbitol dehydrogenase of *Aspergillus niger*, SdhA, is part of the oxido-reductive D-galactose pathway and essential for D-sorbitol catabolism. FEBS Lett 586:378–383

Kubicek CP (1987) Involvement of a conidial endoglucanase and a plasma-membrane-bound beta-glucosidase in the induction of endoglucanase synthesis by cellulose in *Trichoderma reesei*. J Gen Microbiol 133:1481–1487

Kubicek CP (2012) Systems biological approaches towards understanding cellulase production by *Trichoderma reesei.* J Biotechnol Jun 29. [Epub ahead of print]

Kubicek CP, Harman GE (eds) (1998) Trichoderma and gliocladium, vol 1 and 2. Taylor & Francis Ltd., London

Kubicek CP, Mühlbauer G, Grotz M, John E, Kubicek-Pranz EM (1988) Properties of a conidial bound cellulase enzyme system from *Trichoderma reesei*. J Gen Microbiol 134:1215–1222

Kubicek CP, Messner R, Gruber F, Mandels M, Kubicek-Pranz EM (1993) Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose- inhibited β-diglucoside permease. J Biol Chem 268:19364–19368

Kubicek CP, Mikus M, Schuster A, Schmoll M, Seiboth B (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. Biotechnol Biofuels 2:19

Kumar R, Singh S, Singh OV (2008) Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J Ind Microbiol Biotechnol 35:377–391

Limon MC, Pakula T, Saloheimo M, Penttilä M (2011) The effects of disruption of phosphoglucose isomerase gene on carbon utilisation and cellulase production in *Trichoderma reesei* RutC30. Microbl Cell Fact 10:40

Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL (2008) Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. Appl Environ Microbiol 74:6554–6562

Mach-Aigner AR, Grosstessner-Hain K, Pocas-Fonseca MJ, Mechtler K, Mach RL (2010) From an electrophoretic mobility shift assay to isolated transcription factors: a fast genomic-proteomic approach. BMC Genomics 11:644

Mach-Aigner AR, Gudynaite-Savitch L, Mach RL (2011) L-Arabitol is the actual inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). Appl Environ Microbiol 77:5988–5994

Mandels M, Parrish FW, Reese ET (1962) Sophorose as an inducer of cellulase in *Trichoderma viride*. J Bacteriol 83:400–408

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol 26:553–560

Merino ST, Cherry J (2007) Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol 108:95–120

Messner R, Kubicek-Pranz EM, Gsur A, Kubicek CP (1991) Cellobiohydrolase II is the main conidial-bound cellulase in *Trichoderma reesei* and other *Trichoderma* strains. Arch Microbiol 155:601–606

Metz B, de Vries RP, Polak S, Seidl V, Seiboth B (2009) The *Hypocrea jecorina* (syn. *Trichoderma reesei*) *lxr1* gene encodes a D-mannitol dehydrogenase and is not involved in L-arabinose catabolism. FEBS Lett 17:1309–1313

Metz B, Seidl-Seiboth V, Haarmann T, Kopchinskiy A, Lorenz P, Seiboth B, Kubicek CP (2011) Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*. Eukaryot Cell 10:1527–1535

Mishra P, Singh A (1993) Microbial pentose utilization. Adv Appl Microbiol 39:91–152

Mojzita D, Penttilä M, Richard P (2010) Identification of an L-arabinose reductase gene in *Aspergillus niger* and its role in L-arabinose catabolism. J Biol Chem 285:23622–23628

Mojzita D, Koivistoinen OM, Maaheimo H, Penttila M, Ruohonen L, Richard P (2012) Identification of the galactitol dehydrogenase, LadB, that is part of the oxido-reductive D-galactose catabolic pathway in *Aspergillus niger*. Fungal Genet Biol 49:152–159

Nakari-Setälä T, Paloheimo M, Kallio J, Vehmaanperä J, Penttilä M, Saloheimo M (2009) Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. Appl Environ Microbiol 75:4853–4860

Noguchi Y, Tanaka H, Kanamaru K, Kato M, Kobayashi T (2011) Xylose triggers reversible phosphorylation of XlnR, the fungal transcriptional activator of xylanolytic and cellulolytic genes in *Aspergillus oryzae*. Biosci Biotechnol Biochem 75(5):953–959

Pail M, Peterbauer T, Seiboth B, Hametner C, Druzhinina I, Kubicek CP (2004) The metabolic role and evolution of L-arabinitol 4-dehydrogenase of *Hypocrea jecorina*. Eur J Biochem 271:1864–1872

Pakula TM, Salonen K, Uusitalo J, Penttilä M (2005) The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus *Trichoderma reesei*. Microbiology 151:135–143

Penttilä M, Limon C, Nevalainen H (2004) Molecular biology of *Trichoderma* and biotechnological applications. In: Arora DK (ed) Handbook of fungal biotechnology, vol 20, 2nd edn. Marcel Dekker, New York, pp 413–427

Pettersson H, Pettersson G (2001) Kinetics of the coupled reaction catalysed by a fusion protein of β-galactosidase and galactose dehydrogenase. Biochim Biophys Acta 1549:155–160

Portnoy T, Margeot A, Seidl-Seiboth V, Le Crom S, Ben Chaabane F, Linke R, Seiboth B, Kubicek CP (2011) Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. Eukaryot Cell 10:262–271

Reyes-Dominguez Y, Bok JW, Berger H, Shwab EK, Basheer A, Gallmetzer A, Scazzocchio C, Keller N, Strauss J (2010) Heterochromatic marks are associated with the repression of secondary metabolism clusters in *Aspergillus nidulans*. Mol Microbiol 76:1376–1386

Richard P, Putkonen M, Väänänen R, Londesborough J, Penttilä M (2002) The missing link in the fungal L-arabinose catabolic pathway, identification of the L-xylulose reductase gene. Biochemistry. 41(20):6432–6437

Roberts CF (1970) Enzyme lesions in galactose non-utilizing mutants of *Aspergillus nidulans*. Biochim Biophys Acta 201:267–283

Rubio-Texeira M (2005) A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. FEMS Yeast Res 5:1115–1128

Saloheimo A, Aro N, Ilmen M, Penttilä M (2000) Isolation of the *ace1* gene encoding a Cys(2)-His(2) transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei*. J Biol Chem 275:5817–5825

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61:263–289

Schmoll M, Schuster A, Silva Rdo N, Kubicek CP (2009) The G-alpha protein GNA3 of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*) regulates cellulase gene expression in the presence of light. Eukaryot Cell 8:410–420

Schuster A, Tisch D, Seidl-Seiboth V, Kubicek CP, Schmoll M (2012) Roles of protein kinase A and adenylate cyclase in light-modulated cellulase regulation in *Trichoderma reesei*. Appl Environ Microbiol 78:2168–2178

Seibel C, Gremel G, do Nascimento SR, Schuster A, Kubicek CP, Schmoll M (2009) Light-dependent roles of the G-protein alpha subunit GNA1 of *Hypocrea jecorina* (anamorph *Trichoderma reesei*). BMC Biol 7:58

Seiboth B, Hakola S, Mach RL, Suominen PL, Kubicek CP (1997) Role of four major cellulases in triggering of cellulase gene expression by cellulose in *Trichoderma reesei*. J Bacteriol 179:5318–5320

Seiboth B, Hofmann G, Kubicek CP (2002a) Lactose metabolism and cellulase production in *Hypocrea jecorina*: the *gal7* gene, encoding galactose-1-phosphate uridylyltransferase, is essential for growth on galactose but not for cellulase induction. Mol Genet Genomics 267:124–132

Seiboth B, Karaffa L, Sandor E, Kubicek C (2002b) The *Hypocrea jecorina gal10* (uridine 5'-diphosphate-glucose 4-epimerase- encoding) gene differs from yeast homologues in structure, genomic organization and expression. Gene 295:143–149

Seiboth B, Hartl L, Pail M, Kubicek CP (2003) D-xylose metabolism in *Hypocrea jecorina*: loss of the xylitol dehydrogenase step can be partially compensated for by *lad1*-encoded L-arabinitol-4-dehydrogenase. Eukaryot Cell 2:867–875

Seiboth B, Hartl L, Pail M, Fekete E, Karaffa L, Kubicek CP (2004) The galactokinase of *Hypocrea jecorina* is essential for cellulase induction by lactose but dispensable for growth on D-galactose. Mol Microbiol 51:1015–1025

Seiboth B, Hartl L, Salovuori N, Lanthaler K, Robson GD, Vehmaanperä J, Penttilä ME, Kubicek CP (2005) Role of the *bga1*-encoded extracellular β-galactosidase of *Hypocrea jecorina* in cellulase induction by lactose. Appl Environ Microbiol 71:851–857

Seiboth B, Gamauf C, Pail M, Hartl L, Kubicek CP (2007) The D-xylose reductase of *Hypocrea jecorina* is the major aldose reductase in pentose and D-galactose catabolism and necessary for beta-galactosidase and cellulase induction by lactose. Mol Microbiol 66:890–900

Seiboth B, Aghcheh RK, Phatale PA, Linke R, Hartl L, Sauer DG, Smith KM, Baker SE, Freitag M, Kubicek CP (2012) The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. Mol Microbiol 84(6):1150–1164

Seidl V, Gamauf C, Druzhinina IS, Seiboth B, Hartl L, Kubicek CP (2008) The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. BMC Genomics 9:327

Sellick CA, Campbell RN, Reece RJ (2008) Galactose metabolism in yeast-structure and regulation of the Leloir pathway enzymes and the genes encoding them. Int Rev Cell Mol Biol 269:111–150

Sestak S, Farkas V (1993) Metabolic regulation of endoglucanase synthesis in *Trichoderma reesei*: participation of cyclic AMP and glucose-6-phosphate. Can J Microbiol 39:342–347

Sternberg D, Mandels GR (1979) Induction of cellulolytic enzymes in *Trichoderma reesei* by sophorose. J Bacteriol 139:761–769

Strauss J, Horvath HK, Abdallah BM, Kindermann J, Mach RL, Kubicek CP (1999) The function of CreA, the carbon catabolite repressor of *Aspergillus nidulans*, is regulated at the transcriptional and post-transcriptional level. Mol Microbiol 32:169–178

Stricker AR, Grosstessner-Hain K, Würleitner E, Mach RL (2006) Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. Eukaryot Cell 5:2128–2137

Stricker AR, Mach RL, de Graaff LH (2008a) Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). Appl Microbiol Biotechnol 78:211–220

Stricker AR, Trefflinger P, Aro N, Penttilä M, Mach RL (2008b) Role of Ace2 (Activator of Cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. Fungal Genet Biol 45:436–445

Vaheri M, Leisola M, Kaupinnen V (1979) Transglycosylation products of the cellulase system of *Trichoderma reesei*. Biotechnol Lett 1:41–46

Wilson DB (2009) Cellulases and biofuels. Curr Opin Biotechnol 20:295–299

Zeilinger S, Ebner A, Marosits T, Mach R, Kubicek CP (2001) The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the *cbh2* (cellobiohydrolase II-gene) activating element. Mol Genet Genomics 266:56–63

Chapter 19
Microbiologically Produced Carboxylic Acids Used as Building Blocks in Organic Synthesis

Andreas Aurich, Robert Specht, Roland A. Müller, Ulrich Stottmeister, Venelina Yovkova, Christina Otto, Martina Holz, Gerold Barth, Philipp Heretsch, Franziska A. Thomas, Dieter Sicker, and Athanassios Giannis

Abstract Oxo- and hydroxy-carboxylic acids are of special interest in organic synthesis. However, their introduction by chemical reactions tends to be troublesome especially with regard to stereoselectivity. We describe herein the biotechnological preparation of selected oxo- and hydroxycarboxylic acids under "green" conditions and their use as promising new building blocks. Thereby, our biotechnological goal was the development of process fundamentals regarding the variable use of renewable raw materials, the development of a multi purpose bioreactor and application of a pilot plant with standard equipment for organic acid production to minimize the technological effort. Furthermore the development of new product isolation procedures, with the aim of direct product recovery, capture of products or single step operation, was necessary. The application of robust and approved microorganisms, also genetically modified, capable of using a wide range of substrates as well as producing a large spectrum of products, was of special importance. Microbiologically produced acids, like 2-oxo-glutaric acid and 2-oxo-D-gluconic acid, are useful educts for the chemical synthesis of hydrophilic triazines, spiro-connected heterocycles, benzotriazines, and pyranoic amino acids. The chiral intermediate of the tricarboxylic acid cycle, (2*R*,3*S*)-isocitric acid, is another promising compound. For the first time our process provides large quantities of enantiopure trimethyl (2*R*,3*S*)-isocitrate which was used in subsequent chemical transformations to provide new chiral entities for further usage in total synthesis and pharmaceutical research.

