ADVANCES IN BIOCHEMICAL ENGINEERING/BIOTECHNOLOGY 120

Series Editor T. Scheper
Volume Editors C. Wittmann · R. Krull

Biosystems Engineering I

Creating Superior Biocatalysts

Springer

120
Advances in Biochemical Engineering/Biotechnology

Series Editor: T. Scheper

Advances in Biochemical Engineering/Biotechnology

Series Editor: T. Scheper

Recently Published and Forthcoming Volumes

Biosystems Engineering I
Volume Editors: Wittmann, C., Krull, R.
Vol. 120, 2010

Whole Cell Sensing Systems II
Volume Editors: Belkin, S., Gu, M.B.
Vol. 118, 2010

Whole Cell Sensing Systems I
Volume Editors: Belkin, S., Gu, M.B.
Vol. 117, 2010

Optical Sensor Systems in Biotechnology
Volume Editor: Rao, G.
Vol. 116, 2009

Disposable Bioreactors
Volume Editor: Eibl, R., Eibl, D.
Vol. 115, 2009

Engineering of Stem Cells
Volume Editor: Martin, U.
Vol. 114, 2009

Biotechnology in China I
From Bioreaction to Bioseparation and Bioremediation
Volume Editors: Zhong, J.J., Bai, F.-W., Zhang, W.
Vol. 113, 2009

Bioreactor Systems for Tissue Engineering
Volume Editors: Kasper, C., van Griensven, M., Poertner, R.
Vol. 112, 2008

Food Biotechnology
Volume Editors: Stahl, U., Donalies, U. E. B., Nevoigt, E.
Vol. 111, 2008

Protein – Protein Interaction
Volume Editors: Seitz, H., Werther, M.
Vol. 110, 2008

Biosensing for the 21st Century
Volume Editors: Renneberg, R., Lisdat, F.
Vol. 109, 2007

Biofuels
Volume Editor: Olsson, L.
Vol. 108, 2007

Green Gene Technology
Research in an Area of Social Conflict
Volume Editors: Fiechter, A., Sautter, C.
Vol. 107, 2007

White Biotechnology
Volume Editors: Ulber, R., Sell, D.
Vol. 105, 2007

Analytics of Protein-DNA Interactions
Volume Editor: Seitz, H.
Vol. 104, 2007

Tissue Engineering II
Basics of Tissue Engineering and Tissue Applications
Volume Editors: Lee, K., Kaplan, D.
Vol. 103, 2007

Tissue Engineering I
Scaffold Systems for Tissue Engineering
Volume Editors: Lee, K., Kaplan, D.
Vol. 102, 2006

Cell Culture Engineering
Volume Editor: Hu, W.-S.
Vol. 101, 2006

Biotechnology for the Future
Volume Editor: Nielsen, J.
Vol. 100, 2005

Gene Therapy and Gene Delivery Systems
Volume Editors: Schaffer, D.V., Zhou, W.
Vol. 99, 2005

Biosystems Engineering I

Creating Superior Biocatalysts

Volume Editors:
Christoph Wittmann · Rainer Krull

With contributions by

R. Biedendieck · K. Buchholz · B. Bunk · E. Franco-Lara · T. Fürch · A. Ghaderi · H. Gröger · W. Hummel · D. Jahn · M. Jahn · H.-J. Jördening · D. Klein-Marcuschamer · S.Y. Lee · J. Nielsen · M. Papini · J.H. Park · M. Salazar · J. Seibel · G.N. Stephanopoulos · A. Weckbecker · C. Wittmann · V.G. Yadav

Springer

Editors
Prof. Dr. Christoph Wittmann
Technische Universität Braunschweig
Biochemical Engineering Institute
Gaußstraße 17
38106 Braunschweig
Germany
c.wittmann@tu-braunschweig.de

Prof. Dr. Rainer Krull
Technische Universität Braunschweig
Biochemical Engineering Institute
Gaußstraße 17
38106 Braunschweig
Germany
r.krull@tu-braunschweig.de

ISSN 0724-6145 e-ISSN 1616-8542
ISBN 978-3-642-14230-7 e-ISBN 978-3-642-14231-4
DOI 10.1007/978-3-642-14231-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010933955

Cover design: WMXDesign GmbH, Heidelberg, Germany

Printed on acid-free paper

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

Series Editor

Prof. Dr. T. Scheper

Institute of Technical Chemistry
University of Hannover
Callinstraße 3
30167 Hannover, Germany
scheper@iftc.uni-hannover.de

Volume Editors

Prof. Dr. Christoph Wittmann

Technische Universität Braunschweig
Biochemical Engineering Institute
Gaußstraße 17
38106 Braunschweig
Germany
c.wittmann@tu-braunschweig.de

Prof. Dr. Rainer Krull

Technische Universität Braunschweig
Biochemical Engineering Institute
Gaußstraße 17
38106 Braunschweig
Germany
r.krull@tu-braunschweig.de

Editorial Board

Honorary Editors

Advances in Biochemical Engineering/ Biotechnology Also Available Electronically

Advances in Biochemical Engineering/Biotechnology is included in Springer's eBook package *Chemistry and Materials Science*. If a library does not opt for the whole package the book series may be bought on a subscription basis. Also, all back volumes are available electronically.

For all customers who have a standing order to the print version of *Advances in Biochemical Engineering/Biotechnology*, we offer the electronic version via SpringerLink free of charge.

If you do not have access, you can still view the table of contents of each volume and the abstract of each article by going to the SpringerLink homepage, clicking on "Chemistry and Materials Science," under Subject Collection, then "Book Series," under Content Type and finally by selecting *Advances in Biochemical Bioengineering/Biotechnology*

You will find information about the

– Editorial Board
– Aims and Scope
– Instructions for Authors
– Sample Contribution

at springer.com using the search function by typing in *Advances in Biochemical Engineering/Biotechnology*.

Color figures are published in full color in the electronic version on SpringerLink.

Aims and Scope

Advances in Biochemical Engineering/Biotechnology reviews actual trends in modern biotechnology.

Its aim is to cover all aspects of this interdisciplinary technology where knowledge, methods and expertise are required for chemistry, biochemistry, microbiology, genetics, chemical engineering and computer science.

Special volumes are dedicated to selected topics which focus on new biotechnological products and new processes for their synthesis and purification. They give the state-of-the-art of a topic in a comprehensive way thus being a valuable source for the next 3-5 years. It also discusses new discoveries and applications.

In general, special volumes are edited by well known guest editors. The series editor and publisher will however always be pleased to receive suggestions and supplementary information. Manuscripts are accepted in English.

In references *Advances in Biochemical Engineering/Biotechnology* is abbreviated as *Adv. Biochem. Engin./Biotechnol.* and is cited as a journal.

Special volumes are edited by well known guest editors who invite reputed authors for the review articles in their volumes.

Impact Factor in 2009: 4.165; Section “Biotechnology and Applied Microbiology”: Rank 23 of 150

Attention all Users of the "Springer Handbook of Enzymes"

Information on this handbook can be found on the internet at springeronline.com

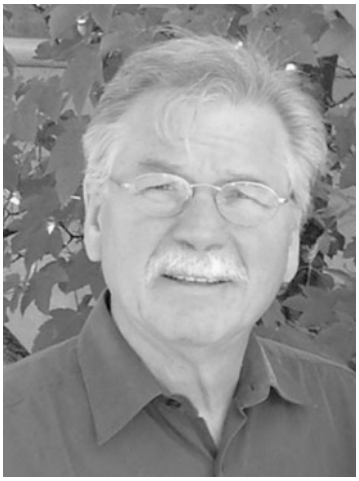
A complete list of all enzyme entries either as an alphabetical Name Index or as the EC-Number Index is available at the above mentioned URL. You can download and print them free of charge.

A complete list of all synonyms (more than 57,000 entries) used for the enzymes is available in print form (ISBN 978-3-642-14015-0) and electronic form (ISBN 978-3-642-14016-7).

Save 15%

We recommend a standing order for the series to ensure you automatically receive all volumes and all supplements and save 15% on the list price.

Laudatio



This special volume "Biosystems Engineering" is dedicated to Professor Dr.-Ing. Dietmar Christian Hempel on the occasion of his 65th birthday, whereby the different contributions display an excellent reflection of his research during the past 30 years, bridging engineering and life sciences.

Prof. Hempel, born in Königsberg, studied Construction and Process Engineering in Dortmund and Berlin (1963–1971). After his PhD work on "Heterogeneous catalytic fixed bed reactors" he became the head for "Reaction and Biochemical Engineering" at the Bayer AG in Leverkusen, where he started to apply chemical engineering principles to biological systems (1975–1980). This was obviously stimulating his later research as Professor of Technical Chemistry and Chemical Engineering at the University of Paderborn (1980–1994) and as Professor and Director of the Institute of Biochemical Engineering at the Technische Universität Braunschweig (1994–2009) where he kept his research interest on biological systems in technical applications – covering various aspects process engineering of biological and biochemical processes. As impressive outcome of his career, Dietmar Christian Hempel authored and co-authored about 300 publications and supervised 70 PhD students as a "Doktorvater". From 2001 to 2008 he was head and speaker of the still successfully continued interdisciplinary DFG-collaborate research centre SFB 578 "Development of biotechnological processes by integrating genetic and engineering methods – From gene to product". This collaborate research centre displays a landmark in the German biochemical engineering community and serves as important link between cellular biology and bioprocess engineering.

Preface

The special volume "Biosystems Engineering" reflects an emerging field of applied research that aims at a system-level understanding of biological systems toward their targeted design and improvement. To obtain system-wide insight, interdisciplinary approaches are applied, which integrate expertise from biologists, engineers, and computer scientists, who have developed powerful analytical methods, modelling concepts, and information technologies for this challenging task. With the increasing need for sustainable production of fuels and chemicals from renewable resources, a major focus in this area is on the optimization of biotechnological production strains and processes. Of specific importance hereby is the consideration of the microorganism as a part of the bioprocess in its entirety.

In this regard, the different contributions of this volume provide an outstanding review of novel tools, methods, and concepts in the field of biosystems engineering, a still young, but yet powerful direction of interdisciplinary research. Part I "Creating Superior biocatalysts" of this volume highlights how system-wide analysis, modelling and understanding of gene regulation, metabolic networks, and fluxes open a new era for rational design and optimization of superior production strains, one of the key pre-requisites for bio-based production of pharmaceuticals, chemicals, materials, and fuels. This includes the introduction of systems and synthetic metabolic engineering approaches as well as industrial application examples. Part II "Linking Cellular Networks and Bioprocesses" illustrates concepts and tools which are based on models and experiments to investigate metabolic networks in correlation with the cell environment. In addition to the introduction of fundamental strategies introducing thermodynamic principles as well as different modelling approaches to metabolic network simulation, different examples directly address the link of cells and bioreactors providing fascinating and valuable insights into the interaction between cellular metabolism and process environment.

Braunschweig, Summer 2010

Christoph Wittmann
Rainer Krull

Contents

Integration of Systems Biology with Bioprocess Engineering: L-Threonine Production by Systems Metabolic Engineering of *Escherichia Coli* 1
Sang Yup Lee and Jin Hwan Park

Analysis and Engineering of Metabolic Pathway Fluxes in *Corynebacterium glutamicum* 21
Christoph Wittmann

Systems Biology of Industrial Microorganisms 51
Marta Papini, Margarita Salazar, and Jens Nielsen

***De Novo* Metabolic Engineering and the Promise of Synthetic DNA** 101
Daniel Klein-Marcuschamer, Vikramaditya G. Yadav, Adel Ghaderi, and Gregory N. Stephanopoulos

Systems Biology of Recombinant Protein Production in *Bacillus megaterium* 133
Rebekka Biedendieck, Boyke Bunk, Tobias Fürch, Ezequiel Franco-Lara, Martina Jahn, and Dieter Jahn

Extending Synthetic Routes for Oligosaccharides by Enzyme, Substrate and Reaction Engineering 163
Jürgen Seibel, Hans-Joachim Jördening, and Klaus Buchholz

Regeneration of Nicotinamide Coenzymes: Principles and Applications for the Synthesis of Chiral Compounds 195
Andrea Weckbecker, Harald Gröger, and Werner Hummel

Index 243

Adv Biochem Engin/Biotechnol (2010) 120: 1–19
DOI: 10.1007/10_2009_57

Published online: 6 February 2010

Integration of Systems Biology with Bioprocess Engineering: L-Threonine Production by Systems Metabolic Engineering of *Escherichia Coli*

Sang Yup Lee and Jin Hwan Park

Abstract Random mutation and selection or targeted metabolic engineering without consideration of its impact on the entire metabolic and regulatory networks can unintentionally cause genetic alterations in the region, which is not directly related to the target metabolite. This is one of the reasons why strategies for developing industrial strains are now shifted towards targeted metabolic engineering based on systems biology, which is termed systems metabolic engineering. Using systems metabolic engineering strategies, all the metabolic engineering works are conducted in systems biology framework, whereby entire metabolic and regulatory networks are thoroughly considered in an integrated manner. The targets for purposeful engineering are selected after all possible effects on the entire metabolic and regulatory networks are thoroughly considered. Finally, the strain, which is capable of producing the target metabolite to a high level close to the theoretical maximum value, can be constructed. Here we review strategies and applications of systems biology successfully implemented on bioprocess engineering, with particular focus on developing L-threonine production strains of *Escherichia coli*.

Keywords L-threonine, Systems biology, Systems metabolic engineering, Bioprocess engineering

S.Y. Lee (✉), J.H. Park
Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea
BioProcess Engineering Research Center, KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea
email: leesy@kaist.ac.kr
S.Y. Lee
Department of Bio and Brain Engineering and Bioinformatics Research Center, KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea
Department of Biological Sciences, KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea

Contents

1 Introduction 2
2 Use of Omics in Metabolic and Bioprocess Engineering 2
3 L-Threonine 7
4 Integration of Systems Biology with Bioprocess Development for L-Threonine Production 12
5 Conclusions 15
References 16

1 Introduction

Systems biology is changing the way biological systems are studied. Genomics, transcriptomics, proteomics, metabolomics, and fluxomics are generating system-wide data and information on genomic contents, transcripts, proteins, metabolites, and metabolic fluxes, respectively. When combined with metabolic, gene regulatory, and signaling network analysis, these data can be used to decipher the current metabolic and regulatory status over much wider picture. Also, mathematical modeling and simulation became an essential tool for better understanding the biological system under investigation at the systems-level. The systems-level analysis of microorganisms using genome-wide experimental and computational methods has been proven to be beneficial for elucidating the cell physiology at the whole cell level, consequently providing us with new strategies for metabolic engineering [1]. Furthermore, midstream and downstream processes are considered together during the upstream strain development so that the entire bioprocess can be optimized. This approach has been employed as a general strategy to develop biotechnologically important microorganisms, which is termed systems metabolic engineering [2, 3].

Microorganisms that are capable of efficiently producing bioproducts, particularly amino acids, have traditionally been constructed by random mutagenesis, which can cause unwanted alterations in cellular physiology. Metabolic engineering allows overcoming such problems by specifically altering the metabolic, regulatory, and cellular machinery in a desired fashion [4–6]. When combined with systems biology tools, metabolic engineering becomes an even more powerful strategy for the enhanced production of target metabolites at a yield near the theoretical maximum [7, 8]. In this chapter, the general strategy for the integration of systems biology with bioprocess engineering is reviewed. Also, some recent successful examples are showcased with particular focus on the development of L-threonine overproducing strain of *Escherichia coli*.

2 Use of Omics in Metabolic and Bioprocess Engineering

In this section, the strategies for metabolic and bioprocess engineering based on the omics technologies are described and their successful examples of applications are summarized.

2.1 Genome

Genomic information available for industrially important microorganisms can be usefully employed to develop strains for the enhanced production of desired bioproducts. In particular, comparative genome analysis allows identification of genes to be modified for a desirable metabolic phenotype. This approach was successfully demonstrated by Ohnishi et al. [9] who carried out genome comparison studies between a wild-type *Corynebacterium glutamicum* strain and an L-lysine-producing strain obtained by random mutation and selection. They were able to identify mutations in the *gnd* gene that are suspected to contribute to the efficient production of L-lysine. They introduced the identified mutation into the AHP-3 strain to make an APG-4 strain, which allowed 15% increase in L-lysine production [9]. Further improvement was performed by introducing *mqo* mutation selected on the basis of genome comparison. The final strain allowed production of 95 g L^{-1} L-lysine by fed-batch culture [10]. In the case of *Mannheimia succiniciproducens*, gene knockout studies were conducted to understand its anaerobic fermentative metabolism and consequently to develop a metabolically engineered strain capable of producing succinic acid without by-product formation [11]. Likewise, the biosynthetic pathway of L-methionine in *C. glutamicum* was elucidated by genome-wide analysis combined with targeted gene deletion and homologous complementation [12].

Genome-wide profiling analysis based on the complete genome sequence can also be used to identify the functions of certain genes encoding enzymes responsible for the bioproduct to be overproduced. Using this method, McHardy et al. [13] discovered potential aminotransferase encoding genes and experimentally characterized them, resulting in the discovery of 11 new aminotransferases participating in the biosynthesis of branched-chain amino acids and phenylalanine in *C. glutamicum*. Two genes involved in the biosynthesis of synechoxanthin were identified by comparative genomics, and their functions were verified by insertional inactivation [14]. The *cruE* gene encodes β-carotene desaturase/methyltransferase, which converts β-carotene to renierapurpurin. The *cruH* gene encodes an enzyme that is responsible for the hydroxylation/oxidation of the C-18 and C-18′ methyl groups of renierapurpurin. Based on these findings, a complete pathway for synechoxanthin biosynthesis was proposed.

In another approach, a reduced-genome strain can be constructed by deleting nonessential genes, while preserving those required for good growth profiles and protein production [15, 16]. As a conclusion, full genome sequence provides invaluable information for genome-scale engineering, thereby leading to the development of strains showing significantly improved performance. This perspective is strongly supported by the fact that recent DNA sequencing technologies allow sequencing of up to one billion bases in a single day at low cost [17].

2.2 Transcriptome

Gene expression profiling by DNA microarray captures mRNA transcription level for thousands of genes of multiple strains under different conditions simultaneously, and thus makes it possible to analyze cell physiology and global regulation at system-wide transcript level. This tool can be used to analyze the cell physiology and to select target genes to be engineered as well. For example, the target genes for the enhanced production of recombinant protein have been successfully identified by DNA microarray experiments by comparing the transcriptome profiles before and after induction during the high cell density culture [18]. Overexpression of the *prsA* (encoding phosphoribosyl pyrophosphate synthetase) and the *glpF* (glycerol transporter) genes, which were selected among the downregulated genes, allowed a significant increase in IFG-I_f production (from 1.8 to 4.3 g L^{-1}). Transcriptome analysis can also be used to characterize certain genes playing critical roles in carbon metabolism. As an example, induction of two genes encoding lactate permease and lactate dehydrogenase during L-glutamate production using lactate as a carbon source in *C. glutamicum* was identified by transcriptome profiling [19]. Recently, enhanced production of L-lysine was achieved by introducing useful mutations identified by transcriptome analysis into a defined L-lysine producer. To identify the amino acid biosynthetic genes that were induced, transcriptome analysis of an L-lysine producer mutant, *C. glutamicum* B-6, was carried out [20]. Among the genes showing different transcript levels, the *leuC* gene was selected as a useful mutation, and subsequently introduced into the defined L-lysine producer, AHD-2 (*hom*59 and *lysC*311), resulting in 14% increase in L-lysine production [21]. In another study, transcriptome analysis was used to identify target genes to be engineered in L-lysine producing *C. glutamicum*, in which upregulated genes including those encoding a methyltransferase and ammonium uptake system were overexpressed, resulting in 40% increase in L-lysine production [22]. More recently, novel factors enhancing heterologous protein secretion in yeast were identified by transcriptome profiling [23]. Overexpression of the significantly upregulated genes involved in protein transport, folding, and exocytosis improved both specific production rate and volumetric productivity of an antibody fragment by ca. 2.5-fold.

Another recent report indicates that transcriptome analysis is an effective way of initially screening the genes targets for molecular breeding. As a demonstration of this approach, a xylitol-producing strain was constructed by inserting the genes encoding the NADPH-dependent D-xylose reductase and D-xylose permease into the *E. coli* chromosome. Comparative transcriptome analysis of xylitol-producing and nonproducing conditions for the recombinant strain revealed that xylitol production caused down-regulation of 56 genes. From this study, the *yhbC* gene was selected as a candidate cause for the suppression of NADPH supply and thus deleted, which resulted in a 2.7-fold increase in xylitol production [24]. Exemplary results mentioned here demonstrate that novel gene targets for strain improvement can be identified based on the global gene expression analysis.

Global transcription machinery engineering (gTME) through the mutagenesis of the transcription factors allows it possible to perturb globally the transcriptome, which consequently allows alteration of the cellular phenotype towards a desired trait [25]. Transcriptome profiling was also used to investigate the novel regulation system in a metabolic pathway of the aminosugar antibiotic, neotrehalosadiamine [26]. It was found that loss of non-PTS type glucose transporter, GlcP resulted in a dramatic activation of neotrehalosadiamine biosynthesis genes. Further analysis might suggest a new gene target to be engineered to increase neotrehalosadiamine biosynthesis.

2.3 Proteome

While conventional biochemical studies focus on a single protein or simple macromolecular complexes, proteomics takes a much broader, more comprehensive and systematic approach to the investigation of biological systems in the context of proteins. Thus, proteome analysis can also be a powerful tool for systems-level analysis and engineering of the biological system. For example, the overexpression of the *ppsA* gene (encoding phage shock protein A), which was suggested to be a good target for strain improvement by proteome profiling, increased the yield of soluble antibody by 50% [27]. Similarly, comparative proteome analysis of *M. succiniciproducens* was performed to understand growth-associated physiological changes of this bacterium [28]. As a result, proteome reference map of *M. succiniciproducens* was established, and targets to be manipulated for the strain improvement were suggested for enhanced succinic acid production. PutA, which is involved in acetate formation, and OadA, which converts oxaloacetate to pyruvate, were expected to be good targets to be deleted to increase succinic acid formation with reduced by-product formation. In another recent paper, comparison of the extracellular proteomes of *E. coli* BL21 (DE3) and W3110 strains was performed during their high cell density cultivation [29]. Among 204 protein spots identified by 2-DE and MS analysis, only 32 proteins were conserved in the two strains; the two strains showed very different patterns in their excreted proteins. The extracellular proteome reference map obtained from this study should be useful in identifying new targets for enhanced excretory production of recombinant proteins as demonstrated in the chapter.

2.4 Metabolome and Fluxome

Metabolome represents the entire set of low-molecular-weight metabolites that are present in a cell and/or outside the cell under particular conditions [30]. Metabolite profiling allows comparative analysis of physiological states under different conditions; e.g., a wild-type vs its mutants, and different culture conditions. Fluxome, the entire set of metabolic fluxes, can also be used for understanding the cellular

metabolic capacities/activities under various conditions. Combined metabolome and fluxome analyses were performed on L-lysine producing strain of *C. glutamicum* to understand the metabolites production and metabolic flux distribution, respectively. ^{13}C isotopomer labeling experiments were carried out to determine metabolic flux distribution, which suggested that the high TCA cycle flux was beneficial for L-lysine production. Also, it was suggested to reduce the secretion of dihydroxyacetone and glycerol, two major byproducts, by metabolic engineering to enhance L-lysine production [31]. Metabolic flux analysis (MFA) of L-lysine producing *C. glutamicum* cultured on fructose identified fructose 1,6-bisphosphatase as an amplification target to achieve increased NADPH supply by increasing flux through the pentose phosphate pathway [32]. This strategy was then successfully implemented by Becker et al. [33, 34] who engineered the NADPH metabolism in *C. glutamicum* and subsequently developed better strains for L-lysine production. In another example, notable improvement of succinic acid production in *E. coli* was achieved using constraints-based flux analysis [35]. Based on the combined application of in silico optimization and MFA, three potential targets for the strain improvement, the glucose phosphotransferase transport system (PTS), the pyruvate carboxylase, and the glyoxylate shunt, were identified for genetic modifications. Using the engineered strain, a high yield of 1.29 mol succinate per mol glucose was achieved.

2.5 Combined Omics Analysis

As genome, transcriptome, proteome, metabolome, and fluxome provide information at different hierarchical levels of cellular activities, it is natural to think of performing combined omics analysis to identify target genes to be engineered. The first example of such combined analysis appeared for the case of combined transcriptome and proteome analysis of *E. coli* cells under high cell density cultivation [36]. One of the interesting findings was that the specific protein biosynthetic capacity of a cell decreases as the cell density increases, which was confirmed by both transcriptome and proteome profiling. Similarly, combined transcriptome and metabolome analysis has been used to develop an *Aspergillus* strain overproducing lovastatin [37]. Through the combined analysis, the target genes were identified and manipulated accordingly, resulting in an improved production of lovastatin by more than 50%. Krömer et al. [38] carried out combined transcriptome, metabolome, and fluxome analysis of L-lysine producing *C. glutamicum* to understand metabolic characteristics. It was found that a decrease in glucose uptake rate caused the metabolic shift from cell growth towards L-lysine biosynthesis.

In another example, comparative genome comparison was combined with constraints-based flux analysis to develop metabolic engineering strategies for the enhanced production of succinic acid by *E. coli* [1, 39, 40]. Comparative genome analysis of *E. coli* and *M. succiniciproducens* predicted five candidate genes to be manipulated for overproducing succinic acid in *E. coli*. Then, constraints-based flux

analysis was carried out to find an optimal combination of the selected genes that makes the strain achieve the maximum biomass and succinic acid production capability upon their knockout. Disruption of the *ptsG*, *pykF*, and *pykA* genes significantly enhanced the succinic acid production [40].

Alper et al. [41, 42] employed constraints-based flux analysis and genome-wide transposon library searching for the overproduction of lycopene in *E. coli*. Although knockout of gene targets predicted by constraints-based flux analysis allowed enhanced production of lycopene, it was below the stoichiometrically maximum value. This limit was overcome by ideally combining them with gene targets identified from transposon methods, some of which led to a mutant strain that was capable of producing lycopene up to 8.5-fold higher than recombinant *E. coli* control strain. In another study, combined transcriptome and MFA has been used to examine the results of rational genetic modification in pantothenate producing stain of *C. glutamicum* [43]. Successful redirection of the carbon flux towards pantothenate biosynthesis by promoter down-mutation was demonstrated by MFA. According to transcriptome profiling, genes involved in amino acid biosynthesis were found to be induced during the production of pantothenate.

It is expected that the method to integrate multiple high-throughput omics data would become more sophisticated, providing more insights into the biological system. Recent multi-level omics analyses of *E. coli* indicate how the cell flexibly regulates itself at various omics levels, depending on the environmental and genetic perturbations, so as to maintain its metabolic integrity [44]. Although this report is limited to the monitoring of the cell, it does provide possible implications of metabolic engineering, which is more precisely described in the following sections in the context of tailored metabolic engineering.

3 L-Threonine

3.1 Biosynthetic Pathways and Regulations Involved

L-Threonine is one of the three major amino acids produced by fermentation processes [45]. Currently, more than 4,000 tons of L-threonine are produced annually by fermentation [46]. In this section, we examine the L-threonine biosynthetic pathway and its regulation, and discuss how the carbon flux can be maximized towards L-threonine biosynthesis by metabolic engineering. The detailed description on L-threonine biosynthetic pathways and regulations involved is shown in Fig. 1.

The L-threonine biosynthetic pathway consists of five enzymatic steps from L-aspartate. *E. coli* has three aspartate kinase isoenzymes, key enzymes which catalyze the first reaction of the L-threonine biosynthetic pathway. The aspartate kinase isoenzymes I, II, and III encoded by the *thrA*, *metL*, and *lysC* genes, respectively, are affected by feedback inhibition by L-threonine, L-methionine, and L-lysine, respectively. *C. glutamicum* has only one aspartate kinase encoded by the *lysC* gene, which is subjected to feedback inhibition by L-lysine and

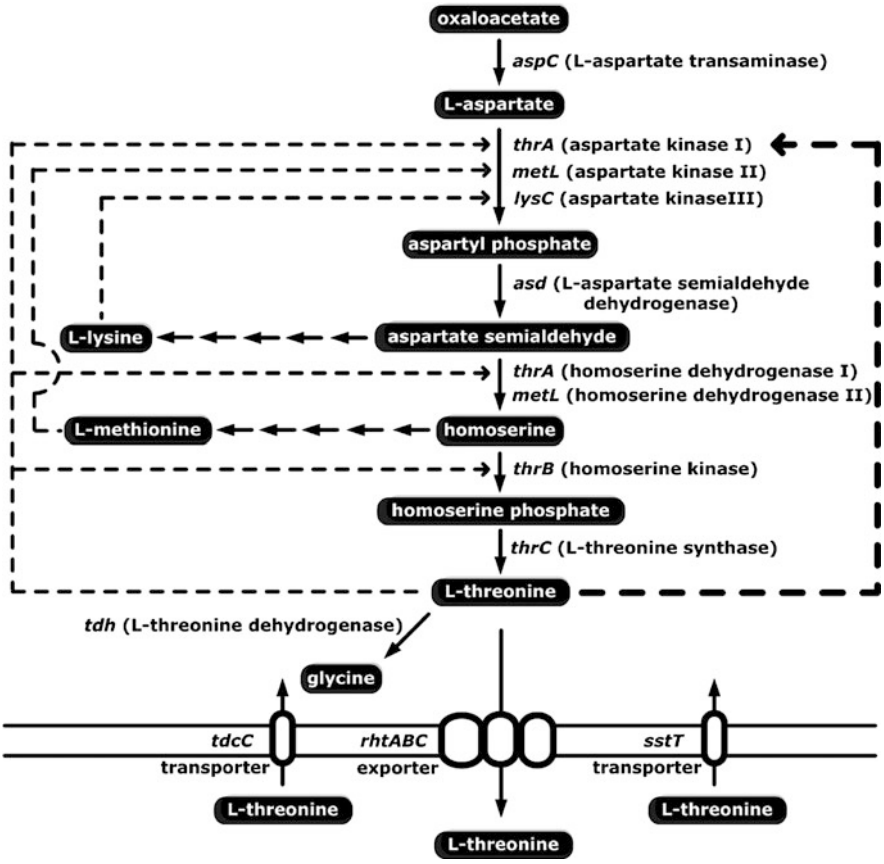


Fig. 1 The biosynthetic pathway of L-threonine, and its regulation in *E. coli*. *Dotted lines* indicate feedback inhibition. *Thick dotted lines* indicate transcriptional attenuation regulation

L-threonine. The homoserine kinase encoded by the *thrB* gene is subject to feedback inhibition by L-threonine. The homoserine dehydrogenase encoded by the *thrA* gene (for *E. coli*) and the *hom* gene (for *C. glutamicum*) is subject to feedback inhibition by L-threonine. The expression of the *thrABC* operon is controlled by transcriptional attenuation mediated by L-threonine and L-isoleucine. This metabolic and regulatory information should be taken into account when designing metabolic engineering strategies for maximizing the carbon flux towards L-threonine biosynthesis.

3.2 Strategies for Targeted Metabolic Engineering

Before systems metabolic engineering is performed, rational metabolic engineering is first carried out to develop a base strain. The following is the procedure taken

towards the development of a base strain for the case of L-threonine producer. First, the feedback inhibition of aspartate kinase isoenzymes and the transcriptional attenuation control for the *thrABC* operon should be removed. Next, the competing pathways for L-lysine and L-methionine biosynthesis should be blocked to make more precursors available for L-threonine biosynthesis. Knocking out of these pathways, however, results in an auxotrophic strain for L-lysine and L-methionine with the consequence that small amounts of these amino acids need to be supplemented for cell growth. Thus, the feedback inhibition and transcriptional attenuation control by these amino acids also need to be removed. Next, the pathway that degrades L-threonine should be removed by deleting the *tdh* gene. Also, the L-threonine exporter system is engineered (amplified) to facilitate the excretion of the overproduced L-threonine.

3.3 Production Strains

The representative L-threonine production strains are described in Table 1. Traditionally, the L-threonine production strains were constructed by random mutation and selection. An amino acid analog-resistant mutant showing normal growth after treating the wild-type strain with mutagens is selected. Then, multiple rounds of mutation and selection are performed until the mutant strain shows the desirable productivity and yield. For example, β-hydroxynorvaline (α-amino-β-hydroxyvaleric acid)-resistant mutants of *Serratia marcescens* capable of producing 13 g L^{-1} of L-threonine were isolated after treating the wild-type strain with *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine [47]. This mutant strain was found to have lost both feedback inhibition and repression of homoserine dehydrogenase (encoded by the *hom* gene) by L-threonine.

The advances in recombinant DNA techniques made it possible to develop L-threonine production strains by limited yet rational metabolic engineering. For example, the *E. coli* L-threonine producing mutant was transformed with plasmid pAJ294 containing the whole L-threonine operon and cultured. The final L-threonine concentration obtained was 13.4 g L^{-1}, which is threefold higher than that obtained without amplification of those genes [48]. In another example, the *hom* (encoding homoserine dehydrogenase), *thrB* (encoding homoserine kinase), and *thrC* (encoding L-threonine synthase) genes of *Brevibacterium lactofermentum* were amplified by plasmid-based overexpression in a *B. lactofermentum* L-threonine producer strain. The resulting strain was able to produce 24.8 g L^{-1} L-threonine, which is 2.1-fold higher than that produced without overexpression of these genes [49].

Obviously, culture conditions can also affect L-threonine production. For example, a high level of dissolved oxygen and lowering the required amino acids are effective for enhancing L-threonine production. By using the optimal conditions, the final L-threonine concentration obtained was 65 g L^{-1}, resulting in a yield of 0.48 g g^{-1} glucose [50]. In another example, when an *S. marcescens* L-threonine producing mutant was grown with urea as a nitrogen source, 55 g L^{-1} L-threonine

Table 1 L-Threonine-producing strains

Strain	Approach	Results	Titer ($g\ L^{-1}$)	References
S. marcescens	Random mutation and selection to construct L-threonine production strain	β-Hydroxynorvaline (α-amino-β-hydroxyvaleric acid)-resistant mutants were obtained	13.0	Komatsubara et al. [47]
E. coli	The whole L-threonine operon was amplified by plasmid-based overexpression in *E. coli* L-threonine producer	A threefold increase in L-threonine production	13.4	Miwa et al. [48]
B. lactofermentum	Three genes (*hom*, *thrB*, *thrC*) were amplified by plasmid-based overexpression in *B. lactofermentum* L-threonine producer	A 2.1-fold increase in L-threonine production	57.7	Ishida et al. [49]
E. coli	Culture conditions for the production of L-threonine were investigated	A high level of dissolved oxygen and lowering the required amino acids are effective for enhancing L-threonine production	65.0	Shimizu et al. [50]
S. marcescens	Ammonia water is more beneficial than urea for the enhanced production of L-threonine as a nitrogen source	A 77.2% increase in L-threonine production	100.0	Masuda et al. [51]
E. coli	An impaired L-threonine uptake system enhanced L-threonine production in *E. coli* L-threonine producer	The inability to take up L-threonine relieved negative effect of remaining feedback inhibition	100.0	Okamoto et al. [53]
E. coli	Genes encoding L-threonine exporter were overexpressed in *E. coli* L-threonine producer	Specific productivity was increased by 290% and 200% with overexpression of *rhtB* and *rhtC* genes, respectively	31.2	Kruse et al. [54]
E. coli	The mutant aspartokinase III encoded by *lysC* was used to enhance L-threonine production	A 30.9% increase in L-threonine production	14.4	Ogawa-Miyata et al. [57]
E. coli	The effect of biotin and air condition on L-threonine production was investigated	Supply of biotin and oxygen-enriched air improved	80.2	Lee et al. [52]

(*continued*)

Table 1 (continued)

Strain	Approach	Results	Titer (g L^{-1})	References
		L-threonine production		
E. coli	Combined genome, transcriptome and proteome analysis were performed on wild-type and mutant strains to elucidate underlying mechanism for overproduction of L-threonine	The two significant mutations in *thrA* and *ilvA* genes were identified essential for overproduction of L-threonine	21.0	Lee et al. [58]
E. coli	Construction of L-threonine producing *E. coli* based on systems metabolic engineering	Production of a high yield of 0.393 g L-threonine per g glucose by batch culture, and 82.4 g L^{-1} L-threonine by fed-batch culture	82.4	Lee et al. [7]
E. coli	Construction of L-threonine producing strain using a reduced-genome strain of *E. coli*	An 83% increase in L-threonine production	40.1	Lee et al. [59]

was produced after 72 h of cultivation (0.764 g L^{-1} h^{-1}). However, when ammonia water was used as a nitrogen source, 65 g L^{-1} L-threonine was produced after 48 h of cultivation (1.354 g L^{-1} h^{-1}), resulting in a 77.2% increase in L-threonine productivity [51]. It was identified that a certain amount of ammonium ion formed from urea disappeared due to heat decomposition, and thus cell growth was retarded, resulting in a decrease in L-threonine production. In some cases, better results are obtained by adding growth factor into the medium. When biotin was added as a growth factor, the L-threonine concentration obtained with the L-threonine producer *E. coli* mutant MT201 was 52.0 g L^{-1}, which is 3.27-fold higher than that obtained without the addition of biotin [52]. L-Threonine production was further enhanced by changing culture conditions. By high cell density culture of MT201 strain in a medium supplied with oxygen-enriched air, 80.2 g L^{-1} L-threonine could be produced [52]. The supply of oxygen-enriched air might have overcome the problem of the accumulation of organic acids which was caused by oxygen-limited condition during the high cell density culture [52].

As mentioned earlier, L-threonine production can be enhanced by engineering the export or uptake system. An efficient L-threonine producer strain of *E. coli* KY10935, which was derived from the wild-type strain by multiple rounds of random mutation and selection, was able to produce 100 g L^{-1} L-threonine after 77 h cultivation [53]. In this strain, the two key enzymes in the L-threonine biosynthesis (homoserine dehydrogenase and homoserine kinase) were identified to be still inhibited by much lower intracellular concentrations of L-threonine than

the externally accumulated levels. This result strongly suggests that the transport system responsible for L-threonine uptake in this strain would have been destroyed, and thus these feedback-sensitive enzymes were able to continue producing L-threonine to an impressively high level. In another *E. coli* L-threonine producer mutant, the specific L-threonine productivity was increased by up to 290% and 200% by the overexpression of the L-threonine exporter genes, *rhtB* and *rhtC*, respectively [54]. The *rhtA* gene of *E. coli* was also identified as a new gene encoding L-threonine exporter [55]. In *C. glutamicum*, the *thrE* gene was identified to be involved in L-threonine export [56].

Removal of negative regulations can also enhance L-threonine production. By removing L-lysine-mediated feedback inhibition of aspartokinase III encoded by the *lysC* gene, L-threonine production could be increased by 30.9% (11.0–14.4 g L^{-1}) in *E. coli* mutant strain [57].

4 Integration of Systems Biology with Bioprocess Development for L-Threonine Production

4.1 Paradigm Shift Towards Systems Metabolic Engineering for Amino Acid Production

The availability of the complete genome sequence of industrially useful microorganisms offers a foundation for the development of more superior strains by employing various genome-scale technologies such as genomics, transcriptomics, proteomics, metabolomics, and fluxomics. When strains are developed by random mutation and selection as described above, it is often difficult to improve further the strain because of our inability to elucidate the consequence of all the mutations including those that are not beneficial for the production of L-threonine. It is notable that the alterations in metabolic and regulatory networks are difficult to understand even after the complete genome sequence of the mutant strain is determined. To overcome these problems, it is necessary to design specifically metabolic engineering strategies with sufficient consideration of the characteristics of the strain to be engineered at the systems-level and the midstream to downstream processes all together. This is "systems metabolic engineering," and allows true integration of systems biology with bioprocess engineering; this will result in 100% genetically defined strain. For this reason, strategies for developing L-threonine producers are now in transition towards systems metabolic engineering from random mutagenesis. Of course, one should not be restricted to 100% rational (systems-level) metabolic engineering as some physiological traits of interest can still be better acquired by random mutagenesis. Flexible mind setting is important to develop truly optimized industrial strains. Nevertheless, systems metabolic engineering will be an essential paradigm in industrial strain development. The three representative examples employing these strategies are described in the next section.

4.2 Case Study 1: Combined Genome, Transcriptome and Proteome Analysis on a Parent Strain and an L-Threonine-Overproducing Mutant Strain

To expand further the metabolic capacity and to extract more accurate metabolic information from the strains developed by classical means, combined omics analyses have been carried out, thereby providing potential targets for further engineering [58]. Combined genome, transcriptome, and proteome analysis was carried out between the parent strain *E. coli* W3110 and the classically constructed L-threonine-overproducer *E. coli* TF5015 to understand the regulatory mechanisms of L-threonine production and the physiological changes in the mutant strain. Genes involved in the glyoxylate shunt, the tricarboxylic acid cycle, and amino acid biosynthesis, were found to be significantly upregulated, whereas the ribosomal protein genes were downregulated. Furthermore, two important mutations in the *thrA* (*thrA345*) and *ilvA* (*ilvA97*) genes were identified to be essential for the overproduction of L-threonine. This is a good example of the utilization of combined omics analysis to elucidate metabolic and regulatory information of the L-threonine producing mutant strain.

4.3 Case Study 2: Systems Metabolic Engineering of E. coli for L-Threonine Production

More recently, L-threonine overproducing strain of *E. coli* has been successfully developed by taking systems metabolic engineering approaches including transcriptome analysis and in silico flux response analysis [7]. The detailed strategy is shown in Fig. 2. After all the negative regulations such as feedback inhibition and transcriptional attenuation regulation were removed by genome engineering, further improvement was made based on transcriptome profiling. The desired levels of two anaplerotic enzymes, phosphoenolpyruvate carboxylase (PPC) and isocitrate lyase (ICL), which were selected as targets for overexpression from transcriptome data, were examined by in silico flux response analysis. The optimized coamplification of PPC and ICL increased the L-threonine production by 51.4%, suggesting that these effects are additive. Further improvement was obtained through transporter and exporter engineering. The fed-batch fermentation process was also improved by reducing acetate production with the help of in silico flux response analysis. The final engineered *E. coli* strain was able to produce L-threonine with an impressively high yield of 0.393 g L-threonine per gram of glucose, and 82.4 g L^{-1} L-threonine by fed-batch culture. This is the first example of finely modulating the metabolic fluxes to desired levels based on the in silico genome-scale flux analysis using the target genes identified by transcriptome profiling. These results also suggest that an industrially competitive strain can be successfully constructed by systems metabolic engineering.