A. Aurich • R. Specht • R.A. Müller • U. Stottmeister (✉)
Helmholtz-Centre for Environmental Research – UFZ,
Permoserstr. 15, 04318 Leipzig, Germany
e-mail: ulrich.stottmeister@ufz.de

V. Yovkova • C. Otto • M. Holz • G. Barth
Institute of Microbiology, Dresden University of Technology, Dresden, Germany

P. Heretsch • F.A. Thomas • D. Sicker • A. Giannis
Institute of Organic Chemistry, University of Leipzig, Leipzig, Germany

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_19,

Keywords Biotechnical cultivation • *Gluconobacter oxydans* • (2*R*,3*S*)-Isocitric acid • Microorganism • 2-Oxo-D-gluconic acid • 2-Oxoglutaric acid • Strain improvement • *Yarrowia lipolytica*

Abbreviations

CA	Citric acid
EDBM	Electrodialysis by bipolar membranes
ICA	(2*R*,3*S*)-Isocitric acid
2-OGA	2-Oxoglutaric acid
2-OGcA	2-Oxo-D-gluconic acid
SF	Shaking flask
STR	Stirred tank reactor
TCA cycle	Tricarboxylic acid cycle

19.1 Introduction

Oxo- and hydroxy-carboxylic acids are highly versatile chemical species and therefore of special interest in organic synthesis. However, the introduction of both the oxo- and the hydroxy group into a carboxylic acid by chemical reactions tends to be troublesome and often lacks efficiency especially with regard to possible introduction of new stereocenters.

Interestingly, oxo- and hydroxy-carboxylic acids are intermediate products of the metabolism of microbial, animal and plant cells under aerobic and anaerobic conditions (Table 19.1).

Little is known about the use of microbiologically produced acids as building blocks in organic synthetic chemistry (Stottmeister et al. 2005). Our team of biotechnologists and chemists intends to optimize the biotechnological preparation of selected oxo- and hydroxy-carboxylic acids under "green" conditions and to use them as building blocks in organic synthesis and pharmaceutical chemistry. In the course of anabolism and catabolism of living cells carboxylic acids were formed by oxidation steps from carbohydrates, hydrocarbons and many other carbon compounds, serving as carbon and energy sources.

The tricarboxylic acid cycle (TCA cycle, Krebs cycle, citric acid cycle) is the central primary metabolic process in energy regeneration and production of metabolites in all living cells. It is further present in "complete" TCA cycle with ten participating acids or modified as "incomplete" TCA cycle, sometimes with shortcuts by anaplerotic sequences.

The participating hydroxy-, oxo-, and unsaturated acids with the chain length from two to six carbon atoms react in six different chemical reaction types catalyzed

Table 19.1 Some oxo- and hydroxy-carboxylic acids formed in cell metabolism

Trivial name	Chemical name (IUPAC)	Formula, [CAS-RN]	pK_a	Stereoisomers	Production
Glyoxylic acid	Oxoethanoic acid	$C_2H_2O_3$ [298-12-4]	3.32	None	Synthetical
Lactic acid	2-Hydroxy-propanoic acid	$C_3H_6O_3$ [50-21-5], L: [79-33-4], D: [10326-41-7], D/L: [598-82-3]	3.86	(+)-(*R*)-(−)-(*S*)-	Synthetical and microbiological
Pyruvic acid	2-Oxopropanoic acid	$C_3H_4O_3$ [127-17-3]	2.49	None	Synthetical and microbiological
Malonic acid	Propanedioic acid	$C_3H_4O_4$ [141-82-2]	2.83;5.69	None	Synthetical
Succinic acid	Butanedioic acid	$C_4H_6O_4$ [110-15-6]	4.19;5.24	None	Synthetical and microbiological
Maleic acid	(*Z*)-Butenedioic acid	$C_4H_4O_4$ [110-16-7]	2.00;6.26	None	Synthetical
Fumaric acid	(*E*)-Butenedioic acid	$C_4H_4O_4$ [110-17-8]	3.02;4.39	None	Synthetical and microbiological
Malic acid	Hydroxybutane-dioic acid	$C_4H_6O_5$ [6915-15-7]	3.405.11	(+)-(*R*)-(−)-(*S*)-	Synthetical and microbiological
Tartaric acid	2,3-Dihydroxy-butanedioic acid	$C_4H_6O_6$ [526-83-0]	2.98;4.34	(+)-(2*R*,3*R*)-(−)-(2*S*,3*S*)-*meso*-	Microbiological and synthetical
Citric acid	2-Hydroxy-propane-1,2,3-tricarboxylic acid	$C_6H_8O_8$ [77-92-9]	3.13;4.76;6.40	None	Microbiological
Isocitric acid	1-Hydroxy-propane-1,2,3-tricarboxylic acid	$C_6H_8O_8$ [320-77-4]	3.29;4.71;6.40	(+)-(2*R*,3*S*)-	Microbiological
2-Oxo-glutaric acid	2-Oxopentane-dioic acid	$C_5H_6O_5$ [328-50-7]	2.47;4.68	None	Microbiological and synthetical
2-Oxo-D-gluconic acid	D-*arabino*-2-Hexulosonic acid	$C_6H_{10}O_7$ [669-90-9]	2.66	None	Microbiological and synthetical

by a multi enzyme complex. The acids of the TCA cycle are detectable not only in cells, but also in the aqueous solution after excretion in large amounts. From the physiological point of view the resulting external pH-decrease supports the cell survival (yeast, bacteria) or is part of the energy and nutrient regulation under cell stress conditions.

Citric acid (CA) is produced in bulk quantities exclusively by biotechnological processes (about 1 million tons/year). Our interest was focussed on the 2-oxo-carboxylic group, like 2-oxoglutaric acid (2-OGA) and 2-oxo-D-gluconic acid (2-OGcA). (2*R*,3*S*)-Isocitric acid (ICA) is another promising compound that can be isolated from the chiral pool. It is known that organic acids are biotechnologically produced by standard cultivation technologies under anaerobic or aerobic conditions.

In general, the biological processes show a lot of advantages:

- use of renewable feed stocks;
- low reaction temperature range (up to 35 °C), no pressure (sterilisation of vessels no more than 2 atm);
- formation of defined chemical species in one step, which otherwise need multi step syntheses;
- stereo- and regiospecific reactions with high selectivity;
- avoiding of unwanted side-products;
- high potential for reaction optimisation/for novel reactions by new molecular biological tools;

However, analysing historical and recently used technologies, we summarize the weak points of biological reactions:

- reaction time is long;
- productivity is low;
- the final concentration in the cultivation broth is low;
- the water demand and waste water production are high;
- the energy demand for the cultivation step (mass transfer) is high;
- the energy demand for product isolation (from aqueous solution) is high;
- the stability of micro organisms especially of high performance strains is often uncertain;
- genetically optimised strains have often restrictions in availability;
- presets of the cultivation, like sterilisation, are energy consuming.

Regarding environmental aspects, the use of biological reactions does not imply a priori environmental friendly reactions. It is necessary to consider the whole process of raw material application, biosynthesis, product recovery and chemical synthesis as unit.

Our goal was the development of process fundamentals (Fig. 19.1) regarding

- the variable use of renewable raw materials (plant oils, carbohydrates);
- development of a variable bioreactor (“multi purpose” reactor; Fig. 19.2);
- development and application of a pilot plant “standard equipment” for organic acid production to minimize the technological effort;

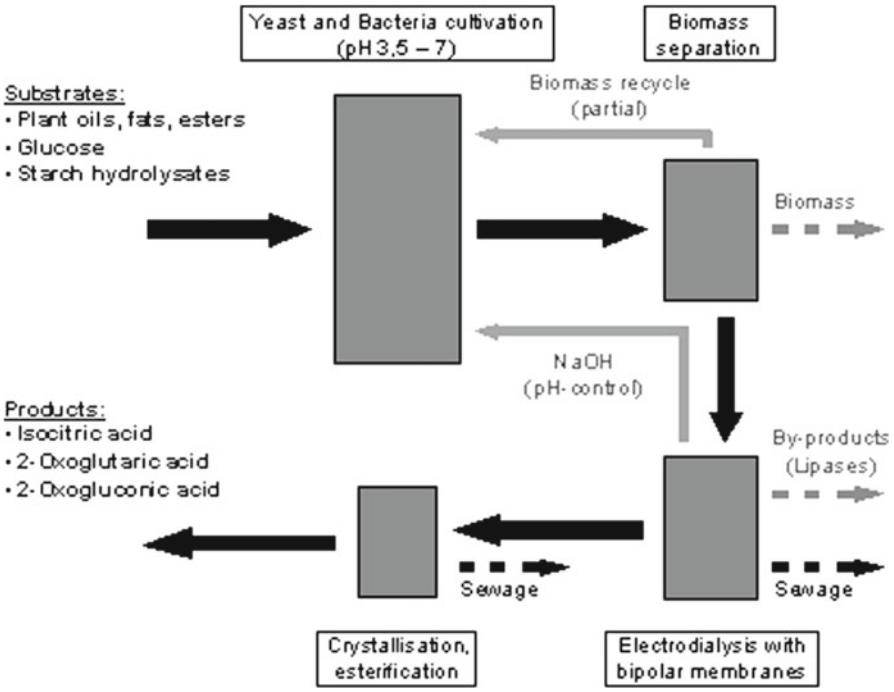


Fig. 19.1 Process flow scheme for a multi-substrate multi-product strategy with yeast or bacteria as microorganisms

Fig. 19.2 Archive of the Helmholtz--Centre for Environmental Research UFZ

- development of new product isolation procedures, with the aim of direct product recovery, direct capture of products or single step operation.

The application of robust and approved microorganisms, capable of using a wide range of substrates as well as producing a large spectrum of products, was of special importance.

19.2 Development of the Standard Equipment for Cultivation, Purification and Product Isolation for Organic Acids

We could achieve one of our main goals by developing a biotechnological process configuration with the following constituents: an aerated stirred tank reactor (STR), which is fitted with control and regulation units for pH, pO_2, temperature, measurements of O_2 and CO_2 in the exhausted gas, biomass separation stage and electrodialysis stage for organic acid separation.

This process flow (Fig. 19.1) allows multi substrate application (carbohydrates, triglycerides) for the formation of multiple products, standardization in cultivation, product recovery and decreased waste arisings resulting from inner cycles.

Electrodialysis by bipolar membranes (EDBM) is a very important step for the purification of the cultivation products and has a positive influence on the efficiency of the whole process. The biomass free cultivation broth can be purified by EDBM. The sodium salts are converted into the free acids with high selectivity and purity. Only a low alkaline content guarantees the fast and complete conversion in the following chemical esterification of the organic acids. Furthermore, the sodium hydroxide solution produced in the EDBM can be used for pH-regulation during the cultivation and is an important economic factor (Ammon 1996; Huang et al. 2007; Moresi and Sappino 2000; Novalic et al. 1995; Rapp 1995; Tongwen and Weihua 2002).

19.3 New Process Design for Biosynthesis of Oxo- and Hydroxycarboxylic Acids

19.3.1 2-Oxoglutaric Acid

As an intermediate of the TCA cycle 2-oxoglutaric acid (2-OGA) – as well as CA – can be exocellularly exuded by microorganisms (Table 19.2). The intracellular energy regulation under conditions of growth limitation and unlimited substrate uptake of the cells grown are important. A precondition for this process is the limitation of reproductive cell growth by nutrient exhaustion from the culture broth. In the case of 2-OGA, thiamine limitation is also necessary because of the thiamine

Table 19.2 Formation of 2-oxoglutaric acid (2-OGA) by microorganisms

Microorganism	Carbon source	Conditions	2-OGA (g L^{-1})	Yield (g g^{-1})	References
Arhthrobacter paraffineus	*n*-Paraffins	STR	60	0.74	Tabaka et al. (1969)
Candida lipolytica	*n*-Paraffins	STR	48	0.60	Tsugawa et al. (1969)
C. lipolytica (diploid)	*n*-Paraffins	STR	185	0.80	Maldonado et al. (1973)
Yarrowia lipolytica H222-27-11	*n*-Paraffins	STR, fed-batch	195	0.90	Weißbrodt et al. (1988)
Y. lipolytica N1	Ethanol	STR, fed-batch	49	0.42	Chernyavskaya et al. (2000), Il'chenko et al. (2003)

dependence of the 2-OGA-dehydrogenase complex. Inactivity of this enzyme causes a metabolic overflow of 2-OGA. In practice, thiamine limitation can be realised only by thiamine auxotrophic strains, and in the absence of complex suppliers such as yeast extract. Suitable substrates for 2-OGA production with *Yarrowia (Y.) lipolytica* are carbon sources that are metabolised via the intermediate acetyl-CoA. In this way, accumulation of the undesired by-product pyruvic acid is minimised.

19.3.1.1 Microbial Synthesis of 2-Oxoglutaric Acid and Process Optimisation

Our strain *Y. lipolytica* H222-27-11 is able to synthesize 2-OGA from purified *n*-paraffin with a chain length of C_{12}–C_{18} under aerobic conditions with controlled thiamine limitation (not higher than 1 μg L^{-1}), and to excrete up to 195 g L^{-1} 2-OGA into the cultivation broth after feeding with *n*-paraffins (Weißbrodt et al. 1988). The substrate related yield is 90%, and the productivity is 1.3 g L^{-1} h^{-1}. The content of other carboxylic acids after 150 h of cultivation was lower than 5%, indicating the selectivity of the process. This bioprocess results in final concentrations in the cultivation broth that are higher than in the commercial CA process. Experiments using substrates from renewable feedstock in shaking flasks (SF) indicate that triglycerides or fatty acids are potentially efficient carbon sources for 2-OGA production with *Y. lipolytica* H 222-27-11 (Sakaguchi et al. 1960). Ethanol can also serve as a carbon source (Chernyavskaya et al. 2000; Il'chenko et al. 2003).

Based on this experiments further research was accomplished with different vegetable oils, and we proved the feasibility, under the condition that thiamine free oils were used to ensure a strict limitation.