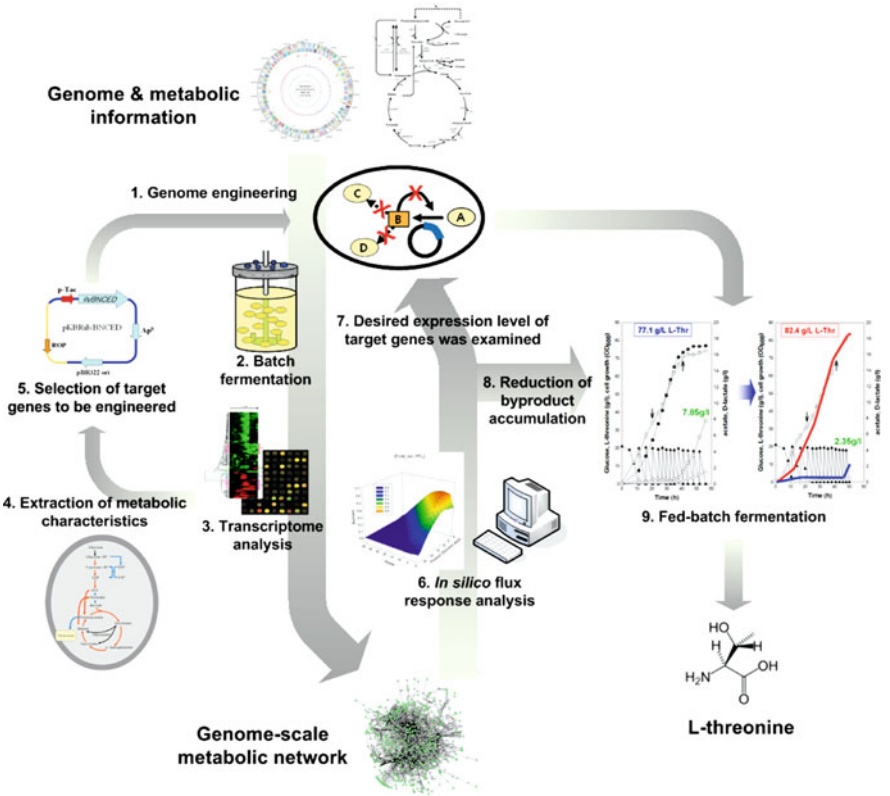


Fig. 2 Strategies for the development of L-threonine production strain by systems metabolic engineering. The L-threonine-producing base strain was constructed based on known metabolic and regulatory information. Further target genes to be engineered were identified by transcriptome profiling, and their desired expression levels were examined by in silico flux response analysis. During the fed-batch fermentation of the engineered strain, acetic acid accumulated considerably and lowered L-threonine production. To solve this problem, in silico flux response analysis was performed, and the *acs* gene was selected as the target gene to be overexpressed. The final engineered strain was able to produce 82.4 g L^{-1} L-threonine by fed-batch culture

4.4 Case Study 3: Construction of L-Threonine Production Strain Using a Reduced-Genome Strain of E. coli and Its Metabolic Characterization by Transcriptome Analysis

As mentioned in Sect. 2.1, a reduced-genome strain, constructed by deleting nonessential genes, was found to be beneficial for protein production [16]. A recent report describes the construction of the L-threonine production strain using a reduced-genome strain of *E. coli* [59]. Posfai et al. [15] constructed a reduced-genome strain of *E. coli* (MDS42), which shows genetic stability and robust

metabolic performance, by removing nonessential genes. The MDS42 strain was engineered to produce L-threonine. For this purpose, a feedback-resistant L-threonine operon (*thrA*BC*) was overexpressed. Next, the pathway that competes with L-threonine formation was removed by deleting the *tdh* (encoding L-threonine dehydrogenase) gene. Then *tdcC* and *sstT* genes encoding L-threonine transporter were deleted to prevent uptaking extracellular L-threonine into the cell. Finally, a mutant L-threonine exporter (*rhtA23*) was introduced into the MDS42. The resulting MDS-205 strain was able to produce 40.1 g L^{-1} L-threonine, which is 83% higher than that obtained with a wild-type *E. coli* strain MG1655 engineered with the same L-threonine-specific modifications (MG-105). To understand the altered global gene expression levels in MG-105 and MDS-205, comparative transcriptome analysis was performed. Most of the genes involved in the central metabolism and L-threonine biosynthesis were found to be upregulated in MDS-205. Increased L-threonine production in MDS42 might result from a decrease in the metabolic burden due to genome reduction. The result obtained in this report demonstrates that a reduced-genome strain of *E. coli* can serve as an efficient host strain for the production of L-threonine and probably other useful bioproducts; however, the performance will be highly dependent on the way of reducing genome and the product of interest.

5 Conclusions

The importance of systems biology in strain development is becoming more accepted by the community with the emergence of increasing numbers of successful examples. By combining multiple omics and computational approaches, systems biology provides valuable information for a thorough understanding of cellular physiology, which significantly contributes to the designing of the strain. Omics analysis including transcriptome and proteome profiling allows identification of gene targets to be engineered based on the comparison of the protein and mRNA contents obtained under different genotypic and/or environmental conditions. There is, however, no general rule for selecting the gene targets out of comparative transcriptome or proteome profiling studies. Researchers need to analyze carefully the data obtained under different conditions. The target genes to be amplified can often be found from the downregulated genes. However, they can sometimes be found from the upregulated genes as well. The selection of strategy should be based on rational thinking that covers cell growth, product formation, reduced by-product formation, etc., all together. This means that even the omics analysis still requires some trial-and-error type approaches to select the target genes to be engineered. Development of a rational way of selecting the target genes to be engineered is being actively pursued by many researchers around the world. Computational approaches have gone through remarkable breakthroughs in metabolic engineering and systems biology, particularly with the development of genome-scale metabolic models. Of course, genome-scale metabolic models that

have been employed in metabolic engineering are not perfect as they do not cover all the metabolic reactions and lack most regulatory constraints. It is expected to be improved through upgrading the metabolic reactions and including regulatory circuits as our knowledge advances [60]. In conclusion, we are starting to see the bright future of applying systems biology to metabolic and bioprocess engineering as successfully demonstrated by the development of several industrial strains. Systems metabolic engineering will likely become an essential practice in developing strains in the near future.

Acknowledgments Our work described in this chapter was supported by the Korea–Australia Collaborative Research Project on the Development of Sucrose-Based Bioprocess Platform (N02071165) from the Ministry of Knowledge Economy and by the Korean Systems Biology Program from the Ministry of Education, Science and Technology (No. M10309020000-03B5002-00000). Further support by the World Class University (WCU) program from the Ministry of Education, Science and Technology, LG Chem Chair Professorship and Microsoft is appreciated.

References

1. Lee SY, Lee DY, Kim TY (2005) Systems biotechnology for strain improvement. Trends Biotechnol 23:349–358
2. Park JH, Lee SY (2008) Towards systems metabolic engineering of microorganisms for amino acid production. Curr Opin Biotechnol 19:454–460
3. Park JH, Lee SY, Kim TY et al (2008) Application of systems biology for process development. Trends Biotechnol 26:404–412
4. Bailey JE (1991) Toward a science of metabolic engineering. Science 252:1668–1675
5. Lee SY, Papoutsakis ET (1999) Metabolic engineering. Marcel Dekker, New York
6. Stephanopoulos G, Aristidou AA, Nielsen J (1998) Metabolic engineering: principles and methodologies. Academic, San Diego, USA
7. Lee KH, Park JH, Kim TY et al (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149
8. Park JH, Lee KH, Kim TY et al (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 USA 104:7797–7802
9. Ohnishi J, Katahira R, Mitsuhashi S et al (2005) A novel *gnd* mutation leading to increased L-lysine production in *Corynebacterium glutamicum*. FEMS Microbiol Lett 242:265–274
10. Ikeda M, Ohnishi J, Hayashi M et al (2006) A genome-based approach to create a minimally mutated *Corynebacterium glutamicum* strain for efficient L-lysine production. J Ind Microbiol Biotechnol 33:610–615
11. Lee SJ, Song H, Lee SY (2006) Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. Appl Environ Microbiol 72:1939–1948
12. Ruckert 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 Biotechnol 104:213–228
13. McHardy AC, Tauch A, Rückert C et al (2003) Genome-based analysis of biosynthetic aminotransferase genes of *Corynebacterium glutamicum*. J Biotechnol 104:229–240
14. Graham JE, Bryant DA (2008) The biosynthetic pathway for synechoxanthin, an aromatic carotenoid synthesized by the euryhaline, unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. J Bacteriol 190:7966–7974

15. Posfai G, Plunkett G III, Feher T et al (2006) Emergent properties of reduced-genome *Escherichia coli*. Science 312:1044–1046
16. Sharma SS, Blattner FR, Harcum SW (2007) Recombinant protein production in an *Escherichia coli* reduced genome strain. Metab Eng 9:133–141
17. Pop M, Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. Trends Genet 24:142–149
18. Choi JH, Lee SJ, Lee SJ et al (2003) Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. Appl Environ Microbiol 69:4737–4742
19. Stansen C, Uy D, Delaunay S et al (2005) Characterization of a *Corynebacterium glutamicum* lactate utilization operon induced during temperature-triggered glutamate production. Appl Environ Microbiol 71:5920–5928
20. Hayashi M, Ohnishi J, Mitsuhashi S et al (2006) Transcriptome analysis reveals global expression changes in an industrial L-lysine producer of *Corynebacterium glutamicum*. Biosci Biotechnol Biochem 70:546–550
21. Hayashi M, Mizoguchi H, Ohnishi J et al (2006) A *leuC* mutation leading to increased L-lysine production and rel-independent global expression changes in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 72:783–789
22. Sindelar G, Wendisch VF (2007) Improving lysine production by *Corynebacterium glutamicum* through DNA microarray-based identification of novel target genes. Appl Microbiol Biotechnol 76:677–689
23. Gasser B, Sauer M, Maurer M et al (2007) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl Environ Microbiol 73:6499–6507
24. Hibi M, Yukitomo H, Ito M et al (2007) Improvement of NADPH-dependent bioconversion by transcriptome-based molecular breeding. Appl Environ Microbiol 73:7657–7663
25. Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9:258–267
26. Inaoka T, Satomura T, Fujita Y et al (2008) Novel gene regulation mediated by overproduction of secondary metabolite neotrehalosadiamine in *Bacillus subtilis*. FEMS Microbiol Lett. doi:10.1111/j.1574-6968.2008.01450.x
27. Aldor IS, Krawitz DC, Forrest W et al (2005) Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. Appl Environ Microbiol 71:1717–1728
28. Lee JW, Lee SY, Song H et al (2006) The proteome of *Mannheimia*, a capnophilic rumen bacterium. Proteomics 6:3550–3566
29. Xia X-X, Han M-J, Lee SY et al (2008) Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. Proteomics 8:2089–2103
30. Oldiges M, Lütz S, Pflug S et al (2007) Metabolomics: current state and evolving methodologies and tools. Appl Microbiol Biotechnol 76:495–511
31. Wittmann C, Kiefer P, Zelder O (2004) Metabolic fluxes in *Corynebacterium glutamicum* during lysine production with sucrose as carbon source. Appl Environ Microbiol 70:7277–7287
32. Kiefer P, Heinzle E, Zelder O et al (2004) Comparative metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* cultured on glucose or fructose. Appl Environ Microbiol 70:229–239
33. Becker J, Klopprogge C, Herold A et al (2007) Metabolic flux engineering of L-lysine production in *Corynebacterium glutamicum*-over expression and modification of G6P dehydrogenase. J Biotechnol 132:99–109
34. Becker J, Klopprogge C, Zelder O et al (2005) Amplified expression of fructose 1, 6-bisphosphatase in Corynebacterium glutamicum increases in vivo flux through the pentose phosphate pathway and lysine production on different carbon sources. Appl Environ Microbiol 71:8587–8596

35. Wang Q, Chen X, Yang Y et al (2006) Genome-scale *in silico* aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. Appl Microbiol Biotechnol 73:887–894
36. Yoon SH, Han M-J, Lee SY et al (2003) Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnol Bioeng 81:753–767
37. Askenazi M, Driggers EM, Holtzman DA et al (2003) Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nat Biotechnol 21:150–156
38. Krömer JO, Sorgenfrei O, Klopprogge K et al (2004) In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J Bacteriol 186:1769–1784
39. Hong SH, Kim JS, Lee SY et al (2004) The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. Nat Biotechnol 22:1275–1281
40. Lee SJ, Lee D-Y, Kim TY et al (2005) 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
41. Alper H, Jin YS, Moxley JF et al (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab Eng 7:155–164
42. Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23:612–616
43. Hüser AT, Chassagnole C, Lindley ND et al (2005) Rational design of a *Corynebacterium glutamicum* pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. Appl Environ Microbiol 71:3255–3268
44. Ishii N, Nakahigashi K, Baba T et al (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 316:593–597
45. Takors R, Bathe B, Rieping M et al (2007) Systems biology for industrial strains and fermentation processes-example: amino acids. J Biotechnol 129:181–190
46. Ikeda M (2003) Amino acid production processes. Adv Biochem Eng Biotechnol 79:1–35
47. Komatsubara S, Kisumi M, Murata K et al (1978) Threonine production by regulatory mutants of *Serratia marcescens*. Appl Environ Microbiol 35:834–840
48. Miwa K, Tsuchida T, Kurahashi O (1983) Construction of L-threonine overproducing strains of *Escherichia coli* K-12 using recombinant DNA techniques. Agric Biol Chem 47:2329–2334
49. Ishida M, Kawashima H, Sato K et al (1994) Factors improving L-threonine production by a three L-threonine biosynthetic genes-amplified recombinant strain of *Brevibacterium lactofermentum*. Biosci Biotech Biochem 58:768–770
50. Shimizu E, Oosumi T, Heima H et al (1995) Culture conditions for improvement of L-threonine production using a genetically self-cloned L-threonine hyperproducing strain of *Escherichia coli* K-12. Biosci Biotech Biochem 59:1095–1098
51. Masuda M, Takamatsu S, Nishimura N et al (1992) Improvement of nitrogen supply for L-threonine production by a recombinant strain of *Serratia marcescens*. Appl Biochem Biotechnol 37:255–265
52. Lee MH, Lee HW, Park JH et al (2006) Improved L-threonine production of *Escherichia coli* mutant by optimization of culture conditions. J Biosci Bioeng 101:127–130
53. Okamoto K, Kino K, Ikeda M (1997) Hyperproduction of L-threonine by an *Escherichia coli* mutant with impaired L-threonine uptake. Biosci Biotech Biochem 61:1877–1882
54. Kruse D, Krämer R, Eggeling L et al (2002) Influence of threonine exporters on threonine production in *Escherichia coli*. Appl Microbiol Biotechnol 59:205–210
55. Livshits VA, Zakataeva NP, Aleshin VV et al (2003) Identification and characterization of the new gene *rhtA* involved in threonine and homoserine efflux in *Escherichia coli*. Res Microbiol 154:123–135

56. 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
57. Ogawa-Miyata Y, Kojima H, Sano K (2001) Mutation analysis of the feedback inhibition site of aspartokinase III of *Escherichia coli* K-12 and its use in L-threonine production. Biosci Biotech Biochem 65:1149–1154
58. Lee JH, Lee DE, Lee BU et al (2003) Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. J Bacteriol 185:5442–5451
59. Lee JH, Sung BH, Kim MS et al (2009) Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb Cell Fact 8:2
60. Akesson M, Forster J, Nielsen J (2004) Integration of gene expression data into genome-scale metabolic models. Metab Eng 6:285–293

Adv Biochem Engin/Biotechnol (2010) 120: 21–49
DOI: 10.1007/10_2009_58

Published online: 6 February 2010

Analysis and Engineering of Metabolic Pathway Fluxes in *Corynebacterium glutamicum*

Christoph Wittmann

Dedicated to Prof. Dr. Dietmar Hempel on the occasion of his 65th birthday

Abstract The Gram-positive soil bacterium *Corynebacterium glutamicum* was discovered as a natural overproducer of glutamate about 50 years ago. Linked to the steadily increasing economical importance of this microorganism for production of glutamate and other amino acids, the quest for efficient production strains has been an intense area of research during the past few decades. Efficient production strains were created by applying classical mutagenesis and selection and especially metabolic engineering strategies with the advent of recombinant DNA technology. Hereby experimental and computational approaches have provided fascinating insights into the metabolism of this microorganism and directed strain engineering. Today, *C. glutamicum* is applied to the industrial production of more than 2 million tons of amino acids per year. The huge achievements in recent years, including the sequencing of the complete genome and efficient post genomic approaches, now provide the basis for a new, fascinating era of research – analysis of metabolic and regulatory properties of *C. glutamicum* on a global scale towards novel and superior bioprocesses.

Keywords ^{13}C metabolic flux analysis, Anaplerosis, Genome-scale model, Glutamate, Lysine, Metabolic engineering, NADPH, Systems biology, Systems metabolic engineering, Tryptophan

Contents

1 Introduction 22
2 Metabolism of *Corynebacterium glutamicum* 23
2.1 Nutritional Requirements and Assimilatory Pathways 23
2.2 Central Carbon Metabolism 23

C. Wittmann
Institute of Biochemical Engineering, Technische Universität Braunschweig, Gaußstrasse 17, 38106, Braunschweig, Germany
e-mail: c.wittmann@tu-braunschweig.de

2.3 Biosynthetic Pathways Towards Biotechnology Products 25
2.4 Genome Sequencing and Metabolic Network Reconstruction 27
3 In Silico Metabolic Network Analysis 28
3.1 Genome-Scale Reconstruction of the Metabolic Network 28
3.2 Stoichiometric Modeling 29
3.3 Dynamic Modeling Approaches 32
4 Analysis of Metabolic Fluxes 32
4.1 Tools and Concepts 33
4.2 Metabolic Fluxes in *C. glutamicum* 34
5 Metabolic Pathway Engineering 38
5.1 Metabolic Engineering of Lysine Production 38
5.2 Metabolic Engineering of Aromatic Amino Acid Production 40
5.3 Metabolic Engineering of Glutamate Production 41
5.4 Utilization of Alternative Substrates 41
5.5 Global Strain Engineering Through Applied Systems Biology 42
5.6 Towards Novel Products 43
References 44

1 Introduction

C. glutamicum was discovered about 50 years ago. Stimulated by the increasing demand for L-glutamate as a flavor enhancer in human nutrition, a screening program in Japan led to the isolation of a soil microorganism, later classified as *C. glutamicum*, which was able to accumulate L-glutamate in the growth medium [1, 2]. Subsequent analysis revealed that glutamate secretion could be triggered by a limited supply of biotin, opening the possibility for industrial production. The application potential of *C. glutamicum* in biotechnology soon increased with successful manipulation of key regulatory properties that allowed accumulation of other amino acids such as lysine, threonine, arginine, or ornithine [3]. These findings initiated intensive research over the past decades. One of the key areas of research was the development of efficient production strains of *C. glutamicum* for which different strategies were pursued. This included classical mutagenesis and selection. With the advent of recombinant DNA technology, targeted genetic engineering of *C. glutamicum* became available and was successfully used to derive improved production properties. Today, *C. glutamicum* is applied to the industrial production of glutamate as flavor enhancer in human nutrition (1.5 million tons per year), [4], lysine (850,000 tons per year) [5] as well as tryptophan (10,000 tons per year) as additive in animal feed [6]. The improvement of existing production strains focuses on higher yield and productivity, better stress tolerance, or a broader substrate spectrum. In addition, current research also aims at the creation of efficient processes for biobased production of novel products involving other amino acids, organic acids, or biofuels. In this regard the present chapter highlights the major contributions dealing with investigation and optimization of the metabolic network of *C. glutamicum* for biotechnology. Hereby, concepts and approaches for systems biology analysis and engineering of metabolic pathways of *C. glutamicum* are presented and illustrated with actual examples.

2 Metabolism of *Corynebacterium glutamicum*

The extensive biochemical and physiological analysis of *C. glutamicum* during the past as reviewed in a recent handbook on *C. glutamicum* [7] provides a rich source of information on many of the enzymes and pathways present in this organism.

2.1 Nutritional Requirements and Assimilatory Pathways

C. glutamicum can use a multitude of organic compounds as sole carbon source or via coutilization. Sugars such as glucose, fructose, sucrose, or mannose are taken up by a phosphotransferase system [8, 9]. For glucose, an alternative uptake system involving intracellular phosphorylation of glucose by a glucokinase has been identified [10]. In addition, *C. glutamicum* can grow on different organic acids such as gluconate [11], lactate, [12], acetate [13], propionate [14], or citrate [15]. For growth and also for amino acid overproduction, cells further require nitrogen and sulfur as major elements in addition to carbon. Suitable nitrogen sources are ammonium or organic compounds such as urea or amino acids [16]. For the uptake of ammonium, two alternative systems are present. At high ammonium levels, assimilation is mainly catalyzed by the glutamate dehydrogenase (GDH), whereas a highly affine system comprising glutamine synthetase and GDH is active under low ammonium levels, although at higher energy expenditure [17]. Among suitable sulfur sources, inorganic sulfate is most common. Its assimilation, however, involves a high demand of redox power for reduction to the biologically compatible sulfide [18]. Other sulfur sources assimilated by *C. glutamicum* are cysteine, sulfonates, or sulfonate esters [18, 19]. With respect to industrial application the most important raw materials are based on molasses (Asia) and starch (America, Europe) applied together with inorganic salts [5]. Thus, the major nutrients utilized are glucose, sucrose and fructose (carbon), ammonia (nitrogen), and sulfate (sulfur).

2.2 Central Carbon Metabolism

Today the metabolic network of the central metabolism of *C. glutamicum* involving glycolysis, pentose phosphate pathway (PPP), TCA cycle as well as anaplerotic and gluconeogenetic reactions is well known (Fig. 1). Different enzymes are involved in the interconversion of carbon between TCA cycle (malate/oxaloacetate) and glycolysis (pyruvate/phosphoenolpyruvate). For anaplerotic replenishment of the TCA cycle, *C. glutamicum* exhibits pyruvate carboxylase [20] and phosphoenolpyruvate (PEP) carboxylase as carboxylating enzymes. Malic enzyme [21] and PEP carboxykinase [22, 23] catalyze decarboxylation reactions from the TCA cycle

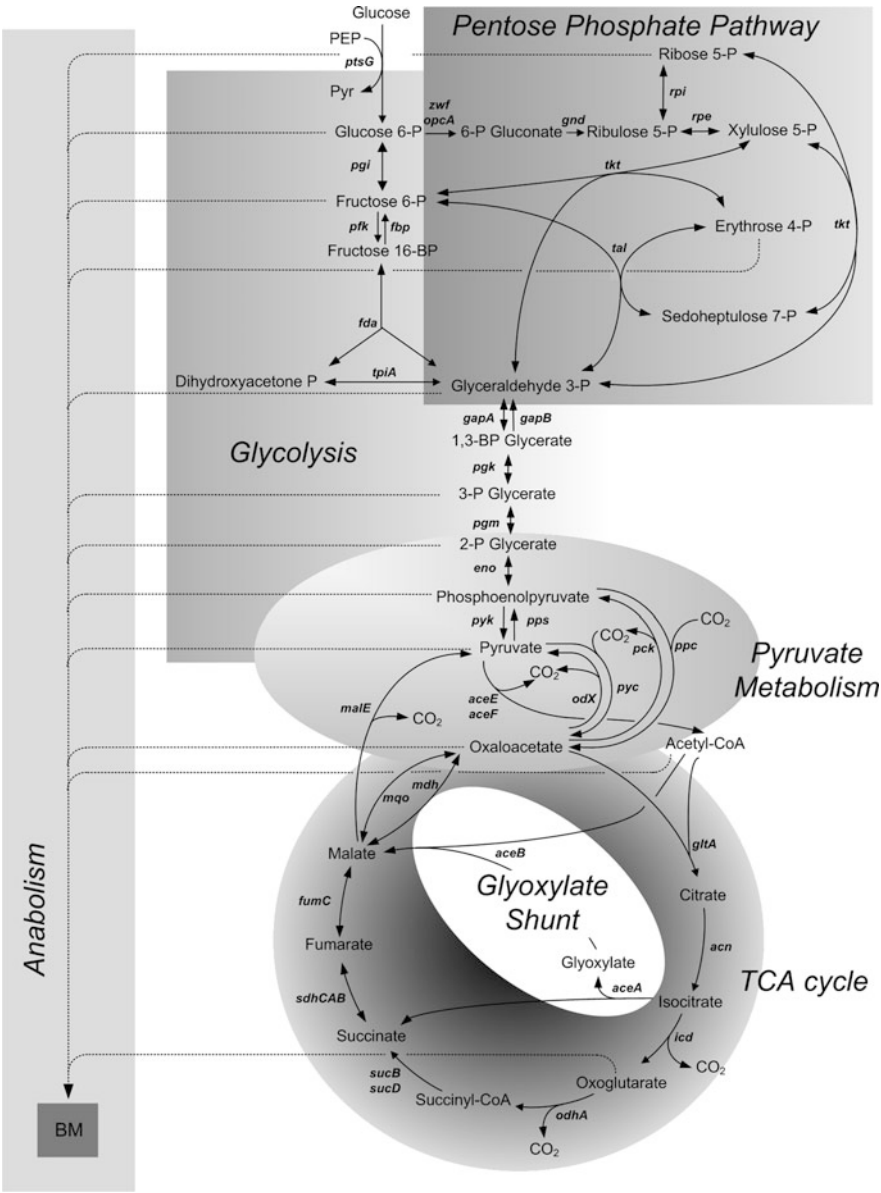


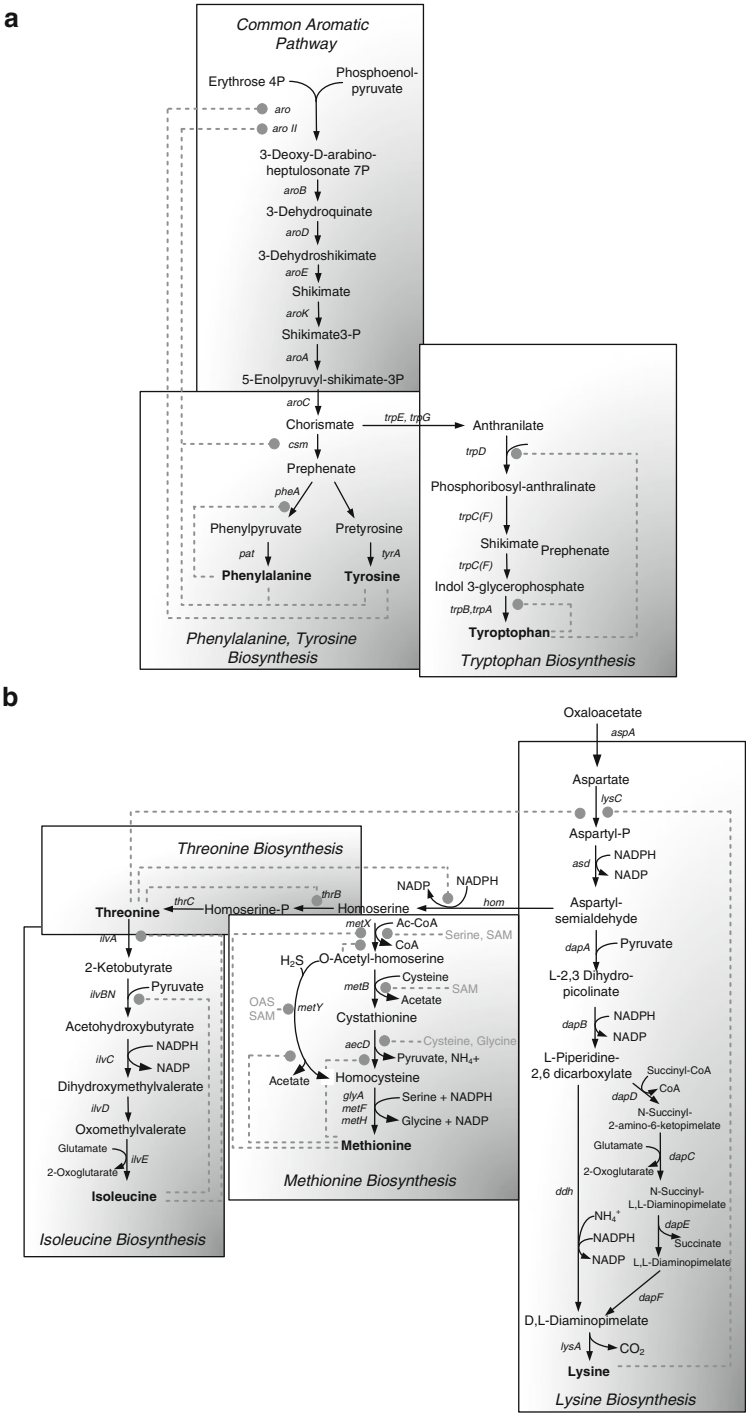
Fig. 1 Central metabolic pathways in *C. glutamicum*

towards glycolysis. As additional gluconeogenetic enzymes, oxaloacetate decarboxylase [24] and PEP synthetase [25] have been proposed. It has been proposed that cyclic cooperation of these carboxylating and decarboxylating enzymes around the pyruvate node is involved in the regeneration of excess ATP [22, 26, 27].

Different reactions are linked to the supply of NADPH. The major enzymes are glucose 6-phosphate dehydrogenase [28, 29] and 6-phosphogluconate dehydrogenase [28, 30] in the oxidative part of the PPP and the TCA cycle enzyme isocitrate dehydrogenase [31]. In selected cases, NADPH supply might involve malic enzyme [32, 33]. In addition to the central catabolism, anabolic routes have also been elucidated. Through rigorous analysis of the cellular composition, detailed information on the anabolic precursor demand was obtained [34]. These analyses were mainly driven by approaches for metabolic flux analysis in growing *C. glutamicum*, where anabolic reactions drain carbon, cofactors, and energy from central catabolism and thus have to be considered. In total, about (16.4 mmol NADPH) (g biomass)$^{-1}$ is required for anabolism. Considering a biomass yield of 0.5 (g dry biomass) (g glucose)$^{-1}$, which is achieved by *C. glutamicum* under aerobic conditions, this results in 1.7 mol NADPH (mol glucose)$^{-1}$ that have to be generated by the NADPH forming reactions in the PPP and the TCA cycle. This substantial anabolic NADPH requirement competes with demand for production pathways, e.g., towards lysine or methionine.

2.3 Biosynthetic Pathways Towards Biotechnology Products

Among the most important amino acids for industrial production are glutamate, the aromatic amino acids, and amino acids belonging to the aspartate family. The biosynthesis of these amino acids is closely linked to the central metabolism (Fig. 2). It is strictly controlled at several steps in *C. glutamicum* as is typically the case in microbial systems [35]. Obviously, the required precursor metabolites, cofactors and energy have to be supplied in appropriate amounts by the central catabolic routes, thus competing with the cellular requirements for growth. Particular problems with the synthesis of amino acids from the aspartate family are the long biosynthetic pathways and the highly connected network involving intermediates and reactions that are shared at the same time by different biosynthetic routes. Therefore, it is not surprising that multiple regulatory steps are required to ensure the balanced synthesis of all these metabolites for cellular demands (Fig. 2). An important branch point occurs at the level of aspartate semialdehyde, where the biosynthetic pathways separate. Upstream of this node is a key point for flux control of lysine biosynthesis. The responsible enzyme is aspartokinase, which catalyzes the formation of aspartylphosphate from aspartate. It is subjected to feedback inhibition by lysine and threonine [36, 37]. Downstream of this node, *C. glutamicum* exhibits a dual pathway for lysine biosynthesis, providing an increased flexibility in response to changing environmental conditions [38, 39]. The synthesis of methionine involves complex networks of regulatory interactions and metabolic pathways. Even with over-expressing of almost all genes involved in methionine biosynthesis via deletion of the central repressor McbR, *C. glutamicum* does not over-produce this amino acid as shown recently, pointing at other (regulatory) mechanisms that are limiting [40]. Due to extended research, complex and efficient


a
Common Aromatic Pathway
Erythrose 4P
Phosphoenol-pyruvate
aro
aro II
3-Deoxy-D-arabino-heptulosonate 7P
aroB
3-Dehydroquinate
aroD
3-Dehydroshikimate
aroE
Shikimate
aroK
Shikimate3-P
aroA
5-Enolpyruvyl-shikimate-3P
aroC
Chorismate
trpE, trpG
Anthranilate
csm
Prephenate
trpD
Phosphoribosyl-anthralinate
pheA
trpC(F)
Phenylpyruvate
Pretyrosine
Shikimate
Prephenate
pat
tyrA
trpC(F)
Phenylalanine
Tyrosine
Indol 3-glycerophosphate
trpB,trpA
Tyroptophan
Phenylalanine, Tyrosine Biosynthesis
Tryptophan Biosynthesis
b
Oxaloacetate
aspA
Aspartate
lysC
Aspartyl-P
asd
NADPH
NADP
Aspartyl-semialdehyde
hom
NADP
NADPH
Threonine Biosynthesis
thrB
thrC
Threonine
Homoserine-P
Homoserine
dapA
Pyruvate
L-2,3 Dihydro-picolinate
dapB
NADPH
NADP
L-Piperidine-2,6 dicarboxylate
Succinyl-CoA
dapD
CoA
N-Succinyl-2-amino-6-ketopimelate
Glutamate
dapC
2-Oxoglutarate
ddh
N-Succinyl-L,L-Diaminopimelate
NH4+
NADPH
NADP
dapE
Succinate
L,L-Diaminopimelate
dapF
D,L-Diaminopimelate
lysA
CO2
Lysine
Lysine Biosynthesis
ilvA
2-Ketobutyrate
ilvBN
Pyruvate
Acetohydroxybutyrate
ilvC
NADPH
NADP
Dihydroxymethylvalerate
ilvD
Oxomethylvalerate
Glutamate
ilvE
2-Oxoglutarate
Isoleucine
Isoleucine Biosynthesis
metX
Ac-CoA
CoA
Serine, SAM
H2S
O-Acetyl-homoserine
Cysteine
metB
SAM
Acetate
OAS
SAM
metY
Cystathionine
aecD
Cysteine, Glycine
Pyruvate, NH4+
Acetate
Homocysteine
glyA
metF
metH
Serine + NADPH
Glycine + NADP
Methionine
Methionine Biosynthesis

feedback inhibition mechanisms have recently been identified in addition to extended transcriptional control [41, 42]. A similar complex picture is yielded for biosynthesis of the aromatic amino acids, where enzyme inhibition by feedback regulation and negative transcriptional control are involved [6, 43].

2.4 Genome Sequencing and Metabolic Network Reconstruction

These previous studies contributed significantly to our current understanding of the physiology of *C. glutamicum*. They each focused on only certain aspects of metabolism and investigated the corresponding reactions as isolated parts, so that an integrated view on metabolism of *C. glutamicum* as a functional network of highly interconnected reactions could not be provided. Such a holistic view, however, is highly desirable to understand the underlying complex system of metabolic and regulatory networks and derive ideas and concepts on how to engineer them in order to achieve superior production strains. As described above, the synthesis of industrial amino acids competes with cellular metabolism for building blocks and cofactors. As an example, pyruvate as precursor for lysine is potentially involved in more than 150 metabolic reactions of *C. glutamicum*. Keeping this in mind, we need a global picture of metabolism to understand the link between all these reactions and find the most efficient strategies to modify the flux to achieve optimum network performance. Metabolic flux analysis displays the first systems oriented approach to unravel the physiology of *C. glutamicum* since it combines experimental data with metabolic network models and allows determining absolute fluxes through larger networks comprising 100–200 reactions of central carbon metabolism [44]. A further milestone towards systems level understanding was the sequencing of the *C. glutamicum* ATCC 13032 genome by at least three different biotechnological companies (BASF, Degussa, and Kyowa Hakko). The circular genome comprises about 3000 genes with a total size of 3.3 kb [45–47]. Subsequent annotation has helped greatly in elucidating the genetic repertoire. Today, the genomic information of *C. glutamicum* is publicly available in bioinformatics data bases. Recommendable databases are provided by KEGG (www.genome.jp) or BioCyc (www.biocyc.com) and contain graphical information on genes, proteins, reactions, and pathways which is very useful when linking metabolites, enzymes and reactions to associate metabolism. The availability of genome information enabled new post genome technologies accelerating strain engineering by systems biology approaches. These comprise experimental and computational tools such as transcriptomics [48–51], proteomics [52–54], or in

Fig. 2 Metabolic pathways in *C. glutamicum* for biosynthesis of the aromatic amino acids tryptophan, tyrosine, and phenylalanine (**a**) and amino acids belonging to the aspartate family including lysine, methionine, threonine, and isoleucine (**b**). Metabolic regulation by feedback inhibition is indicated by *dotted lines*

silico pathway modeling [55, 56]. Meanwhile, sequence information is also available for other closely related species providing a detailed overview on the corynebacterial pan-genome and its metabolic pathways. Sequenced strains include the thermotolerant *C. efficiens* [57], *C. glutamicum* R [58], and the pathogens *C. diphteriae* [59], *C. jeikeium* [60], and *C. kroppenstedtii* [61].

3 In Silico Metabolic Network Analysis

The engineering approach to analysis and design is using a mathematical or computer model. For target identification, modeling approaches are very useful to extract useful information on metabolic networks, their regulation and capacity. Models can be applied to predict phenotype behavior in response to different environmental or genetic perturbations, integrate complex data sets towards systems oriented understanding of network function, or design organisms with optimal network structure and activity. Overall, a wide variety of computational methods exploiting metabolic models have been developed and applied to *C. glutamicum*, yielding valuable insights into its metabolism and providing a sound basis for computer-assisted design in metabolic engineering. The type and complexity of the model hereby depends on the aim of the study. Although the ultimate goal of modeling is the development of dynamic models for the complete simulation of cellular systems, the success of such approaches has been severely hampered by the lack of kinetic information [62]. However, it is possible to assess accurately the theoretical capabilities and operative modes of metabolic systems using stoichiometric models.

3.1 Genome-Scale Reconstruction of the Metabolic Network

With the availability of annotated genome sequences, it has become possible to reconstruct genome-scale biochemical reaction networks for microorganisms. Genome-scale models have been reconstructed for almost 20 bacterial species so far [63–65]. Such models bridge the gap between genome-derived biochemical information and metabolic phenotypes and enable straightforward in silico experiments with whole-cell metabolism. Also for *C. glutamicum*, a genome-scale stoichiometric model was recently created [55]. As for other microorganisms, the construction was carried out in several steps. First a crude model consisting of mass balances for catabolic reactions was assembled from information on the present metabolic pathways and their stoichiometry available in KEGG and BioCyc. Subsequently, the anabolic reactions involved in polymer and biomass synthesis were defined whereby the information was collected from different literature sources. Finally the model was cured and completed by adding missing reactions and integrating information on reaction reversibility. The *C. glutamicum* model was validated against data found in the literature under different conditions such as different biomass production burdens

Table 1 Statistical data on the genome-scale stoichiometric model of *Corynebacterium glutamicum* ATCC 13032 constructed on basis of the genome with 3002 open reading frames [55]

Metabolic reactions	446
Total	213
Biochemical evidence available	209
Clear functional annotation	22
Tentative functional annotation	2
Putative functional annotation	
Metabolites	411
Total	356
Internal metabolites	55
External metabolites (substrates or products)	

and growth on different carbon sources. The model comprises 446 reactions and 411 metabolites (Table 1). Overall, it displays a systematic verification and compilation of data from various sources concerning the metabolic network of *C. glutamicum* and provides a useful basis for in silico studies as well a first step for future integration of complex data sets in systems biology approaches.

3.2 Stoichiometric Modeling

New genome wide stoichiometry based modeling of metabolic pathways is now possible. The stoichiometric data and models available provide the basis for various in silico applications differing in the methodology and algorithms used. Major contributions for *C. glutamicum* have been obtained from flux balancing [66] and elementary (flux) mode analysis [67]. In short, flux balance analysis is a constraint based method to analyze stoichiometric networks. Through introduction of specific biological constraints the set of various possible solutions for the typically underdetermined metabolic system can be limited to a certain solution space. Within this space, optimal steady-state solutions can be calculated by minimizing or maximizing for objective functions such as growth or overproduction of a metabolite using linear programming techniques [68]. Accordingly, using flux balance analysis, a single solution is found to the optimization problem [69]. Hereby, flux balance analysis can highlight the most efficient pathway through the network in order to achieve the particular objective function. Elementary flux mode analysis systematically enumerates all independent minimal pathways through a network, each a unique elementary mode, that are stoichiometrically and thermodynamically feasible [70]. All possible steady-state flux distributions through the metabolic network are nonnegative linear combinations of the set of elementary modes [71]. This can be exploited to extract key properties from metabolic networks such as maximum

network capacity, optimal pathways, network robustness, or phenotype prediction in response to environmental or genetic perturbations [56, 72, 73].

3.2.1 Capacity of the Network of *C. glutamicum* for Lysine Production

For *C. glutamicum* stoichiometric network modeling has been utilized to assess the capacity for the production of lysine as one of the major industrial products [5, 11, 55, 74, 75]. The theoretical maximum molar yield of *C. glutamicum* for lysine production obtained by elementary flux mode analysis is 0.82 mol mol^{-1} [5]. Under these conditions of zero growth, PPP and malic enzyme supply the required NADPH, whereas the TCA cycle is shut off and lysine is exclusively formed via the dehydrogenase branch (Fig. 3a). This mode would require a metabolic cycle of pyruvate carboxylase, malate dehydrogenase, and malic enzyme acting as a transhydrogenase to convert NADH into NADPH. Biochemical evidence for such a cycle has not been obtained so far, but a similar cyclic pathway involving pyruvate carboxylase, malate dehydrogenase, and PEP carboxykinase has recently been shown to operate in vivo in *C. glutamicum* [22, 27]. Omitting the transhydrogenase-like cycle from the network results in a slightly lower maximum lysine yield of 0.75 mol mol^{-1} [77]. Admittedly, all these scenarios imply zero growth which can hardly be realized in a real fermentation. Due to this, the achievable optimum can be expected to be somewhat lower than the value calculated here, but should still be significantly higher than the yield achieved in practice, which is in the range of about 40–50%. Further insight into lysine production has been obtained from flux balance analysis characterizing different in silico mutants and metabolic scenarios [55]. The model described experimental observations from flux analysis of *C. glutamicum* fairly well, but failed in certain rather important aspects. The reason is that biochemical regulation affecting the fluxes cannot be accounted for and reactions with identical overall stoichiometry cannot be resolved using such stoichiometric approaches.