The best results were achieved with rapeseed oil in a STR. After 262 h of cultivation and a feed of 120 g L^{-1} oil, 115 g L^{-1} 2-OGA were detected in the medium.

According to the state of knowledge, this is the highest concentration of 2-OGA produced from vegetable oil. The substrate related yield is 96% with a productivity of 0.61 g L^{-1} h^{-1}.

After an accumulation during the growth phase of the yeast, a partial re-assimilation of pyruvic acid after complete consumption of nitrogen source was detected. The final concentration of byproducts was less than 5 g L^{-1}. A plot of such a process is shown in Figure 19.3 and Table 19.3 (Specht 2004).

19.3.1.2 Product Isolation, Product Recovery

The isolation by bipolar electrodialysis was applicable and gave an alkali free product. The precipitation via hydrazones (see later) was conducted directly from the cultivation broth. In this way, the process was remarkably faster and simpler than separation and crystallisation prior to chemical reactions. Some more details are provided in Sect. 19.5.1.

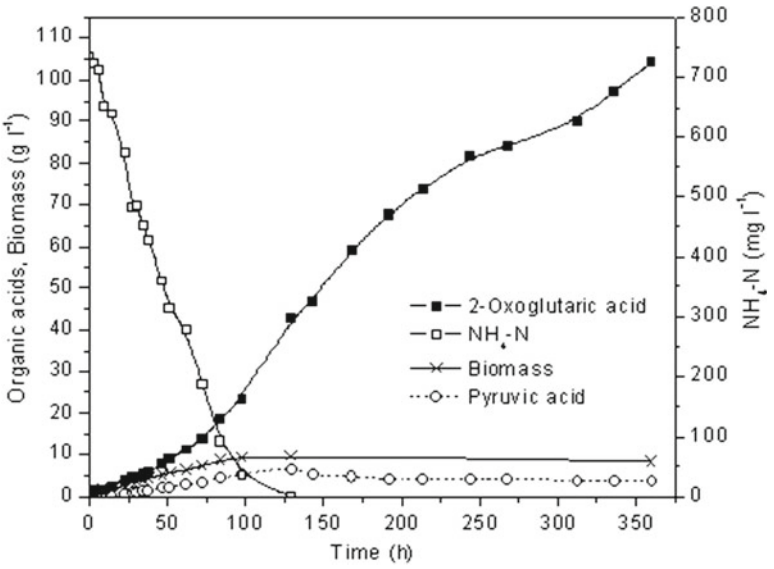


Fig. 19.3 Time course of 2-OGA formation from rapeseed oil by *Yarrowia lipolytica* H222-27-11: 2 L-STR, 30 °C, pH 5.0 (growth phase), after 48 h cultivation pH 3.5 adjusted with 30% NaOH, pO_2-content 50%, stirrer speed 1200 rpm, substrate: 40 g L^{-1} rapeseed oil and 80 g L^{-1} feeding after 15 h; 2.5 g L^{-1} NH_4Cl, 1 μg L^{-1} thiamine

Table 19.3 New substrates for 2-OGA–cultivation: plant oils

Carbon source	Conditions	2-OGA (g L^{-1})	Yield (g g^{-1})	References
Sunflower oil	SF, batch	68	0.61	Specht (2004)
Olive oil	SF, batch	65	0.58	Specht (2004)
Rapeseed oil	SF, batch	71	0.64	Specht (2004)
Rapeseed oil	STR, fed-batch	115	0.96	Specht (2004)

19.3.2 2-Oxo-D-Gluconic Acid

Many bacteria and fungi are able to oxidize glucose and gluconic acid in aqueous solution to 2-oxogluconic acid (2-OGcA). In previous patents and papers, the following microorganisms have been described to be able to oxidize both substrates: *Gluconobacter* (*G.*) *cerinus*, *G. oxydans*, *Pseudomonas* (*P.*) *fluorescens*, *P. aeruginosa*, *Serratia marcescens*, *Klebsellia aerogenes*, *Cyanococcus chromospirans*, and *Erwinia sp.*

However, the papers published before the 1960s should be considered critically, because the reliable analytical specification of the various oxidation products of glucose was only possible after introduction of high performance liquid chromatography (HPLC) and ion-exchange chromatography (IC).

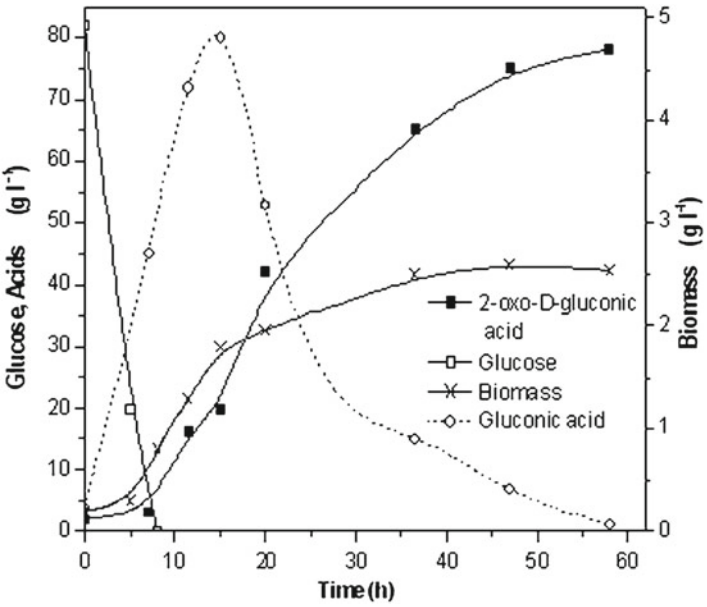


Fig. 19.4 Time course of 2-OGcA formation from glucose by *Gluconobacter oxydans* CCM 1804: 10 L STR; 30 °C; pH 5.5; pO_2 40–100%; medium composition: glucose 80 g L^{-1}, yeast extract 1.0 g L^{-1}

19.3.2.1 Microbial Synthesis

A time course of 2-OGcA production by *Gluconobacter oxydans* CCM1804 is shown in Figure 19.4. The glucose added was completely oxidized to D-gluconic acid. After the exhaustion of glucose the transformation of D-gluconic to 2-OGcA started.

The final concentration was generally about 100 g L^{-1}, the yield is between 0.85 and 0.90 g g^{-1}. The productivity has been calculated to be 3.2 g L^{-1} h^{-1} (Stubbs et al. 1940). The use of glucose feeding techniques seems to be important for the increase of the process performance, with low concentrations of glucose, inhibition of the oxidation process was observed resulting in low 2-OGcA concentrations. On the other hand, glucose concentrations higher than 120 g L^{-1} caused a decrease of the substrate related yield. Unwanted byproducts such as 5-OGcA and 2,5-OGcA were detected (Misenheimer et al. 1965). The highest productivity calculated by these authors was 7.50 g L^{-1} h^{-1} with a yield of 0.95–1.0 g g^{-1}.

19.3.2.2 Process Optimisation

Our novel process avoids the disadvantageous situation with glucose as substrate (mentioned above) (Stottmeister et al. 2005). The process is based on the oxidation of D-gluconic acid by *Serratia marcescens* (Table 19.4). The combination of this

Table 19.4 Comparison: Formation of 2-oxo-D-gluconic acid (2-OGcA) production by glucose oxidation and D-gluconic acid oxidation

Microorganism	Carbon source	Conditions	2-OGcA (g L^{-1})	Yield (g g^{-1})	Productivity (g L^{-1} h^{-1})	Reference
Gluconobacter oxidans CCM 1804	Glucose	Batch	100	0.85	2.8	Stottmeister et al. (1990)
Serratia marcescens IMET 11312	Gluconic acid	Batch	185	0.95	5.1	Stottmeister et al. (1990)

procedure with the Vogelbusch process (Vogelbusch 2012) was mandatory. The rapid glucose oxidation to D-gluconic acid by *Acetobacter methanolicus* without any byproducts can be followed by the *S. marcescens* oxidation step in the same bioreactor after sterilisation. The direct use of D-gluconate as a substrate is desirable because of the low price of the technical grade product. Formation of other oxidation products was not observed. The final concentration was 185 $g\,L^{-1}$ 2-OGcA, the yield 0.95 $g\,g^{-1}\,L^{-1}$ and the productivity 5.1 $g\,L^{-1}\,h^{-1}$.

19.3.2.3 Product Recovery, Direct Capture of Product

The biological oxidation and the formation of a biomass- and cation-free concentrated aqueous solution is a very simple procedure in our standard production unit. However, the attempts to isolate the crystalline free 2-OGcA failed, because of the formation of syrup. The methyl ester of the β-pyranose configuration is crystalline, stable and was the basis for all chemical reactions.

However, if desired, certain reagents are also suitable for a direct reaction with the aqueous electrodialysed cultivation solution containing free 2-oxo-D-gluconic acid. For instance, substituted hydrazines can react to form hydrazones that precipitate, which can be filtered and further processed. In such a case, it is advantageous that the water of the cultivation process has not to be removed by distillation as in the synthesis of the methyl ester. In this way, we were able to achieve the "direct capture of products", the first chemical reaction started directly with the purified cultivation broth. More details are provided in Sect. 19.5.2.

19.3.3 (2R,3S)-Isocitric Acid

(2*R*,3*S*)-Isocitric acid (ICA) (synonymic: D_S-*threo*-isocitric acid) has the constitution of a1-hydroxy-propane-1,2,3-tricarboxylic acid and is a chiral intermediate of the TCA cycle and glyoxylate pathway, which are part of metabolism of all living cells. Three stereoisomers of ICA with the same constitution exist: D_S-*erythro*-isocitric acid, L_S-*erythro*-isocitric acid, L_S-*threo*-isocitric acid (Vickery 1962), but only the D_S-*threo*-isocitric acid is naturally occuring.

19.3.3.1 Microbial Synthesis of (2*R*,3*S*)-Isocitric Acid

In 1960, Sakaguchi was the first author who claimed that the fungus *Penicillium purpurogenum* var. *rubrisclerotium* Thom No. 1148 excretes "isocitric acid" (L_S-*erythro*-ICA, possibly incorrect assigned) on the basis of glucose as substrate (Sakaguchi et al. 1960). The author proposed the application in beverages because of the fresh flavor and the good solubility of ICA. The cultivation time was 8 days to yield 85 $g\,L^{-1}$. The presence of byproducts was not described.

Contrary to fungi, yeasts were described to form ICA (by a mutant of *Candida ravautii*), also from glucose, in 5 days. Some yeast strains of *Candida catenulata* are able to utilize*n*-alkanes for the formation of ICA.

In the 1980's, intensive research focussed on the substitution of the classical *Aspergillus niger* process for CA production because of many technological inconveniences. The investigation of yeast with enhanced process characteristics, substrate variability and technological advantages led to detailed knowledge of the enzymatic regulation of the CA formation. The ratio of CA to ICA was of particular interest, affecting the crystallisation of the final product. The goal was to minimize the isocitric acid ratio by use of strain mutants and technological variations. As a result of these efforts, the ratio CA:ICA = 1:0.08 was described (Stottmeister et al. 1982).

During their work, Finogenova et al. discovered that mutants of the yeast *Yarrowia lipolytica* were able to produce ICA from *n*-alkanes, and from ethanol as main product, calculated the sum of the formed acids (Finogenova et al. 1989, 1991). Finogenova et al. also described the possibility to influence the ratio of ICA to CA by phenotypical optimisation (Finogenova et al. 1991, 2002). The increase of pH from 4.5 to 6.0 in the cultivation broth, high oxygen concentration (60–95% of saturation) and increase of the iron concentration support the increased formation of ICA to a ratio of ICA:CA = 2:1.

19.3.3.2 Process Optimisation

Our results were summarized recently, and opened allowed the application of "unconventional" renewable substrates like plant oil and other triglycerides in biosynthesis in our group and by Kamzolova et al. (Heretsch et al. 2008; Kamzolova et al. 2007, 2008).

One of the targets was ICA overproduction by different strains of *Yarrowia lipolytica* (Table 19.5). It was possible to increase the concentration of the ICA at a high productivity (1.14 $g\,L^{-1}\,h^{-1}$) by use of sunflower and rape seed oil. However, the part of CA in the total amount of acids was higher (Kamzolova et al. 2008).

We were able to achieve the concentration of 93 $g\,L^{-1}$ using our robust wild type strain *Yarrowia lipolytica* EH59 in our standard bioreactor facility with sunflower oil as substrate. (Heretsch et al. 2008). The product formation in a typical ICA/CA-cultivation by *Yarrowia lipolytica* is shown in Figure 19.5.

After exhaustion of ammonium nitrogen at 10 h, the production of biomass stopped and the secretion of CA and ICA started. Sunflower oil was added again after 45, 75, and 105 h. The cultivation was stopped after 167 h, producing a ICA concentration of 93 $g\,L^{-1}$ and CA concentration of 82.8 $g\,L^{-1}$, which is equivalent to a 1.14:1 ratio of ICA:CA and a substrate related yield of 0.65 $g\,g^{-1}$ ICA.

At the end of cultivation the total organic acid concentration of 175.8 $g\,L^{-1}$ was determined. The best ICA:CA ratio with 1:0.75 was calculated after 50 h of the production. The final ratio was 1:0.88.