3.2.2 Rational Design of *C. glutamicum* for Methionine Production

Elementary flux mode analysis has been further applied to study possible routes for the production of methionine in *C. glutamicum* [56]. The maximum theoretical methionine yield on glucose was calculated as 0.59 mol mol^{-1} (Fig. 3b). Also here, a detailed insight into the metabolic network is obtained. The PPP flux supporting the high yield is even higher than that for optimal lysine production, which reflects the enormous amount of 8 NADPH required per methionine. Additional simulations identified promising genetic targets for improved production such as heterologous expression of a transhydrogenase or of a glycine cleavage system. Moreover, it could be shown that the supply of reduced sulfur is beneficial for high carbon yield. The most effective sulfur source was methanethiol, allowing an almost complete conversion of glucose into methionine with a yield of 0.91 C-mol

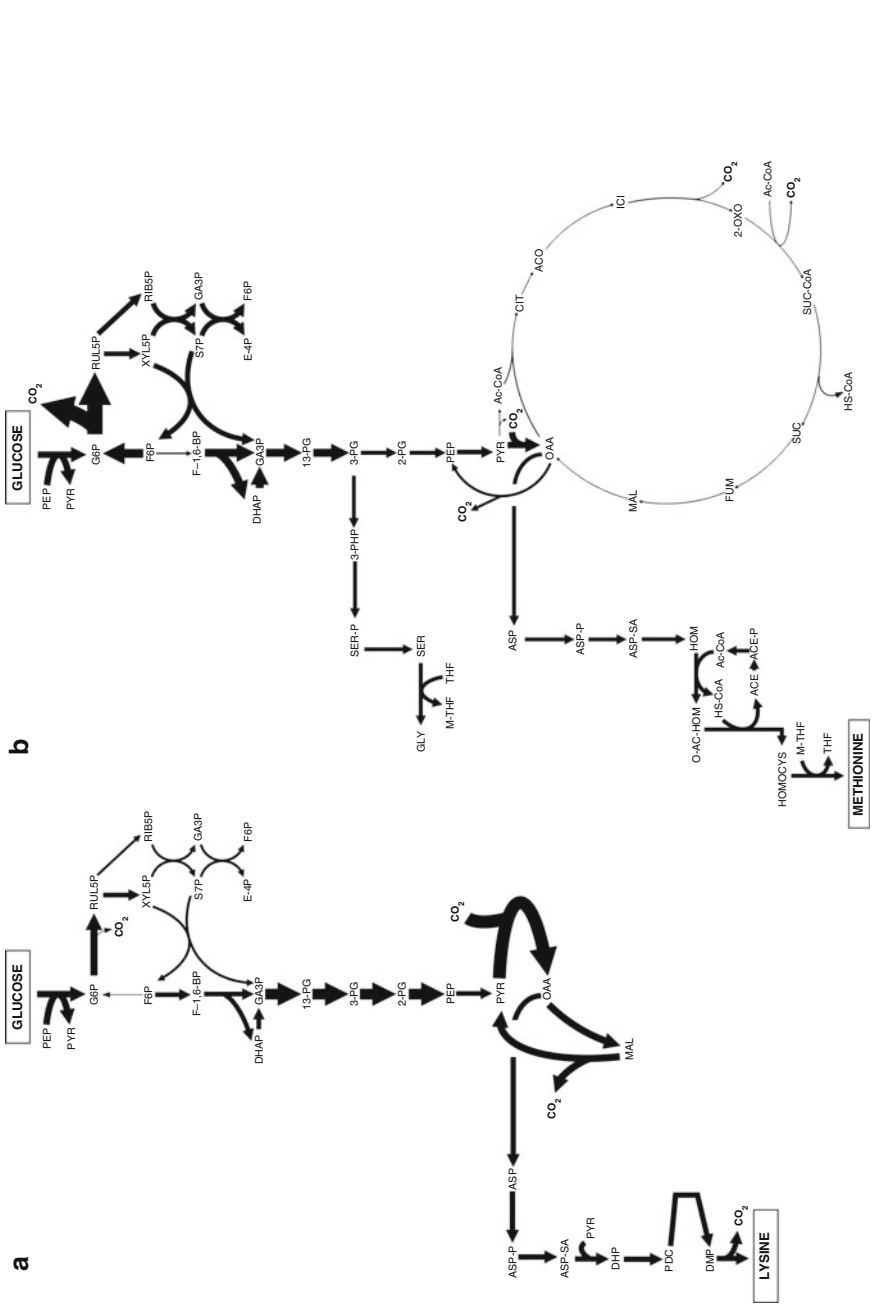


Fig. 3 Metabolic flux distribution of *C. glutamicum* reflecting maximal theoretical carbon yield of lysine (**a**) and methionine (**b**) from glucose. All fluxes are given as relative molar fluxes to the glucose uptake, whereby the relative flux of a reaction is reflected by the thickness of the corresponding arrow. The data are taken from recent simulation studies [5, 76]

C-mol^{-1}. Summarizing, stoichiometric modeling studies for *C. glutamicum* provide a sound basis for strain or process optimization. By calculating the theoretical capacity of an organism for a novel product they also allow useful estimates on the general economical feasibility [56].

3.3 Dynamic Modeling Approaches

Dynamic metabolic network models theoretically allow a most comprehensive and directed optimization of a pathway or of a whole cell. Therefore the ultimate goal of systems biology is the development of dynamic models for the complete simulation of cellular systems. Such dynamic models, however, require the understanding of all interactions influencing the reaction rates in the network. We are still far from that. For *C. glutamicum* detailed dynamic models, based on mechanistic equations for the participating enzymes and concentration measurements of the pathway intermediates involved, have been developed at least for the biosynthesis of lysine [78] and valine [79]. Through metabolic control analysis, such models allow the prediction of bottlenecks and of optimal factors, e.g., enzyme concentrations to achieve increased flux. Admittedly, for most pathways of *C. glutamicum* such kinetic information is still not available. This can be partly overcome by the application of power law or lin-log kinetics in the underlying equations [79, 80]. Dynamic models are also needed for dynamic labeling experiments for flux analysis as has been shown for *E. coli* [81]. The current progress in method development for metabolomics of *C. glutamicum* will surely stimulate the future generation of dynamic models for such organisms by providing extended data sets on intracellular metabolite concentrations required to derive in vivo kinetics [82–85]. For many years, dynamic macroscopic models have been used to describe dynamic phenomena of growth and production and derive control or feeding strategies for optimized bioprocesses. A recent example deals with fed-batch production of valine [79]. Such overall phenomenological models are surely useful but do not provide any detailed description of network activities [86].

4 Analysis of Metabolic Fluxes

In recent years, powerful approaches were developed which allow the quantification of small molecule fluxes through metabolic networks, i.e., in vivo reaction rates. Among the achievements obtained from flux analysis are the identification of novel pathways, the elucidation of metabolic control, identification of targets for rational strain engineering in biotechnology, and first insights into design principles of metabolic networks in systems biology studies. Concerning *C. glutamicum*, metabolic flux analysis provided a fascinating view of its metabolic pathways and offered new possibilities for rational strain engineering as recently reviewed [34, 87].

4.1 Tools and Concepts

4.1.1 Conventional Metabolic Flux Analysis by Stoichiometric Balancing

First flux estimates through larger parts of its metabolism were based on constraining assumed reaction networks with measurement of uptake and production rates [74]. A number of studies utilized stoichiometric balancing to assess the flexibility of the metabolic network [11, 75] and to investigate the influence of environmental conditions such as dissolved oxygen level [88], salt content [89], or nutrient status [12, 90, 91]. However, this conventional approach cannot yield reliable information about parallel or bidirectional reactions and has to rely on balances for NADH or NADPH, which may not be accurate [34]. Moreover, it is limited to derive new conclusions since the results are strongly based on the taken assumptions and not on data [92].

4.1.2 State-of-the-Art ^{13}C Metabolic Flux Analysis and Current Developments

Recent developments combining stoichiometric networks with additional intracellular information from ^{13}C-isotopomer analysis have overcome these limitations and display the state-of-the-art in metabolic flux analysis. Hereby, the feeding of ^{13}C labeled substrates to the growing cells leads to a distribution of the ^{13}C label throughout the metabolic network. The resulting labeling pattern in the intermediary metabolites and other cellular constituents depends on the well defined and well known carbon transfer of the involved reactions and their particular fluxes. As such, the labeling patterns contain the key information about the fluxes and can be used for their estimation. For label quantification, MS or NMR techniques are available, whereby MS has evolved to the preferred method due to its high accuracy, sensitivity, speed, and robustness [93, 94]. The most common method hereby exploits the ^{13}C pattern in proteinogenic amino acids as a rich source of labeling information for the determination of steady-state fluxes in growing cells [95, 96]. Respirometric flux analysis, conceptually based on sole labeling measurement of CO_2 extended the application of flux analysis to nongrowing cells [97–99]. Additionally, labeling measurement of intracellular metabolites has also received increasing attraction since it offers the possibility to resolve fluxes under dynamic conditions [81, 100]. After a decade of intense research and development, such ^{13}C-based flux methods can routinely track steady-state fluxes in microbes [92]. Convenient software tools can be used to calculate the intracellular fluxes from the labeling data utilizing either global parameter fitting from the entire labeling data set [101] or the estimation of local flux ratios from selected labeling patterns [102].

4.2 Metabolic Fluxes in *C. glutamicum*

With respect to metabolic fluxes, *C. glutamicum* is probably the most extensively studied organism so far [34]. Metabolic flux analysis in this organism has become a fundamental tool of strain engineering, since it provides otherwise not accessible key data on cellular function and regulation. The first pioneering studies of *C. glutamicum* using stable isotopes to elucidate metabolic properties were initiated soon after its isolation [103]. Focusing each on selected reaction, these labeling studies identified the dual lysine pathway in *C. glutamicum* [104] or provided first estimates on central pathways, i.e., citric acid cycle, anaplerosis, and glyoxylate shunt [105]. The major findings on metabolic fluxes and their regulation in *C. glutamicum* result from the fully integrated approaches comprising stoichiometric and isotopomer balancing with ^{13}C labeling and thus are quite recent (Table 2).

4.2.1 Metabolic Fluxes of Precursor Metabolism

The synthesis of relevant amino acids such as lysine, threonine, methionine, or glutamate demands for precursor compounds stemming from the TCA cycle. This has stimulated intensive research on fluxes involved in replenishment of the TCA cycle, involving pyruvate carboxylase, PEP carboxylase as carboxylating enzymes, as well as malic enzyme and PEP carboxykinase that catalyze decarboxylation reactions from the TCA cycle towards glycolysis. Metabolic flux analysis unraveled the metabolism around the pyruvate node in great detail (Fig. 4a). It could be shown that carboxylating and decarboxylating enzymes are active at the same time, forming a cyclic pathway for interconversion of C_4 metabolites of the TCA cycle and C_3 metabolites of glycolysis [27]. This cyclic pathway is highly flexible, as cells can redistribute the flux depending on the metabolic burden (Fig. 4b–d). The role of this cycle has been attributed to regeneration of excess ATP under certain conditions [121, 122] or equilibration of metabolite levels around the pyruvate node [123]. Additionally, a contribution to NADPH metabolism has been suggested [12, 124]. Moreover, flux studies identified the enzymes around the pyruvate node as key targets for metabolic engineering of the precursor supply in *C. glutamicum*. An increase of flux through the lysine pathway is linked to an increase of flux through the carboxylating enzymes, providing the lysine precursor oxaloacetate (Fig. 5a). In contrast, the flux through the decarboxylating enzymes, withdrawing oxaloacetate, is negatively correlated with lysine production (Fig. 5b). In light of these findings, amplification of the carboxylating and deletion of the decarboxylating enzymes became a promising strategy to enhance the anaplerotic net flux for improved production.

Beyond lysine, the close link of anaplerotic replenishment of the TCA cycle and production could be also observed for the formation of glutamate, demanding for the TCA cycle precursor α-ketoglutarate (Fig. 5c). Regarding the supply of α-ketoglutarate as product precursor, flux analysis revealed that the specific activity of isocitrate dehydrogenase (ICDH) and GDH did not show large changes

Table 2 Recent key contributions of ^{13}C metabolic flux analysis to *Corynebacterium glutamicum*. The compilation of studies is partly adapted from an actual review on metabolic flux analysis in bioprocess development [87]

Product	Substrate	Scope of the study	References
Lysine	Glucose	Production process monitoring using a ^{13}C-sensor reactor coupled to the production process	[106]
Lysine	Glucose	Conceptual work on systematic quantification of metabolic fluxes using MS	[107]
Lysine	Glucose	Investigation of metabolic fluxes in phosphoglucose isomerase mutants	[108]
Lysine	Glucose Fructose	Serial flux mapping of fed-batch cultures using the sensor reactor approach	[109]
Lysine	Glucose	Comparative metabolic flux analysis on glucose and fructose as carbon source	[110]
Lysine	Glucose	Systems biology profiling of a production process with transcriptome, metabolome, and fluxome analysis	[76]
Lysine	Sucrose	Metabolic fluxes in sucrose-grown cells	[111]
Lysine	Glucose	Metabolic flux analysis at a miniaturized scale enabling high-throughput flux analysis	[112]
Lysine	Glucose	Investigation of the effects of amplified FBPase expression on metabolic fluxes	[113]
Lysine	Glucose	Investigation of the effects of a novel *gnd* mutation on lysine production	[30]
Lysine	Glucose	In vivo flux estimation using respirometric metabolic flux analysis with sole labeling analysis of CO_2	[114]
Lysine	Glucose	Flux profiling in response to feedback deregulation of lysine biosynthesis	[115]
Lysine	Glucose	Investigation of the effects of overexpression and modification of G6P dehydrogenase	[116]
Glutamate	Glucose	Metabolic flux analysis of glutamate production	[117]
Glutamate	Glucose	Investigation of the roles of anaplerotic pathways in glutamate overproduction	[119]
Methionine	Glucose	Identification and flux analysis through of a novel pathway for isoleucine biosynthesis	[40]
Methionine	Glucose	Flux analysis under conditions of oxidative stress resulting from deletion of the global repressor McbR	[119]
Lysine	Glucose	Flux response to deletion of pyruvate kinase	[120]

throughout the fermentation, while that of the α-ketoglutarate dehydrogenase complex (ODHC) significantly decreased upon induction of glutamate production, clearly leading to flux redistribution (Shirai et al. 2005). The results suggest that ODHC plays the largest role in controlling flux at the key branch point of α-ketoglutarate from the view of metabolic flux.

4.2.2 Metabolic Fluxes of NADPH Metabolism

The synthesis of most amino acids demands NADPH so that efficient supply of this cofactor is highly relevant in production processes. In *C. glutamicum*, glucose 6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, isocitrate

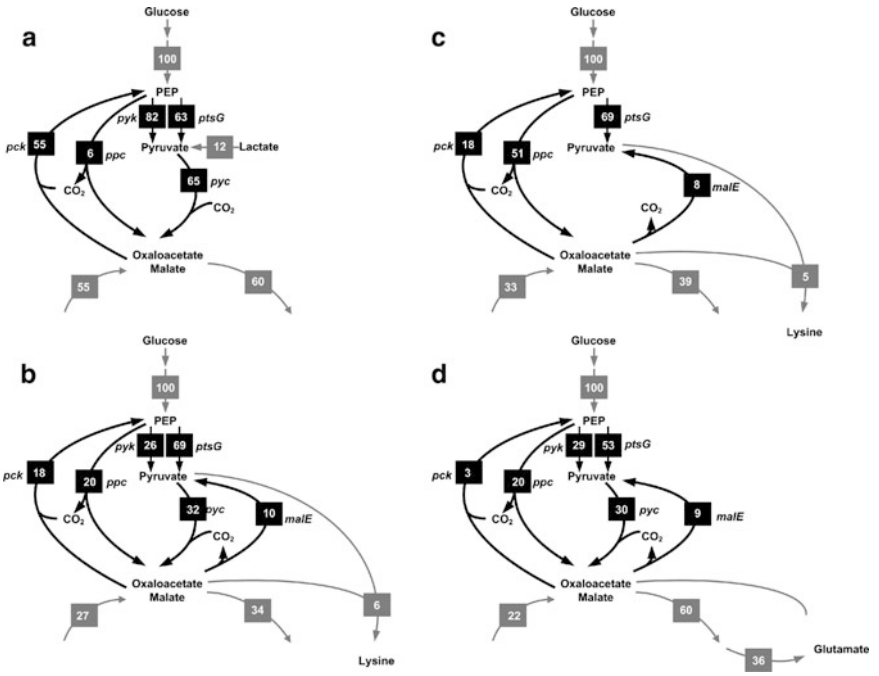


Fig. 4 Metabolic fluxes around the pyruvate node of *C. glutamicum* under different physiological conditions: Growth of the wild type *C. glutamicum* ATCC 13032 in continuous culture on glucose [27] (**a**), batch cultivation of the lysine producing strain *C. glutamicum* ATCC 13032 lysC^{fbr} (**b**), its pyruvate kinase deficient variant *C. glutamicum* ATCC 13032 lysC^{fbr} Δpyk [120] (**c**), and glutamate production in batch culture by *C. glutamicum* AJ-1511 (ATCC13869) on glucose [118] (**d**). All fluxes are given as relative molar fluxes to entry flux into the PEP pool

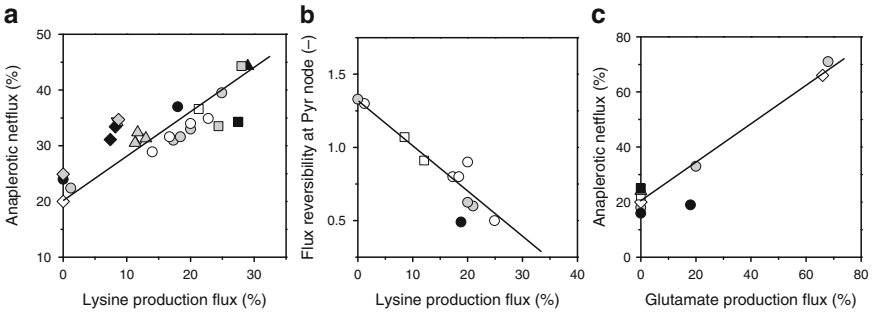


Fig. 5 Metabolic flux analysis of precursor supply and production of lysine (**a**, **b**) and glutamate (**c**) in different *C. glutamicum* strains investigated under various cultivation conditions by ^{13}C flux analysis: [112] (*white circles*), [125] (*gray circles*), [126] (*black circles*), [127] (*white squares*), [111] (*gray squares*), [111] (*black squares*), [128] (*white triangles*), [116] (*gray triangles*), [26] (*black triangles*), [129] (*white diamonds*), [115] (*gray diamonds*), [120] (*black diamonds*), [76] (*white hexagons*), and [118] (*gray hexagons*). The flux reversibility at the pyruvate node is defined as the ratio of the back flux to the anaplerotic net flux [130]

dehydrogenase, and malic enzyme catalyze NADPH generating reactions, whereas NADPH consuming reactions comprise growth with a stoichiometric demand of 16.4 mmol NADPH (g biomass)$^{-1}$ [34] and overproduction. From ^{13}C metabolic flux studies of lysine producing strains under different physiological conditions, remarkable insights into function and regulation of the NADPH metabolism of *C. glutamicum* could be obtained. The NADPH metabolism of *C. glutamicum* is highly flexible, adjusting to the physiological growth state [126], the nutrient status [13, 32, 111], or the genetic background [108, 131]. In most cases this results in an apparent NADPH excess, pointing at so far unassigned reactions which consume NADPH. Possible, but not validated, candidates are NADPH oxidase [132] or cyclic fluxes around the pyruvate node involving malic enzyme [27, 124]. Towards higher production, this apparent excess diminishes and turns into an apparent NADPH limitation [5]. Extrapolating these findings to industrial producers with much higher yields makes a limitation of production by NADPH very likely and strongly suggests the pathways involved in NADPH supply as promising targets for metabolic engineering of *C. glutamicum*. Hereby, stoichiometric investigation of the lysine network almost 20 years ago predicted that an increased lysine yield is linked to an increased flux through the PPP [133]. The importance of the PPP for efficient lysine production was later shown by metabolic flux analysis. Studies, investigating different strains under different conditions, revealed a close link of lysine production flux with the flux through the PPP (Fig. 6a). In contrast to the PPP, the contribution of isocitrate dehydrogenase decreases with increasing lysine production (Fig. 6b). The role of malic enzyme is still not completely clear. Its contribution to NADPH supply has been demonstrated for growth on fructose [32] as well as in a pyruvate kinase deficient lysine production strain [120].

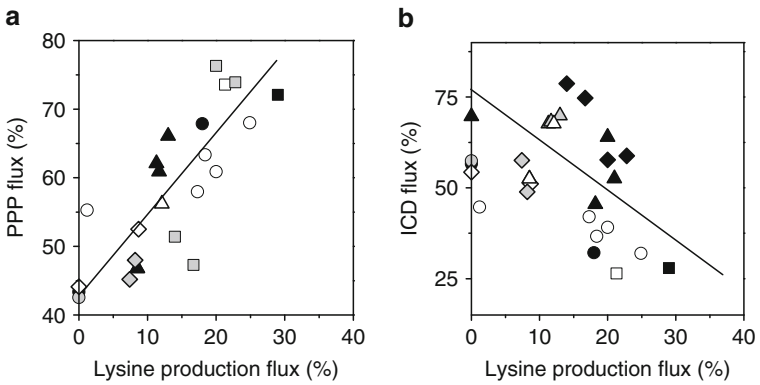


Fig. 6 Metabolic flux analysis linking lysine production with NADPH metabolism involving the pentose phosphate pathway (**a**) and isocitrate dehydrogenase (**b**) in different *C. glutamicum* strains investigated under various cultivation conditions by ^{13}C flux analysis: [125] (*white circles*), [129] (*gray circles*), [126] (*black circles*), [127] (*white squares*), [112] (*gray squares*), [126] (*black squares*), [128] (*white triangles*), [116] (*gray triangles*), [26, 76] (*black triangles*), [115] (*white diamonds*), and [120] (*gray diamonds*)

However, overexpression of the malic enzyme gene did not result in improved lysine production independent from the carbon source tested [134].

5 Metabolic Pathway Engineering

In *C. glutamicum* the biosynthesis of amino acids is strictly regulated through feedback inhibition. Therefore, overproduction requires, first of all, the removal of all metabolic control mechanisms. Moreover, the amplification of biosynthetic pathway genes is a logical step of strain improvement, especially in the presence of transcriptional repression. Moreover, balanced supply of precursor compounds and cofactors has to be addressed to achieve efficient production of the amino acid. For amino acids, these supporting pathways are part of the highly interconnected network of central carbon metabolism, so that modifications will likely interfere with growth related metabolic reactions. Due to these various hurdles, classical mutagenesis and selection did not yield high levels of production, although efforts have been continuing now for almost four decades [6]. The development of recombinant DNA technologies, allowing targeted genetic modification, has initiated intensive research towards rational optimization of *C. glutamicum*, resulting in remarkable progress in production efficiency [30, 135, 136]. In the following, the advances in metabolic pathway engineering of *C. glutamicum* are highlighted. Due to the strong impact of raw material costs, optimization of the conversion yield was at the focus of many studies in order to improve the economy of the production process [137].

5.1 Metabolic Engineering of Lysine Production

5.1.1 Metabolic Engineering of Lysine Biosynthesis

A number of studies have addressed the optimization of lysine production by direct modification of enzymes of the biosynthetic pathway (Fig. 7). The high relevance of engineering these enzymes is underlined by the fact that today every single gene of the lysine biosynthetic pathway is covered with patents by the major industrial players in the field [137]. The release of aspartate kinase, controlling the pathway flux, from feedback inhibition is one of the most important targets. Different point mutations in the *lysC* gene, i.e., its regulatory β-subunit, result in feedback resistant enzyme variants and increase lysine production [36, 138, 140]. Also, overexpression of aspartate kinase is beneficial for production [141]. Another relevant target is the *dapA* gene, encoding dihydrodipicolinate synthase. Amplified expression, increasing lysine was realized using plasmids [138, 142–144] as well as mutation of the promoter sequence [145]. Also, overproduction of diaminopimelate epimerase (DapF) and succinyl-aminoketopimelate transaminase (DapC), two enzymes of the succinylase branch, was beneficial for lysine formation [137]. Following the

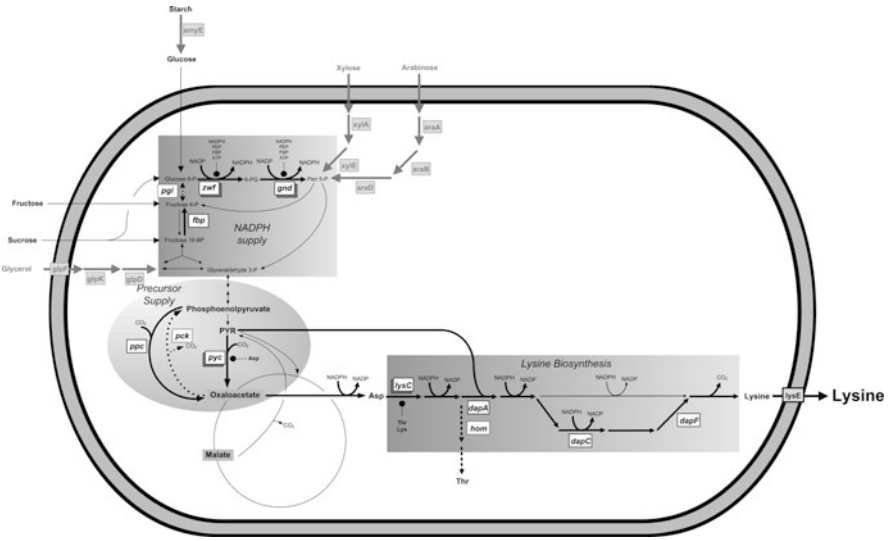


Fig. 7 Metabolic engineering strategies for optimization of lysine production in *C. glutamicum*. Highlighted are modifications for improved supply of precursors and NADPH, for increased flux through lysine biosynthesis and secretion and for extension of the substrate spectrum. The optimization comprises decrease of flux (*thick dotted arrow*), via deletion or attenuation of genes (*white box with dotted line*), increase of flux (*thick solid arrow*) via amplification of genes (*white box with solid line*), modification of metabolic control for feedback regulated enzymes (*dark grey box with solid line*) and the introduction of new reactions (*grey arrows*) via heterologous expression of foreign genes (*white box with grey solid line*)

discovery of the lysine exporter (LysE), the subsequent overexpression of the *lysE* gene resulted in an increased lysine secretion rate [146–148]. The recently performed expression of *lysE* from *C. glutamicum* in a *Methylophilus methylotrophus* lysine producing strain was also shown to improve lysine production from methanol by this organism [149].

5.1.2 Metabolic Engineering of Precursor Supply

The importance of anaplerotic enzymes for supply of the lysine precursor oxaloacetate stimulated intensive metabolic engineering activities (Fig. 7). Among the most striking findings was the identification of pyruvate carboxylase as the major anaplerotic enzyme in *C. glutamicum* [27]. Subsequently, overexpression of its gene has been shown to improve lysine production [150]. Knowing the importance of pyruvate carboxylase for lysine production, the point mutation P458S, identified in a classically derived producer, was introduced into the *pyc* gene and also resulted in strong increase of lysine production [136]. Moreover, also overexpression of PEP carboxylase is beneficial for the formation of amino acids of the aspartate family [151]. In order to reduce the back flux from the TCA cycle and the withdrawal of

oxaloacetate to glycolysis, deletion of PEP carboxykinase resulted in a significant improvement of lysine production [22]. For malic enzyme no clear effects could be observed so far. Neither deletion nor overexpression of the corresponding gene influenced the metabolism of *C. glutamicum* on sugars markedly [21, 152].

5.1.3 Metabolic Engineering of NADPH Supply

As revealed by metabolic flux analysis, the PPP is the major pathway for supply of NADPH required in high amounts for lysine biosynthesis. Due to this, different approaches have aimed at increasing flux through the PPP (Fig. 7). Metabolic flux studies on glucose, fructose, and sucrose identified fructose 1,6-bisphosphatase as a non-obvious target [110, 111]. Subsequent amplified expression of the *fbp* gene indeed increased lysine yield on glucose, fructose, and sucrose up to about 40% [128]. Hereby, the mutant with overexpression of fructose 1,6-bisphosphatase exhibited 10% enhanced PPP flux. A second major target approached was glucose 6-phosphate dehydrogenase. Overexpression of the encoding *zwf* gene in the feedback-deregulated lysine producing strain *C. glutamicum* ATCC13032 lysC^{fbr} resulted in increased lysine production on different carbon sources including the two major industrial sugars, glucose and sucrose [116]. In further successful examples, an increased PPP flux was achieved by modifying the regulatory properties of enzymes in the PPP, i.e., partially releasing glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase from allosteric inhibition. The substitution A243T in the *zwf* gene encoding for glucose 6-phosphate dehydrogenase [116, 153] and the substitution S361F in the *gnd* gene encoding for 6-phosphogluconate dehydrogenase [30] both led to significantly increased lysine titer. The modifications cause an increase in the PPP flux which probably resulted from positively changed kinetics of the enzymes. For glucose based processes, disruption of glycolysis via deletion of phosphoglucose isomerase forces the cell to metabolize the substrate completely via the PPP, and indeed results in increased lysine production [108]. This strategy, however, is only applicable for glucose based processes, since other substrates such as sucrose, glycerol, fructose, xylose, or arabinose require an active phosphoglucose isomerase for channeling carbon into the PPP [111, 128].

5.2 Metabolic Engineering of Aromatic Amino Acid Production

Remarkable progress was also achieved for the synthesis of aromatic amino acids in *C. glutamicum* [6]. Effective approaches comprised the modification of terminal pathways leading to removal of undesired control mechanisms, engineering of the PPP for increased supply of precursors, and transport engineering leading to reduced intracellular pools. Among the most striking findings was the overexpression of the *tkt* gene in the PPP encoding transketolase which strongly enhanced tryptophan production by improved supply of the precursor E4P [154, 155]. To avoid accumulation of indole, the last intermediate in the tryptophan pathway, due

to a limiting supply of L-serine, the availability of L-serine was improved by amplification of the *serA* gene for 3-phosphoglycerate dehydrogenase, the key enzyme in the L-serine biosynthesis [156]. A key to increased production was also the deletion of transport systems for aromatic amino acids impairing their uptake from the medium [6].

5.3 Metabolic Engineering of Glutamate Production

Concerning synthesis of glutamate in *C. glutamicum* which demands efficient replenishment of the TCA cycle, pyruvate carboxylase was identified as a major anaplerotic enzyme [118, 150]. The flux through pyruvate carboxylase strongly increased under glutamate production induced by Tween 40 addition, while PEP carboxylase and PEP carboxykinase fluxes remained constant. An impressive set of studies investigated the metabolism within the TCA cycle linked to glutamate production in great detail [4, 117, 157, 158]. Successful strategies to increase glutamate production involved deletion of *odhA*, encoding a key protein of the complex [159] or attenuation of *odhA* expression via the use of antisense RNA [160].

5.4 Utilization of Alternative Substrates

The major industrial carbon sources for fermentation processes with *C. glutamicum* are cane molasses, beet molasses, sucrose, and dextrose [3]. Since the carbon source is the major cost factor in industrial lysine production [137] and prices for molasses and raw sugar have increased significantly in recent years, many attempts have been aimed at extending the substrate spectrum of *C. glutamicum* to exploit alternative raw materials. Today, *C. glutamicum* has been engineered to utilize a large number of different substrates (Fig. 7). The future use of the corresponding raw materials in industrial fermentations of *C. glutamicum*, however, requires further intensive efforts in metabolic and process engineering. So far, key contributions enable the utilization of the pentose sugars which *C. glutamicum* naturally cannot metabolize. These sugars, namely xylose and arabinose, display significant fractions in agricultural residues and other lignocellulosic biomass recently receiving increasing interest as a cheap and most abundant raw material for biobased production [161]. Through heterologous expression of the *E. coli* genes *xylA* and *xylB* *C. glutamicum* was able to consume xylose as sole source of carbon [162]. In substrate mixtures, glucose-mediated regulation still exerts a measurable influence on xylose consumption kinetics. Similarly, arabinose utilization was also achieved, whereby the genes *araA*, *araB*, and *araD* were also derived from *E. coli* [163]. An interesting study recently investigated the tolerance of *C. glutamicum* to toxic compounds such as furfurals or phenols, typically present in lignocellulosic raw materials [164]. These compounds cause significant inhibition of growth. For

growth-arrested production processes this is not a critical point, but open questions remain to be elucidated for the majority of bioprocesses with *C. glutamicum* which typically are fed-batch processes with important phases of growth of biomass to achieve efficient production. Introducing the *S. griseus amy* gene on an expression vector into a lysine producing strain of *C. glutamicum* allowed synthesizing and secreting of α-amylase into the culture broth [165]. Although some high-molecular-weight degradation products remained in the culture broth, the recombinant strain effectively used soluble starch as carbon and energy substrate for growth and also for lysine production. This could allow the direct conversion of starch into desired products and the avoidance of cost intensive hydrolysation pretreatment currently required when feeding starch hydrolysates from corn, wheat, or cassava for production [166]. *C. glutamicum* cannot utilize glycerol, a stoichiometric by-product of biodiesel production. By heterologous expression of *Escherichia coli* glycerol utilization genes, *C. glutamicum* was engineered to grow on glycerol [167]. The engineered strains were able to produce glutamate efficiently as well as lysine. Recent attempts to enable utilization of lactose by heterologous coexpression of genes from *Lactococci* have shown the basic feasibility, but are still linked to suboptimal performance of the obtained strains [168]. At present, metabolic engineering has significantly broadened the substrate spectrum of *C. glutamicum*.

5.5 Global Strain Engineering Through Applied Systems Biology

The experience of the past clearly shows that detailed quantitative knowledge of metabolic physiology is required for rational design of superior production strains [169]. Especially for the optimization of amino acid production by *C. glutamicum*, characterized by a close connection between central metabolism and product biosynthetic pathways, understanding of global metabolic regulation has turned out to be crucial [5]. In this light, systems biology approaches elucidating cell physiology on a global level, displaying powerful strategies. Metabolic flux analysis may be regarded as a first systems approach to *C. glutamicum* since it combines measurement data with metabolic network models of biosynthetic and central metabolic pathways and predicts metabolic engineering strategies based on such systems oriented insights. Hence, these systems biology approaches per se are not new concepts to design and improve *C. glutamicum*. However, the availability of the genome sequence and the rapid progress in postgenomics methods such as transcriptomics, proteomics, fluxomics, and metabolomics today allow one to study its metabolic and regulatory properties on a truly global level, opening a new era of industrial strain improvement [44]. Rational strain improvement was initially done on a gene-by-gene basis but was recently put on the genome level by comparing the genome sequence of producer and wild-type to identify relevant mutations. The success of this strategy, named "genome breeding" was successfully demonstrated for a *C. glutamicum* L-lysine producer [136]. From an engineering perspective this strategy worked, but it is still based on trial and error, so that a large number of

targets identified from the genome sequence have to be tried. Moreover, it requires an already efficient production strain for target identification, which in the case of lysine took decades of development and is not available for other processes. Here, new omics tools from functional genomics can help to identify key regulators and guide rational strain engineering. Impressive progress has been made toward this goal, especially in the frame of systems biological study of *C. glutamicum* [44]. Transcriptome analysis through DNA microarrays [44, 170] allows global expression profiling of *C. glutamicum*. The proof of value of this technique for rational strain engineering has recently been shown, when novel targets for improved lysine production could be identified from selected transcriptome studies [171, 172]. Similarly, the analysis of the proteome, based on 2-D gel electrophoresis [52–54, 173] has also provided some insights into metabolic processes such as nitrogen starvation [174] or the utilization of citrate [15]. Among all systems oriented approaches, ^{13}C metabolic flux analysis has clearly contributed the most to our current detailed picture of the *C. glutamicum* metabolism. Among the achievements obtained from flux analysis are the identification of novel pathways, the elucidation of metabolic control, identification of targets for rational strain engineering in biotechnology, and first insights into design principles of the metabolic network. Hereby, metabolic flux analysis provided quantitative data which directly reflect the phenotype of the investigated *C. glutamicum* strain, whereas other omics approaches often do not allow a direct conclusion on the active pathways determining the phenotypic behavior [76, 175]. Recent developments provide the next level of systems biology studies, the parallel investigation of *C. glutamicum* on levels of gene expression, proteins, metabolites, and fluxes providing important links between the different functional components of cellular physiology. First examples of such systems-oriented studies already reveal a great potential [76, 119, 176]. Such approaches are especially promising for the targeted multidimensional alteration of complex regulatory networks towards better tolerance of production strains to high temperature or salt levels, or extreme pH values [137], but also reveal a great potential for efficient design of novel bioprocesses. Similarly, systems based metabolic engineering approaches have also created impressive progress in developing superior strains of *Escherichia coli*, the second major industrial amino acid producer [177]. As an example, production of valine [178] or threonine [179] using *E. coli* could be substantially optimized.

5.6 Towards Novel Products

In light of the established fermentation on cheap substrates [5], its known genome sequence [47] and the availability of genetic engineering tools *C. glutamicum* is regarded as key candidate to develop production strain for various other products. In addition to the above-mentioned traditional products, *C. glutamicum* has been recently engineered to accumulate a great variety of different compounds (Table 3) Due to impressive work in recent years, *C. glutamicum* is applicable to produce

Table 3 Metabolic engineering of *Corynebacterium glutamicum* for biotechnological applications. The papers listed comprise either review papers summarizing the progress achieved or recent key contributions.

Product	References
Amino acids	
Glutamate	[180]
Isoleucine	[181, 182]
Lysine	[5]
Methionine	[56, 184]
Serine	[184, 185]
Threonine	[186]
Tryptophan	[6]
Tyrosine	[43]
Valine	[187]
Panthothenate	[188]
Diamines	
Cadaverine	[189]
Organic acids	
Succinate	[190]
D-Lactate	[191]
Biofuels	
Ethanol	[192]

efficiently amino acids, organic acids, diamines, or biofuels, and the coming years will surely see a further broadening of its potential.

Acknowledgments The author gratefully acknowledges support by the BMBF-Grant "Biobased Polyamides through Fermentation" (No 0315239A) within the initiative Bioindustry21.