Table 19.5 Formation of (2*R*,3*S*)-isocitric acid (ICA) and citric acid (CA) by microorganisms

Microorganism	Carbon source	Conditions	ICA (g L^{-1})	CA (g L^{-1})	Ratio ICA:CA (g g^{-1})	Yield (g g^{-1})	Reference
Penicillium purpurogenum No. 1148	Glucose	SF	84.5 (Vickery 1962)	n.d.	–	n.d.	Sakaguchi et al. (1960)
Candida ravautii OM-102	Glucose	STR, batch	60.0	0.7	1:0.01	0.50	Oogaki et al. (1983)
Candida catenulata CBS1904	*n*-alkanes	STR, batch	55.0	n.d.	–	0.90	Nakahara et al. (1987)
Y. lipolytica NMM-149	*n*-alkanes	STR, fed-batch	85.0	16.0	1:0.2	1.23	Finogenova et al. (1989)
Y. lipolytica 704	Ethanol	STR, fed-batch	66.0	31.0	1:0.47	0.64	Finogenova et al. (1991)
Y. lipolytica VKM Y-2373	Rapeseed oil	STR, fed-batch	55.4	21.2	1:0.38	0.95	Kamzolova et al. (2007)
Y. lipolytica VKM Y-2373	Sunflower oil	STR, fed-batch	70.0	50.5	1:0.72	0.75	Kamzolova et al. (2008)
Yarrowia lipolytica EH59	Sunflower oil	STR, fed-batch	93.0	82.3	1: 0.88	0.65	Heretsch et al. (2008)

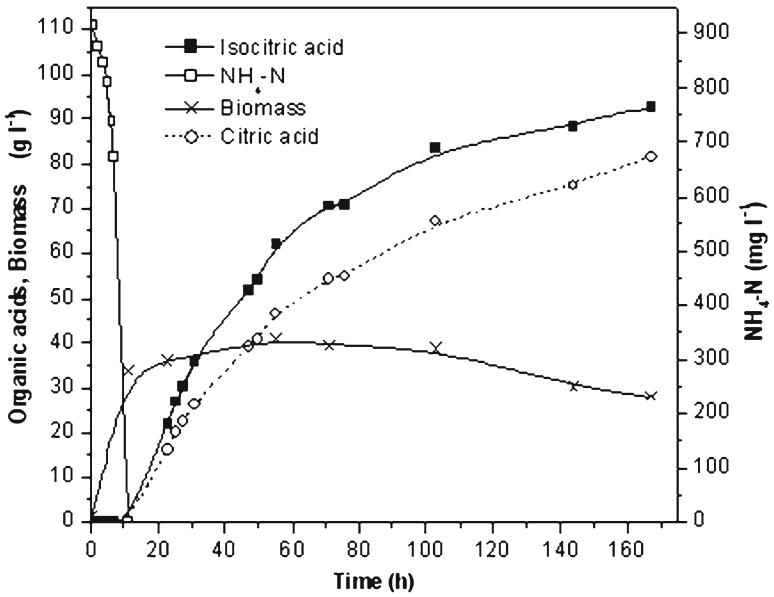


Fig. 19.5 Time course of ICA/CA formation from sunflower oil by *Yarrowia lipolytica* EH59: 15 L-STR, 30 °C, pH 6.0, adjusted with 19 N NaOH, pO_2-content 60%, stirrer speed 800 rpm, substrate: 85 g L^{-1} sunflower oil and 25 g L^{-1} after 40 h, 20 g L^{-1} after 70 h, 15 g L^{-1} after 100 h 3.0 g L^{-1} NH_4SO_4, 1000 μg L^{-1} thiamine

19.3.3.3 Product Isolation and Product Recovery

The yeast cells and residual sunflower oil were separated from the culture solution by cross-flow microfiltration with 0.2 μm pore size. The transformation of the ICA/CA sodium salts from the cell-free culture solution to their corresponding acids and sodium hydroxide was performed using laboratory electrodialysis equipment with bipolar membranes. Details are described in Sect. 19.5.3.

19.4 Metabolic Engineering of Organic Acid Biosynthesis

In the past breeding of industrial useful microbial strains was done mostly by mutagenesis of strains selected before from strain collections. The selected strain should exhibit features of interest for industrial production, like accumulation of certain organic acids. In some cases metabolite analogs were helpful for selection of wanted mutants, like citrate producing fluoroacetate resistant/sensitive mutants after exposure to toxic fluoroacetate. In other cases crosses of strains were done to produce diploid strains with certain features, for example yeast strains for production of

2-OGA. At present, these procedures are more and more replaced by directed remodeling of metabolic pathways by methods of metabolic engineering. There are an increasing number of microorganisms of which the complete genomes are sequenced and a high variability of genetic tools is available. Even an overproduction of the metabolite of industrial interest has not to be a characteristic of the selected basic microbial strain for breeding of a producer strain.

19.4.1 Strain Improvement for 2-Oxoglutaric Acid Production

Improvement of production of 2-OGA has been subject for several studies during last years. A review on microbial production of 2-OGA and metabolic engineering of this process has been published recently (Otto et al. 2011). Although secretion and extracellular accumulation of this organic acid by *Pseudomonas fluorescens* was firstly described in 1946 (Lockwood and Stodola 1946) and afterwards detected as a feature of several other bacteria metabolic engineering of production and secretion of 2-OGA was only done in *Corynebacterium glutamicum* (Verseck et al. 2007, 2009) among bacteria. However, inhibition of glutamate synthesis from 2-OGA by inactivation of glutamate dehydrogenase in *C. glutamicum* resulted only in a moderate increase of 2-OGA from 40 mg L^{-1} to 1.2 g L^{-1}.

In contrast to bacteria more investigations dealt with improvement of 2-OGA production by yeasts, especially *Candida glabrata* (syn. *Torulopsis glabrata*) and *Yarrowia lipolytica*. Wild type strains of *C. glabrata*, which need biotin and thiamine for growth, secretes high amounts of pyruvate (up to about 70 g L^{-1}) and low amounts of 2-OGA (about 1 g L^{-1}). Under certain cultivation conditions secreted amount of 2-OGA could be increased by simultaneous reduction of secreted pyruvate (Huang et al. 2006; Liu et al. 2007). In addition, genetic manipulations were also done to increase the formation of 2-OGA. The overexpression of one of the two acyl CoA synthase encoding genes of *Saccharomyces cerevisiae* resulted in an increase of 2-OGA secretion to about 18 g L^{-1} (Liang et al. 2008). Furthermore, enhancement of the activities of pyruvate dehydrogenase, pyruvate carboxylase or pyruvate decarboxylase by genetic manipulations caused an accumulation of 2-OGA of about 17, 22, and 32 g L^{-1}, respectively. The highest yield of 2-OGA of about 38 g L^{-1} was achieved by a combination of overexpression of the pyruvate decarboxylase encoding gene and the reduction of α-ketoglutarate dehydrogenase activity by addition of hydrogen peroxide (Zhang et al. 2009).

Extensive metabolic remodelling for improvement of 2-OGA production was done with the ascomycetous yeast *Y. lipolytica* during the last years. Several strains of this non-conventional yeast producing high amounts of 2-OGA from different carbon sources have been described (see Sect. 18.3.1; Otto et al. 2011). Also diploid strains were constructed and used for 2-OGA production achieving 185 g L^{-1} on paraffin as carbon source (Maldonado et al. 1973). The highest amounts, about 195 g L^{-1}, have been produced on *n*-alkanes by strain H222-27-11 (also called H355) (Weißbrodt et al. 1988). This hyper-producing *Y. lipolytica* strain was selected as a revertant of the methionine-auxotrophic mutant H222-27 (Weißbrodt et al. 1988).

However, the genetic background of the high productivity of this strain has not been clarified up to now.

In recent years a systematic metabolic engineering was done for improvement of the production of 2-OGA by *Y. lipolytica* (Holz 2011; Otto 2010; Otto et al. 2011, 2012; Yovkova 2011). They have used a uracil-auxotrophic derivative of strain H222-27-11 in most cases as recipient strain for overexpression of enzymes of the TCC by multi-copy insertion of the encoding genes alone or in combination into the genome of this strain. The chosen enzymes and their encoding genes were pyruvate carboxylase (PYC, gene *PYC1)*, NAD^+-dependent isocitrate dehydrogenase subunit 1 (IDH1, gene *IDH1*), NAD^+-dependent isocitrate dehydrogenase subunit 2 (IDH2, gene *IDH2*), $NADP^+$-dependent isocitrate dehydrogenase (IDP, gene *IDP1*), 2-OGA dehydrogenase (KGDH, genes *KGD1*, *KGD2*, *LPD1*), fumarase (FUM, gene *FUM1*), malate dehydrogenase 1 (MDH1, gene *MDH1*) and malate dehydrogenase 2 (MDH2, gene *MDH2*). The resulting transformants were cultivated in 2-OGA producing cultivation media with pharmaceutical grade glycerol or raw glycerol from raps oil refineries as sole carbon source and the spectra of produced organic acids (2-OGA, citric acid (CA), isocitric acid (ICA), succinic acid (SUC), fumaric acid (FUM), malic acid (MAL), pyruvic acid (PYR)) were determined by ion chromatography. The aims of these studies were the improvement of 2-OGA production and the reduction of amounts of other organic acids as by-products. The most important results are summarized presented in the following part. Overexpression of the gene *IDP1* in combination with overexpression of *PYC1, FUM1* or *KGD2* had the strongest effects on increasing of the amount of secreted 2-OGA. About 19–27% more 2-OGA in comparison with strain H222-27-11 have been produced in bioreactor with raw glycerol as carbon source resulting in the highest yield of accumulated 2-OGA of about 186 g L^{-1} after 98 h of cultivation. Overexpression of the gene *FUM1* alone resulted in strong reduction of accumulated amounts of MAL, FUM, and SUC. The highest selectivity of about 98% for 2-OGA was achieved by overexpression of *PYC1* and *IDP1* together (Otto 2010; Yovkova 2011).

19.4.2 Strain Improvement for 2-Oxo-D-Gluconic Acid Production

In contrast to 2-OGA production nothing has been done to improve overproduction of 2-OGcA by remodelling of the metabolic pathway, in spite of the complete genome of the best producer *G. oxydans* is sequenced and the enzyme for synthesis of this compound is known. The sequence of the genome of *G. oxydans* has been published in 2005 by Prust et al. Furthermore, Gupta et al. (1999) have shown that in mutants of *G. oxydans* inhibited in flavin biosynthesis no activity of the flavin-dependent gluconat-2-dehydrogenase (GA-2-DH) was detectable and therefore no 2-OGcA and 2, 5-OGcA was accumulated. The final proof that the enzyme GA-2-DH is responsible for formation of 2-OGcA was done by Elfari et al. (2005). They have shown that inactivation of the GA-2-DH encoding gene resulted in a lack of accumulation of 2-OGcA. The genes encoding enzymes responsible for formation of unwanted by-products 2, 5-OGcA and 5-OGcA in *G. oxydans* are also known (Hölscher et al.

2009). Therefore, construction of strains of *G. oxydans* producing high amounts of 2-OGcA without the accumulation of the unwanted by-products 2, 5-OGcA and 5-OGcA seems to be possible.

19.4.3 Strain Improvement for Production of (2R, 3S)-Isocitric Acid

There are already several fungal species known secreting high amount of (*2R*,*3S*)-isocitric acid as shown in Table 19.5 (Sect. 19.3.3). A problem was the production of CA occurring in most cases as unwanted by-product in relatively high amounts. Several mutants, especially from the yeast *Y. lipolytica*, were already selected (Table 19.5), which produce increased amounts of ICA or reduced amounts of the by-product CA. The first mutants were selected after classical random mutagenesis and their analysis revealed that either reduced activity of isocitrate lyase (ICL) (Finogenova et al. 1986) or increased activity of aconitase (ACO) (Akiyama et al. 1973; Ermakova et al. 1986; Finogenova et al. 1986) caused the increased formation of ICA and reduced accumulation of CA. The first directed approaches for improvement of ICA accumulation and reduction of CA formation were done by Förster et al. (2007) and Holz et al. (2009). Förster et al. (2007) have deleted the ICL encoding gene *ICL1* in the *Y. lipolytica* ICA/CA producing strain H222-S4 and observed only moderate increase of 2–5% of the ICA proportion in comparison with the original strain. Holz et al. (2009) increased the activity of ACO by insertion of multiple copies of the ACO encoding gene *ACO1* in the genome of the same strain and observed an increased proportion of ICA from about 40 to about 70% on sunflower oil but only a moderate increase from about 11 to about 15% on glucose, sucrose or glycerol. These data suggest that the production of ICA may be further increased by directed remodelling of the involved metabolic pathways.