References

1. Kinoshita S, Shigezo U, Shimono M (1957) J Gen Appl Microbiol 3:193
2. Udaka S (1960) J Bacteriol 79:754
3. Ikeda M (2003) Adv Biochem Eng Biotechnol 79:1
4. Schultz C, Niebisch A, Gebel L, Bott M (2007) Appl Microbiol Biotechnol 76:691
5. Wittmann C, Becker J (2007) The L-lysine story: from metabolic pathways to industrial production. In: Wendisch VF (ed) Amino acid biosynthesis – pathways, regulation and metabolic engineering. Springer, Berlin
6. Ikeda M (2006) Appl Microbiol Biotechnol 69:615
7. Eggeling L, Bott M (2005) Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton
8. Dominguez H, Lindley ND (1996) Appl Environ Microbiol 62(10):3878
9. Moon MW, Kim HJ, Oh TK, Shin CS, Lee JS, Kim SJ, Lee JK (2005) FEMS Microbiol Lett 244:259
10. Park SY, Kim HK, Yoo SK, Oh TK, Lee JK (2000) FEMS Microbiol Lett 188:209
11. Vallino JJ, Stephanopoulos G (1994) Biotechnol Prog 10:327
12. Cocaign-Bousquet M, Lindley ND (1995) Enzyme Microb Technol 17:260

13. Wendisch VF, de Graaf AA, Sahm H, Eikmanns BJ (2000) J Bacteriol 182:3088
14. Claes WA, Pühler A, Kalinowski J (2002) J Bacteriol 184:2728
15. Polen T, Schluesener D, Poetsch A, Bott M, Wendisch VF (2007) FEMS Microbiol Lett 273:109
16. Burkovski A (2003) FEMS Microbiol Rev 27:617
17. Meier-Wagner J, Nolden L, Jakoby M, Siewe R, Kramer R, Burkovski A (2001) Microbiology 147:135
18. Lee H-S (2005) Sulfur metabolism and its regulation. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC Press, Taylor and Francis, Boca Raton
19. Rückert C, Koch DJ, Rey DA, Albersmeier A, Mormann S, Pühler A, Kalinowski J (2005) BMC Genomics 6:121
20. Peters-Wendisch PG, Kreutzer C, Kalinowski J, Patek M, Sahm H, Eikmanns BJ (1998) Microbiology 144(Pt 4):915
21. Gourdon P, Baucher MF, Lindley ND, Guyonvarch A (2000) Appl Environ Microbiol 66:2981
22. Riedel C, Rittmann D, Dangel P, Mockel B, Petersen S, Sahm H, Eikmanns BJ (2001) J Mol Microbiol Biotechnol 3:573
23. Jetten M, Sinskey AJ (1993) FEMS Microbiol Lett 111:183
24. Jetten MS, Sinskey AJ (1995) Antonie Van Leeuwenhoek 67:221
25. Jetten MS, Pitoc GA, Follettie MT, Sinskey AJ (1994) Appl Microbiol Biotechnol 41:47
26. Marx A, de Graaf A, Wiechert W, Eggeling L, Sahm H (1996) Biotechnol Bioeng 49(2):111
27. Petersen S, de Graaf AA, Eggeling L, Möllney M, Wiechert W, Sahm H (2000) J Biol Chem 275:35932
28. Moritz B, Striegel K, De Graaf AA, Sahm H (2000) Eur J Biochem 267:3442
29. Ihnen ED, Demain AL (1969) J Bacteriol 98:1151
30. Ohnishi J, Katahira R, Mitsuhashi S, Kakita S, Ikeda M (2005) FEMS Microbiol Lett 242:265
31. Audette GF, Quail JW, Hayakawa K, Bai C, Chen R, Delbaere LT (1999) Acta Crystallogr D Biol Crystallogr 55:1584
32. Dominguez H, Rollin C, Guyonvarch A, Guerquin-Kern JL, Cocaign-Bousquet M, Lindley ND (1998) Eur J Biochem 254:96
33. Kim HM, Heinzle E, Wittmann C (2006) J Microbiol Biotechnol 16:1174
34. Wittmann C, de Graaf A (2005) Metabolic flux analysis in *Corynebacterium glutamicum*. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton
35. Michal G (1999) Biochemical pathways. Wiley, Chichester
36. Kalinowski J, Cremer J, Bachmann B, Eggeling L, Sahm H, Puhler A (1991) Mol Microbiol 5:1197
37. Malumbres M, Martin JF (1996) FEMS Microbiol Lett 143:103
38. Schrumpf B, Schwarzer A, Kalinowski J, Pühler A, Eggeling L, Sahm H (1991) J Bacteriol 173:4510
39. Sonntag K, Eggeling L, De Graaf AA, Sahm H (1993) Eur J Biochem 213:1325
40. Krömer JO, Heinzle E, Schröder H, Wittmann C (2006) J Bacteriol 188:609
41. Rey DA, Pühler A, Kalinowski J (2003) J Biotechnol 103:51
42. Rey DA, Nentwich SS, Koch DJ, Rückert C, Puhler A, Tauch A, Kalinowski J (2005) Mol Microbiol 56:871
43. Ikeda M, Katsumata R (1992) Appl Environ Microbiol 58:781
44. Wendisch VF, Bott M, Kalinowski J, Oldiges M, Wiechert W (2006) J Biotechnol 124:74
45. Haberhauer G, Schröder H, Pompejus M, Zelder O, Kröger B (2001) Patent WO 01/00805
46. Ikeda M, Nakagawa S (2003) Appl Microbiol Biotechnol 62:99
47. 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, Mockel B, Pfefferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A (2003) J Biotechnol 104:5
48. Glanemann C, Loos A, Gorret N, Willis LB, O'Brien XM, Lessard PA, Sinskey AJ (2003) Appl Microbiol Biotechnol 61:61
49. Hayashi M, Ohnishi J, Mitsuhashi S, Yonetani Y, Hashimoto S, Ikeda M (2006) Biosci Biotechnol Biochem 70:546
50. Silberbach M, Burkovski A (2006) J Biotechnol 126:101
51. Huser AT, Becker A, Brune I, Dondrup M, Kalinowski J, Plassmeier J, Puhler A, Wiegrabe I, Tauch A (2003) J Biotechnol 106:269
52. Schluesener D, Fischer F, Kruip J, Rogner M, Poetsch A (2005) Mapping the membrane proteome of *Corynebacterium glutamicum*. Proteomics 5:1317–1330
53. Bendt AK, Burkovski A, Schaffer S, Bott M, Farwick M, Hermann T (2003) Proteomics 3:1637
54. Hermann T, Pfefferle W, Baumann C, Busker E, Schaffer S, Bott M, Sahm H, Dusch N, Kalinowski J, Puhler A, Bendt AK, Kramer R, Burkovski A (2001) Electrophoresis 22:1712
55. Kjeldsen KR, Nielsen J (2009) Biotechnol Bioeng 102:583
56. Krömer JO, Wittmann C, Schröder H, Heinzle E (2006) Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. Metab Eng 8(4):353–369
57. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, Gojobori T (2003) Genome Res 13:1572
58. 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) Microbiology 153:1042
59. Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoysa A, Chillingworth T, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabbinowitsch E, Rutherford KM, Thomson NR, Unwin L, Whitehead S, Barrell BG, Parkhill J (2003) Nucleic Acids Res 31:6516
60. Tauch A, Kaiser O, Hain T, Goesmann A, Weisshaar B, Albersmeier A, Bekel T, Bischoff N, Brune I, Chakraborty T, Kalinowski J, Meyer F, Rupp O, Schneiker S, Viehoever P, Pühler A (2005) J Bacteriol 187:4671
61. Tauch A, Schneider J, Szczepanowski R, Tilker A, Viehoever P, Gartemann KH, Arnold W, Blom J, Brinkrolf K, Brune I, Gotker S, Weisshaar B, Goesmann A, Droge M, Puhler A (2008) J Biotechnol 136:22
62. Varma A, Palsson BO (2001) Nat Biotechnol 12:994
63. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Nat Rev Microbiol 7:129
64. Feist AM, Palsson BO (2008) Nat Biotechnol 26:659
65. Price ND, Reed JL, Palsson BO (2004) Nat Rev Microbiol 2:886
66. Schilling CH, Edwards JS, Letscher D, Palsson BO (2000) Biotechnol Bioeng 71:286
67. Schuster S, Dandekar T, Fell DA (1999) Trends Biotechnol 17:53
68. Edwards JS, Covert M, Palsson B (2002) Environ Microbiol 4:133
69. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO (2004) Trends Biotechnol 22:400
70. Schuster S, Fell DA, Dandekar T (2000) Nat Biotechnol 18:326
71. Gagneur J, Klamt S (2004) BMC Bioinformatics 5:175
72. Carlson R, Fell D, Srienc F (2002) Biotechnol Bioeng 79:121
73. Carlson R, Srienc F (2004) Biotechnol Bioeng 85:1
74. Vallino JJ, Stephanopoulos G (1993) Biotechnol Bioeng 41:633
75. Vallino JJ, Stephanopoulos G (1994) Biotechnol Prog 10:320
76. Krömer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C (2004) J Bacteriol 186:1769
77. Kiss RD, Stephanopoulos G (1991) Biotechnol Prog 7:501
78. Yang C, Hua Q, Shimizu K (1999) J Biosci Bioeng 88:393

79. Magnus JB, Hollwedel D, Oldiges M, Takors R (2006) Biotechnol Prog 22:1071
80. Ruijter GJ, Bax M, Patel H, Flitter SJ, van de Vondervoort PJ, de Vries RP, vanKuyk PA, Visser J (2003) Eukaryot Cell 2:690
81. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W (2007) J Biotechnol 129:249
82. Bolten C, Kiefer P, Letisse F, Portais JC, Wittmann C (2007) Anal Chem 79:3843–3849
83. Borner J, Buchinger S, Schomburg D (2007) Anal Biochem 367:143
84. Strelkov S, von Elstermann M, Schomburg D (2004) Biol Chem 385:853
85. Wittmann C, Krömer JO, Kiefer P, Binz T, Heinzle E (2004) Anal Biochem 327:135
86. Gayen K, Venkatesh KV (2007) J Ind Microbiol Biotechnol 34:363
87. Iwatani S, Yamada Y, Usuda Y (2008) Biotechnol Lett 30:791
88. Dominguez H, Nezondet C, Lindley ND, Cocaign M (1993) Biotechnol Lett 15(5):449
89. Varela C, Agosin E, Baez M, Klapa M, Stephanopoulos G (2003) Appl Microbiol Biotechnol 60:547
90. Cocaign-Bousquet M, Guyonvarch A, Lindley ND (1996) Appl Environ Microbiol 62 (2):429
91. Pons A, Dussap CG, Pequignot C, Gros JB (1996) Biotechnol Bioeng 51(2):177
92. Sauer U (2006) Mol Syst Biol 2:62
93. Wittmann C, Hans M, Heinzle E (2002) Anal Biochem 307:379
94. Wittmann C (2007) Microb Cell Fact 6:6
95. Christensen B, Nielsen J (1999) Metab Eng 1:282
96. Dauner M, Sauer U (2000) Biotechnol Prog 16:642
97. Yang TH, Heinzle E, Wittmann C (2005) Comput Biol Chem 29:121
98. Yang TH, Wittmann C, Heinzle E (2006) Metab Eng 8:417
99. Yang TH, Wittmann C, Heinzle E (2006) Metab Eng 8:432
100. Iwatani S, Van Dien S, Shimbo K, Kubota K, Kageyama N, Iwahata D, Miyano H, Hirayama K, Usuda Y, Shimizu K, Matsui K (2007) J Biotechnol 128:93
101. Wiechert W (2002) J Biotechnol 94:37
102. Fischer E, Zamboni N, Sauer U (2004) Anal Biochem 325:308
103. Shiio I, Otsuka SI, Tsunoda T (1960) J Biochem 47:414
104. Ishino S, Yamaguchi K, Shirahata K, Araki K (1984) Agric Biol Chem 48(10):2557
105. Walker TE, Han CH, Kollman VH, London RE, Matwiyoff NA (1982) J Biol Chem 257:1189
106. Drysch A, El Massaoudi M, Mack C, Takors R, de Graaf AA, Sahm H (2003) Metab Eng 5:96
107. Klapa MI, Aon JC, Stephanopoulos G (2003) Eur J Biochem 270:3525
108. Marx A, Hans S, Mockel B, Bathe B, de Graaf AA (2003) J Biotechnol 104:185
109. Drysch A, El Massaoudi M, Wiechert W, de Graaf AA, Takors R (2004) Biotechnol Bioeng 85:497
110. Kiefer P, Heinzle E, Zelder O, Wittmann C (2004) Appl Environ Microbiol 70:229
111. Wittmann C, Kiefer P, Zelder O (2004) Appl Environ Microbiol 70:7277
112. Wittmann C, Kim HM, Heinzle E (2004) Biotechnol Bioeng 87:1
113. Becker J, Heinzle E, Klopprogge C, Zelder O, Wittmann C (2005) Appl Environ Microbiol 71:8587–8596
114. Yang TH, Wittmann C, Heinzle E (2006) Metab Eng (submitted for publication)
115. Kim HM, Heinzle E, Wittmann C (2006) J Microbiol Biotechnol 16:1174
116. Becker J, Klopprogge C, Herold A, Zelder O, Bolten CJ, Wittmann C (2007) J Biotechnol
117. Shirai T, Nakato A, Izutani N, Nagahisa K, Shioya S, Kimura E, Kawarabayasi Y, Yamagishi A, Gojobori T, Shimizu H (2005) Metab Eng 7:59
118. Shirai T, Fujimura K, Furusawa C, Nagahisa K, Shioya S, Shimizu H (2007) Microb Cell Fact 6:19
119. Krömer JO, Bolten CJ, Heinzle E, Schröder H, Wittmann C (2008) Microbiology 154:3917
120. Becker J, Klopprogge C, Wittmann C (2008) Microb Cell Fact 7:8
121. Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wüthrich K (1997) Nat Biotechnol 15:448

122. Wittmann C, Heinzle E (2001) Biotechnol Bioeng 72:642
123. Sauer U, Eikmanns BJ (2005) FEMS Microbiol Rev 29:765
124. de Graaf AA (2000) Metabolic flux analysis of *Corynebacterium glutamicum*. In: Schügerl K, Bellgard KH (eds) Bioreaction engineering. Springer, Berlin
125. Wittmann C, Heinzle E (2002) Appl Environ Microbiol 68:5843
126. Marx A, Striegel K, de Graaf A, Sahm H, Eggeling L (1997) Biotechnol Bioeng 56(2):168
127. Wittmann C, Heinzle E (2001) Eur J Biochem 268:2441
128. Becker J, Klopprogge C, Zelder O, Heinzle E, Wittmann C (2005) Appl Environ Microbiol 71:8587
129. Sonntag K, Schwinde J, de Graaf A, Marx A, Eikmanns B, Wiechert W, Sahm H (1995) Appl Microbiol Biotechnol 44:489
130. Wittmann C, Heinzle E (2001) Metab Eng 3:173
131. Marx A, Eikmanns BJ, Sahm H, de Graaf AA, Eggeling L (1999) Metab Eng 1:35
132. Matsushita K, Otofuji A, Iwahashi M, Toyama H, Adachi O (2001) FEMS Microbiol Lett 204:271
133. Kiss RD, Stephanopoulos G (1992) Biotechnol Bioeng 39:565
134. Georgi T, Rittmann D, Wendisch VF (2005) Metab Eng 7:291
135. Sahm H, Eggeling L, de Graaf AA (2000) Biol Chem 381:899
136. Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K, Ikeda M (2002) Appl Microbiol Biotechnol 58:217
137. Kelle R, Hermann T, Bathe B (2005) L-Lysine production. In: Eggeling L, Bott M (eds) Handbook of *Cornyebacterium glutamicum*. CRC Press, Taylor and Francis, Boca Raton
138. Cremer J, Eggeling L, Sahm H (1991) Appl Environ Microbiol 57(6):1746
139. Follettie MT, Peoples OP, Agoropoulou C, Sinskey AJ (1993) J Bacteriol 175:4096
140. Sugimoto M, Tanaka A, Suzuki T, Matsui H, Nakamori S, Takagi H (1997) Biosci Biotechnol Biochem 61:1760
141. Jetten MS, Follettie MT, Sinskey AJ (1995) Appl Microbiol Biotechnol 43:76
142. Bonnassie S, Oreglia J, Sicard AM (1990) Nucleic Acids Res 18:6421
143. Eggeling L, Oberle S, Sahm H (1998) Appl Microbiol Biotechnol 49:24
144. Pisabarro A, Malumbres M, Mateos LM, Oguiza JA, Martin JF (1993) J Bacteriol 175:2743
145. Vasicova P, Patek M, Nesvera J, Sahm H, Eikmanns B (1999) J Bacteriol 181:6188
146. Bellmann A, Vrljic M, Patek M, Sahm H, Kramer R, Eggeling L (2001) Microbiology 147:1765
147. Vrljic M, Garg J, Bellmann A, Wachi S, Freudl R, Malecki MJ, Sahm H, Kozina VJ, Eggeling L, Saier MH Jr, Eggeling L, Saier MH Jr (1999) J Mol Microbiol Biotechnol 1:327
148. Vrljic M, Sahm H, Eggeling L (1996) Mol Microbiol 22:815
149. Gunji Y, Yasueda H (2006) J Biotechnol 127:1
150. Peters-Wendisch PG, Schiel B, Wendisch VF, Katsoulidis E, Mockel B, Sahm H, Eikmanns BJ (2001) J Mol Microbiol Biotechnol 3:295
151. Sano K, Ito K, Miwa K, Nakamori S (1987) Agric Biol Chem 51(2):597
152. Netzer R, Peters-Wendisch P, Eggeling L, Sahm H (2004) Appl Environ Microbiol 70:7148
153. Zelder O, Pompejus M, Schröder H, Kröger B, Klopprogge C, Haberhauer G (2005) UP Office, 2005/0014235, A1. US
154. Ikeda M, Katsumata R (1999) Appl Environ Microbiol 65:2497
155. Ikeda M, Okamoto K, Katsumata R (1999) Appl Microbiol Biotechnol 51:201
156. Ikeda M, Nakanishi K, Kino K, Katsumata R (1994) Biosci Biotechnol Biochem 58:674
157. Niebisch A, Kabus A, Schultz C, Weil B, Bott M (2006) J Biol Chem 281:12300
158. Shimizu H, Tanaka H, Nakato A, Nagahisa K, Kimura E, Shioya S (2003) Bioprocess Biosyst Eng 25:291
159. Asakura Y, Kimura E, Usuda Y, Kawahara Y, Matsui K, Osumi T, Nakamatsu T (2007) Appl Environ Microbiol 73:1308
160. Kim J, Hirasawa T, Sato Y, Nagahisa K, Furusawa C, Shimizu H (2009) Appl Microbiol Biotechnol 81:1097–1106

161. Aristidou A, Penttila M (2000) Curr Opin Biotechnol 11:187
162. Kawaguchi H, Vertes AA, Okino S, Inui M, Yukawa H (2006) Appl Environ Microbiol 72:3418
163. Kawaguchi H, Sasaki M, Vertes AA, Inui M, Yukawa H (2008) Appl Microbiol Biotechnol 77:1053
164. Sakai S, Tsuchida Y, Okino S, Ichihashi O, Kawaguchi H, Watanabe T, Inui M, Yukawa H (2007) Appl Environ Microbiol 73:2349
165. Seibold G, Auchter M, Berens S, Kalinowski J, Eikmanns BJ (2006) J Biotechnol 124:381
166. Tateno T, Fukuda H, Kondo A (2007) Appl Microbiol Biotechnol 74:1213
167. Rittmann D, Lindner SN, Wendisch VF (2008) Appl Environ Microbiol 74:6216
168. Barrett E, Stanton C, Zelder O, Fitzgerald G, Ross RP (2004) Appl Environ Microbiol 70:2861
169. Koffas M, Stephanopoulos G (2005) Curr Opin Biotechnol 16:361
170. Wendisch VF (2003) J Biotechnol 104:273
171. Sindelar G, Wendisch VF (2007) Appl Microbiol Biotechnol 76:677
172. Hayashi M, Mizoguchi H, Ohnishi J, Mitsuhashi S, Yonetani Y, Hashimoto S, Ikeda M (2006) Appl Microbiol Biotechnol 72:783
173. Schaffer S, Burkovski A (2005) Proteomics. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton
174. Schmid R, Uhlemann EM, Nolden L, Wersch G, Hecker R, Hermann T, Marx A, Burkovski A (2000) FEMS Microbiol Lett 187:83
175. Schilling O, Frick O, Herzberg C, Ehrenreich A, Heinzle E, Wittmann C, Stulke J (2007) Appl Environ Microbiol 73:499
176. Silberbach M, Schafer M, Huser AT, Kalinowski J, Puhler A, Kramer R, Burkovski A (2005) Appl Environ Microbiol 71:2391
177. Park JH, Lee SY (2008) Curr Opin Biotechnol 19:454
178. Park JH, Lee KH, Kim TY, Lee SY (2007) Proc Natl Acad Sci USA 104:7797
179. Lee KH, Park JH, Kim TY, Kim HU, Lee SY (2007) Mol Syst Biol 3:149
180. Kimura E (2003) Adv Biochem Eng Biotechnol 79:37
181. Colon GE, Nguyen TT, Jetten MS, Sinskey AJ, Stephanopoulos G (1995) Appl Microbiol Biotechnol 43:482
182. Guillouet S, Rodal AA, An GH, Gorret N, Lessard PA, Sinskey AJ (2001) Appl Microbiol Biotechnol 57:667
183. Park SD, Lee JY, Sim SY, Kim Y, Lee HS (2007) Metab Eng 9:327
184. Stolz M, Peters-Wendisch P, Etterich H, Gerharz T, Faurie R, Sahm H, Fersterra H, Eggeling L (2007) Appl Environ Microbiol 73:750
185. Peters-Wendisch P, Stolz M, Etterich H, Kennerknecht N, Sahm H, Eggeling L (2005) Appl Environ Microbiol 71:7139
186. Sahm H, Eggeling L, Eikmanns B, Kramer R (1996) Ann N Y Acad Sci 782:25
187. Blombach B, Schreiner ME, Bartek T, Oldiges M, Eikmanns BJ (2008) Appl Microbiol Biotechnol 79:471
188. Hüser AT, Chassagnole C, Lindley ND, Merkamm M, Guyonvarch A, Elisakova V, Patek M, Kalinowski J, Brune I, Pühler A, Tauch A (2005) Appl Environ Microbiol 71:3255
189. Mimitsuka T, Sawai H, Hatsu M, Yamada K (2007) Biosci Biotechnol Biochem 71:2130
190. Okino S, Noburyu R, Suda M, Jojima T, Inui M, Yukawa H (2008) Appl Microbiol Biotechnol 81:459
191. Okino S, Suda M, Fujikura K, Inui M, Yukawa H (2008) Appl Microbiol Biotechnol 78:449
192. Inui M, Kawaguchi H, Murakami S, Vertes AA, Yukawa H (2004) J Mol Microbiol Biotechnol 8:243

Adv Biochem Engin/Biotechnol (2010) 120: 51–99
DOI: 10.1007/10_2009_59

Published online: 12 February 2010

Systems Biology of Industrial Microorganisms

Marta Papini, Margarita Salazar, and Jens Nielsen

Abstract The field of industrial biotechnology is expanding rapidly as the chemical industry is looking towards more sustainable production of chemicals that can be used as fuels or building blocks for production of solvents and materials. In connection with the development of sustainable bioprocesses, it is a major challenge to design and develop efficient cell factories that can ensure cost efficient conversion of the raw material into the chemical of interest. This is achieved through metabolic engineering, where the metabolism of the cell factory is engineered such that there is an efficient conversion of sugars, the typical raw materials in the fermentation industry, into the desired product. However, engineering of cellular metabolism is often challenging due to the complex regulation that has evolved in connection with adaptation of the different microorganisms to their ecological niches. In order to map these regulatory structures and further deregulate them, as well as identify ingenious metabolic engineering strategies that full-fill mass balance constraints, tools from systems biology can be applied. This involves both high-throughput analysis tools like transcriptome, proteome and metabolome analysis, as well as the use of mathematical modeling to simulate the phenotypes resulting from the different metabolic engineering strategies. It is in fact expected that systems biology may substantially improve the process of cell factory development, and we therefore propose the term Industrial Systems Biology for how systems biology will enhance the development of industrial biotechnology for sustainable chemical production.

Keywords Metabolic engineering, Systems biology, Industrial biotechnology

M. Papini (✉), M. Salazar (✉), and J. Nielsen (✉)
Chalmers University of Technology, Chemical and Biological Engineering – Systems Biology, Kemigården 4, Göteborg 41296, Sweden
e-mail: nielsenj@chalmers.se

Contents

1 Introduction: The Role of Systems Biology in Industrial Biotechnology 52
2 Yeasts and Filamentous Fungi as Cell Factories 57
3 The Systems Biology Toolbox 58
3.1 Genomics 58
3.2 Transcriptomics 60
3.3 Proteomics 61
3.4 Metabolomics 62
3.5 Fluxomics 63
3.6 Metabolic Modeling 64
4 Yeast Cell Factories 67
4.1 Characteristics and Applications of *S. cerevisiae* as Cell Factory 67
4.2 Strain Improvements of Baker and Brewer Yeast 68
4.3 *S. cerevisiae* as a Platform for Chemical Production 71
5 Fungal Cell Factories 76
5.1 Characteristics and Applications of *Aspergillus* Species as Cell Factories 76
5.2 Strain Improvement of the Industrial Production Hosts *A. oryzae* and *A. niger* 77
5.3 *Aspergillus* Species as Platform for Production of Chemicals 78
5.4 Heterologous Protein Production 81
6 Systems Biology as Driver for Industrial Biotechnology 82
6.1 Yeast Examples 83
6.2 *Aspergillus* Examples 87
References 91

1 Introduction: The Role of Systems Biology in Industrial Biotechnology

Microbial fermentations have been used for the production of fermented food and beverages since ancient times and the first microbial fermentation was introduced in the 1920s for the production of citric acid. This paved the way for industrial scale production of penicillin by microbial fermentation during World War II and after that several fermentation processes for the production of other antibiotics were introduced.

Following antibiotics production, microbial industrial fermentations have been used for the production of different metabolites endogenously produced by microorganisms, e.g., enzymes for different applications and large-scale production of bioethanol for fuel usage.

With the introduction of genetic engineering in the 1970s it became possible to produce pharmaceutical proteins like human insulin and human growth hormone (hGH) by microbial fermentations, leading to the development of what is often referred to as the biotech industry.

Before the advent of genetic engineering, hGH was purified from cadavers' pituitary glands, with several drawbacks. It was Genentech, the first biotech company founded in 1976, to clone and express the hGH gene in *Escherichia coli*, opening the way for the production of biopharmaceuticals through recombinant

Table 1 Some of the main biopharmaceuticals available on the market produced through recombinant DNA technologies

Product	Commercial name	Company	Host	Date approved
Anticoagulant (*tPA*)	Tenecteplase	Boehringer Ingelheim	CHO cells	2001
Anticoagulant (*tPA*)	Activase	Genentech	CHO cells	1987
Insulin	Actrapid/Velosulin/ Insulatard	Novo Nordisk	*S. cerevisiae*	2002 (EU)
Insulin	Humulin	Eli Lilly	*E. coli*	1982
Human growth hormone	Somavert	Pfizer	*E. coli*	2003
Human growth hormone	Protopin	Genentech	*E. coli*	1985
Erythropoietin	Aranesp	Amgen	CHO cells	2001
IFN a	Pegasys	Hoffman-La Roche	*E. coli*	2002
IFN b	Rebif	Ares-Serono	CHO cells	1998
Recombinant vaccine against Hep B	Ambirix	GlaxoSmithKline	*S. cerevisiae*	2002
Recombinant vaccine against Hep B	HBVAXPRO	Aventis	*S. cerevisiae*	2001
Monoclonal antibody-based product	Xolair	Genentech	CHO cells	2003
Monoclonal antibody-based product	Bexxar	Corixa/Glaxo Smith Kline	CD20 cells	2003
Rheumatoid arthritis	Enbrel	Amgen	CHO	2000

For the first two compounds the last and the first chronologically available on the market are listed; for all the other products only the last version is commercially available. Table adapted from Primrose and Twyman [247]

fermentation processes. The same company, in 1982, launched the first biotechnological drug on the market: human insulin. Since then several microorganisms have been explored as production systems for recombinant therapeutic proteins but, due to their ability to perform proper protein glycosylation and folding, mammalian cells dominate as cell factories in the field of biopharmaceuticals (see Table 1 for the main therapeutic molecules produced through biotechnology).

The introduction of genetic engineering also allowed the development of microbial cell factories for the production of chemicals through the so-called *metabolic engineering* [1–3]. Initially, even though the technology for expressing heterologous metabolic pathways in a given microbial cell factory was available, it was not possible to develop cell factories that could produce chemicals at a cost of goods that allowed competition with traditional routes using petrochemicals as raw material. Following the developments in genomics, primarily in the health care sector but also in the field of industrial microbiology, a number of new technologies became available, improving our ability to perform far more detailed phenotypic characterization. Furthermore, the development of advanced measurement techniques allowed for the establishment of detailed mathematical models describing complex biological systems, a field often referred to as *systems biology* [4–6].

The progress obtained in genomics and systems biology accounted for the possibility to perform more directed genetic modifications of cell factories,

substantially advancing metabolic engineering. Thus, through metabolic engineering, it is today possible to design, develop and implement very efficient cell factories for the production of different chemicals.

The production of compounds related to the chemical market through biotechnology is also referred to as *industrial biotechnology* [7]. The products of industrial biotechnology can roughly be classified into two main classes: commodities and fine chemicals. Among commodities, or basic chemicals, we find all those products which can be used as starting materials for the production of other basic chemicals, specialty or manufactured goods. Specialty chemicals are more elaborated compounds with high-added value. Inside these classes we can find further "subclasses" such as antibiotics, organic acids, etc. which are listed in Table 2. Some of the products mentioned have been produced by microbial systems for years whilst the production of others became feasible only after the development of genetic engineering techniques [8].

Compared to conventional chemical processes, the use of microbial systems offers the advantages of reduction of wastes as well as minimization of costs and pollutant gas emissions. In addition, the ability of utilizing renewable sources instead of non-renewable ones, turned industrial biotechnology into a key technology for the development of a sustainable chemical industry [9, 10].

In the last two decades, industrial biotechnology reported an impressive increase and several sub-disciplines emerged within this field. The US Department of Energy [11] has estimated the biotech market of small molecules to reach US $100 billion by 2010 and US $400 billion by 2030. Furthermore, a popular report from McKinsey & Co. [12] forecasted a production of chemicals through biotech to represent up to 20% of the total chemical market by 2020.

There are already several examples of chemicals being produced by microbial fermentation of engineered cell factories, whose production through metabolic engineering has been boosted by the use of genomics tools, e.g., 1,3-propanediol used for polymer production, riboflavin used as a vitamin, and 7-aminodeacetoxycephalosporanic acid (7-ADCA) used as a precursor for antibiotics production. Furthermore, in the quest to develop a more sustainable society, the chemical industry is currently developing novel processes for many other fuels and chemicals, e.g., butanol, to be used for fuels, organic acids to be used for polymer production, and amino acids to be used as feed.

Despite the advancements in the integration of systems biology tools into metabolic engineering, there are still a number of hurdles in the optimization of novel cell factories for sustainable production of fuels, chemicals and pharmaceuticals [13]. Several attempts are focused on the development of tools that can facilitate metabolic engineering and hereby speed up the development of novel cell factories. The integration of systems biology tools with metabolic engineering for advancing industrial biotechnology is herein referred to as *industrial systems biology* (Fig. 1) [14]. Industrial systems biology offers the opportunity to introduce new concepts in metabolic engineering and hereby advances the development of cell factories (Fig. 1).

Table 2 Overview of different products of industrial biotechnology

Class of compounds	Product	Organism	Reference
Primary metabolites	Ethanol	*S. cerevisiae* *Z. mobilis*	[248, 249]
	Glycerol	*S. cerevisiae*	[250, 251]
	Malic acid	*S. cerevisiae*	[252]
Organic acids	Citric acid	*A. niger* *Y. lypolitica*	[253, 254]
	Acetic acid	*Acetobacter* *C. formicoaceticum*	[255, 256]
	Lactic acid	*S. cerevisiae* *Lactic acid bacteria*	[158, 257]
	Succinic acid	*M. succiniproducens* *E. coli* *A. succiniproducens* *A. succinogenes*	[258–262]
Amino acids	L-Lysine	*C. glutamicum*	[263–265]
	L-Phenylalanine	*E. coli*	[266]
	L-Tyrosine	*C. glutamicum* *E. coli*	[267–271]
	L-Glutamate	*C. glutamicum*	[272, 273]
Vitamins	L-Ascorbic acid	*G. oxidans*/*E. herbicola* *S. cerevisiae*	[274, 275]
	Riboflavin	*B. subtilis*	[276–278]
Antibiotics	Penicillin	*P. chrysogenum*	[279]
	Erythromicin A	*S. erythraea*	[280]
	Rifampicin	*S. mediterranei*	[281]
	Cephalosporin	*A. chrysogenum*	[282]
	Nystatin	*S. nourseii* *Pseudomonas* spp.	[283]
Non-antibiotics	Avermectin	*S. avermitilis*	[284]
	Bioinsecticides	*B. thuringiensis*	[285, 286]
	Flavonoids	*S. cerevisiae* *E. coli*	[287–290]
	Resveratrol (flavonoid)	*S. cerevisiae* *E. coli*	[291–293]
	β-Carotene	*S. cerevisiae*	[164, 168, 294]
	Lycopene (antioxidant)	*E. coli*	[295, 296]
	Astaxanthin	*E. coli*	[297]
	Taxol precursor (diterpenoid)	*S. cerevisiae*	[298, 299]
	Limonene (monoterpenoids)	*E. coli*	[300]
	Sesquiterpenes (patchoulol, cubebol, valencene)	*S. cerevisiae*	[301]
	Polyketides	*S. cerevisiae*, *Streptomyces* spp. *Bacillus* spp.	[302, 303]
Enzymes	Lipase	*A. niger*	[304]
	Aminopeptidase	*A. oryzae*	[305]
	Human α-proteinase	*A. niger*	[306, 307]
	α-Amylases	*A. oryzae*	[308]
	α-Galactosidase	*A. oryzae*	[309, 310]
	Subtilisin	*Bacillus*	[311, 312]

(*continued*)

Table 2 (continued)

Class of compounds	Product	Organism	Reference
	Glucoamylase	*A. niger*	[313]
	Asparaginase	*A. niger*	[314]
	Proteases	*A. oryzae*, *A. niger*	[206]
	Glucoisomerase	*S. murinus*	[201]
Polymers	1,3-Propanediol	*E. coli*	[315–317]
	Polyhydroxyalkanoates	*E. coli*	[318]
	3-Hydroxypropionic acid	*E. coli*	[319]
Dyes	Indigo	*E. coli*	[320, 321]
Compounds with therapeutic activities	Antibodies	*S. cerevisiae*	[322]
	Antimalaria drug precursor	*S. cerevisiae*	[323]
	IGF-If	*E. coli*	[324]
	Shikimic acid (antiviral)	*E. coli*	[325]

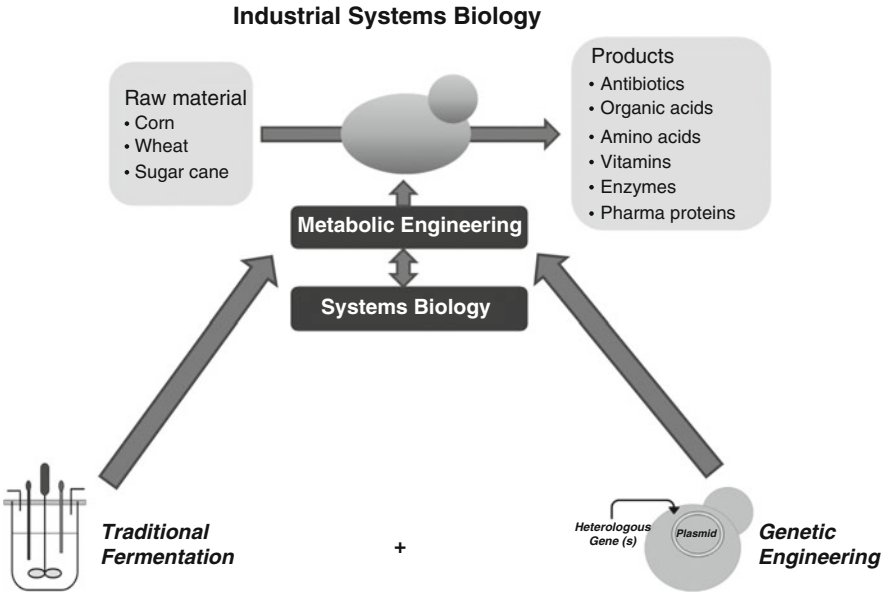


Fig. 1 The integration of systems biology tools with metabolic engineering for advancing industrial biotechnology is herein referred to as *industrial systems biology*. Industrial systems biology offers the opportunity to introduce new concepts in metabolic engineering and advances the development of cell factories: through traditional fermentation for production of natural products and, after the advent of recombinant DNA technology, for the production of recombinant metabolites [1]

Here we will review the impacts of systems biology on industrial biotechnology. Even though we may give examples from many different cell factories, our focus will be on yeast and filamentous fungi, widely used cell factories in industrial biotechnology.

2 Yeasts and Filamentous Fungi as Cell Factories

In 1789 the chemist Lavoisier discovered the microbial conversion of sugar into ethanol and only successively was this ability attributed to yeasts, as explained by Pasteur [15]. It is not surprising that, since then, yeasts have been extensively studied and their capacity to convert sugars into ethanol and carbon dioxide exploited to produce beer, wine, and bread.

Brewing and baking represent the oldest applications of yeasts for industrial purpose and today a wide variety of compounds is produced in yeasts through metabolic engineering. Among the different yeasts species, the most studied and well characterized is *Saccharomyces cerevisiae*, which has been engineered to produce, or to attempt to produce, a wide range of products such as ethanol, organic acids, and complex natural compounds (e.g., isoprenoids, polyketides, etc.). Apart from *S. cerevisiae*, other industrially relevant yeasts can be mentioned. The methylotrophic yeast *Pichia pastoris* is a widely used production system for heterologous proteins, especially after the successful engineering of the human glycosylation pathway [16, 17].

Filamentous fungi and prokaryotic diversity represent other host systems exploited by mankind. Filamentous fungi from the genus *Aspergillus* are outstanding producers of organic acids, capable of converting as much as 95% of the available carbon into organic acids [18]. In addition, *Aspergillus* species are also widely used for production of enzymes [19]. Another important filamentous fungus is *Penicillium chrysogenum*, mainly known for its ability to produce penicillin, but today also used for production of 7-ADCA, a precursor for the production of the antibiotic cephalexin.

It is generally difficult to choose an optimal cell factory for a given application, as each organism has advantages and drawbacks, depending on the product and the kind of process (see Table 3). Some general considerations when choosing a cell

Table 3 Characteristics of the main workhorses of industrial biotechnology

	S. cerevisiae	*P. pastoris*	*Aspergillus* spp.	*E. coli*
Availability of genomic tools	+++	+	++	+++
Online resources/databases	+++	+	++	++
Robust fermentation technologies	+++	++	++	+++
Advanced molecular biology techniques	+++	++	++	+++
Tolerance to low pH	++	++	+++	–
Capability of using complex feedstock (natural capability or status of engineering)	+	+	++	+
Compartmentalization	+	+	+	–
Glycosylation	+	++	+	–
Folding ability	++	++	+	–
Protein secretion in the media	++	++	+	–
Presence of endotoxins/viral proteins/LPS	–	–	–	+
Fast growth	+	+	++	++

The properties of the yeasts *S. cerevisiae* and *P. pastoris* are compared with those of filamentous fungi from *Aspergillus* species and the main prokaryote production host *E. coli*

factory are its metabolic capabilities towards the production of the desired product, the presence of an appropriate intracellular environment for a certain gene and/or product (i.e., an enzyme might be inhibited at low/high pH), and possible sensitivity towards the product to be produced.

3 The Systems Biology Toolbox

The complexity of the cellular machinery is a constant challenge for biological research. Redundancies and complex regulatory circuits within the cells often hamper the effect of introduced genetic modifications. Regulation occurs at all levels of cellular control: at the transcriptional level, at the translational level, at the post-translation level and through metabolite–enzyme interactions. In the order to gain a detailed understanding of cellular processes, a holistic view is required. In the recent years, a large variety of high-throughput experimental technologies have been developed to help us unravel the complex metabolic network of living cells and several studies based on systems-wide measurements have been published. Although this chapter is by no means exhaustive, we attempt to convey our philosophy that the integration of global information from various levels of metabolic hierarchy (*omics* data) is absolutely essential for understanding the relationship between changes in genes expression and the resulting phenotypes. Systems biology represents a way to meet the challenge of integrating diverse sets of *omics* information and the rapid increase in available genomes of industrially relevant microorganisms [20–27] makes the system-wide approach more and more appealing for the future [28].

In this section, we give an overview of the current genomics, transcriptomics, proteomics, metabolomics, fluxomics, and modeling techniques used to optimize the design of cell factories. Extensive compendiums of high-throughput data for a wide range of microorganisms are available and stored in databases together with many accessible bioinformatics tools. Table 4 gives an overview of some relevant online resources.

3.1 Genomics

Since the publication in 1977 of the first complete genome sequence, that of the Bacteriophage fX174, a viral genome with only 5,368 base pairs (bp) [29], the field of genomics has been of growing importance to biological studies. Thanks to the up-to-date sequencing technologies and powerful bioinformatics techniques, this field is now growing rapidly. Recent development of low-cost DNA sequencing techniques has led to an explosive growth in the number of sequenced genomes and will turn manual annotations into a luxury (for a complete review see Otero et al. [14]). Even though genome sequencing projects for simple microbial species started as early as 1980, the genomics era of bacterial species started in 1995 with the

Table 4 Useful databases required for conducting systems biology studies in yeasts and filamentous fungi

Resource	Internet URL
Genomics and functional genomics	
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov
BROAD Institute Database	http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiDownloads.html
Saccharomyces Genome Database (SGD)	http://www.yeastgenome.org/
Joint Genome Institute (JGI)	http://genome.jgi-psf.org/euk_home.html
Transcriptomics	
EMBL-EBI Resources	http://www.ebi.ac.uk/
Gene Ontology	http://www.geneontology.org/GO.function.guidelines.shtml?all
TRANSFAC Database	http://www.gene-regulation.com/cgi-bin/pub/databases/transfac/search.cgi].
Proteomics	
ExPASy	http://expasy.org/tools
Pfam Database	http://pfam.sanger.ac.uk/
Metabolic modeling	
KEGG Database (Kyoto Encyclopedia of Genes and Genomes)	http://www.genome.ad.jp/kegg/
Reactome	http://www.reactome.org/
BRENDA	http://www.brenda-enzymes.info/
UniProt Database	http://www.uniprot.org/

Table adapted from [60, 198]

release of the genome sequence of *Haemophilus influenzae* [30], and following the release of a few other bacterial genomes the *E. coli* genome was released in 1997 [20]. The genome sequence of yeast was released in 1996 [22], but the complete sequence of chromosome III (the first complete chromosome sequence released) was already available in 1992. Genome sequencing of filamentous fungi is generally considered to have started as late as February 2001 [31] with the publication of a draft version of the genome sequence of *Neurospora crassa*. Nearly three decades have passed since the invention of electrophoretic methods for DNA sequencing, often referred to as Sanger sequencing, and its cost-effectiveness has mainly been driven down following the introduction of automation and the numerous refinements of this technology [32]. Recently various novel sequencing technologies have been developed, each aspiring to reduce costs to the point at which the genomes of individual humans could be sequenced as part of routine health care [32]. For example, the human genome project led to a significant decrease in sequencing costs, from USD $10 per finished base to 10 finished bases per USD $1 [33], thus reducing the cost of DNA sequencing by over two orders of magnitude [34]. As discussed lately by Shendure and Ji [34], next-generation DNA sequencing has the potential to accelerate dramatically biological and biomedical research by enabling the comprehensive analysis of genomes, transcriptomes, and interactomes. These practices aim at becoming inexpensive, routine and widespread, rather than

requiring significant production-scale efforts. Several projects have appeared since then, such as the Genomes to Life program launched by the US Department of Energy [35], where the goal was to achieve a basic understanding of thousands of microbes and microbial systems in their native environments and which had a huge demand of computing, technology, data storage, manipulation, and systems-level integration. In addition, such "large-scale biology" represents a significant compilation of resources from the traditional question-driven approach that had been so successful in laying the foundations of molecular biology.

3.2 Transcriptomics

Several DNA arrays have been designed for transcriptome analysis in a variety of platforms ranging from prefabricated oligonucleotide chips (Affymetrix GeneChip technology) to custom-made chips (spotted arrays) where a robot spots cDNA, oligonucleotides, or PCR products on a glass slide or membrane [36]. Several companies manufacture arrays in a wide range of designs for a variety of applications such as Affymetrix [37], NimbleGen [38], Agilent Technologies [39], Fermlab [40], and Febit [41].

Gene expression analysis using DNA microarrays is a powerful tool for dissecting complex regulatory circuits as one can simultaneously measure the expression of thousands of genes giving the opportunity to obtain a snapshot of the total mRNA pool of a living cell or tissue. Nevertheless, because of the often poor correlation between gene transcription and metabolic fluxes, high-throughput transcriptome analysis has not been used for identification of metabolic engineering targets [42]. Microarrays can be used to identify "new genes" involved in a pathway or regulatory networks and therefore to make functional assignments. Another application of DNA microarrays is the detection of mutations in specific genes, allowing the simultaneous screening of many possible mutations within a single gene, an application also known as genotyping [43].

To facilitate the collection and comparison of acquired data, transcriptome databases have been built, e.g., the Gene Expression Omnibus database [44] and ArrayExpress database [45], which are repositories for transcriptomics data mainly in the form of gene-indexed expression profiles from a curated subset of experiments where CEL files obtained from microarrays experiments are stored and allow researchers to conduct their own analysis of the data from preprocessing and normalization up to gene expression index calculation and statistical analysis. Findings from these studies will have broad implications for understanding the cells as a whole. An impressive number of transcriptome studies in a wide variety of industrial microorganisms have been conducted so far, particularly in the yeast *S. cerevisiae* [46, 47], in the bacteria *Bacillus subtilis* [43] and *Streptomyces coelicolor* [48], and in filamentous fungi such as *Aspergillus niger* [49], *Aspergillus oryzae* [50], and *Aspergillus terreus* [51].

3.3 Proteomics

According to Anderson et al. [52] and Joyce et al. [53], proteomics can be defined as the identification and quantification of protein-level measurements. The main goal of proteomics is the quantitative description of protein expression and its changes under the influence of biological perturbations [52]. A number of proteomics methods have been developed so far, such as isotope-coded affinity tags [54], protein arrays [55], two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) combined with mass spectrometry (MS) of excised proteins [56], yeast two-hybrid systems [57], and high affinity epitope tags [58]. However, not all the technologies are suitable for all microorganisms due to several factors such as cell rigidity (i.e., cell-lysing step of filamentous fungi is a limiting step for application of numerous technologies [59]. More recently, the introduction of MS as a robust and sensitive technology for protein analysis had a major impact on the analysis of complex proteome samples [60]. For a comprehensive summary of software solutions for handling proteomics data we refer to Mueller et al. [60]. Furthermore, a thorough list of open-source tools for MS/MS-base proteomics is available in a review from Newvizhskii and coworkers [61].

Yeast proteome studies, such as the one from Zhu and coworkers [55], helped to identify new calmodulin and phospholipid interacting proteins and a common potential binding motif for many of the calmodulin-binding proteins. Currently, microarrays of an entire eukaryotic proteome can be prepared and screened for diverse biochemical activities such as detection of post-translational modifications. Analysis of protein expression in yeast can also be performed using high affinity epitope tagging resulting in fusion proteins expressed under the control of their natural promoters [58]. The application of this methodology helped to uncover transcriptional regulation where protein levels were not detectable by other proteomics techniques.