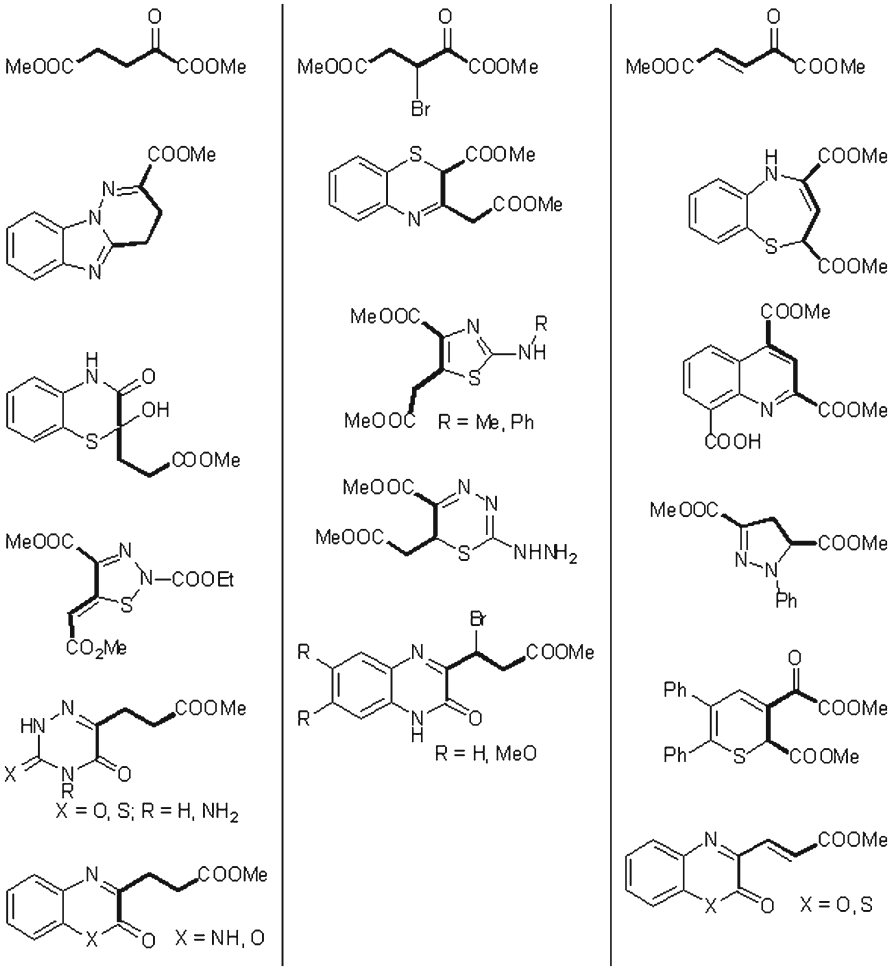
19.5 Synthetic Use of Microbiologically Produced Carboxylic Acids

Three of the carboxylic acids made in our laboratories by cultivation serve as starting materials for organic syntheses, during which their value is raised. We intend to show that many different possibilities for reaction exist as soon as the building block is made readily accessible. For this question it is not of concern if the compound is achiral, like 2-OGA, or chiral, like 2-OGcA and ICA. However, ICA has more or less gone unnoticed by most scientists and media. It is well known as a member of the TCA cycle but apart from that it is not commercially available for syntheses nor has it been frequently used (Internetchemistry 2008; Medicalnewstoday 2008; Physorg 2012).

Throughout our work, we are more than ever convinced that an interdisciplinary cooperation between microbiologists and chemists can open new doors for others, like medicinal chemists, biochemists and pharmacists.

19.5.1 2-Oxoglutaric Acid as Building Block for Heterocycles

This oxo dicarboxylic acid has a larger potential for the synthesis of various types of heterocycles than one may think at first glance. Additionally, it is a suitable precursor to form related reactive derivatives like valuable glutaconic acid esters. Scheme 19.1 gives an overview on different heterocycles synthesised in our group, starting from dimethyl 2-oxoglutarate, dimethyl 3-bromo-2-oxo-glutarate and dimethyl (*E*)-2-glutaconate. Examples are drawn in three columns with the corresponding precursor compound in the first line.



Scheme 19.1 Various heterocyclic skeletons derived from dimethyl 2-oxoglutarate, dimethyl 3-bromo-2-oxoglutarate and dimethyl (*E*)-2-glutaconate

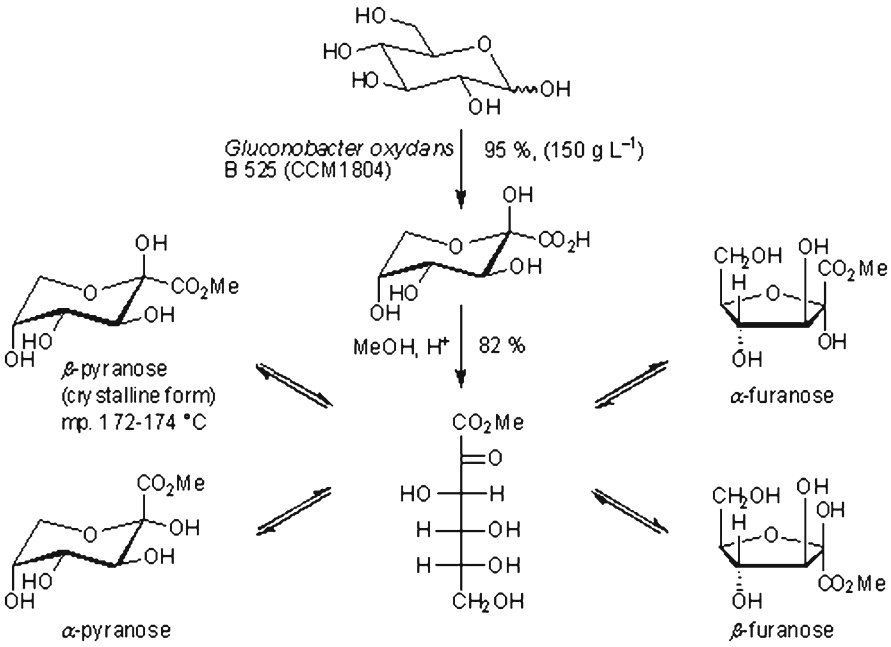
Fragments of the initial C_5 chain are in many cases incorporated via pathways other than the traditional manner (incorporate mainly the carbons C-1 and C-2 by cyclocondensation). In each formula drawn, the C_5 unit is highlighted in bold. In some cases it is possible to start with the precipitation of a heterocycle precursor directly from the cultivation broth, e.g., in case of hydrazones. If this is possible, one can go without expensive workup of 2-OGA from the aqueous solution. In such a case, the requirements of a so called "direct product recovery" or "direct capture of product" are fulfilled. More than 70 novel compounds were prepared and are described in several papers (Blitzke et al. 1993, 1994, 1997; Hartenstein et al. 1993; Khodja et al. 1994; Schwesinger et al. 1992).

19.5.2 2-Oxo-D-Gluconic Acid as Building Block

2-OGcA (D-*arabino*-2-hexulosonic acid), a member of the chiral pool, is a similar suitable building block as 2-OGA. Interestingly, this 2-oxo-aldonic acid is commercially not available and the authors will appreciate any comments about this fact or inquiries, since there is a kg-stock of pure methyl 2-oxo-D-gluconate (2-OGcMe) available. Before our investigations, heterocycles based on 2-OGcA have not been reported, probably due to the fact that just the hemicalcium salt was sold, but only in analytical amounts, a circumstance that we were confronted with in the Sect. 19.5.3 about ICA, too.

What can be expected from incorporating this acid into heterocycles? The product will be chiral and will contain portions of highly desired structures, derived from the chiral pool. This may be important for studying the interaction with biological receptors. From the physicochemical point of view, it will also be equipped with hydrophilicity due to the polyol substitution. Although the biotechnological formation is relatively simple, the isolation of 2-OGcA from the cultivation broth is a challenge. Complete removal of water from the biomass-free and electrodialysed cultivation solution leads to free 2-OGcA, but only as syrup, which complicates its further use as free acid. However, after reflux in methanol, a solution of the methyl ester is obtained, which crystallizes in the β-pyranose configuration in 82% yield. This biotechnological isolation technique is much better than the traditional chemical synthesis starting from D-fructose. Although the latter one is optimized, it involves huge amounts of potassium permanganate, leaves large amounts of chemical waste and gives a lower product yield (Kirrbach et al. 1992). As typical for many saccharides, 2-OGcA also forms multiple structures in solution.

In Scheme 19.2 the mutarotation process is shown. The β-pyranose is the favoured isomer (analogously for the free acid). This equilibrium has already been investigated earlier on a pure synthetic sample (Crawford et al. 1980). The crystalline methyl ester can be easily stored under normal conditions and was the precursor for a variety of reactions described later.

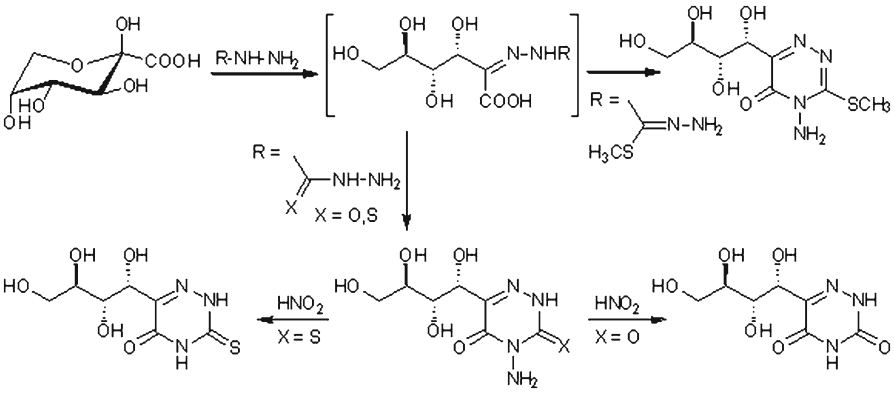


Scheme 19.2 Mutarotation of methyl 2-oxo-D-gluconate formed by esterification

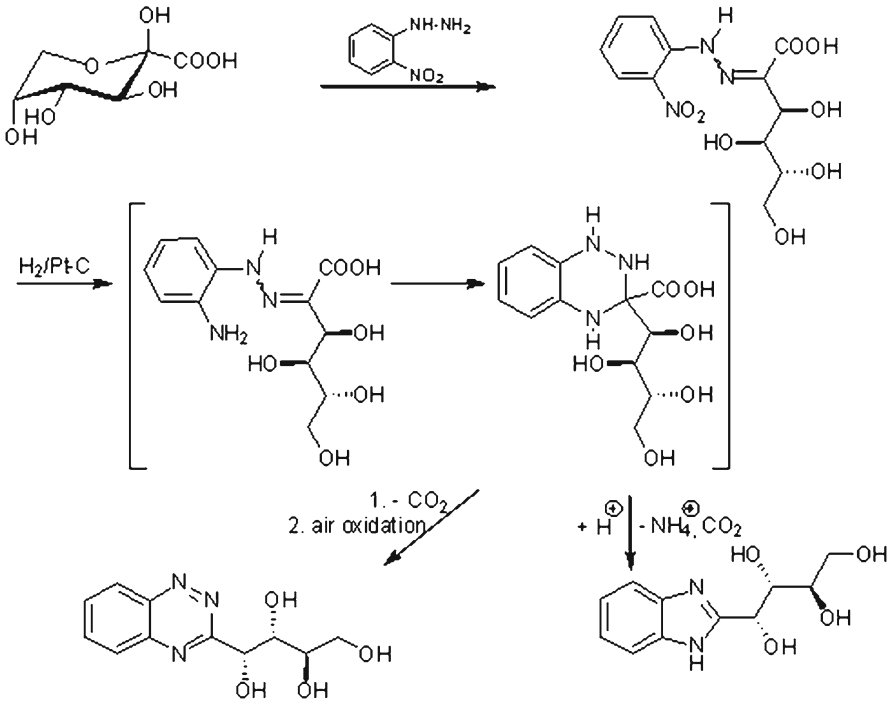
However, certain reagents are also suitable for a direct reaction with the aqueous electrodialyzed cultivation solution containing free 2-OGcA. For instance, substituted hydrazines can react to form hydrazones that precipitate, and can be further processed after filtration. In such a case it is advantageous, that the water of the cultivation process has not to be removed by distillation as in the synthesis of the methyl ester.

We will now summarize some possibilities for the synthesis of saccharide-heterocycle combinations in the following schemes. Note, that the saccharidic part is processed or modified in rather different ways. It can be found attached to heterocycles as an open chain tetritolyl unit. Also, a traditional cyclic sugar part is possible, but attached in a spiro junction and not glycosidic. Finally, the possibility for a hetero analogous modification of the cyclic 2-OGcA into a certain kind of sugar amino acid has been investigated. The anomeric C-atom of the sugar moiety resembles the α-C-atom of an amino acid, in this case. Currently, the synthesis of sugar tensides based upon this oxo acid is under investigation.

Utilizing the original cultivation solution, 2-OGcA is reacted with carbohydrazide, thiocarbohydrazide or *S*-methyl thiocarbohydrazide to yield the corresponding carbo- or thiocarbohydrazones. Their cyclocondensation gives rise to novel 1,2,4-triazinones with a hydrophilic and chiral side chain (Scheme 19.3) (Andersch 1998).