As for all the *omics* technologies, data processing is a central and critical component of a successful proteomics experiment and it is often the most complex and time-consuming step. There have been considerable advances in the field of proteomics informatics in the past few years, mainly encouraged by free open-source software tools [60, 61]. The benefits arising from the availability of data repositories for raw data and results has been shown. For example, in a research community effort, guidelines were developed for using MS in proteomics applications [62]. This has the function of comprising a checklist of information that should be provided when generating a data set that is submitted to a database or when such an experimental step is reported in a scientific publication. Therefore, compatible data standards have now been established and supported both by public databases and by data processing software [62]. MIAPE-MS, is a project where general features, such as ion sources, i.e., electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) and all major components after ion source, are documented. In addition to the data resulting from the procedure, the peaks list generation method, the location of the raw data from which they were

generated, the method by which quantification was performed (where appropriate) and the resulting quantitative data set should also be documented [62].

MS has also lately emerged as a valuable method for proteogenomics annotations that improved prediction of genes. A recent study showed that such a comparative proteogenomics approach (like comparative genomics) allows addressing the problems that remained beyond the reach of the traditional "single proteome" approach in MS [63]. In particular, as suggested by Gupta et al. [63], comparative proteogenomics could deal with the notoriously difficult problem of one-hit-guess that will improve the existing gene prediction tools in genomics, and allow identification of rare post-translational modifications.

3.4 Metabolomics

Metabolomics is a tool to quantify the cell's metabolites, aiming at increasing the understanding of how changes in metabolites levels affect phenotypes [64]. As suggested by Tweeddale [65], the metabolome can be described as all the metabolites comprised within a cell. Here we will refer to metabolome as "*the full catalogue of metabolites generated by a growing organism, detected and quantified at a certain time during their growth or production cycle*" which is consistent with the definition given by Oliver et al. [66]. For a detailed review in metabolome analysis as well as for a compendium of concepts, methods and examples, we refer to some recent reviews [64, 67, 68].

Several methods aimed at analyzing metabolite levels were developed in the early days of biochemistry, but only recently attempts to detect metabolite changes have been made. This has been achieved by simultaneously analyzing a large number of metabolites through high-throughput methods, mainly with the help of analytical techniques such as MS [68] and nuclear magnetic resonance (NMR) [69]. One of the earlier methods used was capillary electrophoresis-ESI MS (CE-ESI-MS), which was developed in 2002 by Soga et al. [70] and later applied to the quantitative analysis of amino acids [71]. In the last decade, metabolite profiling using isotopic tracers has been largely used, and several publications are found in a vast range of species, i.e., in the bacteria *E. coli* [72, 73], in the yeast *S. cerevisiae* [74, 75], and in filamentous fungi [76]. These techniques have been used for functional genomics studies and have provided deeper insight into the microbial metabolism and cellular physiology.

Metabolic footprinting represents a niche within metabolomics because it primarily focuses on the analysis of extracellular metabolites [64]. Mapelli et al. [64] defines the exometabolome as the entire set of low molecular weight compounds present in the extracellular medium. This method has been used to successfully discriminate mutations in genes closely related in the metabolism. In mutations that are silent, metabolite analysis should be among the preferred approaches to detect any genotypic alteration, as suggested by Raamsdonk et al. [69], since metabolite levels might change substantially to compensate for the effect of the mutations. The analysis of metabolites that

are naturally released from microorganisms can provide useful information for fundamental research in the field of functional genomics and strain characterization [64].

Several applications of this technique can be found, for example in the production of antibiotics or biosurfactants, for the identification of quorum sensing metabolites which might contribute to the discovery of novel natural drugs, in the secretion of complex natural products such as antitumor agents and immunosuppressors [77], or for identification of commercially important traits in the brewing [78] and wine industries [79], or as a direct and rapid quality control method able to monitor the specific aroma generated during wine fermentations [64], confirming its potential for industrial biotechnology.

To use metabolic footprinting as a technique for high-throughput applications, benchmark spectra databases with identified peaks are required so that peak patterns obtained from MS or NMR analysis can be rapidly translated into relevant biological information. Common experimental procedures should, ideally, also be established for metabolite analysis [80] such as those existing in proteomics or transcriptomics. Nevertheless, the scientific community has only recently attempted to achieve these tasks. Several databases for identification of metabolomics signals by MS are now available, for instance, BIGG [81], BioCyc [82], MSlib [83], NIST [84], Metlin [85], and HMDB [86] databases. For a more comprehensive list of resources we refer to the review of Werner and coworkers [68].

3.5 *Fluxomics*

The aim of fluxomic studies is the quantification of the carbon fluxes through metabolic networks, where the fluxome can be defined as the ensemble of the metabolic fluxes present in a cell in a certain time. Fluxes represent the final outcome of the cellular regulation at different levels [87], differently from the others *omics* tools, which are based on the detection of only one level of information. Fluxomics provide a global perspective, giving information about how metabolites change in the metabolic network, determining a certain phenotype [88]. The flux at any given reaction in the metabolic network is often an unknown function of enzyme activity, substrate and product concentrations as well as the underlying kinetics. The enzyme activity depends on transcriptional and translational efficiencies in addition to post-transcriptional and post-translational modifications. Therefore, a given flux can be defined as being regulated at the hierarchical (from gene to enzyme activity) and/or at the metabolic level (kinetic dependence of flux on metabolite pools) [89, 90].

Because of the time-dependency of the metabolic fluxes, there are no direct methods for their analysis in-vivo. Nevertheless, intracellular fluxes can be quantified assuming that the intracellular concentration of metabolites is constant at all times (pseudo-steady state assumption). For a given metabolic network, the balance around each metabolite imposes a number of constraints on the system. In general, if there are "J" fluxes and "K" metabolites, then the degree of freedom is "$F = J - K$". Through the measurements of F fluxes, i.e., nutrient uptake, growth

rate and product secretion rates, the remaining fluxes can be calculated applying linear programming [3].

Intracellular fluxes can be estimated more precisely through ^{13}C tracer experiments. Following ^{13}C feeding to a cell it is possible to analyze metabolic products, such as amino acids, and measure ^{13}C enriched patterns, so to be able to reconstruct the flux distribution from the measured data [91]. To obtain flux data from the labeling patterns, two techniques can be applied: NMR [92, 93] and MS [94, 95]. Due to the low intracellular concentration of metabolites, these are often difficult to measure therefore the analysis of the labeling pattern of amino acids in proteins is used as input for flux quantification. Here proteins are hydrolyzed to release labeled amino acids and further analyzed by NMR of GC-MS. Once NMR or MS spectra are recorded, the next step is the quantitative interpretation of the isotopomer data by using mathematical models that describe the relationship between fluxes and the observed isotopomer abundance [96, 97]. Some of the mathematical approaches used include cumulative isotopomer (cumomers) [98], bondomers [99], and fractional labeling [100]. For a more comprehensive review on the methods we refer to Sauer [91].

3.6 Metabolic Modeling

Mathematical models are widely applied in biosciences and different modeling routes can be taken to describe biological systems. The type of model to use depends completely on the objective of the study. Models can be dynamic or static, deterministic or stochastic. Kinetic models are commonly used to study transient states of the cell such as the cell cycle [101] or signal transduction pathways [102], whereas stoichiometric models are generally used when kinetics parameters are unknown and steady state systems is assumed [48, 103].

The progressive annotation of genome sequences has led to the reconstruction of genome-scale metabolic networks. Different versions of metabolic reconstruction have been reconstructed for several microorganisms and the up-to-date available models for *S. cerevisiae* and *Aspergillus* spp. are listed in Table 5. [104]. Metabolic models are reconstructed by collecting the stoichiometry of all the metabolic reactions occurring in a given cell in a stoichiometric matrix. All the reactions included in the model are balanced for metabolites, cofactors, and energy, and reversibility is also assumed. Several reviews on the process of reconstruction of these networks are present [105, 106]. Through the use of flux balance analysis (FBA) and fixing an objective function it is possible to use stoichiometric models to simulate the optimal flux distribution under different conditions and to predict cellular parameters such as growth or product formation [106, 107]. Optimal solutions within this space can be determined with constraints-based optimization [108, 109]. A fair objective function is maximum growth, i.e., experimental and in silico studies with *E. coli* have

Table 5 Reconstructed genome-scale metabolic models of *S. cerevisiae*, *Aspergillus niger*, and *A. oryzae* completed by December 2008

Organism	Genome sequenced	Metabolic network characteristics					Reference
		Total reactions (unique)[a]	Total metabolites (unique)[b]	Total genes (enzymes)[c]	Compartmentalization[d]	Model Id[e]	
S. cerevisiae	*S. cerevisiae* S288C	1,175 (842)	584	708	Cytoplasm, mitochondria, extracellular	iFF708	[107]
	S. cerevisiae S288C	1,489 (1,149)	646	750	Cytoplasm, mitochondria, peroxisome, nucleus, endoplasmic reticulum, Golgi apparatus, vacuole, extracellular	iND750	[219]
	S. cerevisiae S288C	1,038	636	672	Cytoplasm, mitochondria, extracellular	iLL672	[326]
	S. cerevisiae S288C	1,431	1,013	795	Cytoplasm, mitochondria, extracellular	iIN795	[220]
	S. cerevisiae S288C	1,761	1,168	–	Cytoplasm, mitochondria, peroxisome, nucleus, endoplasmic reticulum, Golgi apparatus, vacuole, extracellular	-	[221]
A. niger	*A. niger* CBS 513.88	355	284	20	Cytoplasm, mitochondria, glyoxysome, extracellular	iHD20	[327]
	A. niger CBS 513.88 and *A. niger* ATCC 9029	2,443	2,349	(988)	Cytoplasm, extracellular	iJS988	[328]
	A. niger CBS 513.88 and *A. niger* ATCC 1015	2,240 (1,190)	1,045 (782)	871	Cytoplasm, mitochondria, extracellular	iMA871	[236]
Aspergillus oryzae	*A. oryzae* RIB40	(1,679)	1,040	1,184	Cytoplasm, mitochondria, extracellular	iWV1184	[237]

This overview accounts for a total of three species represented by nine reconstructed metabolic network models

Notes: Genome-scale models report several different parameters characterizing them

[a]*Total reactions*: It includes intracellular, extracellular and transport reactions. Where possible, unique reactions were defined as the total number of reactions absent of any isoenzyme catalyzed reactions, where the reaction stoichiometry was identical

[b]*Total metabolites*: It includes all reactants, products, cofactors, catalysts, and intermediates involved in any stoichiometric reaction. Unique metabolites are defined as those unique in chemical structure, since a fraction of metabolites with identical chemical structure may be found in multiple compartments

[c]*Total genes*: It includes all open reading frames (ORF) producing a gene product that catalyzes a defined stoichiometric reaction. However, several model reconstructions only include the gene product (i.e., enzymes) with no indication of ORF associations

[d]*Compartments* reported in the reconstructed model. However, in several models, the compartments are not mentioned therefore, compartmentalization was assigned by inspection of the model

[e]*Common nomenclature* used for model identification, i-First name-Last name-Number of ORFs represented in the model. Some exceptions are present, and the model name used in the original publication was provided for consistency

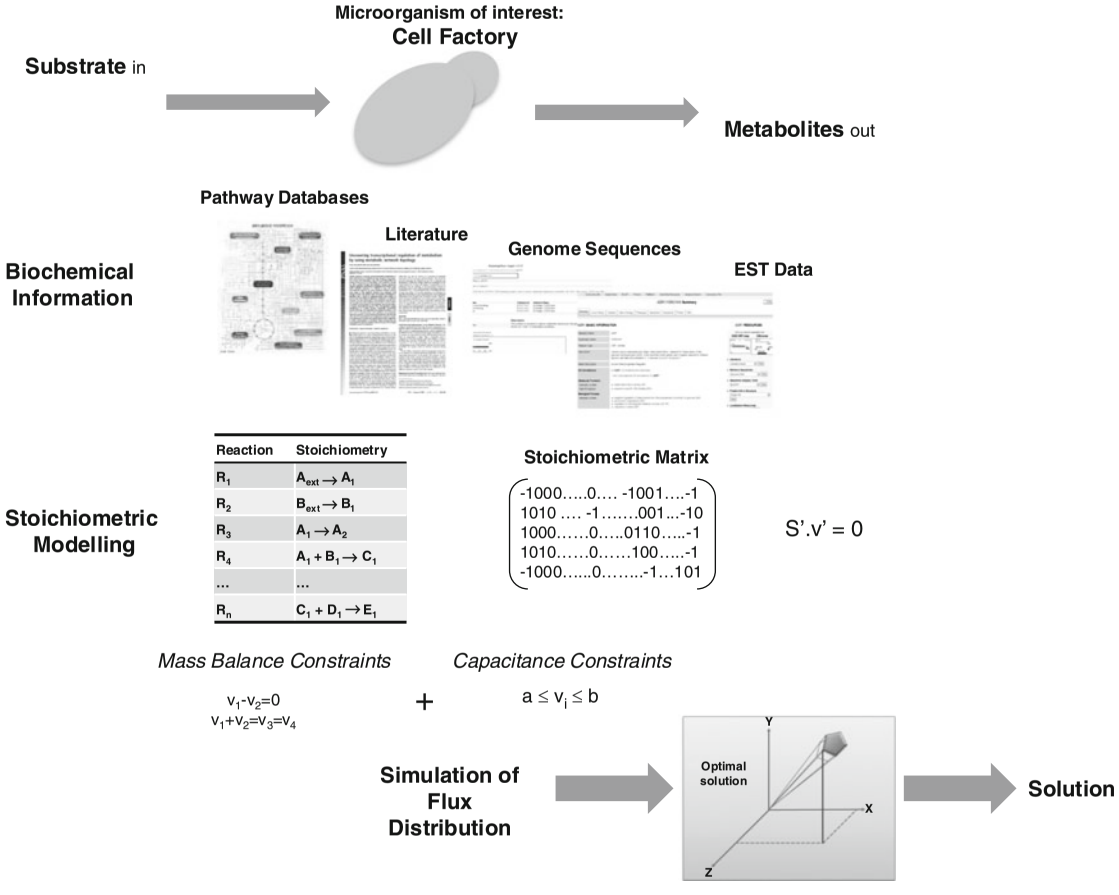


Fig. 2 Simplified workflow for the reconstruction of genome-scale metabolic models and their use in simulating cellular function through flux balance analysis

demonstrated that its metabolic network has evolved for optimization of the specific growth rate [104].

Additionally, genome-scale metabolic models (GSMMs) provide a framework for organizing, integrating, and analyzing *omics* data. A figure showing the basic steps and a simplified workflow [14, 110] for reconstruction of metabolic models is shown in Fig. 2.

In silico models have already produced several interesting results. For instance, the initial models predicted about 60% of knockout growth phenotypes correctly in *Helicobacter pylori* [111], 86% in *E. coli* [112, 113], and about 85% in *S. cerevisiae* [106]. With the genome-scale in silico model of *E. coli*, several experimentally testable predictions have been formulated describing the quantitative relationship between the primary carbon source uptake rate, oxygen uptake rate, and cellular growth rate.

The availability of well annotated genomes have not solely enabled the reconstruction of genome-scale stoichiometric models, but together with the high-throughput metabolomics and fluxomics data [91, 114], as well as methods estimating thermodynamic properties of the systems [115], they have opened the possibility of building genome-scale kinetic models. The development of these models is still in its infancy and there are several hurdles hampering the establishment of such models [116]. For a comprehensive framework we refer to Jamshidi et al. [116], who suggested a workflow for constructing genome-scale kinetic networks as well as the underlying properties in these reconstructions. In the near future, the challenge is to have comprehensive models which account for metabolism, transcription, translation, and regulation, serving as the basis for iterative, prospective computer-aided design of technologically useful products.

4 Yeast Cell Factories

4.1 Characteristics and Applications of S. cerevisiae as Cell Factory

Thanks to its long tradition in baking, brewing, and wine making, the yeast *S. cerevisiae* has been extensively studied and characterized. A deep knowledge in genetics, physiology, and biochemistry has been reached, allowing the establishment of robust molecular biology methodologies [117] and fermentation technologies. After the "genomics revolution," many *omics* tools have been developed using this yeast, and an impressive amount of information about gene expressions levels, metabolites patterns and metabolic fluxes is available and easily accessible. Besides its wide use as cell factory in industrial biotechnology, this yeast also serves as an eukaryal model organism [118–120].

S. cerevisiae is a unicellular microorganism and compared to higher eukaryotes, its genetic manipulation can be performed easily. A wide variety of expression

vectors (centromeric, episomal, low/high copies), dominant or auxotrophic markers, and efficient transformation protocols are available [121]. Furthermore, thanks to the high efficiency of homologous recombination (HR), it is possible to integrate stably genes into its genome. Many molecular biology techniques are routinely applied with this host system; therefore wide collections of deletion mutants have been systematically constructed and physiologically characterized and are commercially available [122]. Additionally, this yeast has been given the GRAS (Generally Regarded as Safe) status, making it an attractive production platform for many applications. Due to its ability to grow on minimal mineral media without particular requirements, and thanks to the long tradition with large-scale fermentations, it is not unexpected that this yeast is among the most exploited cell factories.

As a compartmentalized microorganism *S. cerevisiae* is able, differently from prokaryotes, to perform protein processing and folding, a feature particularly attractive for recombinant expression of mammalian proteins. Nevertheless yeasts and humans have different glycosylation pathways, which can be inconvenient as it can lead to production of protein drugs that cause an immunogenic response. Currently the engineering of humanized glycosylation pathway [123] has been achieved in *P. pastoris* [124], but this technology can also easily be implemented in *S. cerevisiae*. When considering this yeast as a host for recombinant protein production, issues related to hyper-glycosylation and, in some cases, inefficient secretion have to be mentioned. Particularly, the choice of a host for the successful production and secretion of a recombinant protein is totally dependent on the nature of the protein [125].

Besides being applied for the production of recombinant proteins, *S. cerevisiae* is extensively evaluated for its ability to produce commodity chemicals such as, for instance, organic acids. There are several reasons for this: (1) it can tolerate relatively low pH, allowing the recovery of the acid in its carboxylated form and hereby minimizing the associated downstream costs; (2) it has a very high glycolytic flux, which is exploited in bioethanol production and upon redirection of the carbon fluxes it is possible to obtain very high productivities; (3) the engineering of the central carbon metabolism is relatively easy, and several studies have shown that it is also possible to expand the substrate range of *S. cerevisiae* for efficient utilization of pentose sugars like xylose and arabinose. This capability can be directly coupled to the production of chemicals, opening new possibilities for the development of sustainable processes and conferring an additional advantageous feature to this microorganism.

4.2 Strain Improvements of Baker and Brewer Yeast

From the time when microbial systems were first exploited as cell factories, there has been a constant attempt to improve their properties or their performance in terms of yield and productivity. The technology of manipulating and improving microorganisms to enhance their metabolic capabilities for biotechnological applications is referred to as strain improvement [126, 127], and in order to develop optimal yeast strains for industrial applications, several traits have to be improved.

All of them are product and process dependent but generally an increased yield of product on the substrate [128] and an increased growth rate are common properties that *S. cerevisiae* is required to have in order to produce an efficient processes. More specific desired features are the high productivity of ethanol in brewing and wine making, of CO_2 in baking, and minimization of overflow metabolism (ethanol and glycerol) and increased biomass yield on the substrate in applications where product formation is directly coupled to biomass.

To achieve these goals two main routes can be pursued: one is improving the technology of the process through bioprocess engineering [129] and the other is to genetically manipulate the strain. The advances obtained in process development through bioprocess engineering will not be discussed here. Strain improvement has generally been achieved through random mutagenesis and screening through classical genetics techniques. Random mutagenesis can be applied using chemicals (such as ethyl methane sulfonate, EMS) or physical mutagens (e.g., UV light exposure). More recent mutagenesis techniques are those based on error-prone PCR (e.g., using DNA polymerase without proofreading ability). Other advanced techniques which hold relevance, especially when polygenic phenotypes are desired, are global transcriptional machinery engineering and whole genome shuffling [130].

In yeast, classical genetic improvement [131] is mainly applied through sexual breeding (i.e., by mating strains) [132] or through parasexual hybridization techniques [133] such as protoplast fusion. Since these methods represent the oldest protocols for strain improvement, they have been widely used to improve the processes of baking, brewing and wine making. In wine making low sulfide content and optimal flocculation are two desired characteristics. To develop an improved phenotype showing these properties, Romano et al. [134] successfully applied spore conjugation. An example of application of electrofusion techniques is that described by Urano et al. [135] where the authors succeeded in changing the phenotype of a brewing strain from non-flocculating into a highly-flocculating one. Another example of strain improvement using non-recombinant techniques is that described by Higgins et al. [136] on baker's strains. The authors were challenged to combine the two properties, normally separated in wild types, of efficiently leavening both non-sugared and sweet bread dough by applying several rounds of combined sporulation and enrichment under high osmotic conditions. With the aim to improve the properties of baker's yeast, Angelov et al. [137] used chemical mutagenesis (EMS) [138] to generate strains with improved maltose-fermenting capabilities. Van Dijck et al. [139] developed en efficient protocol to isolate mutants with increased tolerance to freezing conditions. The work was performed both on a lab and on an industrial strain but whereas the selective pressure applied on the lab strain involved heat stress treatment, for the industrial strain repetitive freezing treatments were used as selective conditions, selecting afterwards for surviving strains.

Even though all the mentioned approaches involve the use of different techniques, they are all are based on the common procedure of generating mutants and screening for a desired phenotype. Even when these methods are very efficient,

they do show some drawbacks: usually the screening process requires considerable work and no mechanistic understanding of the generated phenotype or correlation with its genotype is present. This makes it impossible to improve the strains further and it is difficult to add/remove properties without altering its general performances [140].

The development of metabolic engineering has of course enabled more directed genetic modifications through the use of recombinant DNA technologies. There are also several examples of the application of metabolic engineering in the field of brewing and baking, even though traditional industries have generally been reluctant to introduce genetically modified microorganisms.

Molasses are among the preferred substrates in baking and brewing applications [141]. They are composed not only of glucose but also of sugars such as galactose, fructose, and sucrose. In *S. cerevisiae* the uptake of glucose prevents the utilization of the other sugars, a phenomenon also known as glucose repression; this leads to increased process time, since only a minor part of the given substrate can be metabolized. Even though glucose repression has been the object of many detailed studies [142], a full understanding of this phenomenon has not been achieved. What is known is that the protein Mig1p plays an essential role, binding to the promoter of several glucose-repressed genes, including galactose metabolism genes *GAL*, melibiase genes *MEL* (necessary for hydrolysis of melibiose into glucose and galactose), maltose metabolism genes *MAL* (maltose permease and maltase), and the invertase *SUC2* (necessary for the hydrolysis of sucrose into glucose and fructose). Several attempts aiming at relieving glucose repression have been made [143]. An obvious target is the deletion of *MIG1* [144]; however, it has been shown that this deletion is not able to eliminate glucose repression completely. Glucose repression represents a limitation also for the efficient utilization of oligosaccharides as raffinose, melibiose, and maltotriose, substrates widely used in brewing applications. To obtain a strain able to metabolize melibiose, the *S. pastorianus* melibiase gene *MEL1* was expressed in *S. cerevisiae* on a multicopy vector or by integration into the genome in the *LEU2* site [145]. Since the melibiase gene is also subjected to glucose repression, the *MEL1* gene was integrated into the genomics loci of *MIG1* and all the copies of *GAL80* were also deleted [146]. This work, performed on a strain with industrial background, led to the successful engineering of a melibiose utilizing strain alleviated in glucose repression on galactose utilization. The obtained phenotype showed an increased ethanol yield of 4% due to efficient utilization of melibiose.

The achievement of optimal flocculation properties in brewer's strain has also been targeted by metabolic engineering. The overexpression of the flocculin gene *FLO1* is an obvious strategy which, unfortunately, leads to premature flocculation [147] and, as briefly mentioned, flocculation should only arise at the beginning of the stationary phase when the fermentation process is almost over. Thanks to the good knowledge of this yeast and to the presence of robust molecular biology tools, the promoter of *FLO1* was replaced with the promoter *HSP30* [148], previously identified as able to turn on the expression only at the beginning of the stationary phase [149].

In baker's yeasts the ability to resist several rounds of freezing and thawing without losing the property of dough leavening confers an obvious advantage. This property, also called cryoresistance, is known to be related to the trehalose content. A strain engineered through the overexpression of neutral-trehalase gene *NTH1* and trehalose synthase gene *TPS1* [150] showed an increased trehalose content even though this property was shown to confer resistance only under non-fermenting conditions [151].

4.3 *S. cerevisiae as a Platform for Chemical Production*

With the development of metabolic engineering the role of *S. cerevisiae* as cell factory became further consolidated. Numerous approaches to engineer *S. cerevisiae* for the production of a wide range of chemicals can be found in the literature. Whereas the most promising are summarized in Table 6, we will describe below just a few examples where metabolic engineering of yeast has been applied in industrial biotechnology.

L-Lactic acid is an organic acid mainly used in the food industry thanks to its probiotic effects but also, to some extent, in pharmaceuticals and cosmetic applications. A new role of L-lactic acid as a mature fine chemical has emerged in the last few years due to its use as monomer in the manufacturing of biodegradable plastics, packaging materials, fibers, and solvents [152]. This last application explains the growth in the market in the last decade: in 2007 the worldwide production of lactic acid was estimated to be around 130,000–150,000 tons/year [153] with an average cost of about 1.59 US$/kg. In the 1980s, lactic acid was synthetically manufactured through the intermediate lactonitrile, but this process had the disadvantage of

Table 6 Principal products of the cell factory *S. cerevisiae* through metabolic engineering

Product	Reference
Insulin	[329, 330]
Various antibody	[331]
Hepatitis B surface antigen	[332]
Antimalaria drug precursor	[323]
Hirudin	[333]
Ethanol	[334, 335]
Lactic acid	[162, 336]
Isoprenoids	[337, 338]
Polyketides	[303]
Polyhydroxyalkanoates	[339, 340]
Butanol	[341]
Carotenoids	[294]
Hydrocortisone	[168]
Taxol intermediate	[299]
Resveratrol	[291, 292]
Vitamin C	[342]

producing both stereoisomers of lactic acid. Currently the manufacture is based on a fermentative processed with homolactic bacteria such as *Lactobacillus lactis*, *Lactobacillus delbrueckii*, *Lactobacillus bulgaricus*, or *Lactobacillus helveticus*. These bacteria, in the presence of glucose and limited oxygen availability, are able to efficiently produce lactic acid with a molar yield close to 2. The fermentation is operated anaerobically through a batch or fedbatch process and usually lasts from 2 to 4 days [154]. To maintain the pH at the optimum for bacteria growth (pH 6) it is necessary to add bases (usually calcium hydroxide or calcium carbonate), resulting in the production of the acid in its dissociated form (pKa of lactic acid is 3.86). The calcium lactate produced has to be filtered, carbon treated, evaporated, and acidified with sulfuric acid to recover the lactic acid, resulting in the formation of huge amounts of calcium sulfate as by-product. Another drawback of this process is represented by the nutritional requirements of many lactic acid bacteria (LAB), which leads to end product inhibition [155]. To tackle this problem, several methods to remove lactic acid during the fermentation have been reported (electrodialysis, ultrafiltration, and microfiltration) which, even if successful, add additional costs to the process. In order to render the process more cost-competitive, several approaches to improve the performances of LAB have been reported, including the attempt to generate acid tolerant mutants by genome shuffling [130].

An obvious advantage would come from using a microorganism tolerant to lower pH. Yeasts do show this feature and *S. cerevisiae* is a particularly suitable candidate as a cell factory for the production of lactic acid [91]. *S. cerevisiae* does not naturally accumulate lactic acid, as pyruvate (which is the precursor in the bacterial pathway) is metabolized either to acetyl-coenzyme A by the pyruvate dehydrogenase complex or to acetaldehyde and further to ethanol by pyruvate decarboxylase (PDC; see Fig. 3). To produce lactic acid it is therefore necessary to engineer yeast by introducing lactic acid dehydrogenase gene *LDH* that allows

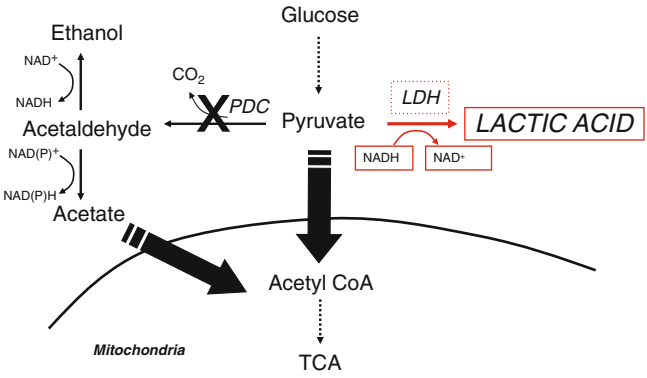


Fig. 3 Metabolic engineering for L-lactic acid production in *Saccharomyces cerevisiae*. The abbreviation of enzymes is described as follows: *LDH* bovine lactate dehydrogenase, *PDC* pyruvate decarboxylase complex encoded by the genes *PDC1*, *PDC5*, and *PDC6*, which were deleted in this strategy

reduction of pyruvate to lactic acid. Porro et al. [156] described the engineering of a *S. cerevisiae* strain expressing a bovine *LDH* gene under the control of a strong promoter. This strain is able to direct the glycolytic fluxes towards the production of lactic acid leading to accumulation of lactic acid in the medium up to 20 g/L. The engineered strain, however, still produces ethanol and it was therefore thought that preventing the onset of ethanol fermentation, e.g., by shutting down PDC activity, could help to re-direct the carbon flux further towards lactic acid formation, thereby increasing the lactate yield. In yeast the deletion of PDC genes *PDC1*, *PDC5*, and *PDC6* results in poor growth on glucose [157], but when a strain with these deletions was transformed with a vector containing the *LDH* gene, the cells regained the ability to grow on glucose. The resultant strain showed reduced ethanol formation but the lactate yield was not substantially improved.

A similar approach has been applied to the Crabtree-negative yeast *Kluyveromyces lactis* [158] which, differently from *S. cerevisiae*, only has one *PDC* gene, showing that in this yeast the pyruvate flux toward ethanol can be fully replaced by lactic acid production. An elegant approach to characterize the physiology of the strain engineered for lactic acid production is described by van Maris et al. [159]. In a study of the strain overexpressing the *LDH* gene and deleted of all the three genes coding for PDC (*PDC1*, *PCD5*, and *PDC6*), it was found that lactate production was strongly dependent on the oxygen availability, consequently the homofermenting strain was not able to grow under anaerobic conditions, a result that can be explained assuming no net ATP gain from lactic fermentation, probably because of an energy requirement for lactate export. Another route that was taken to boost the formation of lactic acid was the overexpression of the lactate transporter codified by the gene *JEN1* [160], in the strain where *LDH* was already overexpressed [161, 162].

Even though the yields obtained in yeast are not yet comparable with those obtained with LAB, *S. cerevisiae* has great potential for the development of a cost-competitive process, leading to the conclusion that this eukaryotic host can represent an alternative production platform to LAB. The production of lactic acid from biomass with LAB has already been the object of investigation [153]; nevertheless, the advances obtained with *S. cerevisiae* in using raw biomass as substrate represent important drivers for exploiting this yeast for chemicals production.

What emerges from the lactic acid case is the hurdle coming from the tight regulation of the metabolism. This factor has a huge impact in the design of metabolic engineering strategies. The manipulation of apparently obvious targets does not usually confer the desired properties and the related effects are often unpredictable.

Besides the engineering of *S. cerevisiae* for organic acid production, through metabolic engineering it is possible to reconstruct entire pathways. In 1994, Yamano et al. [163] reported the reconstruction of a complete secondary metabolic pathway in *S. cerevisiae*, resulting in the ability of the yeast to produce β-carotene and lycopene. Carotenoids are a class of pigments used in the food industry and, due to their antioxidant properties, they have wide commercial interest. The biosynthesis of these compounds does naturally not occur in *S. cerevisiae* and to allow

carotenoids production it was necessary to transfer the pathway from a naturally producing microorganism, in this case the Gram-negative bacteria *Erwinia uredovora* (Fig. 4). By expressing the heterologous genes coding for four enzymatic activities, it was possible to obtain a yeast strain that could produce 103 μg β-carotene/g biomass. This study represents one of the first examples of pathway engineering in yeast, demonstrating the versatility of *S. cerevisiae* for reconstruction of more complex metabolic engineering designs.

Due to the many applications and the industrial relevance of this class of molecules, several studies to identify strategies to increase their production have been conducted. In the recent work by Verwaal et al. [164], *S. cerevisiae* has been engineered with the genes from the β-carotene production pathway from the yeast *Phaffia rhodozyma,* which is naturally able to produce these compounds. When these genes were cloned in *Candida utilis*, the carotenoids yields increased remarkably [165] and this led Verwaal et al. [164] to the idea that the production of β-carotene in *S. cerevisiae* could be improved by conferring properties from another yeast species. Like *S. cerevisiae*, *P. rhodozyma* also produces farnesyl-diphosphate (FPP; see Fig. 4) which is further converted into geranylgeranyl

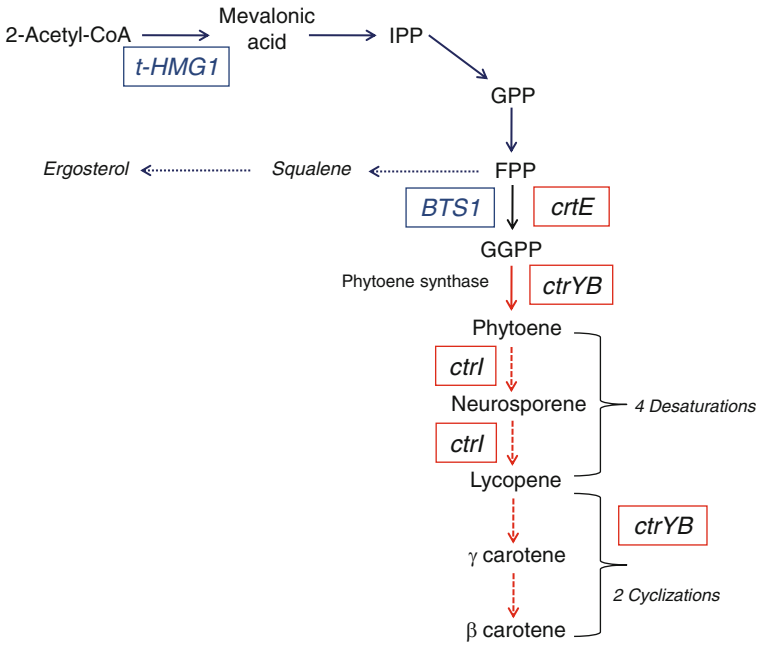


Fig. 4 Reconstruction of the carotenoids biosynthetic pathway in *S. cerevisiae*. Heterologous activities: *t-HMG1* 3 hydroxy-3-methylglutaryl-coenzymeA, *ctrE* geranylgeranyldiphosphate synthase (from *Xanthophyllomyces dendrorhous*), *ctrI* phytoene desaturase, *ctrYB* bifunctional enzyme phytoene synthase and lycopene cyclase. Endogenous yeast activities: *BTS1* geranylgeranyldiphosphate synthase. Metabolites: *IPP* isopentenyl diphosphate, *GPP* geranyl diphosphate, *FPP* farnesyl-diphosphate, *GGPP* geranylgeranyl diphosphate

diphosphate (GGPP). The expression of the two genes *crtI* and *crtYB* should be sufficient for β-carotene production (Fig. 4); nevertheless, it was found that the overexpression of *crtE* increased the total carotenoids levels. Since plasmid expression resulted in unstable expression, the threes genes *crtI*, *crtYB*, and *crtE* were integrated into the genome. Furthermore, it was tested whether an increased precursor formation through the overexpression of the gene coding for the catalytic domain of HMG-CoA reductase (*tHMG1*) could boost β-carotene production. The combination of these strategies resulted in a strain that was able to produce 5.9 mg/g dry weight of β-carotene (approximately 57-fold more than the previous example) [166, 167]. This work shows the potential for further development of *S. cerevisiae* as a platform for carotenoids production.

Another important achievement in the field of pathway engineering is represented by the engineering of *S. cerevisiae* for hydrocortisone production [168]. Several therapeutical proteins are produced in this yeast [169], but hydrocortisone represents the first non-proteinaceous compound with therapeutic applications produced in this yeast. Hydrocortisone is a human steroid hormone, used in the biopharmaceutical market as an anti-inflammatory, abortive or antiproliferative agent. The chemical synthesis involves more than 40 steps and even though the actual production has been substantially simplified it is still very complex and cost-demanding. In an elaborative work, Szczebara et al. [168] reported the reconstruction of the pathway leading to hydrocortisone recruiting enzymatic activities from different sources and successively by adjusting the endogenous metabolism of *S. cerevisiae* to the introduction of the heterologous pathway. For the conversion of the endogenous intermediate ergosterol into hydrocortisone, nine heterologous products were necessary (Fig. 5) (one plant gene and five steps catalyzed by eight mammalian proteins: four members of the P450 superfamily [170, 171], three electron carriers, and 3β-hydroxy steroid dehydrogenase/isomerase (3β-HSD)). Most of them were membrane associated, adding high demand for proper assembly of the pathway. After the reconstruction of the recombinant pathway, the endogenous metabolism was adjusted by removing activities draining intermediates. The study was based on previous work of Duport et al. [172] where the authors engineered this yeast for the formation of progesterone. In this strategy the authors expressed the bovine electron carriers *ADR*, *ADX*, the bovine *P450scc*, the *Arabidopsis thaliana* Δ7Red, and the human *3βHSB* genes. The expression was achieved by integration into the genome of a strain where *ERG5* gene (Δ22-sterol desaturase) was deleted. This work represents an important advance in metabolic engineering of microorganisms for industrial applications [173] and it witnesses the advantage of engineering such complex phenotypes in yeast in virtue of their physiological similarity to mammalians. The fact that yeasts are essentially equipped with the same organelles as mammalian cells is probably not important for processes based on single step conversion, but it becomes a major advantage when expressing pathways catalyzed by enzymes located on different organelles.

The mentioned examples further demonstrate the advantages of working with a host where robust molecular biology tools allow custom-tailed expression of the

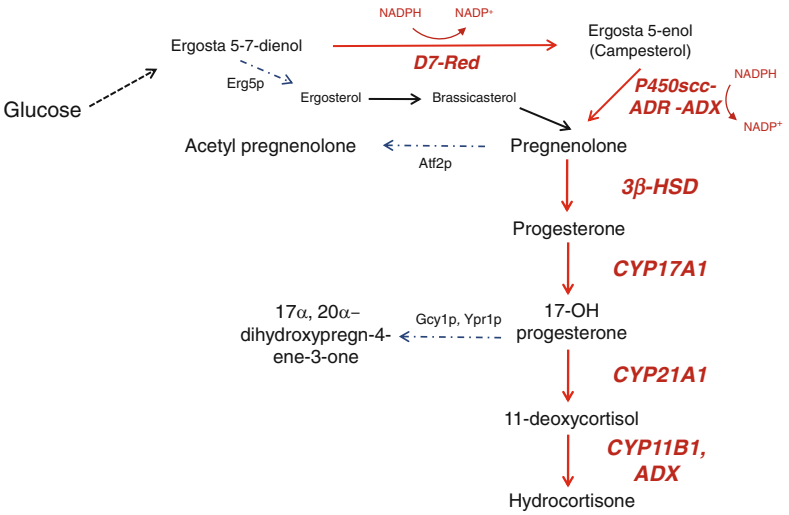


Fig. 5 Reconstructed pathway for hydrocortisone biosynthesis in *S. cerevisiae*. Recombinant enzymatic activities: *D7-Red* D7-sterol reductase gene from *Arabidopsis thaliana*, *P450scc* bovine side chain cleavage cytochrome P450, *ADR* bovine adrenoxin reductase, *ADX* adrenodoxin, *3β-HSD* type II human 3β-hydroxy steroid dehydrogenase/isomerase, *CYP17A1* 17a-steroid hydroxylase, *CYP21A1* 21a-steroid hydroxylase, *CYP11B1* 11b-steroid hydroxylase. *Dashed blue arrows* denote deleted enzymes activities of Erg5p, Atf2p, Gcy1p and Ypr1p

genes to be expressed. Without such an extensive knowledge of the techniques necessary to perform genetic manipulation, these approaches would not have been feasible or at least not to the same extent.

5 Fungal Cell Factories

5.1 *Characteristics and Applications of Aspergillus Species as Cell Factories*

Filamentous fungi are excellent producers of many metabolites that are valuable commodities, e.g., citric acid, itaconic acid and gluconic acid. Apart from metabolites, *Aspergillus* species are also able to secrete a wide range of enzymes, representing a rich source for the production of enzymes that can be used in detergents, food processing, and material handling. Of particular interest is that *Aspergillus* spp. have the capability to degrade a wide range of natural organic substrates including plant materials [174].

Another important feature is the GRAS status already granted by the Food and Drug Administration of the US government to *A. oryzae* and *A. niger* [175]. This is a highly desired classification in the food industry and moreover, as suggested by Baker and Bennett [174], because of the long history in the use of *A. oryzae*

and *A. niger* for the preparation of human foods and beverages, the process for production of new products from these fungi can be easily approved. The main concern of using *Aspergillus* as a production host, as with many filamentous fungi, is the potential production of exometabolites with toxic properties. Consistently, the safety of *A. niger* and *A. oryzae* as production organisms has been widely documented [176, 177] and metabolites profiles produced by several isolates of these strains used in enzyme production have been published [178, 179].

Here we will mainly focus on the use of *Aspergillus* species in biobased industrial processes. However, other filamentous fungi such as *Trichoderma reesei*, due to its high range of secreted lignocellullose degrading enzymes, e.g., cellulases and xylanases, and *Penicillum* sp. or *Rhizopus oryzae,* often used in the production of fermented foods and alcoholic beverages in Indonesia, China and Japan, are also of high industrial significance.

5.2 Strain Improvement of the Industrial Production Hosts *A. oryzae and A. niger*

So far *Aspergillus* cell factories have been mainly exploited for the production of endogenous metabolites and enzymes. Nevertheless, the fermentation technologies are very well developed, and *Aspergillus* species are therefore attractive cell factories for metabolite production. As mentioned in several recent publications, we are moving forward in the analysis of *Aspergillus* spp. mostly due to the availability of genome sequences and the efforts of the fungal community to develop robust technologies for genetic manipulation and several *omics* tools.

Classical strain improvement has been mainly done through chemical mutagenesis or natural mutagenesis caused by selective pressure. Worth mentioning is the first report of *Aspergillus* transformation, published 25 years ago [180]. In filamentous fungi, targeted gene modification is usually hampered by the low transformation efficiencies naturally achieved [181]. Targeting and replacement of gene loci in filamentous fungi, as in other organisms, are supported by the cellular machinery that accomplishes recombination and DNA repair [182]. Especially the rate of homologous recombination in a given host determines the efficiencies in knockout approaches using genetic markers that are flanked by homologous sequences of the gene locus to be replaced. In filamentous fungi, these homologous sequences are generally larger (several hundreds of base pairs) compared to other organisms [181, 183], i.e., bacteria and yeast where a minimal length of 30–50 bp is sufficient to ensure a high yield of HR [184]. There are two main mechanisms of DNA repair, by HR and by non-homologous end joining (NHEJ). In filamentous fungi, repair of DNA damage seems to occur primarily by using the NHEJ pathway, and a DNA fragment that is desired to be integrated in a certain position in the genome is therefore often integrated ectopically impeding the achievement of the desired phenotype. Recently, disruption of some genes involved in the NHEJ pathway,

namely *ku70* and *ku80*, enhanced the gene targeting efficiency in the yeast *K. lactis* [185] and in the filamentous fungus *N. crassa* [186]. This finding was immediately applied to industrially relevant filamentous fungi, for instance, *A. oryzae* and *A. sojae* [187] and later to *A. niger* [188]. In the first case, disruption of *ku70* or *ku80* enhanced the targeting efficiency from 10% to greater than 60%. However, the efficiency was reproducibly lower than that of ku-disruptants of *K. lactis* and *N. crassa* (nearly 100% for both) and suggested the existence of an unknown NHEJ pathway in *A. oryzae* and *A. sojae* [187]. More recently, deletion of DNA ligase IV (LigD), another factor involved in the NHEJ pathway, resulted in a targeting efficiency as high as 100% in *N. crassa* [189]. It has already been applied to *A. oryzae* [190] and it will hopefully be transferred to other industrial cell factories soon. This technology will allow the engineering of these species by allowing the deletion of industrially undesirable traits to improve productivity and safety further. The development of recombinant DNA technologies have given the possibility of introducing targeted mutations (i.e., overexpressing genes or deletion of undesired ones) instead of the random generation and further screening for the desired phenotypes which is very costly and time consuming. One of the most used transformation methods used for filamentous fungi is the protoplast mediated transformation (PMT) method, developed earlier for *S. cerevisiae* and adapted for filamentous fungi. Nevertheless, this method preferentially produces multicopy integration events [191]. Alternative methods for fungal transformation such as electroporation, biolistic transformation and *Agrobacterium* mediated transformation have been developed [192, 193]. These methods are valuable for fungal strains that do not form sufficient amount of protoplasts or for strains where their protoplasts do not regenerate. Overall, as discussed by Meyer et al. [181], individual species have to be considered independently and the most appropriate method identified and optimized for each strain.

In order to conduct industrial strain development, facile methods must be developed to understand genetic control of metabolite production and to identify productive routes for engineering [2, 194].

5.3 *Aspergillus* Species as Platform for Production of Chemicals

The first applications of *Aspergillus* species go back in time more than 2,000 years, when they were used to ferment rice, soybeans, and other plant foods in Asia [174]. Their use for industrial purposes started several centuries later when they were used to convert sugars into production of organic acids used as acidity regulators in the brewing and beverage manufacturing processes. The production of citric acid is based on the studies conducted by Wehmer in 1891 for the production of oxalic acid [195]. However, this fermentation was not promoted to a commercial process because oxalic acid could be produced at a lower cost by other means. Citric acid fermentation by *A. niger* has been studied for nearly a century, achieving yields of approximately 200 g/L and more than 90% conversion of the feedstock sugar after

improvement of the production process through adjustment of the cultivation medium (e.g., tuning the manganese concentration), controlling, to some extent, the morphology through control of agitation and aeration parameters [196]. The now known Pfizer process, originally used to produce citric acid by surface culture fermentation [197], revolutionized and represent the first large-scale submerged culture based process for chemical production.

Among *Aspergillus* spp., *A. oryzae* and *A. sojae* have been the preferred fungi used in the traditional Japanese fermentation industries to produce several products, including soy sauce, sake, bean curd seasoning and vinegar, all known as koji processes [198]. The use of strains from the *A. oryzae*/*sojae* species is one of the earliest examples of solid state fermentation. In 1894, the Japanese scientist Jokichi Takamine placed the first US patent to produce an enzyme in a microbial system; his commercial product was named Taka-diastase, a diastatic (amylolitic) koji enzyme cocktail used for treatment of indigestion. Nowadays, this amylase is used as an ingredient of dried infant grain cereals that can aid digestion [174]. Currently there are five major distributors in Japan supplying *A. oryzae* conidiospores to more than 4,500 sake (Japanese alcoholic beverage, about 1,900 brewers), miso (soybean paste, about 1,200 brewers) and shoyu (soy sauce, about 1,500 brewers) brewers in Japan, excluding several of the biggest soy-sauce companies that produce their own spores [198]. In 1988, *A. oryzae* was used for the first time for the commercial production of a heterologous enzyme, a lipase used for laundry detergents [199].

More recently, *Aspergillus* spp. have been used for production of enzymes used in starch processing, baking, animal feed, and in the paper and pulping industry, as well as for degradation of biomass for bioethanol and bioproducts [200]. For instance, amyloglucosidase (AMG) production occurs at levels far exceeding 10 g/L. This enzyme is manufactured using *A. niger* by companies such as Novozymes [201]. AMG hydrolyses 1,4- and 1,6-alpha linkages in liquefied starch and it is used, for example, in the production of glucose syrups from starch or dextrins. Similarly, alpha-amylase produced by both *A. niger* and *A. oryzae,* is an enzyme which degrades alpha-1,4 bonds in starch. This enzyme is widely used in the brewing industry to liquefy starch added to wort, in the sugar industry to break down the starch present in cane juice, in the starch industry to produce glucose syrups, and for high-speed, high-temperature removal of sizing of textiles before dyeing [202]. Other kind of enzymes endogenously produced by *A. niger* are invertases, pectinases, phytases and proteases (for details see Table 7), which are used in different types of food and feed processing [174, 203]. The three major fermentation industries DSM (The Netherlands), Novozymes (Denmark), and Genencor-Danisco (Denmark) employ *Aspergillus* as their most important production organism. For DSM it is *A. niger*, for Novozymes it is mainly *A. oryzae*, and Genencor-Danisco exploits both organisms [195].

Recently, an *A. oryzae* derived enzyme under the name of AcrylAway was launched by Novozymes [201]. This product is an asparaginase used for reduction of acrylamide levels in dough-based products by up to 90% without changing the appearance or taste of foods such as biscuits, cookies, French fries, crackers,

Table 7 Commercial enzyme preparations produced by *A. niger* and *A. oryzae* for food, feed and technology processing, i.e., detergents industry

Enzyme	Host organism	Donor organism	IUB number	Application		
				Food	Feed	Technology
Aminopeptidase	*A. oryzae*	None	3.4.11.x	Y	N	Y
Amylase (alpha)	*A. oryzae*	None	3.2.1.1	Y	Y	Y
Arabinanase	*A. niger*	None	3.2.1.99	Y	Y	N
Arabinofuranosidase	*A. niger*	*Aspergillus* sp.	3.2.1.55	Y	N	N
Asparaginase	*A. niger*	*Aspergillus* sp.	3.5.1.1	Y	N	N
Carboxypeptidase (serine-type)	*A. niger*	*Aspergillus* sp.	3.4.16.x	Y	N	N
Catalase	*A. niger*	*Aspergillus* sp.	1.11.1.6	Y	N	Y
Cellulase	*A. niger*	None	3.2.1.4	Y	Y	N
Galactosidase (alpha)	*A. niger*	None	3.2.1.22	Y	Y	N
Glucanase (beta)	*A. niger*	None	3.2.1.6	Y	Y	N
Glucoamylase or amyloglucosidase	*A. niger*	None	3.2.1.3	Y	N	Y
Glucose oxidase	*A. niger*	None	1.1.3.4	Y	N	Y
Glucosidase (alpha)	*A. niger*	None	3.2.1.20	Y	N	N
Hemicellulase	*A. niger*	None	–	Y	Y	N
Inulase	*A. niger*	None	3.2.1.7	Y	N	N
Laccase	*A. oryzae*	*Myceliopthora* sp.	1.10.3.2	Y	N	Y
Lactase or galactosidase (beta)	*A. oryzae*	*Aspergillus* sp.	3.2.1.23	Y	N	N
Lipase triacylglycerol	*A. oryzae*	*Candida* sp.	3.1.1.3	Y	N	Y
Mannanase (endo-1.4-beta)	*A. niger*	None	3.2.1.78	Y	Y	N
Pectin lyase	*A. niger*	None	4.2.2.10	Y	Y	N
Pectin methylesterase or Pectinesterase	*A. niger*	*Aspergillus* sp.	3.1.1.11	Y	Y	N
Pentosanase	*A. niger*	None	–	Y	N	N
Peroxidase	*A. oryzae*	*Coprinus* sp.	1.11.1.7	N	N	Y
Phosphatase	*A. niger*	None	3.1.3.2	Y	N	N
Phospholipase A	*A. niger*	*Aspergillus* sp.	3.1.1.4	Y	N	N
Phospholipase B	*A. niger*	*Aspergillus* sp.	3.1.1.5	Y	N	N
Phytase	*A. oryzae*	*Peniophora* sp.	3.1.3.8	Y	Y	N
Polygalacturonase or pectinase	*A. niger*	None	3.2.1.15	Y	Y	Y
Protease (incl. milkclotting enzymes)	*A. oryzae*	None	3.4.2x.x	Y	Y	Y
Tannase	*A. niger*	None	3.1.1.20	Y	Y	N
Transglucosidase	*A. niger*	None	2.4.1.24	N	N	Y
Xylanase	*A. oryzae*	*Thermomyces* sp.	3.2.1.8	Y	Y	N

There are currently at least 32 different commercial enzymes produced by either *A. niger* or *A. oryzae*. Table adapted from the Association of Manufacturers and Formulators of Enzyme Products [208]

bread, among others. Similarly, another asparaginase was developed by DSM, PrentAse, but in this case it is produced by *A. niger* [204]. Since acrylamide is classified as a potential carcinogen, it is important to decrease its consumption by humans. Studies carried out by the FDA as well as other studies using food

consumption data from several countries (Australia, Norway, The Netherlands, Sweden, USA) have estimated an acrylamide exposure of 0.3–0.8 μg/kg body weight per day [205]. Other examples of products manufactured by DSM and exploiting *Aspergillus* species are Brewer's Clarex, which prevents chill-haze in beer [206] and a protease product that eliminates bitterness of protein supplemented sport drinks [207]. A comprehensive list of commercial enzymes produced by *A. niger* and *A. oryzae* species is presented in Table 7, which is adapted from AMFEP (Association of Manufacturers and Formulators of Enzyme Products, Brussels, Belgium) [208].

Secondary metabolites are also produced by *Aspergillus* species. The most important is the cholesterol lowering agent compactin, generally known as lovastatin. The Merck Research Laboratories patented this compound in the 1980s [209] and subsequently became the first statin manufactured for human drug use produced by *A. terreus*. This drug has been produced since then, giving Merck billions of US dollars in revenues [210].

5.4 Heterologous Protein Production

Aspergillus spp., similarly to yeasts, are eukaryotic microorganisms able to perform glycosylation and therefore suitable for recombinant protein production due to their capability of proper protein processing [211]. Nevertheless, while homologous proteins are produced in a 10–50 g/L range, heterologous proteins are often produced at levels 10- to 100-fold lower [212]. This gives an indication that *A. niger* has severe problems in expressing and secreting foreign proteins. To date, even though these subjects have been the target of several studies, a complete map detailing the secretion machinery as well as strategies for its improvement has not been entirely established. According to Guillemette and coworkers [213], one bottleneck appears to be the result of mis-folding of heterologous proteins in the endoplasmic reticulum (ER) during early stages of secretion, with related stress responses in the host, including the unfolded protein response (UPR).

Guillemette's study [213] was the first genome-wide analysis of both transcriptional and translational events following protein secretion stress in *Aspergillus*. This study was conducted in the industrial host *A. niger*. The work revealed that predicted proteins encoded by most of the up-regulated genes function as part of the secretory system, for example, probable chaperones, foldases, glycosylation enzymes, vesicle transport proteins, and ER-associated degradation (ERAD) proteins. The analysis also confirmed the post-transcriptional control of *hacA*, encoding the UPR-mediating transcription factor, and highlighted that differential translation also occurs during ER stress, in particular for genes encoding secreted proteins or proteins involved in ribosomal biogenesis and assembly. This study gave insight into the molecular basis of protein secretion and secretion-related stress in this efficient protein-secreting fungus *A. niger*, and provided a valuable opportunity to identify target genes for manipulation in strain improvement strategies.

The first efforts to produce recombinant proteins in *Aspergillus* were made in the early 1990s. Lactoferrin was the largest heterologous protein and the first mammalian glycoprotein expressed in *Aspergillus* [214]. The host, *A. oryzae,* was engineered to express the human lactoferrin (*hLF*) by placing the cDNA under control of the *A. oryzae* α-amylase promoter and the 3' flanking region of the *A. niger* glucoamylase gene. HLF was expressed and secreted at levels of up to 25 mg/L. It was appropriately N-linked glycosylated and correctly processed at the N terminus. Hence, this expression system appeared to be suitable for the large-scale production and secretion of biologically active mammalian glycoproteins [214].

Recently, another interesting effort of heterologous protein production in *A. oryzae* was conducted by a Japanese research group. Tamalampudi et al. [215] expressed the lipase-encoding B gene from C*andida antarctica* (CALB) in *A. oryzae*. This lipase has been expressed in other hosts such as *E. coli* and *P. pastoris* achieving yields as low as 5.2 mg/L (periplasmic expression) and 44 mg/L (extracellular) respectively for this kind of protein [216]. The resultant recombinant *A. oryzae* lipase was immobilized within biomass support particles (BSPs) made of polyurethane foam and the BSPs were successfully used for the hydrolysis of *para*-nitrophenol butyrate [215]. Additionally, the role of a signal peptide important for improving protein secretion was also evaluated and therefore the expression plasmids were constructed with either the homologous and heterologous protein secretion signal sequences of the triacylglycerol lipase gene *tglA* from *A. oryzae* and the lipase B gene *CALB* from *C. antarctica*. The C-terminal signal peptide tag did not alter the catalytic properties of the lipase enzyme and Western blotting analysis demonstrated the presence of cell wall and membrane bound lipase. Thus, this constitutes a successful example of heterologous protein production at lab scale in *Aspergillus*.

6 Systems Biology as Driver for Industrial Biotechnology

From the examples discussed above it emerges that metabolic engineering usually involves expression of recombinant genes/pathways. Inserting the enzymatic activities necessary for a certain conversion is nevertheless an intuitive approach, dependent on what it is known a priori about the metabolism of the host. Unfortunately, as a consequence of complex regulation in the cell, these strategies are often not successful and, to be able to optimize strategies for the direct engineering of the cell factories, quantitative data that can help to unravel the tight regulation of the cell are required.

A solution to this hurdle was first given by *genomics*, when several genome-wide techniques such as transcriptome and metabolome analysis started to be routinely applied on microbial systems. These techniques, besides requiring significant expertise in data analysis [217], allow the extraction of a vast quantity of information. Unfortunately, the sole presence of this wealth of data is not sufficient to understand the cell behavior from a holistic perspective. To address this issue,

metabolic pathways have to be considered as coregulated and interconnected by all the different cofactors present in the cell [218] and efforts to integrate data coming from the different levels of organization of the cell machinery are necessary [217]. It thus appears clear that the challenge faced today by systems biology is to integrate these data to provide insight into the regulation of the metabolic network, hereby establishing a quantitative link between genotype and phenotype [42].

6.1 Yeast Examples

Being one of the most well characterized production hosts and important eukaryotic models, *S. cerevisiae* was among the first microorganisms to benefit from systems biology. The availability of complete genome sequences has led to the reconstruction of GSMMs for many organisms, with *S. cerevisiae* being the first eukaryote with a reconstructed network [106]. In its first version this model consisted of 1,175 reactions linked by 708 open reading frames. The reconstructed network has been revised and extended several times [219] and it has recently been updated, including a detailed description of lipid metabolism [220]. Additionally, as community effort, a consensus metabolic network reconstruction has been completed, formalizing the "community knowledge" of yeast metabolism. This consensus metabolic network accounts for 1,761 total reactions and 1,168 metabolites [221].

GSMMs provide a good scaffold to integrate *omics* data and they have served as a valid framework for the development of other tools such as the algorithms Opt-Knock [222], Opt-Strain [223], and Opt-Gene [224]. These algorithms can predict the phenotypes resulting from genes insertion/deletion and, even though they have been mainly applied on *E. coli* study cases, these methods have also recently been used on yeast. To describe the architecture of metabolic networks, Patil and Nielsen [225] reconstructed a metabolic graph where the relationship between enzymes and metabolites are mapped. Based on this reconstruction, the authors developed an algorithm to identify reporter metabolites by integrating gene expression data with the topological information from GSMMs [225]. This algorithm was applied to study the physiological response of *S. cerevisiae* to ammonium and amino acids limitation [47]. In this work, reporter metabolites analysis was used to analyze the transcriptional data obtained from different fermentation conditions, leading to identification of spots in the metabolism where relevant changes were occurring, a result that could not be achieved using standard clustering methods, which ignore that gene products act in different metabolic pathways and in different cellular compartments [226].

Another approach to exploit the network topology for integrated analysis of metabolome data is described by Cakir et al. [227]. Metabolome analysis is particularly useful to uncover the connectivity of the metabolic networks in light of the high dependence among gene transcripts, enzyme levels, and fluxes [87]. The authors developed an algorithm to identify reporter reactions and the algorithm was used to study on a wild-type strain the effects of media change and the behavior of a

recombinant strain engineered in the redox metabolism, spotting parts of the metabolism not directly related but nevertheless affected by the perturbations. When transcriptome data were included in this algorithm, even information as to whether a certain reaction is metabolically or hierarchically regulated could be gathered.

Not only can GSMMs be used as scaffold for the analysis of genome-wide data but the simulation of cellular functions is also possible. Using FBA, GSMMs can be used to simulate and predict the behavior of the cell. It has been shown that GSMMs can give good prediction of flux distribution under different physiological conditions once an objective function is fixed [106]. For these reasons, GSMMs can be considered promising tools to explore the possible effects of introduced perturbations (i.e., genetic manipulation) and to point to regions of the metabolism where endogenous adjustments are likely to happen, and hence to identify potential targets for metabolic engineering applications [110]. The prediction performances of GSMMs can be improved by incorporating measured physiological parameters so that the possible resulting flux distribution is constrained [228]. The physiological parameters can be values directly measured during the growth, e.g., uptake rates or enzymes activities, else genome-wide measurements such as transcriptome data. The incorporations of these data can be used to improve the prediction of flux distributions. Exometabolome data can also be combined with genome-scale models [229], giving the possibility to discriminate among different *S. cerevisiae* mutants impaired in respiration.

To date, systems biology based studies have been widely applied to characterize yeast phenotypes generated through metabolic engineering. In this perspective, the cycle of genetic manipulation, systemic study of the resultant engineered strains, and the generation of new data that can be further incorporated in the models to suggest new targets for strain improvement, represent an extremely valuable approach to uncover and improve the abilities to engineer microbial cells to attain desirable phenotypes. In Table 8 some examples of systems biology studies for industrially relevant application are reported according to the tool that has been used. However, besides the studies aimed at characterizing the generated strains, there are just a few studies where systems biology played an active and direct role in suggesting new strategies for the engineering of more efficient cell factories.

One of the most significant examples where systems biology has aided the design of yeast cell factories is for improving bioethanol production. Bioethanol is the largest product of industrial biotechnology with an estimated world production of 46 billion liters per year (data from 2005) [230]. The fermentation process for ethanol production is conducted anaerobically. Under these conditions, glycerol is produced as a by-product to re-oxidize the excess of cytosolic NADH that originates mainly from biomass formation. To increase the production of ethanol, a direct approach is the redirection of the carbon flow towards this product, minimizing the formation of glycerol and other by-products. Several attempts to decrease glycerol production through metabolic engineering of the redox metabolism have been addressed. A significant contribution to the field was given by Nissen et al. [231] who evaluated whether a reduced formation of NADH, together with an increased consumption of ATP, could result in an increased

Table 8 Some of the applications of systems biology for production of industrially relevant compounds in *S. cerevisiae* or characterization of *S. cerevisiae* phenotypes with potential industrial applications

	Study	Industrial interest	Reference
Transcriptome	Improvement of galactose uptake	Brewing, baking application	[343]
	Identification of genes involved in ethanol tolerance	Bioethanol	[344]
	Gene expression during wine fermentation	Wine making	[345]
	Transcriptome data using reporter metabolites for 6-MSA production in recombinant strain	Polyketides	[346]
	Response to altered redox metabolism	Glycerol/bioethanol	[347]
	Temperature effects on industrial and laboratory strain	Wine making	[348]
	Effects of high lactic acid concentration on the physiology	Lactic acid production	[349]
Metabolome	Strain classification	–	[74]
	Analysis of deletion mutant	–	[350]
	Identification of silent phenotypes from gene deletion	–	[69]
	Metabolite profiling under saccharification and fermentation of lab/industrial strains	Bioethanol production	[351]
	Metabolite profiling under very high gravity ethanol fermentation	Bioethanol production	[352]
Proteome	Analysis of strains engineered for growth on xylose	Bioethanol production	[353]
	Comparative proteome analysis of industrial strains in industrial settings	Bioethanol production	[354]
	Transcriptome and proteomic for wine fermentation	Wine making	[355]
	Transcriptome and proteome study the regulation of xylose metabolism	Bioethanol production	[356]
	Proteomic study of glucose repression	–	[357]
Flux analysis	Growth on different media	–	[358]
	Study of the glycerol overproducing strain	Glycerol production	[359]
	Integration of different growth conditions with flux balance model predictions	–	[360]
	Compared flux profiles of recombinant strains for efficient xylose uptake	Bioethanol production	[361]
	Phenotypic characterization of glucose repression mutants	–	[362]
	Integration of flux balance analysis and physiological of respiration deficient mutants	Bioethanol production	[234]
	Malic acid production	Malic acid	[252]
	Elementary flux mode analysis on a recombinant strain	Polyhydroxybutyrate production	[363]
Genome-scale models	Improved ethanol yield	Bioethanol production	[232]
	Reconstructed nutrient-controlled transcriptional regulatory network, coupled to GSMM	–	[364]
	Development of a dynamic flux balance model based on GSMM	Fedbatch ethanol production	[365]

ethanol yield under anaerobic conditions. To verify this hypothesis, the authors engineered the ammonium assimilation pathway, changing the cofactor requirement in amino acid synthesis and hereby reduced the formation of NADH, a strategy that resulted in an increase of 10% in the ethanol yield and a decrease of 38% in glycerol formation.

Taking a more holistic approach, Bro et al. [232] used the reconstructed genome-scale metabolic model of *S. cerevisiae* to screen among different strategies. The screened strategies were based on gene insertion analysis by introducing one reaction (taken from the LIGAND database) at a time into the model. This allowed for in silico screening of the effect of inserting different reactions on the ethanol, glycerol, and biomass yields. As outputs, several potential targets giving high ethanol and low glycerol yields were identified and those thermodynamically unfeasible were manually discarded. The strategy previously chosen by Nissen et al. [231] was also tested in silico, showing good agreement with the experimental data and therefore indicating the good predictability of the model under these simulation conditions. According to the simulations, the best scenario would have allowed for an increase of 10% in the ethanol yield and a complete elimination of glycerol production. This scenario involved heterologous expression of GAPN, an enzyme catalyzing the irreversible oxidation of glyceraldehyde-3-phosphate and $NADP^+$ into 3-phosphoglycerate and NADPH, a conversion that naturally takes place through two steps in *S. cerevisiae*. The prediction of the model was tested in vitro by engineering *S. cerevisiae* with the recombinant gene *gapN* from *Streptococcus mutants* [233]. The performances of the recombinant strain were evaluated in anaerobic glucose-limited batch cultures and the yields from the in vivo data were compared to those obtained in silico. The experimental glycerol yield showed a decrease of 40% whereas the ethanol yield improved by 3%. If the glycerol production rate was constrained to a value similar to that experimentally found, the ethanol and biomass yields predicted by the model fitted quite well with the experimentally obtained value. Additionally, it was shown that the in vivo activity of the GAPN enzyme was lower than the value required by the model in order to reach the optimal conditions, and therefore the model predicted results could possibly be obtained by increasing the GAPN activity. In the same work the GAPN expressing strain was engineered in silico with the activities necessary to metabolize pentose sugars (expressing of xylose reductase and xylitol dehydrogenase). Again, the performances of such a strain were simulated on mixed glucose–xylose and the yields were compared to those experimentally obtained in chemostat cultivations. Again, the experimental data showed good agreement with the results predicted by the GSSM. The described approach is probably one of the few yeast applications where systems biology helped in the direct design of a successful metabolic engineering strategy. This case study represents a good example of how systems biology can potentially contribute to industrial biotechnology, substituting intuitive strategies with non-obvious approaches screened through in silico predictions, minimizing time and increasing the probability of success.

A more recent approach for improved ethanol production is the work performed by Dikicioglu et al. [234], based on the previous observation that ethanol yields

could be increased using respiration deficient nuclear petite mutants [235]. The authors used GSMMs together with FBA to calculate flux distribution of different mutants totally or partially deficient in respiration and compared the predicted fluxes with those obtained experimentally. Once again the reconstructed network model proved to be able to carefully describe the behavior of the different mutants carrying deletion in the respiratory chain. Nevertheless, for mutants deleted in genes coding for regulatory proteins, FBA was not able to provide a successful description, as expected consequence of the total absence of regulatory mechanisms in stoichiometric models.

Even though these results are encouraging, the application of such tools in an industrial setting is still far from reality. The reasons can be found in the fact that these tools require pioneering work in order to be optimized. Another limitation of constrained based modeling is the lack of any kind of regulation explaining the impossibility of being applied under many conditions.

When computing cell behavior, the phenotypes obtained from each round of simulation have continuously to be evaluated and the set conditions improved. The computational procedure usually goes on through several rounds of simulation, verification of the reliability of the phenotypes obtained in silico, constraining of more parameters, and new computation, until satisfactory, biologically-meaningful results are obtained.

6.2 *Aspergillus* Examples

There is substantial industrial interest in the exploitation of *Aspergillus* species as cell factories. Therefore it is obvious to consider the exploitation of the solid knowledge base from systems biology for future design of "improved and better" cell factories. A major hurdle in this exploitation is that many of the high-throughput experimental techniques and bioinformatics tools available for analysis of the datasets are not well suited for fungi, at least not as well as for *S. cerevisiae*.

Compared to the industrial platform *S. cerevisiae*, the availability of molecular biology tools as well as the application of high-throughput technologies on *Aspergillus* species is still in its infancy. Therefore the proliferation of studies using any *omics* technology and, furthermore, the integration of more than one yielding a systems biology approach, is rather limited, especially when applying it to the production of commercially interesting industrial products. Nevertheless, as mentioned earlier, systems biology is becoming an attractive and very promising field in the design of strategies yielding better industrial workhorses by applying ad hoc manipulations not solely to yeasts, but also to *Aspergillus* cell factories. Several genome-scale models for industrially relevant microorganisms have been generated, for instance for the fungi *A. niger* [236] and *A. oryzae* [237]. These reconstructed metabolic networks contain 1,190 and 1,053 unique reactions, respectively, linking 871 genes in the former case and 1,314 genes in the later (for details see Table 5). Using the resources currently available and strongly advocating

Table 9 Transcriptome studies conducted with *Aspergillus* species using different microarray formats

Species	Gene models (%)	Type	Reference
A. nidulans	19.4	cDNA	[366, 367]
A. nidulans	26.0	cDNA	[368]
A. nidulans	79.0	Oligo	[369, 370]
A. nidulans	89.2	Nimblegen	[371]
A. nidulans	30.6	Febit	[372]
A. nidulans	16.7	cDNA	[373]
A. nidulans	26.1	cDNA	[374]
A. nidulans	88.3	Nimblegen	[375, 376]
A. nidulans	99.6	Affymetrix	[377]
A. niger	100.0	Affymetrix	[27, 49, 213, 243, 378]
A. niger	99.3	Affymetrix	[50, 377, 379, 380]
A. fumigatus	96.2	Oligo	[26, 381]
A. oryzae	16.8	cDNA	[382]
A. oryzae	24.4	cDNA	[383]
A. oryzae	97.7	Nimblegen	[384]
A. oryzae	99.7	Affymetrix	[377]
A. oryzae	89.1	Oligo	[385]
A. oryzae	98.6	Oligo	[40]
A. flavus	39.7	Oligo	[386, 387]
A. flavus	39.9	Oligo	[388]
A. flavus	6.0	cDNA	[389, 390]

Table adapted from [391]

coordination of research and development of new resources, there is likely to be a new "golden age" of fungal cell factory design. In Table 9 several transcriptome studies of different *Aspergillus* spp. are reported.

6.2.1 Lovastatin Production

The development of methods to comprehensively assess gene expression, for instance, DNA microarrays, provides the opportunity to correlate patterns of global gene expression with the production of specific metabolites. One of the "best" and first examples in Aspergilli is the study conducted by Askenazi et al. [51] on *A. terreus,* where the integration of transcriptional and metabolite profiles directed the engineering of the lovastatin-producing *A. terreus* strain. This analysis identified specific tools, including promoters for reporter-based selection systems that were employed to improved lovastatin production by *A. terreus*.

An association analysis was performed to determine the gene expression patterns that correlated with the level of lovastatin production as well as other secondary metabolites, e.g., (+)-geodin.

When Askenazi's study was conducted, the genome sequence of *A. terreus* was not available and the study was therefore based on the generation of random genomics fragments with an average size of 2 kb to generate a microarray of approximately 21,000 elements. The array also contained *A. terreus* genes suspected to promote lovastatin production, such as those genes encoding the

lovastatin biosynthetic components, i.e., *lovE* [238, 239] or implicated in modulation of secondary metabolism, i.e., *creA*, *fadA*, *ganB*, among others [240, 241]. The arrays were used to generate profiles for 21 strains producing increased or decreased lovastatin levels. The genomics fragment microarray methodology allowed monitoring genome-wide expression patterns. The overexpression of different genes resulted in distinct patterns of lovastatin production, suggesting that those genes elicited metabolic responses via distinct mechanisms of action. Microarray data analysis tools such as hierarchical clustering of the transcriptional profiling data and principal component analysis (PCA) were applied and helped in the former case to identify strains displaying similar metabolite profiles with relatively similar transcriptional profiles and, in the case of PCA to determine a few linear combinations of genes that explained most of the variance present in the data sets, and hence provide insight into the underlying variables which differentiated the engineered strains.

To gain a more biologically meaningful understanding of the variation among the strains, the association analysis was conducted. In order to perform it, the acquired data included data sets in which the levels of metabolites and global gene expression patterns varied. For a set of engineered strains, where genes previously implicated in lovastatin production were modulated and or modified, the resulting phenotypes were analyzed for secondary metabolite production by high-pressure liquid chromatography and electrospray MS. Secondary metabolites and gene expression values were expressed as ratios of the value from an engineered strain relative to that from the reference strain.

Following sequencing, homology searching and contig analyses were conducted and in many cases similar expression patterns were found for multiple clones with overlapping sequences. A list of sequences that encoded elements with expression patterns that were positively or negatively associated to lovastatin production was generated. Hereby the authors (not surprisingly) found the genes constituting the *A. terreus* lovastatin biosynthetic cluster (64 kb genomics region predicted to encode 18 proteins) positively associated with lovastatin production, for instance, *lovA*, *lovB*, *lovC*, *lovD*, *lovF*, *lvrA*, etc. Interestingly, however, was that LovE, a transcription factor encoded within the cluster and suspected to regulate, at least partially, the coordinate expression of these genes [239], was also found. Additionally, the association analysis identified many genes that encoded proteins either predicted or known to play a role in production of secondary metabolites other than lovastatin, for example, PksM, a non-ribosomal peptide synthetase, and a dimethylallyl-cycloacetyl-L-tryptophan synthase homologous to enzymes required for production of several fungal secondary metabolites such as cyclopiazonic acid [242], which is a mycotoxin produced by several *Penicillium* and *Aspergillus* species i.e., *A. flavus* and *A. versicolor*.

Additionally, the study demonstrated that there are fatty acid metabolism genes that were positively associated with lovastatin production and tended to encode catabolic enzymes that are predicted to promote formation of the polyketide precursors acetyl-CoA and malonyl-CoA, whereas fatty acid metabolism genes that are negatively associated with secondary metabolite production encode anabolic enzymes, i.e., acyl-CoA oxidase, fatty acid desaturase, and fatty acid synthases.

6.2.2 Integrated Genomics Approach for Improving Protein Production

Recently, an integrated genomics approach, led by the Dutch company DSM [243], was developed in order to determine diverse cellular responses of *A. niger* aiming at increasing protein production. Continuous fermentations were conducted with three enzyme overproducing strains and the strains were compared to their isogenic fungal host strains. Integration of transcriptomics and proteomics data was assessed for the identification of key bottlenecks in protein production. Not surprisingly, genes which showed coregulation at the transcriptome and proteome level were selected as potential candidates for strain improvement. Up-regulated proteins included proteins involved in carbon and nitrogen metabolism as well as (oxidative) stress response and proteins involved in protein folding and ERAD. Reduction of protein degradation through removal of the ERAD factor *doaA* combined with the overexpression of the oligosaccharyl-transferase *sttC* in *A. niger* resulted in overproduction of the model protein B-glucuronidase (GUS). Glucuronidases are members of the glycosidase family of enzymes which catalyze breakdown of complex carbohydrates and were selected as model enzymes to evaluate changes in protein production. DoaA is a factor required for ubiquitin-mediated proteolysis [244] and the *sttC* gene is involved in glycosylation of secretory proteins which is important for protein folding [245]; therefore, these genes represented good candidates for modification. To our knowledge, this is one of the first published applications of an integrated genomics approach applied by the industry based on controlled fermentations of *Aspergillus* to improve protein production in a generic way [243].

The production of a heterologous lipase, and a homologous hydrolase and protease by the corresponding production strains, was also analyzed. In the case of heterologous lipase production, due to its foreign nature, all proteins involved in protein folding were up-regulated upon production, whereas no additional proteins involved in protein folding were up-regulated in the homologous (hydrolase and protease) producing strains. This indicates that heterologous lipase production causes problems at the level of protein folding as already discussed by Guillemette et al. [213] in their study on heterologous protein production in *A. niger*, whereas overproduction of a homologous protein, for instance hydrolases and proteases, results in no specific protein folding response.

Furthermore, integration of transcriptomics and proteomics data helped to uncover complex regulatory (post-translational) mechanisms which do not justify simple gene overexpression as the unique mechanism to remove bottlenecks in the process and which would not be revealed by looking at a single level of information. For example, in *S. cerevisiae* it has been demonstrated that invert relationships in mRNA levels and protein levels occur in many cases and are indicators of several factors such as mRNA turnover, and initiation of translation and stability of proteins which are independent processes and are controlled at different layers of regulation [89, 246]. For this reason, straightforward genome modifications are not always the answer for achieving successful strain optimization and a more complete evaluation and integration of the information is required such as the one provided by systems biology giving a more holistic view.

References

1. Bailey JE (1991) Science 252:1668–1675
2. Nielsen J (2001) Appl Microbiol Biotechnol 55:263–283
3. Stephanopoulos G, Aristidou A, Nielsen J (1998) Metabolic engineering: principles and methodologies. Academic, San Diego
4. Jewett MC, Hofmann G, Nielsen J (2006) Curr Opin Biotechnol 17:191–197
5. Raab RM, Tyo K, Stephanopoulos G (2005) Adv Biochem Eng Biotechnol 100:1–17
6. Sanford K, Soucaille P, Whited G et al (2002) Curr Opin Microbiol 5:318–322
7. Otero JM, Olsson L, Nielsen J (2007) Genet Eng News 27:28–31
8. Chotani G, Dodge T, Hsu A et al (2000) Biochim Biophys Acta 1543:434–455
9. Gavrilescu M, Chisti Y (2005) Biotechnol Adv 23:471–499
10. van Berkel R (2000) Cleaner production for process industries. Plenary Lecture, Chemeca, Perth, WA
11. Energy Information Administration. Available at http://www.eia.doe.gov/. Accessed 15 Jan 2009
12. McKinsey (2003) Available at http://www.mckinsey.com/mgi/publications/turkey/index.asp. Accessed 30 Jan 2009
13. Lee SY, Kim HU, Park JH et al (2009) Drug Discov Today 14:78–88
14. Otero JM, Nielsen J (2009) Industrial Systems Biology. Biotechnol Bioeng 105(3): 439–460
15. Barnett JA (2003) Microbiology 149:557–567
16. Hamilton SR, Davidson RC, Sethuraman N et al (2006) Science 313:1441–1443
17. Li HJ, Sethuraman N, Stadheim TA et al (2006) Nat Biotechnol 24:210–215
18. Karaffa L, Kubicek CP (2003) Appl Microbiol Biotechnol 61:189–196
19. Davies RW (1994) Prog Ind Microbiol 29:527–560
20. Blattner FR, Plunkett G, Bloch CA et al (1997) Science 277:1453–1474
21. Galagan JE, Calvo SE, Cuomo C et al (2005) Nature 438:1105–1115
22. Goffeau A, Barrell BG, Bussey H et al (1996) Science 274:546–567
23. Kunst F, Ogasawara N, Moszer I et al (1997) Nature 390:249–256
24. Machida M, Asai K, Sano M et al (2005) Nature 438:1157–1161
25. Martinez D, Berka RM, Henrissat B et al (2008) Nat Biotechnol 26:553–560
26. Nierman WC, Pain A, Anderson MJ et al (2005) Nature 438:1151–1156
27. Pel HJ, de Winde JH, Archer DB et al (2007) Nat Biotechnol 25:221–231
28. Nielsen J, Jewett MC (2006) Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. In: 25th International specialized symposium on Yeasts (ISSY25), Blackwell, Espoo, Finland
29. Sanger F, Air GM, Barrell BG et al (1977) Nature 265:687–695
30. Smith HO, Tomb JF, Dougherty BA et al (1995) Science 269:538–540
31. Galagan JE, Calvo SE, Borkovich KA et al (2003) Nature 422:859–868
32. Shendure J, Mitra RD, Varma C et al (2004) Nat Rev Genet 5:335–344
33. Collins FS, Morgan M, Patrinos A (2003) Science 300:286–290
34. Shendure J, Ji HL (2008) Nat Biotechnol 26:1135–1145
35. Frazier ME, Johnson GM, Thomassen DG et al (2003) Science 300:290–293
36. Gabig M, Wegrzyn G (2001) Acta Biochim Pol 48:615–622
37. Affymetrix. Available at http://www.affymetrix.com/index.affx. Accessed 20 Jan 2009
38. NimbleGen. Available at http://www.nimblegen.com/. Accessed 20 Jan 2009
39. Agilent Technologies. Available at http://www.chem.agilent.com/en-US/products/instruments/dnamicroarrays/pages/default.aspx. Accessed 20 Jan 2009
40. Fermlab. Available at http://www.fermlab.com/motechinfo.html. Accessed 20 Jan 2009
41. Febit. Available at http://www.febit.com/go/en/. Accessed 20 Jan 2009
42. Jewett MC, Oliveira AP, Patil KR et al (2005) Biotechnol Bioprocess Eng 10:385–399
43. Jarmer H, Berka R, Knudsen S et al (2002) FEMS Microbiol Lett 206:197–200

44. Gene Expression Omnibus Database. Available at http://www.ncbi.nlm.nih.gov/geo/. Accessed 20 Jan 2009
45. Array Express Database. Available at http://www.ebi.ac.uk/microarray-as/ae/. Accessed 20 Jan 2009
46. Roberts GG, Hudson AP (2006) Mol Genet Genomics 276:170–186
47. Usaite R, Patil KR, Grotkjaer T et al (2006) Appl Environ Microbiol 72:6194–6203
48. Borodina I, Siebring J, Zhang J et al (2008) J Biol Chem 283:25186–25199
49. Yuan XL, van der Kaaij RM, van den Hondel C et al (2008) Mol Genet Genomics 279: 545–561
50. Vongsangnak W, Salazar M, Hansen K, Nielsen J (2009) Genome-wide analysis of maltose utilization and regulation in aspergilli. Microbiology 155:3893–3902
51. Askenazi M, Driggers EM, Holtzman DA et al (2003) Nat Biotechnol 21:150–156
52. Anderson NL, Anderson NG (1998) Electrophoresis 19:1853–1861
53. Joyce AR, Palsson BO (2006) Nat Rev Mol Cell Biol 7:198–210
54. Gygi SP, Rist B, Gerber SA et al (1999) Nat Biotechnol 17:994–999
55. Zhu H, Bilgin M, Bangham R et al (2001) Science 293:2101–2105
56. Bader GD, Hogue CWV (2002) Nat Biotechnol 20:991–997
57. Auerbach D, Thaminy S, Hottiger MO, Stagljar I (2002) The post-genomic era of interactive proteomics: Facts and perspectives. Proteomics 2:611–623
58. Ghaemmaghami S, Huh W, Bower K et al (2003) Nature 425:737–741
59. Kniemeyer O, Lessing F, Scheibner O et al (2006) Curr Genet 49:178–189
60. Mueller LN, Brusniak MY, Mani DR et al (2008) J Proteome Res 7:51–61
61. Nesvizhskii AI, Vitek O, Aebersold R (2007) Nat Methods 4:787–797
62. Taylor CF, Binz PA, Aebersold R et al (2008) Nat Biotechnol 26:860–861
63. Gupta N, Benhamida J, Bhargava V et al (2008) Genome Res 18:1133–1142
64. Mapelli V, Olsson L, Nielsen J (2008) Trends Biotechnol 26:490–497
65. Tweeddale H, Notley-McRobb L, Ferenci T (1998) J Bacteriol 180:5109–5116
66. Oliver SG, Winson MK, Kell DB et al (1998) Trends Biotechnol 16:373–378
67. Dunn WB, Bailey NJC, Johnson HE (2005) Analyst 130:606–625
68. Werner E, Heilier JF, Ducruix C et al (2008) J Chromatogr B 871:143–163
69. Raamsdonk LM, Teusink B, Broadhurst D et al (2001) Nat Biotechnol 19:45–50
70. Soga T, Ueno Y, Naraoka H et al (2002) Anal Chem 74:2233–2239
71. Soga T, Kakazu Y, Robert M et al (2004) Electrophoresis 25:1964–1972
72. Kaderbhai NN, Broadhurst DI, Ellis DI et al (2003) Comp Funct Genomics 4:376–391
73. Maharjan RP, Seeto S, Ferenci T (2007) J Bacteriol 189:2350–2358
74. Allen J, Davey HM, Broadhurst D et al (2003) Nat Biotechnol 21:692–696
75. Mas S, Villas-Boas SG, Hansen ME et al (2007) Biotechnol Bioeng 96:1014–1022
76. Kouskoumvekaki I, Yang ZY, Jonsdottir SO et al (2008) BMC Bioinformatics 9:59
77. Smedsgaard J, Nielsen J (2005) J Exp Bot 56:273–286
78. Pope GA, MacKenzie DA, Defemez M et al (2007) Yeast 24:667–679
79. Howell KS, Cozzolino D, Bartowsky EJ et al (2006) FEMS Yeast Res 6:91–101
80. Sumner LW, Amberg A, Barrett D et al (2007) Metabolomics 3:211–221
81. BIGG Database. Available at www.bigg.ucsd.edu/. Accessed 20 Jan 2009
82. BioCyc Database. Available at www.biocyc.org/. Accessed 20 Jan 2009
83. MSlib Database. Available at www.ualberta.ca/~gjones/mslib.htm. Accessed 20 Jan 2009
84. NIST Database. Available at www.nist.gov/srd/nist1a.htm. Accessed 20 Jan 2009
85. Metlin Database. Available at www.metlin.scripps.edu. Accessed 20 Jan 2009
86. HMDB Database. Available at www.hmdb.ca/extrIndex.htm. Accessed 20 Jan 2009
87. Nielsen J (2003) J Bacteriol 185:7031–7035
88. Sauer U (2004) Curr Opin Biotechnol 15:58–63
89. Daran-Lapujade P, Rossell S, van Gulik WM et al (2007) Proc Natl Acad Sci USA 104:15753–15758
90. ter Kuile BH, Westerhoff HV (2001) FEBS Lett 500:169–171