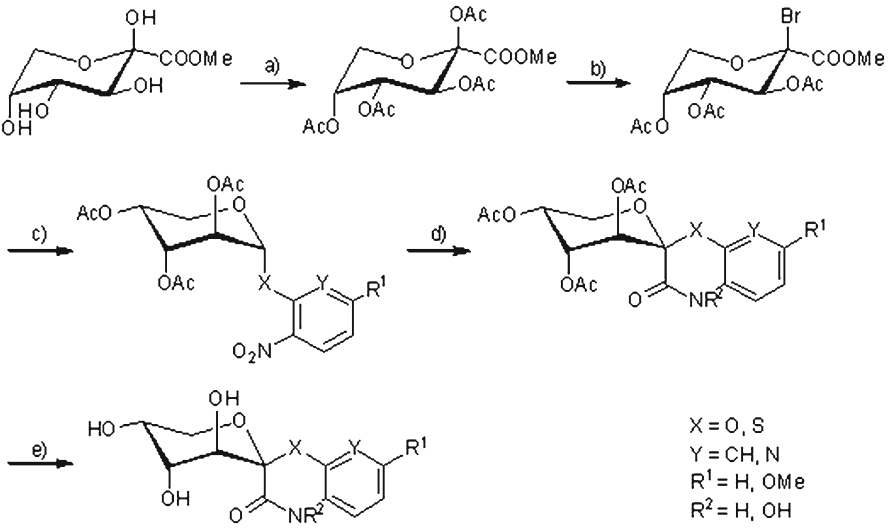


Scheme 19.3 Formation of open chain 6-(D-*arabino*-tetritol-1-yl)-1,2,4-triazine derivatives



Scheme 19.4 Reductive cyclization of substituted 2-nitrophenyl hydrazones to form saccharidic 1,2,4-benzotriazines and benzimidazoles

Similarly, a cultivation broth of 2-OGcA (or alternatively crystalline 2-OGcMe) can be used to prepare functionalized 1,2,4-benzotriazines and benzimidazoles (Scheme 19.4) (Andersch and Sicker 1999). In the first step, 2-nitrophenylhydrazones are precipitated. Depending on the further treatment they react in a sequence of reductive cyclization and elimination/oxidation to form the former or latter of the



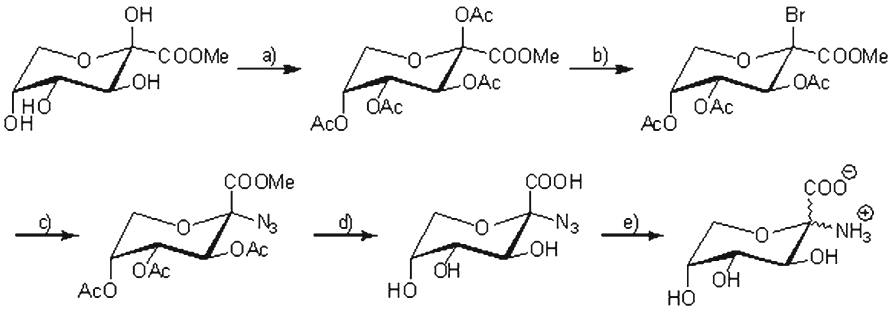
Scheme 19.5 Transformation of methyl 2-oxo-D-gluconate into spiro-connected saccharidic heterocycles by protection, anomeric activation, nucleophilic substitution, reductive cyclisation and deprotection. Reagents and conditions: a) Ac_2O, cat. H_2SO_4, 60 °C, 15 min; b) HBr/AcOH; c) K_2CO_3, acetone,  ; d) H_2/Pt-C in MeOH or pyridine, or Zn/NH_4Cl; e) NaOMe/MeOH, Amberlite IR 120

above heterocycles. (Note, for 1,2,4-triazinones, use of protecting group chemistry is not necessary.) Such 1,2,4-benzotriazines have been claimed as potential anticancer agents (Stottmeister et al. 2001).

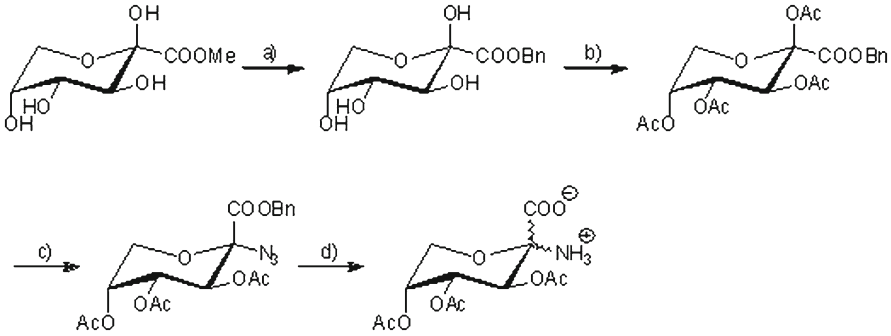
Methyl 2-oxo-D-gluconate is also a suitable precursor for novel spiro-connected heterocycles. In this case, the use of protecting groups is necessary. The first intermediate compound is the β-pyranoid tetraacetate of the methyl ester (Scheme 19.5), which is accessible in 82% under the conditions given (a). Nucleophilic substitution of the 2-acetoxy group with a solution of 33% hydrogen bromide in acetic acid (b) yields 98% of the corresponding glycosyl bromide. This central reactive building block was subjected to nucleophilic substitution reactions by various nucleophiles derived by deprotonation from (substituted) 2-nitrophenols, 2-nitrothiophenols, 2-aminothiophenol or 2-hydroxy-3-nitropyridins. This reaction is *trans*-diastereoselective and leads to protected *ortho*-nitro-aryl-*O*- and *S*-glycosides as glycosidic precursors for heterocyclizations. Such glycosides can be reductively cyclized by catalytic hydrogenation (with influence of the solvent) or zinc dust to yield cyclic hydroxamic acids or lactams. They belong to the following novel spiro connected skeletons: spiro[pyrido[3,2-*b*]oxazin-2,2'-pyran], spiro[1,4-benzothiazin-2,2'-pyran], and spiro[pyrido[3,2-b][1,4]-oxazin-2,2'-pyran] and spiro[1,4-benzoxazine-2,2'-pyrans] (Andersch et al. 1999a, b, c, d).

Finally, we tried to transform 2-OGcA into the heteroanalogue 2-amino-D-*arabino*-hex-2-ulopyranosonic acid by chemical syntheses.

The feature of this compound is that it is a hybrid of a sugar and an α-amino acid with its anomeric centre being also the α-C-atom of the amino acid. Whereas the



Scheme 19.6 Synthesis of 2-amino-D-*arabino*-hex-2-ulopyranosonic acid. Reagents and conditions: a) Ac_2O, cat. H_2SO_4, 60 °C; b) HBr/AcOH; c) NaN_3, DMSO; d) K_2CO_3, MeOH/H_2O, e) H_2/Pd-C

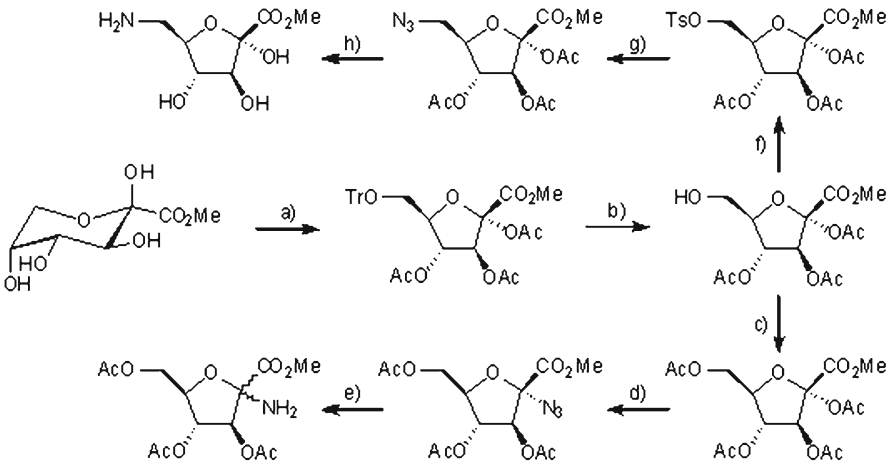


Scheme 19.7 Synthesis of acetyl protected 2-amino-D-*arabino*-hex-2-ulopyranosonic acid. Reagents and conditions: a) BnOH, Amberlite IR 120, 80 °C; b) Ac_2O, AcOH, H_2SO_4, 50 °C; c) $TMSN_3$, $SnCl_4$, $AgClO_4$, RT; d) H_2/Pd-C

heteroanalogue mentioned above is a pyranose derivative, in principle furanose derivatives are conceivable as well when looking at the structural isomerism of methyl 2-oxo-D-gluconate in solution (compare Scheme 19.2). Both goals have been achieved, as shown by the transformations depicted in Schemes 19.6, 19.7, 19.8.

Sugar amino acids as the pyranoid one are rare examples that unite basic properties of two structural worlds, carbohydrates and amino acids in a small molecule (Andersch et al. 2000; Schmidt et al. 2005a, b).

A sequence of trityl protection of the primary alcohol followed by acetyl protection of the secondary alcohols gave rise to a crystalline α-furanose derivative (see Scheme 19.8, middle and bottom line). The 2,3,4-tri-*O*-acetyl-α-D-*arabino*-2-hexulofuranosonic acid methyl ester (middle, right) proved to be a central building block from which two kinds of furanoid amino acid esters could be derived, an α-amino acid ester (lower, left) and an ω-amino acid ester (upper, left).



Scheme 19.8 Syntheses for a furanoid sugar α- and ω-amino acid ester based upon 2-OGcA. Reagents and conditions: a) TrCl, DMAP, Ac_2O, pyridine; b) HCl, CH_2Cl_2; c) Ac_2O, pyridine; d) $TMSN_3$, $SnCl_4$, CH_2Cl_2; e) H_2/PdC; f) TsCl, pyridine; g) NaN_3, DMF; h) i) NaOMe/MeOH, Amberlite IR 120; ii) H_2/Pd-C

*19.5.3 (2*R,3S*)-Isocitric Acid*

This work is based on findings which were made during the production of ICA by cultivation and using this compound in organic syntheses as recently published (Heretsch et al. 2008).

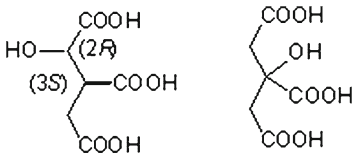
The TCA cycle can be regarded as the biochemical pathway that is of paramount importance to all aerobic organisms. Since it evolved about three billion years ago as a direct consequence of a significant rise in atmospheric oxygen levels, life on earth made an unprecedented evolutionary quantum leap.

Hence, it is not surprising, that all of the eight intermediates in the TCA cycle earned a lot of attention in modern life sciences. Although a naive approach may be to assume the TCA cycle to be a finished chapter in the book of science, a closer look reveals this statement to be misleading. One member managed to escape at least the chemists' access until today: ICA. It is formed by action of the enzyme aconitase from CA via the intermediate *cis*-aconitate (*cis*-aconitic acid [499-12-7]).

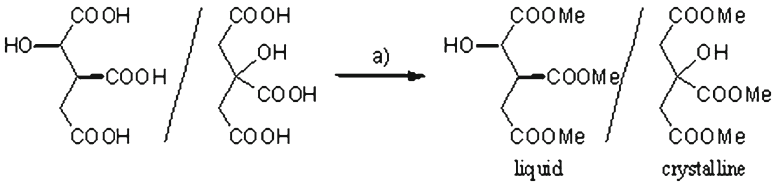
ICA, a chiral α-hydroxy tricarboxylic acid, resembles its well known (and biotechnologically easily accessible) relative CA to such a point that all attempts of isolating ICA from mixtures on a scale that would allow preparative chemical manipulations have proved unsuccessful.

A closer look to the chemical structures provides both disillusionment and hope (Scheme 19.9).

ICA and CA differ if at all only slightly in their physicochemical properties such as molar mass, solubility in a variety of solvents and polarity. All these properties are directly connected to standard purification methods (distillation, extraction and



Scheme 19.9 Structures of (2*R*,3*S*)-isocitric- and citric acid



Scheme 19.10 Esterification of ICA and CA as a separation principle. Reagents and conditions: a) MeOH (excess), 2,2-dimethoxypropane (excess), TMSCl (cat.)

chromatography) – and therefore do not provide any opportunities of separating these chemical species.

In a structurally based approach the hydroxy group attached to the 2-position (ICA) or 3-position (CA) can provide a valuable handle for a separation principle based on a chemical modification. Reducing the polarity generated by three carboxylic acid moieties can be achieved by both, esterification with another alcohol, or intramolecular esterification (lactonisation) with the free hydroxy group the molecule possesses.

The exhaustive esterification of a 1:1 mixture of both acids with an excess of methanol provided the trimethyl esters as a highly viscous oil, from which only the trimethyl citrate precipitated as needles upon standing a prolonged time (3 days) at 4 °C. Then simple filtration gave nearly pure trimethyl isocitrate as the oily supernatant with yields varying from 80 to 88% (with regard to initial ICA concentration in the cultivation broth) (Scheme 19.10).

Although property differences like this are not predictable in general, one can gain some basic insights into these kind of separation based on fractional crystallisation. When carboxylic acids in a molecule have to be transformed to their esters and a crystalline product is desired, one should avoid more complex alcohols than methanol since these alcohols will form esters with a lower melting point than the corresponding methyl ester. A simple example can be gained from the known esters of enantiopure tartaric acid (best known in synthetic chemistry for their use as chiral precatalysts in the Sharpless asymmetric apoxidation procedure (Katsuki and Sharpless 1980)). While tartaric acid dimethyl ester is a crystalline solid (melting point 48–49 °C) its higher analogues tartaric acid diethyl ester and diisopropyl ester are highly viscous oils.

The difference in crystallisation behaviour of trimethyl citrate and trimethyl isocitrate can also be rationalized. The before mentioned position of the hydroxy group makes all difference between CA being a symmetric and therefore achiral molecule, and ICA having a much lower degree of symmetry and therefore being a chiral compound. Since the process of crystallisation is the spatial arrangement of molecules into a crystal lattice of the highest possible symmetry to gain the highest possible amount of stabilisation, which is directly connected to the crystallisation enthalpy, it is not surprising, that a highly symmetrical species like trimethyl citrate easily arranges into a lattice while trimethyl isocitrate does not.

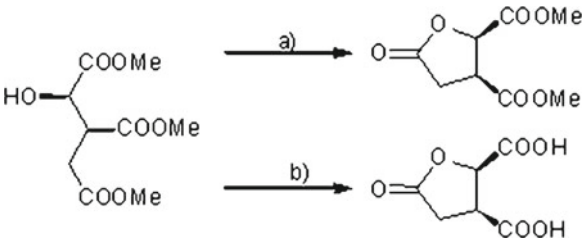
For the first time this separation/purification process provides large quantities of enantiopure trimethyl isocitrate which can now be used in subsequent chemical transformations to raise complexity, access new functionality and finally to provide new chiral entities for the ongoing fields of total synthesis and pharmaceutical research (Manufacturingchemist 2008). A closer consideration of ICA with the eyes of a synthetic chemist uncovers both uniqueness and deficiencies of its chemical structure. This highly oxygenated molecule provides increased functionality in a very close spatial arrangement with two neighbouring stereo centers.

Standard laboratory techniques still suffer from their problem processing compounds of high polarity due to extensive loss of material in aqueous extractions and normal phase chromatography or solubility problems in standard organic solvents. Therefore the transformation of ICA into derivatives of lower polarity is a highly desirable aim.

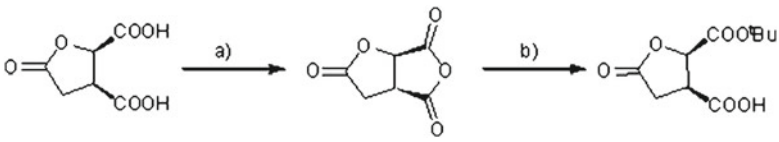
Three carboxylic acid moieties, having only slightly different chemical reactivity profiles, are a problem, if regioselective manipulations are desired. Therefore a chemical way to distinguish between these functionalities should be found first in a chemical transformation approach. All these aims were combined in our research.

On the way to three distinguishable carboxylic acid moieties, first a lactonisation was carried out to yield the five membered lactone. Dependent on the reaction conditions applied in this transformation, one can synthesize both, the lactone dicarboxylic acid or the lactone dimethyl ester, which offer different further processing (see Schemes 19.11, 19.12 and 19.14).

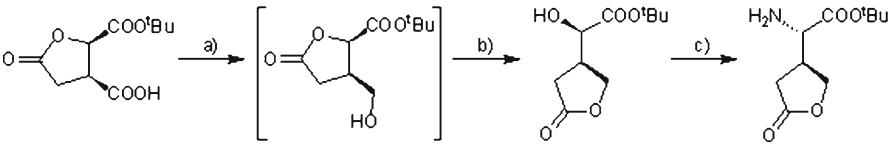
The lactone dicarboxylic acid yielded the lactone anhydride upon treatment with acetic anhydride at elevated temperatures (160 °C), which is itself a readily



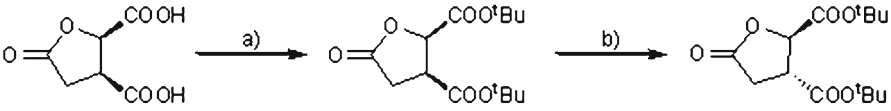
Scheme 19.11 Lactone formations from trimethyl isocitrate. Reagents and conditions: a) toluene, *p*TsOH (cat.), reflux (Dean-Stark); b) HCl (4 M), reflux



Scheme 19.12 Further processing towards three distinguishable carboxylic acid moieties. Reagents and conditions: a) Ac_2O, 160 °C; b) *tert*-BuOH, reflux



Scheme 19.13 Synthesis of a not naturally occurring lactone amino acid. Reagents and conditions: a) $BH_3 \cdot THF$, 0 °C – RT; b) spontaneous; c) i) MsCl, pyridine, CH_2Cl_2; ii) NaN_3, DMF; iii) H_2/Pd-C



Scheme 19.14 Inversion of stereo chemistry at C-3 position of lactone di-*tert.*-butyl ester. Reagents and conditions: a) *iso*-butene, H_2SO_4 (cat.), CH_2Cl_2; b) DBU, CH_2Cl_2, reflux

crystallising substance. It could therefore be used to raise the purity to the point where no contamination with residual CA derivatives is detectable. Since the cyclic anhydride is highly reactive towards any kind of nucleophile, strict exclusion of moisture is necessary in this transformation. Reactions with a variety of hindered and unhindered oxygen or nitrogen based nucleophiles suggested that the anhydride opening proceeded with complete regioselectivity to yield only the derivatives modified in C-2-position and a free acid in C-3-position. When using *tert*-butanol as nucleophilic species an ICA derivative was obtained that had one of its initial carboxylic acids bound in the lactone, another as a *tert*-butyl ester (presenting an orthogonal protective group with regard to the lactone) and the third still present in the free carboxylic acid in C-3-position. Hence the initial aim of three distinguishable carboxylic acid moieties was reached without long and tedious reaction sequences.

This newly created species proved to be of high value due to its easy transformation into nitrogen containing molecules. The lactone amino acid does not occur naturally and was generated by reduction of the free carboxylic acid with borane tetrahydrofurane complex, followed by the spontaneous rearrangement of the

primary alcohol into the secondary alcohol presumably due to release of ring strain generated by two adjacent substituents in a *syn* fashion in the former one. Nucleophilic substitution with azide after transforming the alcohol into a mesylate as a leaving group yielded the azido ester, which was finally hydrogenated to yield the free amino functionality (Scheme 19.13).

Finally, the possibility of creating other stereo isomers of ICA was investigated. Since ICA has two stereo centers it can occur as one of four stereoisomers. By inverting the stereochemistry of one of these two stereocenters, a diastereoisomer of naturally occurring ICA is synthesized commonly termed as *allo*-isocitric acid.

This task could be achieved by thermodynamically driven epimerisation of the C-3 stereocenter of the lactone di-*tert*-butyl ester (itself available in one step from lactone diacid) under carefully controlled alkaline conditions (Scheme 19.14).

In conclusion, the search for new chiral building blocks for organic and pharmaceutical chemistry does not always need to be carried out at remote sites of metabolism. Sometimes even in the very heart of aerobic metabolism gold can be mined. As shown in this last case study the building block ICA is not only of practical value because of its use in several further transformations, but also presents a capstone in the chemical reconstruction of TCA cycle.

19.6 Concluding Remarks

With 2-OGA, 2-OGcA and ICA, three compounds were microbiologically produced in large quantities (kilogram scale) and used in chemical syntheses. With these building blocks in hand, we were able to synthesize a large number of novel compounds with different structures and promising new applications. All of these compounds were produced with achieving environmental efforts like minimization of waste production, energy demand, process cycles and use of renewable resources. In case of 2-OGA and ICA the application of plant oils lead to a successful outcome.

Furthermore, some new steps and ideas should be mentioned:

- use of the “multi purpose” equipment for cultivation and product recovery for organic acid production;
- use of stable microorganisms (“work horses”) from strain collections;
- phenotypic and genotypic optimization of the cultivation process up to product concentrations and selectivities, which are of interest for processing in technical scale, development of new steps of product isolation, realising e.g. the idea of “direct product recovery”.

From the point of view of organic chemistry, the breakthrough of this approach consists in making ICA available on kilogram scale for the first time. This compound is a most promising chiral building block for bioactive molecules.

References

Akiyama S, Suzuki T, Sumino Y, Nakao Y, Fukuda H (1973) Induction and citric acid productivity of fluoroacetate-sensitive mutant strains of *Candida lipolytica*. Agr Biol Chem 37:879–888

Ammon K (1996) Bioprozessentwicklung am Beispiel der enzymatischen Vorstufensynthese von Vitamin C. PhD Thesis, Universität Dortmund, Dortmund

Andersch J (1998) On the synthesis of *N*-heterocycles starting from biotechnologically obtained 2-Oxo-D-gluconic acid. PhD Thesis, Universität Leipzig, Leipzig

Andersch J, Sicker D (1999) Reductive cyclization of carbohydrate 2-nitrophenyl-hydrazones to chiral functionalized 1,2,4-benzotriazines and benzimidazoles. J Heterocycl Chem 36:589–594

Andersch J, Sicker D, Wilde H (1999a) Methyl D-arabino-hex-2-ulopyranosonate as a building block for spiro[1,4-benzoxazine-2,2'-pyrans]. Carbohydr Res 316:85–94

Andersch J, Sicker D, Wilde H (1999b) Synthesis of spiro[pyrido[3,2-b][1,4]oxazin-2,2'-pyrans] based upon methyl D-arabino-2-hexulopyranosonate. Tetrahedron Lett 40:57–58

Andersch J, Sicker D, Wilde H (1999c) Synthesis of spiro[1,4-benzothiazin-2,2'-pyrans] starting from methyl D-arabino-2-hexulosonate. J Heterocycl Chem 36:457–460

Andersch J, Sicker D, Wilde H (1999d) Synthesis of spiro[pyrido[3,2-b][1,4]oxazin-2,2'-pyrans] based upon methyl D-arabino-2-hexulosonate. J Heterocycl Chem 40:57–58

Andersch J, Hennig L, Wilde H (2000) *N*-Glycosidation of D-arabino-hex-2-ulosonic acid. Carbohydr Res 329:693–697

Blitzke T, Hartenstein H, Sicker D, Wilde H (1993) Synthesen auf der Basis von 2-Oxoglutarsäure. III. Synthese und Reaktionen von (*E*)-2-Oxoglutaconsäuredimethylester. J Prakt Chem 335(8):683–686

Blitzke T, Greif D, Kempe R, Pink M, Pulst M, Sicker D, Wilde H. (1994) Synthesen auf der Basis von 2-Oxoglutarsäure. IV. Regioselektive Synthese substituierter2-(2-Methoxycarbonyl-2*H*-thiopyran-3-yl)glyoxylsäuremethylester. J Prakt Chem 336: 163–165

Blitzke T, Sicker D, Wilde H (1997) Organic syntheses based on 2-oxoglutaric acid. V. Syntheses of novel 2*H*-1,4-benzothiazines and a 2,5-dihydro-1,5-benzothiazepine. J Heterocycl Chem 34(2):453–455

Chernyavskaya OG, Shishkanova NV, Il'chenko AP, Finogenova TV (2000) Synthesis of α-ketoglutaric acid by *Yarrowia lipolytica* yeast grown on ethanol. Appl Microbiol Biotechnol 53(2):152–158

Crawford TC, Andrews GC, Faubl H, Chmurny GN (1980) The structure of biologically important carbohydrates. A carbon-13 nuclear magnetic resonance study of tautomeric equilibriums in several hexulosonic acids and related compounds. J Am Chem Soc 102(7):2220–2225

Elfari M, Ha SW, Bremus C, Merfort M, Khodaverdi V, Herrmann U, Sahm H, Görisch H (2005) A *Gluconobacter oxydans* mutant converting glucose almost quantitatively to 5-keto-D-gluconic acid. Appl Microbiol Biotechnol 66:668–674

Ermakova IT, Shishkanova NV, Melnikova OF, Finogenova TV (1986) Properties of *Candida lipolytica* mutants with the modified glyoxylate cycle and their ability to produce citric and isocitric acid. I. Physiological, biochemical and cytological characteristics of mutants grown on glucose or hexadecane. Appl Microbiol Biotechnol 23(5):372–377

Finogenova TV, Shishkanova NV, Ermakova IT, Kataeva IA (1986) Properties of *Candida lipolytica* mutants with the modified glyoxylate cycle and their ability to produce citric and isocitric acid. II. Synthesis of citric and isocitric acid by Candida lipolytica mutants and peculiarities of their enzyme systems. Appl Microbiol Biotechnol 23:378–383

Finogenova TV, Shishkanova NV, Illarionowa WI, Losinov AB, Karklin RY, Peltzmane IZ, Jeschov VA (1989) Strain *Candida lipolytica* HMM-149 as producer of isocitric acid. USSR Inventor's Certificate SU915466, dated 23.06.1989

Finogenova TV, Shishkanova NV, Fausek EA, Eremina SS (1991) Biosynthesis of isocitric acid from ethanol by yeasts. Appl Microbiol Biotechnol 36:231–235

Finogenova TV, Kamzolova SV, Dedyukhina EG, Shishkanova NV, Il'chenko AP, Morgunov IG, Chernyavskaya OG, Sokolov AP (2002) Biosynthesis of citric and isocitric acids from ethanol by mutant *Yarrowia lipolytica* N 1 under continuous cultivation. Appl Microbiol Biotechnol 59(4–5):493–500

Förster A, Jacobs K, Juretzek T, Mauersberger S, Barth G (2007) Overexpression of the ICL1 gene changes the product ratio of citric acid production by *Yarrowia lipolytica*. Appl Microbiol Biotechnol 77:861–869

Gupta A, Felder M, Verma V, Cullum J, Qazi GN (1999) A mutant of *Gluconobacter oxydans* deficient in gluconic acid dehydrogenase. FEMS Microbiol Lett 179:501–506