91. Sauer U (2006) Mol Syst Biol 2:62
92. Marx A, de Graaf AA, Wiechert W et al (1996) Biotechnol Bioeng 49:111–129
93. Sauer U, Hatzimanikatis V, Bailey JE et al (1997) Nat Biotechnol 15:448–452
94. Gombert AK, dos Santos MM, Christensen B et al (2001) J Bacteriol 183:1441–1451
95. Klapa MI, Quackenbush J (2003) Biotechnol Bioeng 84:739–742
96. Wiechert W, de Graaf AA (1997) Biotechnol Bioeng 55:101–117
97. Wiechert W, Siefke C, de Graaf AA et al (1997) Biotechnol Bioeng 55:118–135
98. Wiechert W, Mollney M, Isermann N et al (1999) Biotechnol Bioeng 66:69–85
99. van Winden WA, Heijnen JJ, Verheijen PJT (2002) Biotechnol Bioeng 80:731–745
100. Christensen B, Gombert AK, Nielsen J (2002) Eur J Biochem 269:2795–2800
101. Chen KC, Csikasz-Nagy A, Gyorffy B et al (2000) Mol Biol Cell 11:369–391
102. Klipp E (2007) Yeast 24:943–959
103. Borodina I, Nielsen J (2005) Curr Opin Biotechnol 16:350–355
104. Ibarra RU, Edwards JS, Palsson BO (2002) Nature 420:186–189
105. Edwards JS, Covert M, Palsson B (2002) Environ Microbiol 4:133–140
106. Famili I, Forster J, Nielson J et al (2003) Proc Natl Acad Sci USA 100:13134–13139
107. Forster J, Famili I, Fu P et al (2003) Genome Res 13:244–253
108. Bonarius HPJ, Schmid G, Tramper J (1997) Trends Biotechnol 15:308–314
109. Stephanopoulos G (1998) Biotechnol Bioeng 58:119–120
110. Patil KR, Akesson M, Nielsen J (2004) Curr Opin Biotechnol 15:64–69
111. Schilling CH, Covert MW, Famili I et al (2002) J Bacteriol 184:4582–4593
112. Edwards JS, Ibarra RU, Palsson BO (2001) Nat Biotechnol 19:125–130
113. Covert MW, Palsson BO (2002) J Biol Chem 277:28058–28064
114. Breitling R, Vitkup D, Barrett MP (2008) Nat Rev Microbiol 6:156–161
115. Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Biophys J 92:1792–1805
116. Jamshidi N, Palsson BO (2008) Mol Syst Biol 4
117. Hinnen A, Meyhack B, Heim J (1989) Biotechnology 13:193–213
118. Mustacchi R, Hohmann S, Nielsen J (2006) Yeast 23:227–238
119. Petranovic D, Nielsen J (2008) Trends Biotechnol 26:584–590
120. Vemuri GN, Aristidou A (2005) Microbiol Mol Biol Rev 69:197–216
121. Romanos MA, Scorer CA, Clare JJ (1992) Yeast 8:423–488
122. Giaever G, Chu AM, Ni L et al (2002) Nature 418:387–391
123. Chigira Y, Oka T, Okajima T et al (2008) Glycobiology 18:303–314
124. Hamilton SR, Gerngross TU (2007) Curr Opin Biotechnol 18:387–392
125. Gasser B, Mattanovich D (2007) Biotechnol Lett 29:201–212
126. Parekh S, Vinci VA, Strobel RJ (2000) Appl Microbiol Biotechnol 54:287–301
127. Rowlands RT (1984) Enzyme Microb Technol 6:3–10
128. Attfield PV, Bell PJL (2003) Genetics and classical genetic manipulation of industrial yeasts. In: Functional Genetics of Industrial Yeasts. de Winde JH, ed, Topics in Current Genetics, (S Hohmann, Series Ed), 2, pp 17–55. Springer-Verlag, Heidelberg and New York
129. Weuster-Botz D, Hekmat D, Puskeiler R et al (2007) Adv Biochem Eng Biotechnol 105: 205–247
130. Patnaik R, Louie S, Gavrilovic V et al (2002) Nat Biotechnol 20:707–712
131. Spencer JFT, Spencerd DM, Reynolds N (1988) J Basic Microbiol 28:321–333
132. Demain AL, Solomon NA (1986) Manual of industrial microbiology and biotechnology. American Society for Biotechnology, Washington, DC
133. Lindegren CC, Lindegren G (1943) Proc Natl Acad Sci USA 29:306–308
134. Romano P, Soli MG, Suzzi G et al (1985) Appl Environ Microbiol 50:1064–1067
135. Urano N, Sato M, Sahara H et al (1993) J Biotechnol 28:249–261
136. Higgins VJ, Bell PJL, Dawes IW et al (2001) Appl Environ Microbiol 67:4346–4348
137. Angelov AI, Karadjov GI, Roshkova ZG (1996) Food Res Int 29:235–239
138. Sherman F, Fink GR, Hicks JB (1986) Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

139. van Dijck P, Gorwa MF, Lemaire K et al (2000) Int J Food Microbiol 55:187–192
140. Dequin S (2001) Appl Microbiol Biotechnol 56:577–588
141. Donalies UEB, Nguyen HTT, Stahl U et al (2008) Adv Biochem Eng Biotechnol 111:67–98
142. Klein CJL, Olsson L, Nielsen J (1998) Microbiology 144:13–14
143. Klein CJL, Olsson L, Ronnow B et al (1996) Appl Environ Microbiol 62:4441–4449
144. Olsson L, Larsen ME, Ronnow B et al (1997) Appl Environ Microbiol 63:2366–2371
145. Liljeström-Suominen PL, Joutsjoki V, Korhola M (1988) Appl Environ Microbiol 54: 245–249
146. Rønnow B, Olsson L, Nielsen J et al (1999) J Biotechnol 72:213–228
147. Bidard F, Blondin B, Dequin S et al (1994) Curr Genet 25:196–201
148. Riou C, Nicaud JM, Barre P et al (1997) Yeast 13:903–915
149. Verstrepen KJ, Bauer F, Michiels C et al (1999) Eur Brew Conv Mon 28:30
150. Londesborough J (1995) US Patent 5,422,254
151. Lewis JG, Learmonth RP, Attfield PV et al (1997) J Ind Microbiol Biotechnol 18:30–36
152. Auras R, Harte B, Selke S (2004) Macromol Biosci 4:835–864
153. John RP, Nampoothiri KM, Pandey A (2007) Appl Microbiol Biotechnol 74:524–534
154. Datta R, Henry M (2006) J Chem Technol Biotechnol 81:1119–1129
155. Singh SK, Ahmed SU, Pandey A (2006) Proc Biochem 41:991–1000
156. Porro D, Brambilla L, Ranzi BM et al (1995) Biotechnol Prog 11:294–298
157. Hohmann S (1991) J Bacteriol 173:7963–7969
158. Porro D, Bianchi MM, Brambilla L et al (1999) Appl Environ Microbiol 65:4211–4215
159. van Maris AJA, Winkler AA, Porro D et al (2004) Appl Environ Microbiol 70:2898–2905
160. Andrade RP, Casal M (2001) Fungal Genet Biol 32:105–111
161. Branduardi P, Sauer M, De Gioia L et al (2006) Microb Cell Fact 5
162. Porro D, Bianchi MM, Ranzi BM et al (1999) Yeast strains for the production of lactic acid. PCT WO 99/14335
163. Yamano S, Ishii T, Nakagawa M et al (1994) Biosci Biotechnol Biochem 58:1112–1114
164. Verwaal R, Wang J, Meijnen JP et al (2007) Appl Environ Microbiol 73:4342–4350
165. Miura Y, Kondo K, Saito T et al (1998) Appl Environ Microbiol 64:1226–1229
166. Polakowski T, Stahl U, Lang C (1998) Appl Microbiol Biotechnol 49:66–71
167. Shimada H, Kondo K, Fraser PD et al (1998) Appl Environ Microbiol 64:2676–2680
168. Szczebara FM, Chandelier C, Villeret C et al (2003) Nat Biotechnol 21:143–149
169. Gerngross TU (2004) Nat Biotechnol 22:1409–1414
170. Directory of P450-containing Systems. Available at http://www.icgeb.org/~p450srv/. Accessed 20 Jan 2009
171. Werck-Reichhart D, Feyereisen R (2000) Genome Biol 1:rev 3003.3001–rev 3003.3009
172. Duport C, Spagnoli R, Degryse E et al (1998) Nat Biotechnol 16:186–189
173. Dumas B, Brocard-Masson C, Assemat-Lebrun K et al (2006) Biotechnol J 1:299–307
174. Baker SE, Bennett JW (2008) An overview of the genus Aspergillus. In: Goldman GH, Osmani SA (eds) The Aspergilli: genomics, medical aspects, biotechnology, and research methods, 1st edn. CRC, Taylor & Francis, Boca Raton, Florida
175. Perrone G, Susca A, Cozzi G et al (2007) Stud Mycol 59:53–66
176. Barbesgaard P, Heldthansen HP, Diderichsen B (1992) Appl Microbiol Biotechnol 36: 569–572
177. Schuster E, Dunn-Coleman N, Frisvad JC et al (2002) Appl Microbiol Biotechnol 59: 426–435
178. Blumenthal CZ (2004) Regul Toxicol Pharmacol 39:214–228
179. van Dijck PWM, Selten GCM, Hempenius RA (2003) Regul Toxicol Pharmacol 38: 27–35
180. Tilburn J, Scazzocchio C, Taylor GG et al (1983) Gene 26:205–221
181. Meyer V (2008) Biotechnol Adv 26:177–185
182. Bird D, Bradshaw R (1997) Mol Gen Genet 255:219–225
183. Krappmann S, Sasse C, Braus GH (2006) Eukaryot Cell 5:212–215

184. Hua SB, Qiu MS, Chan E et al (1997) Plasmid 38:91–96
185. Kooistra R, Hooykaas PJJ, Steensma HY (2004) Yeast 21:781–792
186. Ninomiya Y, Suzuki K, Ishii C et al (2004) Proc Natl Acad Sci USA 101:12248–12253
187. Takahashi T, Masuda T, Koyama Y (2006) Mol Genet Genomics 275:460–470
188. Meyer V, Arentshorst M, El-Ghezal A et al (2007) J Biotechnol 128:770–775
189. Ishibashi K, Suzuki K, Ando Y et al (2006) Proc Natl Acad Sci USA 103:14871–14876
190. Mizutani O, Kudo Y, Saito A et al (2008) Fungal Genet Biol 45:878–889
191. de Groot MJA, Bundock P, Hooykaas PJJ et al (1998) Nat Biotechnol 16:839–842
192. Michielse CB, Arentshorst M, Ram AFJ et al (2005) Fungal Genet Biol 42:9–19
193. Ruiz-Diez B (2002) J Appl Microbiol 92:189–195
194. Nielsen J (1998) Curr Opin Microbiol 1:330–336
195. van Dijck PWM (2008) The importance of Aspergilli and regulatory aspects of *Aspergillus* nomenclature in biotechnology. In: Varga J, Samson RA (eds) Aspergillus in the genomic era. Wageningen Academic, Wageningen, The Netherlands
196. Chotani GK, Dodge TC, Gaertner AL et al (2007) Industrial biotechnology: discovery to delivery. In: Kent JA (ed) Kent and Riegel's handbook of industrial chemistry and biotechnology, 11th edn. Springer-Verlag, New York
197. Currie JN (1917) J Biol Chem 31:15–37
198. Machida M, Yamada O, Gomi K (2008) DNA Res 15:173–183
199. Christensen T, Woeldike H, Boel E et al (1988) Biotechnology 6:1419–1422
200. Archer DB, Peberdy JF (1997) Crit Rev Biotechnol 17:273–306
201. Novozymes (2008) Available at http://www.novozymes.com/en/MainStructure/Productfinder/. Accessed 20 Jan 2009
202. NCBE. Available at http://www.ncbe.reading.ac.uk/ncbe/materials/enzymes/termamylliquid.html. Accessed 20 Jan 2009
203. Goodfrey T, West S (1996) Industrial enzymology. Stockton, Macmillan, New York, London
204. DSM (2008) Available at http://www.dsm.com/en_US/html/dfs/news_items/preventase_approved_in_Singapore_and_NL.htm. Accessed 20 Jan 2009
205. FDA (2006) The 2006 exposure assessment for acrylamide. Available at http://www.cfsan.fda.gov/~dms/acryexpo/acryex5.htm. Accessed 20 Jan 2009
206. Lopez M, Edens L (2005) J Agric Food Chem 53:7944–7949
207. Edens L, Dekker P, Van der Hoeven R et al (2005) J Agric Food Chem 53:7950–7957
208. AMFEP. Association of manufacturers and formulators of enzyme products. Available at http://www.amfep.org/list.html. Accessed 20 Jan 2009
209. Alberts-Schonberg G, Alberts AW (1980) Hypocholesteremic fermentation products and process of preparation. US Patent 4,231,938
210. Tobert JA (2003) Nat Rev Drug Discov 2:517–526
211. Punt PJ, van Biezen N, Conesa A et al (2002) Trends Biotechnol 20:200–211
212. van den Hondel CAMJJ, Punt PJ, van Gorcom RFM (1991) Heterologous gene expression in fungi. In: Bennet JW, Lause LL (eds) More genetic manipulations of filamentous fungi. Academic, Orlando, FL
213. Guillemette T, van Peij N, Goosen T et al (2007) BMC Genomics 8:158
214. Ward PP, Lo JY, Duke M et al (1992) Biotechnology 10:784–789
215. Tamalampudi S, Talukder M, Hama S et al (2007) Appl Microbiol Biotechnol 75:387–395
216. Larsen MW, Bornscheuer UT, Hult K (2008) Protein Expr Purif 62:90–97
217. Stephanopoulos G, Alper H, Moxley J (2004) Nat Biotechnol 22:1261–1267
218. Hermann T (2004) Curr Opin Biotechnol 15:444–448
219. Duarte NC, Herrgard MJ, Palsson BO (2004) Genome Res 14:1298–1309
220. Nookaew I, Jewett MC, Meechai A et al (2008) BMC Syst Biol 2:15
221. Herrgard MJ, Swainston N, Dobson P et al (2008) Nat Biotechnol 26:1155–1160
222. Burgard AP, Pharkya P, Maranas CD (2003) Biotechnol Bioeng 84:647–657
223. Pharkya P, Burgard AP, Maranas CD (2004) Genome Res 14:2367–2376

224. Patil KR, Rocha I, Forster J et al (2005) BMC Bioinformatics 6:308
225. Patil KR, Nielsen J (2005) Proc Natl Acad Sci USA 102:2685–2689
226. Grotkjaer T, Winther O, Regenberg B et al (2006) Bioinformatics 22:58–67
227. Cakir T, Patil KR et al (2006) Integration of metabolome data with metabolic networks reveals reporter reactions. Molecular Systems Biology 2(50):1–11
228. Akesson M, Forster J, Nielsen J (2004) Metab Eng 6:285–293
229. Cakir T, Efe C, Dikicioglu D et al (2007) Biotechnol Prog 23:320–326
230. Otero JM, Panagiotou G, Olsson L (2007) Adv Biochem Eng Biotechnol 108:1–40
231. Nissen TL, Kielland-Brandt MC, Nielsen J et al (2000) Metab Eng 2:69–77
232. Bro C, Regenberg B, Forster J et al (2006) Metab Eng 8:102–111
233. Boyd DA, Cvitkovitch DG, Hamilton IR (1995) J Bacteriol 177:2622–2627
234. Dikicioglu D, Pir P, Onsan ZI et al (2008) Appl Environ Microbiol 74:5809–5816
235. Hutter A, Oliver SG (1998) Appl Microbiol Biotechnol 49:511–516
236. Andersen MR, Nielsen ML, Nielsen J (2008) Mol Syst Biol 4:13
237. Vongsangnak W, Olsen P, Hansen K et al (2008) BMC Genomics 9:14
238. Hutchinson CR, Kennedy J, Park C et al (2000) Antonie Van Leeuwenhoek 78:287–295
239. Kennedy J, Auclair K, Kendrew SG et al (1999) Science 284:1368–1372
240. Bailey C, Arst HN (1975) Eur J Biochem 51:573–577
241. Hicks JK, Yu JH, Keller NP et al (1997) EMBO J 16:4916–4923
242. Christensen BE, Kaasgaard S (2000) World Patent Application, WO 00/39322
243. Jacobs DI, Olsthoorn MMA, Maillet I et al (2009) Fungal Genet Biol 46:S141–S152
244. Rabinovich E, Kerem A, Frohlich KU et al (2002) Mol Cell Biol 22:626–634
245. Yan Q, Lennarz WJ (2002) J Biol Chem 6:47692–47700
246. de Groot MJL, Daran-Lapujade P, van Breukelen B et al (2007) Microbiology 153:3864–3878
247. Primrose S, Twyman R (2006) Principles of gene manipulation and genomics. Blackwell, Oxford
248. Bai FW, Anderson WA, Moo-Young M (2008) Biotechnol Adv 26:89–105
249. Zanin GM, Santana CC, Bon EPS et al (2000) Appl Biochem Biotechnol 84–6:1147–1161
250. Bulthuis BA, Gatenby AA, Haynie SL et al (2002) Method for the production of glycerol by recombinant organisms. E. I. du Ponte de Nemours and Company, Wilmington, DE
251. Geertman JM, van Maris AJ, van Dijken JP et al (2006) Metab Eng 8:532–542
252. Zelle RM, de Hulster E, van Winden WA et al (2008) Appl Environ Microbiol 74:2766–2777
253. Berovic M, Legisa M (2007) Biotechnol Annu Rev 13:303–343
254. Forster A, Aurich A, Mauersberger S et al (2007) Appl Microbiol Biotechnol 75:1409–1417
255. Huang YL, Mann K, Novak JM et al (1998) Biotechnol Prog 14:800–806
256. Raspor P, Goranovic D (2008) Crit Rev Biotechnol 28:101–124
257. Ishida N, Suzuki T, Tokuhiro K et al (2006) J Biosci Bioeng 101:172–177
258. Chatterjee R, Millard CS, Champion K et al (2001) Appl Environ Microbiol 67:148–154
259. Lee PC, Lee SY, Hong SH et al (2002) Appl Microbiol Biotechnol 58:663–668
260. Lee PC, Schmidt-Dannert C (2002) Appl Microbiol Biotechnol 60:1–11
261. Lee SJ, Lee DY, Kim TY et al (2005) Appl Environ Microbiol 71:7880–7887
262. Liu YP, Zheng P, Sun ZH et al (2008) J Chem Technol Biotechnol 83:722–729
263. Georgi T, Rittmann D, Wendisch VF (2005) Metab Eng 7:291–301
264. Koffas MAG, Jung GY, Stephanopoulos G (2003) Metab Eng 5:32–41
265. Ohnishi J, Mitsuhashi S, Hayashi M et al (2002) Appl Microbiol Biotechnol 58:217–223
266. Yakandawala N, Romeo T, Friesen AD et al (2008) Appl Microbiol Biotechnol 78:283–291
267. Berry A (1996) Trends Biotechnol 14:250–256
268. Bongaerts JM, Kramer M, Muller U et al (2001) Metab Eng 3:289–300
269. Lutke-Eversloh T, Stephanopoulos G (2007) Appl Microbiol Biotechnol 75:103–110
270. Qi WW, Vannelli T, Breinig S et al (2007) Metab Eng 9:268–276
271. Qi Y, Zhao RX, Cao HX et al (2002) J Cell Biochem 86:79–89
272. Shimizu H, Tanaka H, Nakato A et al (2003) Bioprocess Biosyst Eng 25:291–298
273. Kimura E (2003) Adv Biochem Eng Biotechnol 79:37–57

274. Branduardi P, Fossati T, Sauer M et al (2007) PLoS One 2:e1092
275. Manoj K (2002) Production of ascorbic acid. US Patent 6,358,715
276. Perkins JB, Sloma A, Hermann T et al (1999) J Ind Microbiol Biotechnol 22:8–18
277. Stahmann KP, Revuelta JL, Seulberger H (2000) Appl Microbiol Biotechnol 53:509–516
278. Zhu YB, Chen X, Chen T et al (2007) FEMS Microbiol Lett 266:224–230
279. Robin J, Jakobsen M, Beyer M et al (2001) Appl Microbiol Biotechnol 57:357–362
280. Chen Y, Deng W, Wu JQ et al (2008) Appl Environ Microbiol 74:1820–1828
281. Banerjee UC, Saxena B, Chisti Y (1992) Biotechnol Adv 10:577–595
282. Elander RP (2003) Appl Microbiol Biotechnol 61:385–392
283. Fjaervik E, Zotchev SB (2005) Appl Microbiol Biotechnol 67:436–443
284. Yoon YJ, Kim ES, Hwang YS et al (2004) Appl Microbiol Biotechnol 63:626–634
285. Pearson D, Ward OP (1988) J Appl Bacteriol 65:195–202
286. Rowe GE, Margaritis A (2004) Biotechnol Bioeng 86:377–388
287. Leonard E, Yan YJ, Koffas MAG (2006) Metab Eng 8:172–181
288. Ralston L, Subramanian S, Matsuno M et al (2005) Plant Physiol 137:1375–1388
289. Ro DK, Douglas CJ (2004) J Biol Chem 279:2600–2607
290. Yan YJ, Chemler J, Huang LX et al (2005) Appl Environ Microbiol 71:3617–3623
291. Becker JVW, Armstrong GO, Van der Merwe MJ et al (2003) FEMS Yeast Res 4:79–85
292. Beekwilder J, Wolswinkel R, Jonker H et al (2006) Appl Environ Microbiol 72:5670–5672
293. Zhang YS, Li SZ, Li J et al (2006) J Am Chem Soc 128:13030–13031
294. Yamano Y, Yoshizawa M, Ito M (1999) J Nutr Sci Vitaminol 45:49–62
295. Alper H, Miyaoku K, Stephanopoulos G (2005) Nat Biotechnol 23:612–616
296. Yoon SH, Park HM, Kim JE et al (2007) Biotechnol Prog 23:599–605
297. Wang CW, Oh MK, Liao JC (1999) Biotechnol Bioeng 62:235–241
298. DeJong JM, Liu YL, Bollon AP et al (2006) Biotechnol Bioeng 93:212–224
299. Engels B, Dahm P, Jennewein S (2008) Metab Eng 10:201–206
300. Carter OA, Peters RJ, Croteau R (2003) Phytochemistry 64:425–433
301. Asadollahi MA, Maury J, Møller K et al (2007) Biotechnol Bioeng 99:666–677
302. Kealey JT, Liu L, Santi DV et al (1998) Proc Natl Acad Sci USA 95:505–509
303. Wattanachaisaereekul S, Lantz AE, Nielsen ML et al (2008) Metab Eng 10:246–254
304. Prathumpai W, Flitter SJ, McIntyre M et al (2004) Appl Microbiol Biotechnol 65:714–719
305. Kusumoto KI, Matsushita-Morita M, Furukawa I et al (2008) J Appl Microbiol 105:1711–1719
306. Chill L, Trinh L, Azadi P et al (2008) Biotechnol Bioeng 102:828–844
307. Talabardon M, Yang ST (2005) Biotechnol Prog 21:1389–1400
308. Carlsen M, Nielsen J (2001) Appl Microbiol Biotechnol 57:346–349
309. de Vries RP (2008) A genomic look at physiology and extracellular enzymes of Aspergillus in relation to utilization of plant matter. In: Varga J, Samson RA (eds) Aspergillus in the genomic era. Wageningen Academic, The Netherlands
310. Shankar SK, Mulimani VH (2007) Bioresour Technol 98:958–961
311. Westers L, Westers H, Quax WJ (2004) Biochim Biophys Acta 1694:299–310
312. Zukowski MM, Stabinsky Y, Levitt M (1990) Subtilisin analogs. US Patent 4,914,031
313. Pedersen H, Beyer M, Nielsen J (2000) Appl Microbiol Biotechnol 53:272–277
314. DSM. DSM nutritional products. Available at http://www.dsm.com/en_US/html/dnp/prod_-vit_b2.htm. Accessed 20 Jan 2009
315. Cameron DC, Altaras NE, Hoffman ML et al (1998) Biotechnol Prog 14:116–125
316. Nakamura CE, Whited GM (2003) Curr Opin Biotechnol 14:454–459
317. Tong IT, Liao HH, Cameron DC (1991) Appl Environ Microbiol 57:3541–3546
318. Suriyamongkol P, Weselake R, Narine S et al (2007) Biotechnol Adv 25:148–175
319. Suthers PF, Cameron DC (2001) Production of 3-hydroxypropionic acid in recombinant organisms. PCT WO 01-16346
320. Ensley BD, Ratzkin BJ, Osslund TD et al (1983) Science 222:167–169
321. Weyler W, Dodge TC, Lauff JJ et al (1999) Microbial production of indigo. US Patent 5,866,396

322. Werner RG (2004) J Biotechnol 113:171–182
323. Ro DK, Paradise EM, Ouellet M et al (2006) Nature 440:940–943
324. Choi JH, Lee SJ, Lee SY (2003) Appl Environ Microbiol 69:4737–4742
325. Chandran SS, Yi J, Draths KM et al (2003) Biotechnol Prog 19:808–814
326. Blank LM, Kuepfer L, Sauer U (2005) Genome Biol 6:R49
327. David H, Akesson M, Nielsen J (2003) Eur J Biochem 270:4243–4253
328. Sun JB, Lu X et al (2007) Metabolic peculiarities of *Aspergillus niger* disclosed by comparative metabolic genomics. Genome Biol 8(9)
329. Kjeldsen T (2000) Appl Microbiol Biotechnol 54:277–286
330. Markussen J, Fiil N, Hansen MT et al (1990) DNA-sequence encoding biosynthetic insulin precursors and process for preparing the insulin precursors and human insulin. US Patent 4,916,212
331. Joosten V, Christien L, van den Hondel C et al (2003) Microb Cell Fact 2:1
332. Valenzuela P, Medina A, Rutter WJ (1982) Nature 298:347–350
333. Kim MD, Rhee SK, Seo JH (2001) J Biotechnol 85:41–48
334. Chemier JA, Fowler ZL, Koffas MA et al (2009) Adv Enzymol Relat Areas Mol Biol 76:151–217
335. van Maris AJA, Winkler AA, Kuyper M et al (2007) Adv Biochem Eng Biotechnol 108: 179–204
336. Saitoh S, Ishida N, Onishi T et al (2005) Appl Environ Microbiol 71:2789–2792
337. Maury J, Asadollahi MA, Moller K et al (2005) Adv Biochem Eng Biotechnol 100:19–51
338. Maury J, Asadollahi MA, Moller K et al (2008) FEBS Lett 582:4032–4038
339. Carlson R, Srienc F (2006) J Biotechnol 124:561–573
340. Zhang B, Carlson R, Srienc F (2006) Appl Environ Microbiol 72:536–543
341. Steen EJ, Chan R, Prasad N et al (2008) Microb Cell Fact 7:36
342. Sauer M, Branduardi P, Valli M et al (2004) Appl Environ Microbiol 70:6086–6091
343. Bro C, Knudsen S, Regenberg B et al (2005) Appl Environ Microbiol 71:6465–6472
344. Hirasawa T, Yoshikawa K, Nakakura Y et al (2007) J Biotechnol 131:34–44
345. Pizarro F, Vargas FA, Agosin E (2007) Yeast 24:977–991
346. Wattanachaisaereekul S (2007) Transcriptome data using reporter metabolites for 6-MSA production in recombinant strain. Technical University of Denmark, Kgs Lyngby, Denmark
347. Bro C, Regenberg B, Nielsen J (2003) Biotechnol Bioeng 85:269–276
348. Pizarro FJ, Jewett MC, Nielsen J et al (2008) Appl Environ Microbiol 74:6358–6368
349. Abbott DA, Suir E, van Maris AJA et al (2008) Appl Environ Microbiol 74:5759–5768
350. Bundy JG, Papp B, Harmston R et al (2007) Genome Res 17:510–519
351. Devantier R, Scheithauer B, Villas-Bôas SG et al (2005) Biotechnol Bioeng 90:703–714
352. Villas-Bôas SG, Åkesson M, Nielsen J (2005) FEMS Yeast Res 5:703–709
353. Salusjarvi L, Poutanen M, Pitkanen JP et al (2003) Yeast 20:295
354. Cheng JS, Qiao B, Yuan YJ (2008) Appl Microbiol Biotechnol 81:327–338
355. Zuzuarregui A, Monteoliva L, Gil C et al (2006) Appl Environ Microbiol 72:836–847
356. Salusjarvi L, Kankainen M et al (2008) Regulation of xylose metabolism in recombinant *Saccharomyces cerevisiae*. Microbial Cell Factories 7(18):1–16
357. Usaite R, Wohlschlegel J, Venable JD et al (2008) J Proteome Res 7:266–275
358. Dos Santos MM, Thygesen G, Kötter P et al (2003) FEMS Yeast Res 4:59–68
359. Kleijn RJ, Geertman JM, Nfor BK et al (2007) FEMS Yeast Res 7:216–231
360. Snitkin ES, Dudley AM et al (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. Genome Biol 9(9):RI 40
361. Grotkjaer T, Christakopoulos P, Nielsen J et al (2005) Metab Eng 7:437–444
362. Raghevendran V, Gombert AK, Christensen B et al (2004) Yeast 21:769–779
363. Carlson R, Fell D, Srienc F (2002) Biotechnol Bioeng 79:121–134
364. Herrgard MJ, Lee BS, Portnoy V et al (2006) Genome Res 16:627–635
365. Hjersted JL, Henson MA, Mahadevan R (2007) Biotechnol Bioeng 97:1190–1204

366. Pocsi I, Miskei M et al (2005) Comparison of gene expression signatures of diamide, H_2O_2 and menadione exposed Aspergillus nidulans cultures - linking genome-wide transcriptional changes to cellular physiology. Bmc Genomics 6(182):1–18
367. Sims AH, Robson GD, Hoyle DC et al (2004) Fungal Genet Biol 41:199–212
368. Malavazi I, Savoldi M, Di Mauro SMZ et al (2006) Eukaryot Cell 5:1688–1704
369. Breakspear A, Momany M (2007) Eukaryot Cell 6:1697–1700
370. Malavazi I, Savoldi M, da Silva Ferreira ME et al (2007) Mol Microbiol 66:74–99
371. Bok J, Hoffmeister D, Maggio-Hall L et al (2006) Chem Biol 13:31–37
372. Mogensen J, Nielsen H, Hofmann G et al (2006) Fungal Genet Biol 43:593–603
373. Ray A, Macwana S, Ayoubi P et al (2004) BMC Genomics 5:22
374. Sims AH, Gent ME, Lanthaler K et al (2005) Appl Environ Microbiol 71:2737–2747
375. David H, Hofmann G, Oliveira A et al (2006) Genome Biol 7:R108
376. David H, Ozcelik I, Hofmann G et al (2008) BMC Genomics 9:163
377. Andersen MR, Vongsangnak W, Panagiotou G et al (2008) Proc Natl Acad Sci USA 105:4387–4392
378. Jorgensen TR, Goosen T, van den Hondel CAMJJ et al (2009) BMC Genomics 10:44
379. Andersen MR, Lehmann L, Nielsen J (2009) Systemic analysis of the response of Aspergillus niger to ambient pH. Genome Biology 10(5):R47
380. Salazar M, Vongsangnak W, Panagiotou G, Andersen MR, Nielsen J (2009) Uncovering transcriptional regulation of glycerol metabolism in Aspergilli through genome-wide gene expression data analysis. Mol Genet Genomics 282(6):571–586
381. Perrin R, Fedorova N, Bok J et al (2007) PLoS Pathog 3:e50
382. Maeda H, Sano M, Maruyama Y et al (2004) Appl Microbiol Biotechnol 65:74–83
383. Masai K, Maruyama J, Sakamoto K et al (2006) Appl Microbiol Biotechnol 71:881–891
384. Kimura S, Maruyama J, Takeuchi M et al (2008) Biosci Biotechnol Biochem 72:499–505
385. Tamano K, Sano M, Yamane N et al (2008) Fungal Genet Biol 45:139–151
386. Chang P, Wilkinson J, Horn B et al (2007) Appl Microbiol Biotechnol 77:917–925
387. Price M, Yu J, Nierman W et al (2006) FEMS Microbiol Lett 255:275–279
388. Kim J, Yu J, Mahoney N et al (2008) Int J Food Microbiol 122:49–60
389. O'Brian G, Fakhoury A, Payne G (2003) Fungal Genet Biol 39:118–127
390. Price M, Conners S, Tachdjian S et al (2005) Fungal Genet Biol 42:506–518
391. Andersen MR, Nielsen J (2009) Fungal Genet Biol 46:S180–S190

Adv Biochem Engin/Biotechnol (2010) 120: 101–131
DOI: 10.1007/10_2009_52

Published online: 26 February 2010

De Novo Metabolic Engineering and the Promise of Synthetic DNA

Daniel Klein-Marcuschamer*, Vikramaditya G. Yadav*, Adel Ghaderi, and Gregory N. Stephanopoulos

Abstract The uncertain price and tight supply of crude oil and the ever-increasing demand for clean energy have prompted heightened attention to the development of sustainable fuel technologies that ensure continued economic development while maintaining stewardship of the environment. In the face of these enormous challenges, biomass has emerged as a viable alternative to petroleum for the production of energy, chemicals, and materials owing to its abundance, inexpensiveness, and carbon-neutrality. Moreover, the immense ease and efficiency of biological systems at converting biomass-derived feedstocks into fuels, chemicals, and materials has generated renewed interest in biotechnology as a replacement for traditional chemical processes. Aided by the ever-expanding repertoire of microbial genetics and plant biotechnology, improved understanding of gene regulation and cellular metabolism, and incessantly accumulating gene and protein data, scientists are now contemplating engineering microbial cell factories to produce fuels, chemical feedstocks, polymers and pharmaceuticals in an economically and environmentally sustainable way. This goal resonates with that of metabolic engineering – the improvement of cellular properties through the intelligent design, rational modification, or directed evolution of biochemical pathways, and arguably, metabolic engineering seems best positioned to achieve the concomittant goals of environmental stewardship and economic prolificity.

Improving a host organism's cellular traits and the potential design of new phenotypes is strongly dependent on the ability to effectively control the organism's genetic machinery. In fact, finely-tuned gene expression is imperative for achieving an optimal balance between pathway expression and cell viability, while avoiding cytotoxicity due to accumulation of certain gene products or metabolites. Early attempts to engineer a cell's metabolism almost exclusively relied on merely

G.N. Stephanopoulos (✉), D. Klein-Marcuschamer*, V.G. Yadav*, and A. Ghaderi
Bioinformatics and Metabolic Engineering Laboratory, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: gregstep@mit.edu

*denotes equal authorship

deleting or over-expressing single or multiple genes using recombinant DNA, and intervention targets were predominantly selected based on knowledge of the stoichiometry, kinetics, and regulation of the pathway of interest. However, the distributive nature of metabolic control, as opposed to the existence of a single rate-limiting step, predicates the controlled expression of multiple enzymes in several coordinated pathways to achieve the desired flux, and, as such, simple strategies involving either deleting or over-expressing genes are greatly limited in this context. On the other hand, the use of synthetic or modified promoters, riboswitches, tunable intergenic regions, and translation modulators such as internal ribosome entry sequences, upstream open reading frames, optimized mRNA secondary structures, and RNA silencing have been shown to be enormously conducive to achieving the fine-tuning of gene expression. These modifications to the genetic machinery of the host organism can be best achieved via the use of synthetic DNA technology, and the constant improvement in the affordability and quality of oligonucleotide synthesis suggests that these might well become the mainstay of the metabolic engineering toolbox in the years to come. The possibilities that arise with the use of synthetic oligonucleotides will be delineated herein.

Keywords Gene circuits, Metabolic control, Oligonucleotide synthesis, Regulatory engineering, Synthetic biology

Contents

1 Introduction 102
2 A Brief History of Oligonucleotide Synthesis 104
3 Molecular Engineering for the Control of Gene Expression 107
3.1 Re-Engineering Translational Control 108
3.2 Modulating the Transcriptional Machinery of the Cell 115
4 Exploring New Possibilities with Hitherto Elusive Genes 116
5 Gene Circuit Engineering Using Synthetic DNA Technology 117
6 Synthetic Libraries: Beyond Conventional Pathway Engineering 121
7 The Next Frontier in Metabolic Engineering: The Road to Synthetic Organisms 124
8 Conclusions 126
References 126

1 Introduction

Mankind's desire to raise crops and animals with superior physical traits dates back to prehistoric times, and the science (or art) of selective breeding has indeed matured during the course of man's habitation of this planet. Inasmuch, metabolic engineering, rudimentarily defined, is a modern, more potent avatar of selective breeding. More precisely, while manipulation of an organism's hereditary content to effect desirable changes in its phenotype does constitute an important aspect of the science, metabolic engineering goes much beyond mere genetic manipulation

and trait evaluation. Traditionally defined, it is the directed improvement of cellular properties through the intelligent design or rational modification of biochemical pathways [1, 2]. Thus, pragmatic metabolic engineering necessitates a precise understanding of metabolic control, which in turn is only obtained via rigorous quantification of metabolic fluxes. This combination of molecular biology, especially recombinant DNA techniques, and flux quantification tools forms the crux of present-day metabolic engineering [2]. Although seemingly contradictory to the classical definition of metabolic engineering, combinatorial metabolic engineering and directed evolution have also proven to be useful tools for crafting novel phenotypes.

Since its formal inception nearly two decades ago, metabolic engineering has achieved sterling success in the development of novel microbial strains for use in sustainable and cost-competitive bioprocesses. Some notable examples include the production of bulk chemicals such as propanediol, ethanol, and biopolymers such as poly(hydroxybutyrate) and other poly(hydroxyalkanoates), as well as fine chemicals such as synthetic drug intermediates, lycopene, and lysine [1]. Metabolic engineering has also found use in the production of glycosylated proteins [3] and other biotherapeutics [4]. While most of this early success almost exclusively relied on introducing a particular enzyme or set of enzymes into a host cell [5], the recent explosion in the volume of gene and protein data significantly improved understanding of cellular metabolism and genetic regulation. Additionally, advances in microbial genetics and plant biotechnology have emboldened metabolic engineers to take on grander and more pressing challenges such as energy and human health. Monumental as these challenges may seem, the incessant technological developments – most notably innovations in gene sequencing [6], *de novo* oligonucleotide synthesis [7], in silico enzyme design and protein engineering [8], "omics" tools [9], and synthetic biology [10, 11] – provide several reasons for metabolic engineers to remain optimistic.

An emerging theme in metabolic engineering – and one that could dominate the landscape in the years to follow – is the scientific community's increasing belief that solutions to many of mankind's challenges already exist in nature in one form or another, and that utilizing nature's warehouse via engineering heterologous metabolic pathways in productive hosts to manufacture fuels and chemicals is a compelling route to a sustainable future. Unfortunately, present techniques to utilize the rich biodiversity are almost exclusively based on lengthy and error-prone recombinant DNA methodologies, and this has undoubtedly hampered the productivity of metabolic engineers. Additionally, the growth in the volume of biological data has been accompanied by a commensurate rise in the number of hypotheses postulated to explain the observed phenomena and, consequently, the need to construct genetic elements, pathways and engineered cells to test these hypotheses is acute [7]. These trends have justifiably raised the profile of *de novo* DNA synthesis as being arguably suited to have a significant impact on future metabolic engineering endeavours owing to its ability to rapidly construct and test genetic designs. Synthetic oligonucleotides open up the possibility of synthesizing altogether novel genes, pathways and their variants (i.e., libraries) under precise control of user-specified signals to tune their expression, to list a few of the uses that

will be described herein. However, *de novo* oligonucleotide synthesis has yet to reach technological maturation. As the cost of synthesis drops further and its efficiency improves, so will its pertinence rise. This chapter will attempt to explain why the authors believe that economical and efficient DNA synthesis would usher in a new paradigm in metabolic engineering, preceded by a brief history of oligonucleotide synthesis to showcase this technology's phenomenal rate of development. As shall be evident later, and when full maturation arrives, the capabilities of metabolic engineering will exponentiate.

2 A Brief History of Oligonucleotide Synthesis

Much of the history presented herein has been adapted from Daniel Brown's seminal work, *A Brief History of Oligonucleotide Synthesis* that appears as a chapter in the *Methods in Molecular Biology* series [12]. While this presentation is self-contained, it is highly condensed and most of the mechanistic details of the syntheses have been deliberately omitted (an illustrative timeline is given in Fig. 1). Readers are encouraged to refer to Brown's original treatise to gain a wholesome appreciation of the chemical mechanisms that defined this pivotal technology through its evolution.

It might come as a surprise to many that the original exponents of oligonucleotide synthesis did not see – or claim – any obvious biological applications [13]. Then, it was viewed more as an exotic research interest of organic chemists interested in phosphorylation chemistry than a serious scientific enterprise. Several overt difficulties associated with the actual synthesis technique contributed to this outlook. Not only was phosphorylation poorly understood, but synthesizing oligonucleotides involved the sequential addition and removal of, then, largely ineffective protecting groups. Moreover, separating the synthesized oligonucleotides was cumbersome, at best, and an efficient enzymatic route to synthesize polynucleotides was already in widespread use at the time. Despite these difficulties, some visionary chemists did take up the challenge and much of the early work in oligonucleotide synthesis was centred on phosphorylation.

By the 1950s, the synthesis of sugar phosphates using phosphoryl chloride was gaining recognition and chemists had now turned their attention to synthesizing diesters, especially phospholipids. However, diester synthesis proceeded via a highly reactive phosphorochloridate intermediate that eventually yielded several undesirable by-products, and this prompted chemists to investigate alternate synthesis schemes. Of the several avenues that were proposed, the use of phosphorus and polyphosphate chemistry for coenzyme synthesis became quite popular, and ensuing studies on nucleoside polyphosphate synthesis by A. R. Todd's group at the University of Cambridge during the 1950s and 1960s, besides identifying aryl groups to be effective protecting agents, suggested that rapid deprotection of the side chains of the phosphate species following the esterification reaction served to reduce non-specific phosphoryl reactions, thereby improving product yields.