Hartenstein H, Blitzke D, Sicker D, Wilde H (1993) Synthesen auf der Basis von 2-Oxoglutarsäure. II. Zur Synthese von Heterocyclen durch Reaktionen von 3-Brom-2-oxoglutarsäuredimethylester mit Binucleophilen. J Prakt Chem 335(2):176–180

Heretsch P, Thomas F, Aurich A, Krautscheid H, Sicker D, Giannis A (2008) Syntheses with a chiral building block from the citric acid cycle: (2*R*,3*S*)-isocitric acid by fermentation of sunflower oil. Angew Chem Int Ed 47:1958–1960

Hölscher T, Schleyer U, Merfort M, Bringer-Meyer S, Görisch H, Sahm H (2009) Glucose oxidation and PQQ-dependent dehydrogenases in *Gluconobacter oxydans*. J Mol Microbiol Biotechnol 176:6–13

Holz M (2011) Gentechnische Optimierung der Hefe *Yarrowia lipolytica* zur biotechnologischen Produktion von Succinat. PhD Thesis, Dresden University of Technology, Dresden

Holz M, Otto C, Kretzschmar A, Yovkova V, Aurich A, Pötter M, Marx A, Barth G (2009) Overexpression of alpha-ketoglutarate dehydrogenase in *Yarrowia lipolytica* and its effect on production of organic acids. Appl Microbiol Biotechnol 89:1519–1526

Huang HJ, Liu LM, Li Y, Du GC, Chen J (2006) Redirecting carbon flux in *Torulopsis glabrata* from pyruvate to alpha-ketoglutaric acid by changing metabolic co-factors. Biotechnol Lett 28:95–98

Huang C, Xu T, Zhang Y, Xue Y, Chen G (2007) Application of electrodialysis to the production of organic acids: state-of-the-art and recent developments. J Memb Sci 288(1–2):1–12

Il'chenko AP, Chernyavskaya OG, Shishkanova NV, Finogenova TV (2003) Biochemical characterization of the yeast *Yarrowia lipolytica* overproducing carboxylic acids from ethanol: nitrogen metabolism enzymes. Microbiol 72(4):418–422

Internetchemistry (2008) Kilogram quantities at last! http://www.internetchemie.info/news/2008/feb08/isoctric-acid.html. Accessed 20 Feb 2012

Kamzolova SV, Finogenova TV, Lunina YN, Perevoznikova OA, Minachova LN, Morgunov IG (2007) Characteristics of the growth on rapeseed oil and synthesis of citric and isocitric acids by *Yarrowia lipolytica* yeasts. Microbiol 76:20–24

Kamzolova SV, Finogenova TV, Morgunov IG (2008) Microbiological production of citric and isocitric acids from sunflower oil. Food Technol Biotechnol 46:51–59

Katsuki T, Sharpless KB (1980) The first practical method for asymmetric epoxidation. J Am Chem Soc 102(18):5974–5976

Khodja M, Schwesinger H, Sicker D, Wilde H (1994) Synthesis of 3,4-dihydro-pyridazino[1,6-*a*] benzimidazoles by reductive cyclization of dialkyl 2-(2-nitrophenyl-hydrazono) glutarates. Heterocycles 37:401–411

Kirrbach S, Schnelle R-R, Stottmeister U, Hauptmann S, Mann G, Wilde H, Sicker D (1992) Optimierte Synthesen für Methyl- und Natrium-2-oxo-D-gluconat als Synthesebausteine. J Prakt Chem 334:537–539

Liang N, Wang M, Liu L, Du G, Chen J (2008) Enhancing alpha ketoglutaric acid production in *Torulopsis glabrata*: increase of acetyl-CoA availability. Wei Sheng Wu Xue Bao 48:874–878

Liu L, Li Y, Zhu Y, Du G, Chen J (2007) Redistribution of carbon flux in *Torulopsis glabrata* by altering vitamin and calcium level. Metab Eng 9:21–29

Lockwood LB, Stodola FH (1946) Preliminary studies on the production of α-ketoglutaric acid by *Pseudomonas fluorescens*. J Biol Chem 164:81–83

Maldonado P, Desmarquest J-P, Gaillardin C, Binet D (1973) Process for getting diploid *Candida lipolytica* strains for α-ketoglutarate fermentation. US Patent 3930946, Institute Francaise du Petrole, France

Manufacturingchemist (2008) White biotechnology for greener chemistry. http://www.manufacturingchemist.com/technical/article_page/White_biotechnology_for_greener_chemistry/40815. Accessed 20 Feb 2012

Medicalnewstoday (2008) Isocitric acid from fermentation of sunflower oil: a new building block for pharma? http://www.medicalnewstoday.com/articles/96266.php. Accessed 20 Feb 2012

Misenheimer TJ, Anderson RF, Lagoda AA, Tyler DD (1965) Production of 2-ketogluconic acid by *Serratia marcescens*. Appl Environ Microbiol 13:393–396

Moresi M, Sappino F (2000) Electrodialytic recovery of some fermentation products from model solutions: techno-economic feasibility study. J Memb Sci 164:129–140

Nakahara T, Kaimaktchiev AC, Oogaki-Chino M, Uchida Y, Tabuchi T (1987) Isocitric acid production from *n*-alkanes by *Candida catenulata*. Agric Biol Chem 51:2111–2116

Novalic S, Jagschits F, Okwor J, Kulbe KD (1995) Behaviour of citric acid during electrodialysis. J Memb Sci 108:201–205

Oogaki M, Inoue M, Kaimaktchiev AC, Nakahara T, Tabuchi T (1983) Production of isocitric acid from glucose by *Candida ravautii*. Agric Biol Chem 48:789–795

Otto C (2010) Gezielte Beeinflussung der Ausbeute und des Nebenproduktspektrums der α-Ketoglutarat-Synthese in Yarrowia lipolytic mittels "metabolic engineering". PhD Thesis, Dresden University of Technology, Dresden

Otto C, Yovkova V, Barth G (2011) Overproduction and secretion of α-ketoglutaric acid by microorganisms. Appl Microbiol Biotechnol 92:689–695

Otto C, Yovkova V, Aurich A, Mauersberger S, Barth G (2012) Variation of the by-product spectrum during α-ketoglutaric acid production from raw glycerol by overexpression of fumarase and pyruvate carboxylase in *Yarrowia lipolytica*. Appl Microbiol Biotechnol. doi:10.1007/s00253-012-4085-1

Physorg (2012) Isocitric acid from fermentation of sunflower oil – a new building block for pharma? http://www.physorg.com/news121343388.html. Accessed 20 Feb 2012.

Prust C, Hoffmeister M, Liesegang H, Wiezer A, Fricke WF, Ehrenreich A et al (2005) Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. Nat Biotechnol 23:195–200

Rapp HJ (1995) Die Elektrodialyse mit bipolaren Membranen Theorie und Anwendung. PhD Thesis, Universität Stuttgart, Stuttgart

Sakaguchi K, Abe S, Beppu T (1960) Method of producing *allo*-isocitric acid by fermentation. US Patent 2949404 16.08.1960

Schmidt S, Teich L, Khodja M, Sicker D (2005a) Synthesis of pyranoid sugar α-amino acid derivatives from 2-oxo-D-gluconic acid. Lett Org Chem 2:165–171

Schmidt S, Wilde H, Hunger J, Sicker D (2005b) Synthese furanoider Zuckeraminosäuren ausgehend von fermentativ gewonnener 2-Oxo-D-gluconsäure. Z Naturforsch B: J Chem Sci 60:1168–1174

Schwesinger H, Dalski A, Sicker D, Wilde H, Mann G (1992) Synthesen auf der Basis von 2-Oxoglutarsäure.I.Synthesesubstituierter2-Nitrophenylhydrazonevon2-Oxodicarbonsäureestern und Untersuchung ihres Tautomerie- und Absorptionsverhaltens. J Prakt Chem 334:257–264

Specht R (2004) Produktion von 2-Oxoglutarsäure aus Pflanzenölen unter der Verwendung der Hefeart Yarrowia lipolytica. Diplomarbeit, Technische Universität Dresden, Dresden

Stottmeister U, Hoppe K (1991) Organische Genuss-Säuren. In: Lebensmittel-Biotechnologie. H. Ruttloff (Ed.) Akademie Verlag Berlin, pp 516–557

Stottmeister U, Behrens U, Weißbrodt E, Barth G, Franke-Rinker D, Schulze E (1982) Nutzung von Paraffinen und anderen Nichtkohlenhydrat-Kohlenstoffquellen zur mikrobiellen Citronensauresynthese. Z Allg Mikrobiol 22:399–424

Stottmeister U, Puschendorf K, Thiersch A, Berger R, Düresch R, Richter H, Schmidt JST, Iske U, Jechorek M, Uhlig H (1990) Verfahren zur Gewinnung von 2-Oxogluconsäure mittels Bakterien, DD 278 362 A1 IBT/UFZ

Stottmeister U, Schoenfelder M, Wilde H, Sicker D, Andersch J (2001) Novel 3-substituted 1,2,4-benzotriazines. A method for their production and the use thereof for the treatment and prophylaxis of tumours, EP20000985213.

Stottmeister U, Aurich A, Wilde H, Andersch J, Schmidt S, Sicker D (2005) White biotechnology for green chemistry: fermentative 2-oxocarboxylic acids as novel building blocks for subsequent chemical syntheses. J Ind Microbiol Biotechnol 32:651–664

Stubbs JJ, Lockwood LB, Roe ET, Tabenkin B, Ward GE (1940) Ketogluconic acid from glucose. Ind Eng Chem 32:1626–1631

Tabaka K, Kimura K, Yamaguchi Y (1969) Process for producing L-glutamic acid and α-ketoglutaric acid. US Patent 3450599, Kyowa Hakko Kogyo, Tokyo

Tongwen X, Weihua Y (2002) Citric acid production by electrodialysis with bipolar membranes. Chem Eng Process 41(6):519–524

Tsugawa, R., Nakase, T., Yamashita, K (1969) Production of α-ketoglutaric acid by fermentation of hydrocarbons, US Patent 3616213 Ajinomoto Co. Inc., Tokyo

Verseck S, Karau A, Weber M (2007) Fermentative Herstellung von alpha-Ketoglutarsäure. DE 10 2007 051 451.6, Evonik Degussa GmbH, Essen

Verseck S, Karau A, Weber M (2009) Fermentative production of alpha-ketoglutaric acid. WO2009053489, Evonik Degussa GmbH, Essen

Vickery HB (1962) A suggested new nomenclature for the isomers of isocitric acid. J Biol Chem 237:1739–41

Vogelbusch (2012) Biocommodities. http://www.vogelbusch-biocommodities.com/en/index.php. Accessed 20 Feb 2012

Weißbrodt E, Barth G, Weber H, Stottmeister U, Düresch R, Richter P (1988) Production of 2-oxoglutaric acid by yeasts. Patent DD 267 999 IBT/UFZ, dated 06.01.1988

Yovkova V (2011) Genetische Optimierung der Hefe Yarrowia lipolytica für die biotechnologische Gewinnung von α-Ketoglutarsäure. PhD Thesis, Dresden University of Technology, Dresden

Zhang D, Liang N, Shi Z, Liu L, Chen J, Du G (2009) Enhancement of α-ketoglutarate production in *Torulopsis glabrata*: redistribution of carbon flux from pyruvate to α-ketoglutarate. Biotechnol Bioproc Eng 14:134–139

About the Editors

Xiaoyuan Wang is Professor of Microbiology in the State Key Laboratory of Food Science and Technology at Jiangnan University. He received his Ph.D. from University of London, and completed postdoctoral training at University of Texas Houston Medical School. Then he worked as a Research Associate at Duke University Medical School until 2007. During the years Dr. Wang has been using molecular genetics and biochemical approaches to study the structure, function, and biosynthesis of key molecules in bacteria. In 2007, Dr. Wang joined Jiangnan University and was appointed Cheung Kong Chair Professor by the Ministry of Education of China. His research interests focus on microbial metabolic engineering and synthetic microbiology.

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5,

Jian Chen is Professor of Fermentation Engineering in the School of Biotechnology at Jiangnan University. He received his Ph.D. from Jiangnan University, and then worked as visiting professor in Osaka University, Inha University, and Tokyo Institute of Technology. Professor Chen has been using process optimization technologies and genetic engineering methods to improve the industrial biotechnology, especially in the bio-production of nutritional chemicals, such as amino acids, vitamins, and organic acids. In 2012, Prof. Chen became the Fellow of International Union of Food Science and Technology for his contribution to the production of functional food additives by biotechnology. His research interests include metabolic engineering of industrial microorganisms and process optimization of industrial biotechnology.

Peter J. Quinn was Professor of Biochemistry at King's College London during 1989–2009 and is currently Emeritus Professor of Biochemistry. He has held visiting professorships at Pittsburgh, Nagoya and Tsinghua Universities. His primary research interest is biological membranes and their constituents. The approach in his research has been to apply a range of biophysical methods including real-time synchrotron x-ray diffraction, neutron scattering, differential scanning calorimetry, freeze-fracture electron microscopy, nuclear magnetic resonance spectroscopy, laser flash photolysis and Fourier transform infrared spectroscopy to address questions concerned with relationships between biomembrane structure and function. Professor Quinn received his Ph.D. from the University of Sydney and was awarded a D.Sc. from the University of London in 1980.

Index

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5,

R

S