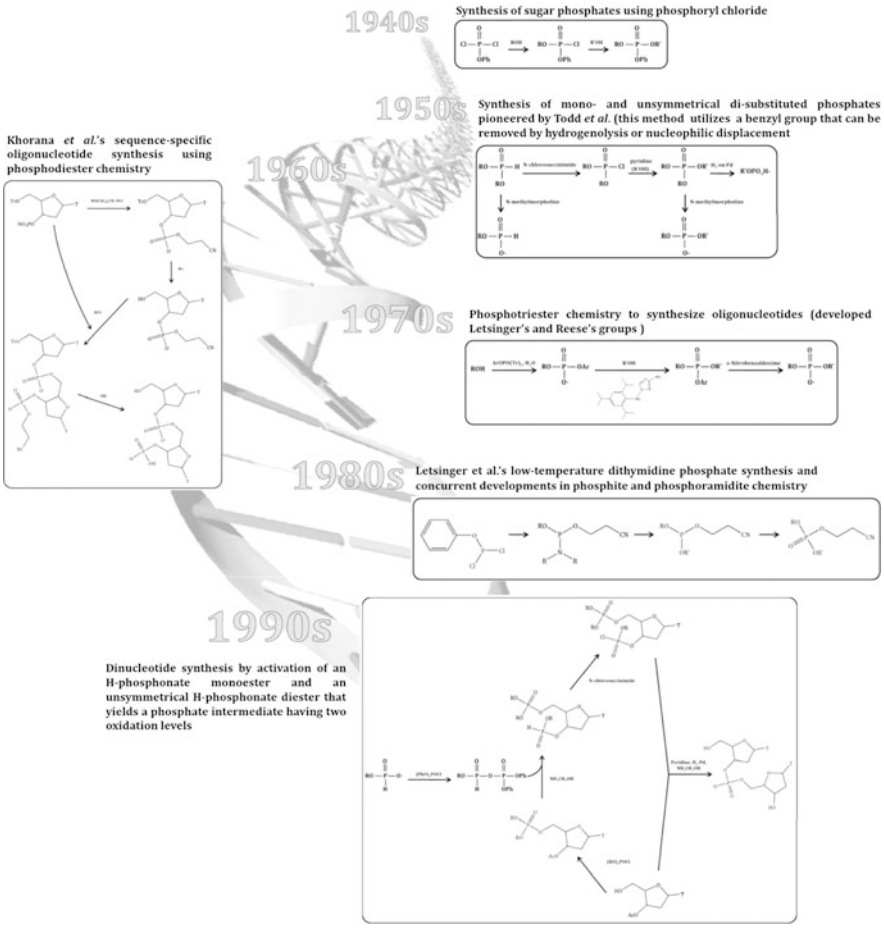


Fig. 1 A brief timeline of oligonucleotide synthesis

Accordingly, protocols for activating monoesters such as adenosine-5′-phosphate would soon be developed and these breakthroughs would eventually pave the way for H. Khorana and his colleagues' pioneering work on phosphodiester chemistry and sequence-defined synthetic oligonucleotides at the University of Wisconsin, Madison.

By the 1960s, base (mainly adenine, guanine, and cytosine) protection using acyl groups, and nucleoside-5′-*O* protection using dimethoxytrityl groups, would become well-established techniques. These discoveries coincided with a precipitous increase in interest to decipher the genetic code, and, in due course, total synthesis of di-, tri-, and tetranucleotides, their subsequent enzymatic conversion to polydeoxyribo-, and later to polyribonucleotides, with repeating sequences, would soon be realized. One of the early applications of these polyribonucleotides was to synthesize proteins in vitro, and, as the 1970s drew to a close, Khorana would

successfully synthesize two entire tRNA genes and a transcribable fMet-tRNA sequence. Soon, by the mid-1980s, Yu Wang and his group at the Shanghai Institute of Organic Chemistry would synthesize yeast alanine tRNA, including all the other minor ribonucleoside residues.

Concurrent with the aforementioned breakthroughs in phosphodiester chemistry, several research groups around the world, notably R.L. Letsinger's group at Northwestern University and C.B. Reese's at King's College in London, were actively pursuing phosphotriester synthesis. These researchers aimed to address two fundamental issues: finding neutral and more adequate protecting groups for the phosphoryl side-chains to enable the efficient formation and rapid deprotection of the triester products [14], and activating the triester-forming coupling reaction without favouring the formation of undesirable symmetrically substituted products [15]. A solution to these problems would eventually be proposed by J. Catlin and F. Cramer of the Max Planck Institute for Experimental Medicine, who, in the early 1970s, formulated a method to protect nucleoside hydroxyl groups to allow chain extension in either or both directions, and, consequently, achieve block condensations more efficiently. A few years after Catlin and Cramer, C. Daub and E. van Tamelen at Stanford University developed a technique that utilized thiolate-displaceable methyl groups as protecting agents – a method that, in the years to come, would become the cornerstone of phosphotriester synthesis, the method of choice for synthesizing oligonucleotides.

Besides advancing phosphotriester synthesis, R.L. Letsinger made another, equally significant contribution to oligonucleotide synthesis. Previously, liquid phase oligomer building blocks could only be enzymatically linked to each other to yield longer chains. Letsinger's work on nucleotide chain extension on solid supports – largely inspired by R.B. Merrifield's work on solid-phase peptide synthesis at the Rockefeller Institute during the 1960s – ushered in a new paradigm in oligonucleotide synthesis. Automated oligonucleotide synthesis now became a reality and HPLC could be used to purify the final products.

By the 1980s, the genomic revolution had swept biology and the demand for oligonucleotides as probes, primers, and linkers, as well as experiments to study controlled mutagenesis and protein-nucleic acid interactions, suddenly rocketed. Also during this decade, the last piece of the puzzle would finally fall in place as M. Caruthers and his research group at the University of Colorado at Boulder optimized a method for automated chain assembly utilizing phosphite and phosphoramidite chemistry. Their work was particularly telling as it yielded several routes to modify the basic oligomer structure. Immediately following Caruthers's breakthrough, *de novo* oligonucleotide synthesis received another boost when J. Stawinski and co-workers at the University of Stockholm and another research group at Genentech Inc. concurrently demonstrated the applicability of H-phosphonate chemistry for versatile, high-yield, automated oligomer assembly.

The 1980s also witnessed the development of the polymerase chain reaction (PCR) by Kary Mullis. Despite its seminal contributions to molecular biology, PCR has been deliberately omitted from this review. After all, PCR is template-dependent, and a key argument of this discussion is that there are inherent

advantages of being unconstrained by a particular sequence, even when the sequence is partially or completely derived from nature.

Prolific developments in machine-based oligoribonucleotide synthesis based on the phosphoramidite chemistry would occur during the 1990s, and RNA synthesis, long considered inherently more demanding, was well on its way to becoming as routine as DNA synthesis. In the years since, DNA and RNA syntheses technologies have continued to mature, and the past decade has witnessed these technologies becoming increasingly sophisticated, cheaper, and faster. Despite these rapid developments, several challenges remain for *de novo* DNA synthesis. First, the chain lengths of synthetic oligonucleotides have remained relatively short. However, recent work at the J. Craig Venter Institute [16] is a significant step toward synthesis of large genomic sequences and the problem of short chain lengths might well be short-lived. Another oft-overlooked problem is one of scale. Solution-based, phosphotriester synthesis methods could address this concern [12], and large- to very-large-scale synthesis of deoxyoligonucleotides for gene therapy using this methodology is already on the rise [7]. Additionally, the use of ink-jet printing, photolabile protecting groups, and photo-initiated acid deprotection is fast emerging as an established technique to synthesize oligonucleotides [7], and as the synthesis density (defined as the number of unique oligonucleotides that can be synthesized per unit area) on these microchips improves, so will the scale and throughput of this technology.

Finally, the recent emergence of several companies offering economical synthesis services and their visible popularity within the life sciences community seem to suggest that the price constraints typically associated with *de novo* DNA synthesis are fast receding (the price today ranges from approximately \$0.1 to >\$1.50 per base, depending on the length, complexity, and required sequence accuracy). The elusive and almost unimaginable goal of designing bench-top oligonucleotide synthesis and assembly machines akin to PCR thermocyclers remains. Such instruments that do not require their operators to be highly-trained chemists and offer the same ease and convenience of a PCR thermocycler could change the very face of synthetic biology and metabolic engineering, as shall be made evident in the ensuing discussions.

3 Molecular Engineering for the Control of Gene Expression

When it was first conceived nearly two decades ago, metabolic engineering aimed to systematically improve the properties of a cell by intelligently designing and/or rationally modifying its native biochemical pathways, while ensuring that its vital metabolic functions remain uncompromised. At about the same time, genetic engineering was fast realizing maturity and, consequently, techniques utilizing recombinant DNA formed the bedrock of early metabolic engineering. Accordingly, most of the early attempts to engineer a cell's metabolism typically involved deleting or over-expressing single or multiple genes using recombinant DNA based

on knowledge of the stoichiometry, kinetics, and regulation of a particular pathway [17]. However, as the body of knowledge on biochemical pathways steadily grew, metabolic engineering would correctly emphasize the distributive control of a metabolic pathway as opposed to the existence of a single rate-limiting step. Experimental tools for modulation of gene expression provided an alternative to either complete deletion or burdensome over-expression of enzymes in the pathway of interest [18, 19]. This essentially implied that any attempt to re-engineer successfully the cell's native metabolism or introduce a heterologous pathway necessitated the balanced expression of several enzymes. Thus, metabolic engineering would eventually focus on the development of tools for finely tuning genetic expression to strike an optimal balance between pathway expression and cell viability, while specifically avoiding deleterious effects such as impaired growth, metabolic imbalance, and cytotoxicity due to accumulation of certain gene products or metabolites [17, 19, 20].

Numerous developments in cell and molecular biology and genetics have vastly improved the metabolic engineering toolbox. Several genetic and metabolic control mechanisms have been elaborated in recent years [21–25] and this information promises to improve vastly the ability to fine-tune pathway fluxes. Already, tools such as synthetic [26] or modified promoters [27] and tunable intergenic regions (TIGRs) [28] have yielded many promising results, and potential use of translation modulators such as the 5′ CAP structure, the 3′ poly-A tail, internal ribosome entry sequences (IRESs), upstream open reading frames (uORFs) [21], optimized mRNA secondary structures [23], and RNA silencing [24, 29] could significantly expand the capabilities of metabolic engineering. Synthetic DNA will allow full exploitation of these elements for building multi-enzymatic, precisely controlled pathways. Some of these possibilities will be delineated in the ensuing discussions.

3.1 Re-Engineering Translational Control

3.1.1 Optimization of mRNA Sequence and Structure for Effective Control of Translation

The translation of mRNA to protein concludes the gene expression cascade and links the proteome to the genome. Consequently, control of translation can be a direct and effective means to modulate the proteome [21, 30, 31]. In addition to transcript interactions with protein regulators, translation is also modulated by structural features or regulatory sequences appearing within the mRNA molecules. The 7-methylguanylate triphosphate nucleotidyl caps at the 5′ end, poly-A tails, uORFs, and IRESs are examples of structures that affect the rate and efficiency of translation in eukaryotes [21].

Translation can be divided into three distinct steps – initiation, elongation, and termination – and modifying any one of these steps should, in theory, affect translation. The most basic explanation for the structure-induced control of

initiation of translation is the formation of hairpins or knots within the mRNA on account of complementary base pairing within the transcript itself or due to interactions with proteins or smaller trans-acting RNAs. Such secondary structures either block or favour ribosome access to the initiation site, subsequently affecting initiation of translation [21, 23]. It should be noted, however, that the mechanisms of translation initiation differ in prokaryotes and eukaryotes. While the polycistronic structure of mRNA is a major determinant of translation control in prokaryotes, this phenomenon is entirely absent in eukaryotes.

In prokaryotes, several structural elements affect the initiation and rate of translation. The binding of the AUG start codon to the UAC anti-codon in fMet-tRNA sets the reading frame for translation by the smaller ribosomal subunit, and modifications to the start codon and its surrounding sites could drastically affect the rate of initiation of translation [21]. While modifications to the start codon always lowered the initiation rate [23], A/U-rich or G-poor initiation sites exhibited elevated rates of translation [32]. It is speculated that A/U-rich initiation sites inhibit the formation of secondary structures. It has also been shown that complementary binding of unique, purine-rich sequences about four to five nucleotides long and upstream of the start codon, called Shine-Dalgrano (SD) sequences, to sequences near the 3′ end of the 16S rRNA affects initiation of translation (SD sequences are the prokaryotic-specific version of ribosome binding sites, or RBS). Modifying the length and nucleotide content of the SD sequence, and its position relative to the start codon (generally five to eight nucleotides upstream) could substantially improve the rate of initiation of translation, especially if secondary structures that inhibit initiation are known to exist within the transcript [23].

At this stage, it might be intuitive to the reader that manipulating the ribonucleotide sequence of 5′ non-coding regions could potentially lead to modifications in the structure of the mRNA transcripts, and deliberate inclusion or exclusion of secondary structures such as hairpin loops could drastically alter the rate of translation. If one were able to predict the effect of manipulating the ribonucleotide sequence *a priori*, one could, in essence, synthesize a gene with these desired modifications and expect optimal expression of the encoded proteins. Although a clear cause-and-effect relationship is yet to be established, the improved understanding of the structural basis of translational control has been accompanied by incessant improvements in the ability to predict the formation of potentially translation-influencing secondary structures on the mRNA transcripts using bioinformatics. Numerous algorithms have been developed in recent years to predict effectively the secondary structures of mRNA transcripts [33], and several software packages utilizing these algorithms are freely available on the internet [33–35]. Assisted by these methods, one could synthesize a collection of gene constructs, each exhibiting a structural peculiarity that putatively modulates the rate of translation, and the several permutations and combinations of structural features could be tested. This capability, when realized, would be immensely powerful and could yield a universal method to fine-tune gene expression in engineered cells.

The polycistronic structure of prokaryotic mRNA presents another intriguing prospect. In some cases, a downstream cistron lacking a strong (or exposed) RBS is

translated with varying efficiency upon translation of the upstream cistron, resulting in non-equimolar production of proteins expressed by the polycistronic genes (with the expression of downstream gene being weaker). Generally, as the ribosomes progress through the transcript, base-pairing disruption occurs, thereby opening RBS-obscuring secondary structures and allowing translation of the downstream cistron. The proximity of cistrons also permits coupled translation when the second RBS cannot recruit ribosomes independently [23]. Coupled translation of two or more cistrons provides a mechanism for controlling gene expression by changing the order of the cistrons and the structure and strength of their RBSs. Inequitable translation could also be corrected via use of unique TIGRs [20, 28]. In fact, depending on one's choice of the intergenic sequence, the succeeding genes could be translated to a greater extent compared to their preceding genes (in contrast to the case of coupled translation). The TIGR approach involves the use of three regions – two of which are variable sequences that form hairpin loops when transcribed, and a third region that incorporates one of many possible RNase E sites. The hairpin-forming sequences flank the RNase E site, and when transcription occurs, the nascent mRNA is cleaved at the RNase E site to yield two new transcripts whose stabilities are individually modulated by the hairpin structures (a more detailed discussion on the influence of secondary structures on mRNA stability has been included at the end of this section). As with the SD sequences, modifying the length and nucleotide content of the TIGRs could greatly affect the efficiency of translation; numerous other features could also be included in the TIGRs to improve translation. Some of these include IRESs, mismatched bulges, and asymmetries, among others.

Structural peculiarities also influence the translation of eukaryotic transcripts. However, unlike prokaryotic genes, eukaryotic genes are individually transcribed to monocistronic mRNA. This implies that the molecular mechanisms of translation manifested in eukaryotes are markedly different from their prokaryotic counterparts and, as a consequence, different structural features from the ones outlined previously would have to be incorporated into eukaryotic transcripts to control and modulate their translation effectively. One such structure is the 5′ 7-methylguanylate triphosphate nucleotidyl (m7G) cap. The complex formed on account of interactions between the m7G cap, its neighbouring regions, and eIF4E protein strongly augments ribosomal binding to the translation initiation sites [21, 23] and variations to the 5′ terminal sequence could greatly alter the translational efficiency [36]. Another structural feature known to play a role in the control of mRNA translation in eukaryotes is the unique nucleotide sequence that encompasses the start codon. This feature, often referred to as the context for initiation, has an optimal sequence that is about 10 nucleotides-long (inclusive of the start codon), and mutations to specific sites within the context for initiation were shown to affect greatly initiation of translation [37]. Using oligonucleotide synthesis, one could construct several contexts and test the effect of point mutations on initiation of translation. In addition to the m7G cap and the context for initiation, the 5′ untranslated region (UTR) is also known to elicit some degree of translation control. In mammalian cells, the 5′ UTR that constitutes the leader sequence of

the transcripts is considerably GC-rich. As a result, the leader sequences of mammalian transcripts exhibit a higher density of secondary structures [23]. Yeast transcripts, in contrast, have AU-rich leader sequences [38]. Based on the preceding discussion, one could speculate that the presence of secondary structures in the leader sequences of the transcripts will duly affect their rate of translation. Recent investigations into the mechanisms of re-initiation of translation have also revealed some interesting conclusions. It was observed that some structural features on the mRNA transcripts facilitate re-initiation of its translation more than others and, accordingly, translation capacity, defined simply as the number of times a transcript is translated, is also a function of the nucleotide sequence [23]. Thus, rational modifications to the transcript sequence to promote or eliminate formation of secondary structures assisted by synthetic DNA technology could improve control of gene expression.

The stability of the mRNA transcripts is another determinant of translational efficiency. It has been postulated that mRNAs are subjected to alternative decay processes based on their structural characteristics and the kinetics of these decay processes vary across time-scales [22]. By incorporating selected stabilizing structural elements within the mRNA transcripts, one could, in theory, improve mRNA lifetime. By coupling this phenomenon to other structural-based properties such as translational throughput and re-initiation of translation, one could effectively tune gene expression to optimize throughput of a pathway.

The list of examples detailed above is by no means exhaustive. Several additional examples of the structure-based control of translation, such as repressor protein binding sequences on the mRNA transcripts and ribozymes [39], are among others that could be listed. The intentions of this chapter are not to enumerate examples of how structural characteristics of mRNA influence its translation. What should be evident from the preceding discussion is that mRNA structure is a critical determinant of translation, and synthetic DNA technology offers the possibility to alter and probe the effects of mRNA sequence on its structure and the resulting translation efficiency. Further developments in this area will be immensely valuable to metabolic engineering as it will enable practitioners to fine-tune the fluxes of desired metabolic pathways through modulation of protein levels.

3.1.2 Codon Optimization for Efficient Gene Expression

As was evident in the previous section, the structural characteristics of the mRNA transcripts play a significant role in translation. An overwhelming majority of the structurally dependent mechanisms of translation control dealt with inhibition or improvement of ribosomal access to the initiation sites or altering the rate at which ribosomes scan across the transcripts. This section aims to complement the previous discussion by considering a more fundamental aspect of translation – codon usage.

There are 64 possible combinations of A, T (U), C, and G, and 20 distinct amino acids. This disparity between the number of codons and the number of the

amino acids that they encode generates considerable redundancy in codon specificity. In other words, each amino acid could be encoded by as many as three codons. Organisms tend to exhibit a preference for certain codons over others; for example, arginine is encoded by CGC, AGG, and AGA, but in *Escherichia coli*, AGG and AGA are categorized as "rare codons" and contribute to only 0.14% and 0.21% of all instances of arginine [40]. As a result, heterologous gene expression in this host could be severely constrained if the desired genes exhibit an organism-bias with several occurrences of the AGA codon. Some common impediments to heterologous expression of a gene with several rare codons include ribosome stalling (a phenomenon wherein onset of a rare codon during translation arrests or impedes ribosomal scanning of the transcript, subsequently reducing ribosomal loading at downstream initiation codons) and frame shifting (an event that occurs when a ribosome encounters a rare codon and in an attempt to circumvent stalling, mistranslates the transcript). In some cases, early termination of translation has also been reported [41].

Organism-biased codon usage can be remedied by manipulating the interplay between a codon and its encoded amino acid. It is evident that several codons encode a particular amino acid, but only one amino acid is encoded by a particular codon. This redundancy is, thus, unidirectional in this context. Consequently, one could either increase the availability of the amino acid corresponding to the rare codon or change the transcript sequence to reflect the codon preferences of the host (analogous to market forces, one can either manipulate the supply or the demand sides).

It has been shown that the ability of a codon to be translated is directly correlated to its corresponding tRNA pool sizes [40, 42]. Thus, one could modulate the efficacy of codon translation by altering its corresponding tRNA pool size [41–43]. In the example provided earlier, by increasing the tRNA levels of AGA, one could increase its frequency of translation into arginine. tRNA levels could be increased in several ways, but the most routinely utilized technique involves transforming commercially available tRNA transcribing plasmids into the host [41, 42]. However, this method is fairly rudimentary and could be onerous if several rare codons appear within the mRNA transcript. Moreover, increasing the abundance of tRNA could unintentionally modify expression levels of endogenous genes [41] and replication of the additional plasmids would further burden the cell unnecessarily.

In view of these drawbacks, modifying the sequence of the exogenous gene to conform its codon usage to that of the host may be a more suitable alternative. This solution, colloquially termed codon optimization, requires *de novo* gene synthesis whenever rare codons appear in more than a handful of instances. By synthesizing genes with rare codons replaced by those that are abundantly expressed in that host, one could circumvent most of the problems associated with codon bias. In a recent study, 30 human genes were targeted for heterologous expression in *E. coli* [41]. Three variants for each gene were expressed in the bacterium – the native form of the gene, a codon-optimized synthetic sequence, and the native form of the gene augmented by plasmids encoding tRNA for the rare codons. It was observed that the codon-optimized synthetic sequences were, on average, expressed at significantly higher levels compared to the other two variants. It is believed that the use of

Fig. 2 This graphic illustrates the application of codon-optimization to a DNA sequence. The sequence chosen encodes (**a**) the first 17 amino acids of *Thermomyces lanuginosus* xylanase, encoded by the *xyn-A1* gene. The native DNA sequence (**b**) was reported by Yin et al. [104]. We used the OPTIMIZER software, available on-line [105], to generate a codon-optimized sequence for a high level of expression in *Escherichia coli*. Optimized sequence generated using (**c**) the one AA-one codon algorithm differed from that generated using (**d**) a randomized optimization based on codon usage probabilities. Codons shown in *red* indicate a different usage from the native sequence; *underlined codons* represent those which were assigned different codon usage by the two optimization algorithms chosen

codon-optimized sequences for heterologous gene expression alleviates the cell's translational load and, consequently, is more efficient than simply manipulating tRNA pool sizes. Over-expression of rare tRNAs is not always futile: it is beneficial and favoured when several genes with the same codon bias have to be expressed. In such cases, the expense of synthesizing large amounts of DNA may marginalize the benefit of codon-optimization and increasing tRNA pool sizes is preferred [41].

Several bioinformatic tools have been developed in recent years to facilitate codon optimization [44]. Some of these, such as the one offered by DNA 2.0, are freely available on the internet. Many of these algorithms simply replace all codons with synonymous codons associated with the highest level of expression (Fig. 2). However, critics of this method argue that merely replacing all codons with highly-expressed synonyms could yield many undesirable effects in those organisms that do not exhibit strong usage biases for some intermediately utilized codons. This has prompted the development of codon optimization algorithms based on codon usage frequencies instead of the trivialized one AA-one substitution routine described above (Fig. 2) [45].

Progress in DNA synthesis capabilities as well as in silico codon optimization algorithms have ensured that procurement of codon-optimized sequences is no longer a limiting step in biological experimentation. However, knowledge about codon usage is presently confined to only a handful of organisms, and if metabolic engineering is to break new ground, expansion of this information base is imperative.

3.1.3 RNA Silencing for Balanced Gene Expression

Another emerging theme in biology that is envisioned to enrich metabolic engineering is RNA silencing. The ensuing discussion also illustrates how synthetic DNA technology will be central to the fruitful implementation of the principles of RNA silencing.

RNA silencing, or post-transcriptional gene silencing, was first observed, quite serendipitously, in plants and fungi little over a decade ago [46] when successive attempts to over-express endogenous genes by introducing transgenic copies merely annulled expression of both sets of genes [47]. Since this fortuitous discovery, RNA silencing has steadily transitioned from being an intellectual curiosity to a way of analyzing the effects of single gene knockouts, then to a well-established methodology, and finally to a tool for implementing widespread genetic changes in a variety of hosts [46].

RNA silencing can occur via a variety of mechanisms, and some of these are more amenable to metabolic engineering than others. All silencing pathways are triggered by 21–27 nucleotide-long RNA fragments collectively known as small RNAs, and include small interfering RNAs (siRNAs), repeat-associated siRNAs (rasiRNAs), and micro RNAs (miRNAs). These small RNAs are produced by a multi-domain ribonuclease III enzyme, aptly referred to as DICER, in response to regulatory processes. Once produced, small RNAs mediate the activity of RNA-dependent RNA polymerases to guide the assembly of protein complexes such as the RNA-induced silencing complex (RISC) or the RNA-induced initiation of transcriptional gene silencing (RITS) complex – the main agents of the gene silencing machinery [47]. In addition, small RNAs also hybridize to the mRNA transcripts by complementary base pairing and, in the process, inhibit their translation and subsequently initiate their degradation by the silencing machinery [29]. The complementary binding that occurs between small RNA and its target transcript is often referred to as a sense/anti-sense (S/A) pair [29].

The ability to tightly control gene expression via reliable RNA silencing is particularly pertinent to metabolic engineering. Unfortunately, the complexity of the silencing machinery encumbers the triggering of gene silencing using conventional genetic engineering protocols. As a result, applications involving RNA silencing have been largely inconspicuous in metabolic engineering. However, recent studies have suggested that the sequence and structure of siRNAs greatly influence their ability to enter the RISC and, consequently, impact their function [47–49]. S/A pairing has also been used in prokaryotes successfully (see below). Thus, it could be hypothesized that synthetic RNAs with desired sequences could be utilized to trigger selectively the silencing processes, thereby controlling gene expression.

RNAi and S/A pairing promise to be significant additions to the metabolic engineering toolbox, as they provide supplementary ways of manipulating gene expression. Multiple combinations of over-expression and attenuation, each varying in their gene targets and the extent of modulation, could then be realized and, consequently, a wider landscape of pathway states could be investigated. For example, anti-sense RNA has been used to alter fluxes in *Clostridium*, a genus that has been

extensively used for the production of biofuels [50, 51]. RNAi-based anti-apoptosis engineering to minimize cell death in bioreactors is another excellent example of the applicability of RNAi to metabolic engineering [52]. Apoptosis is a highly-regulated cascade involving pro- and anti-apoptotic proteins, and disturbing the fine balance that exists between these disparately functioning protein groups through gene silencing could drastically improve cell viability, thereby improving bioprocess yields. Other potential metabolic engineering applications of gene silencing include inhibition of protease expression in recombinant protein producing strains.

A recently developed methodology to control the spacing, timing, and degree of gene expression using light-activated RNA interference [53] underscores the growing significance of synthetic DNA technology to metabolic engineering. In this methodology, siRNA complementary to the mRNA target strand is first synthesized and then modified by addition of a photolabile group. Functionalizing the siRNA strand renders it momentarily inactive but, following transformation into the host, the siRNA strand is activated by irradiation at a predefined wavelength to release the photolabile group. Thus, one is able to induce gene silencing in a highly controllable manner. Additionally, the development of bioinformatic algorithms [15, 48] and robust design rules [54, 55] to predict the efficacy of siRNA strands have further ensconced the role of synthetic siRNA-induced gene silencing in future metabolic engineering endeavours.

That the triad of structural control of translation, codon optimization, and RNA silencing has been recognized by mainstream metabolic engineering is testament to the potential of synthetic DNA technology as an invaluable tool in biotechnology. As the cost of oligonucleotide synthesis continues to decline and gene synthesis machines become widely available, several more applications, some of which are outlined in subsequent sections, could be envisaged.

3.2 Modulating the Transcriptional Machinery of the Cell

Probing the host's metabolic landscape necessitates selective variation of the expression levels of the genes of interest – a concept that is well beyond the scope of simple gene deletions or over-expressions. While the manipulation of translational activity seems to be an effective route, this aim can also be achieved by modulating the transcriptional machinery of the cell. The RNA polymerase complex that lies at the heart of transcription forms an obvious target for such perturbations and one approach that could profusely benefit from the use of synthetic DNA is promoter engineering. Briefly, promoter engineering involves generating a library of promoters of varying strengths in order to encompass the entire spectrum of gene expression levels, and though the traditional method for generating such libraries – namely, random mutagenesis of a constitutive promote using error-prone PCR – is highly effective, the use of synthetic DNA, suitably corroborated by bioinformatics tools, could not only generate novel promoters, but also vastly accelerate the screening of superior promoter sequences. This concept has been further elaborated in the Sect. 6.

4 Exploring New Possibilities with Hitherto Elusive Genes

Of the numerous applications of synthetic DNA technology that promise to revolutionize metabolic engineering, few compare in scope and impact to the ability to access genetic information from intractable sources. Difficulties in exploiting certain genes arise either when the organism possessing the gene is difficult to handle or when, owing to several causes, these genes are difficult to isolate from the host. Pathogens represent one class of organisms that are extremely difficult to handle for obvious reasons. However, despite the considerable environmental and health risks posed by pathogens, it is incontrovertible that access to the genetic information of these organisms would be instrumental in the formulation of effective therapies against them. Many pathogens also harbour genes that code for interesting enzymatic activities, a case-in-point being those of the hyaluronic acid (HA)-producing pathway in *Streptococcus*. Though the genetic information necessary to produce HA has been isolated from *Streptococcus* species and cloned into species such as *Bacillus subtilis* [56], *E. coli* [57], and *Lactococcus lactis* that are generally regarded as being safe [58], this scheme still requires handling of pathogens for isolation of genomic DNA. Synthetic DNA raises the possibility for investigators to access genes of harmful species without courting the threat of infection [59], and also offers other advantages such as reduced costs associated with forgoing specialized, biohazard-compliant equipment, customized laboratory space for handling pathogens, and meticulous practices to mitigate the risk of infection. It is important to note that the sequences of these elusive genes are known beforehand. All the examples provided in this discussion are based on this assumption.

A related example of synthetic DNA enabling the ready acquisition of elusive genes is the isolation of genes from prohibitively slow-growing cultures such as those of filamentous fungi and plant cells. Not only are the genetic engineering protocols specific to these organisms rather inadequately developed, several of these organisms take months to achieve the necessary cell densities, and in the absence of robust and reliable cloning protocols that require very little genetic material as inputs (as is the case for organisms such as *E. coli*), synthetic DNA offers a convenient and immensely faster route to procure these elusive genes. Additionally, synthetic DNA is an excellent alternative for DNA from some organisms that is hard to secure for ethical or conservationist reasons – humans, polar bears, and blue whales are some notable examples that fall in this category.

It is a well-established fact that regulatory mechanisms within a cell control its genomic integrity and all of its metabolic activities. In fact, such complex regulatory circuits are an evolutionary response aimed at improving cell fitness and survivability. Thus, it should be unsurprising that regulatory mechanisms and, occasionally, peculiar cellular physiologies and morphologies also render the genomes of some organisms difficult to isolate. For example, some organisms behave very differently under laboratory conditions compared to their natural environments. That is the case of the amicronucleate *Tetrahymena thermophila*, which can only be observed in coexistence with species containing micronuclei and

cannot be isolated in the laboratory [60]. This implies that genomic explorations in the amicronucleate are either confined to field experiments or must rely on synthetically-derived DNA. In the latter case, one could conduct genome sequencing work in the natural environments (this method has been lately referred to as metagenomics) and later synthesize these genes for subsequent experimentation. This approach was recently sensationalized by J. Craig Venter and his colleagues at JCVI when they embarked on a months-long voyage to profile the biodiversity of the world's oceans. This escapade and the tomes of subsequent research that it inspired have yielded several insights [61–64] and some groups have been very successful at using these approaches [65, 66]. However, despite its promise (and hype), tapping nature's diversity could offer little, if any benefit if the exotic genes one unearths using metagenomics cannot be studied in further detail owing to material and protocol limitations that were previously outlined in this section. Furthermore, isolation and amplification of one of the many identified sequences may be more difficult than *de novo* synthesis. Exploiting metagenomic data with the aid of synthetic DNA could have significant implications for metabolic engineering as it opens up the possibility of engineering hitherto unimaginable biosynthetic pathways and improving strain characteristics. For example, genes that enable microorganisms to not only survive, but thrive in such demanding environments as hydrothermal vents could potentially be expressed in laboratory strains to improve their survivability in the unnatural and stressful confines of a bioreactor.

DNA synthesis technologies are also enabling researchers to access genes of extinct organisms. While these projects are intellectually fascinating to palaeontologists, some readers might wonder about their scientific value to biotechnology. The recent emergence of the H5N1 avian influenza virus heightened concerns about the risk of a possible global pandemic and raised several questions about humanity's preparedness to mitigate such an occurrence. By studying the synthetic analogues of the genes of the influenza virus that decimated the globe in 1918, researchers will be better equipped to engineer better vaccines. In another palaeogenomic study of biotechnological significance, collagen fragments were recently extracted from a 500,000-year-old mammoth specimen and a 68 million-year-old *Tyrannosaurus rex* specimen [67]. That this unique type of collagen was able to remain intact for such a long period presents an opportunity in the field of biomaterials, and synthetic analogues of the genes will undoubtedly aid in producing this once-extinct type of collagen for further testing in the laboratory.

5 Gene Circuit Engineering Using Synthetic DNA Technology

Regulation of expression, particularly transcription, is commonly coordinated by several nucleotide sequences associated with the gene in question. Some of these elements include promoters, operators, terminators as well as structurally-diverse intergenic regions (see the Sect. 3.1.1). The gene and its associated elements

(which constitute an operon) frequently interact with other elements within the same or a different operon. It is partly the collection of these interactions that yields the complex regulatory networks commonly observed in biological systems. It thus follows that manipulating the elements of an operon could influence gene regulation and, as shall be addressed in the following pages, synthetic DNA offers unparalleled capability for harnessing the utility of gene networks for cellular engineering.

Of all the operon elements mentioned previously, the promoter is arguably the most influential. It exhibits a strong control over the cell's transcriptional machinery by recruiting the RNA polymerase either directly or through the action of transcription factors. Thus, the promoter is an obvious candidate for manipulation. Promoter (re-)engineering – modifying the promoter sequence to modulate transcriptional efficiency – has been effectively demonstrated in several recent studies based on synthetic oligonucleotides [68–70]. The generation and applications of synthetic oligonucleotide libraries to promoter engineering have been addressed in greater detail in the Sect. 6.

The operator sequence of an operon is also an important agent in the control of gene expression. Regulatory proteins interact with the operator to regulate initiation of transcription by either blocking or improving access of the polymerase to the promoter site. These regulatory proteins may be activated or inactivated by physical or chemical stimuli, and their versatility raises the possibility of constructing very interesting, and perhaps useful, regulatory networks. Operators, like promoters, could also be re-engineered using oligonucleotide synthesis. Moreover, by synthesizing constructs that combine several of these operon elements instead of limiting oneself to just one specific feature, researchers could traverse a broader genetic landscape and, consequently, many more complex regulatory schemes could be developed [71].

Operator elements can be stitched together to form multipart circuits with interesting properties. The genetic toggle switch is one such illustration [72]. A toggle switch, genetic or otherwise, involves two mutually repressing elements, i.e., each element represses the other element. Accordingly, a genetic toggle switch is made up of two cistrons, each preceded by a different promoter. Each cistron encodes a repressor protein that inhibits the expression of the other cistron. Finally, one of the promoters is also assigned control over the expression of a suitable reporter (generally GFP). In the study by Gardner et al., a P_{TRC-2} promoter (IPTG inducible and repressed by *lacI*) controls expression of *tetR* and a FACS-optimized GFP mutant (the reporter). Expression of *lacI*, in turn, is controlled by the $P_{L\text{-}tetO\text{-}1}$ promoter (aTc inducible and repressed by *tetR*). A schematic of the interplay between these cistrons is presented in Fig. 3a. Such a genetic device could be used to switch metabolism selectively and, consequently, is very relevant to metabolic engineering.

Another interesting synthetic regulatory element combining different operon elements is an oscillatory network of transcriptional regulators dubbed the "repressilator" [73]. Such a network consists of three operons, each exhibiting temporal repression of one of the other two operons via expression of a promoter-repressing

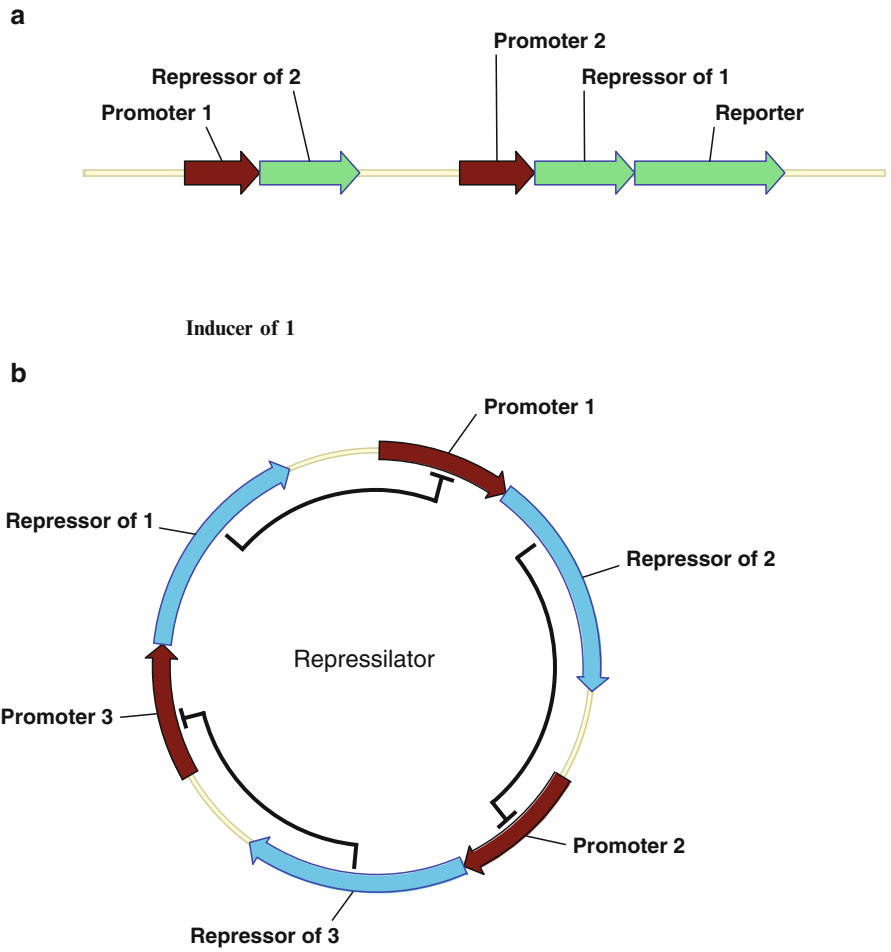


Fig. 3 These graphics illustrate the regulatory relationships between the elements of different operons in synthetic networks. Note that only promoters, regulatory proteins, and reporters are shown; each operon is assumed to have a terminator, though not illustrated. The binary switch synthetic network (**a**) can toggle the expression of a reporter by addition of an inducer of either promoter [72]. The repressilator (**b**) is a synthetic network that gives rise to oscillating levels of regulatory protein. By coupling the expression of a reporter to one of these regulatory proteins, the oscillatory nature of this network may be observed [73]

protein. Additionally, each operon cannot repress the expression of the protein that represses it. Thus, repression of the operons is cyclical. The repressilator designed by Elowitz and Leibler consisted of operons encoding *lacI*, *tetR*, and *λ cI*. These operons used the λ P_R (repressed by *λ cI*), $P_{L\text{-}lacO\text{-}1}$ (repressed by *lacI*) and $P_{L\text{-}tetO\text{-}1}$ (repressed by *tetR*) promoters, respectively. A schematic of this network is shown in Fig. 3b. To ensure that the lifetimes of the repressors were on similar scales, each

repressor was marked with a C-terminal tag recognized by *E. coli* proteases. The oscillatory phenotype was observed using a fourth operon which controlled the expression of modified GFP under the control of the *tetR*-repressible $P_{L\text{-}tetO\text{-}1}$ promoter. In the context of metabolic engineering, a similar construct could be used to generate dynamic pools of intracellular metabolites to overcome thermodynamic limitations imposed on the flux through a particular pathway.

Another recent study involving the use of synthetic oligonucleotides availed of a combinatorial approach to probe characteristics of an assortment of networks generated from interactions between simple regulatory elements [74]. Three repressor proteins (*lacI*, *tetR*, and λ *cI*) and five promoters (P_1^L, P_2^L, P^T, P_+^λ, and P_-^λ) were used in this study. P_1^L and P_2^L – promoters of the lac operon – are repressed by *lacI* and induced by IPTG. The P^T promoter is repressed by *tetR* and induced by aTc, whereas λ *cI* represses P_-^λ and activates P_+^λ accordingly. Cistrons for each of the proteins were assembled together with each of the five promoters, and a protease-recognizable sequence was added to the 3′-end of each construct to modulate their half lives. A collection of plasmids containing random assortments of any 3 of the 15 constructs was then generated. Incidentally, some of the regulatory schemes that could be generated by this method include the NOR, NAND, and NOT IF logical circuits. Each plasmid also contained an operon expressing GFP under the control of a P_-^λ promoter. The functional properties of the synthetic regulatory networks so generated were characterized by exposing the networks to IPTG and aTc and then measuring the fluorescence pattern of GFP. As expected, it was observed that the NOR circuit expresses GFP in the absence of both inducers, while the NAND circuit fails to express GFP unless both inducers are present. The NOT IF circuit expresses GFP in the presence of a single inducer, consequently giving rise to two possible such schemes. This methodology reveals the richness of regulatory interactions that could be achieved using a simple set of operon elements. Encouragingly, several more regulatory patterns could be obtained by either expanding the number of regulatory elements used to construct these circuits or combining a greater number of operons within a single scheme.

The development of novel regulatory devices using synthetic operons, such as the toggle switch, repressilator, and the logical circuits, attests to the possibility of designing complex regulatory schemes by associating a small number of genetic elements. However, procuring some of these regulatory networks using conventional recombinant DNA techniques would be a cumbersome and time-intensive endeavour. Synthetic alternatives alleviate some of the challenges associated with generating such intricate networks. Aided by gene circuit design software such as BioJADE [71] and Gene Designer [75], researchers can now frame several regulatory schemes in silico to suit their demands, and then synthesize their constructs for experimental characterization. Though several more developments in *in silico* design tools are still needed, the challenge of implementing and proving these schemes *in vitro* or *in vivo* is largely overcome by DNA synthesis technologies. This is especially important given that constructing untested genetic circuits is, in contrast to the electronic counterpart, principally a troubleshooting task.

6 Synthetic Libraries: Beyond Conventional Pathway Engineering

One of the main motivations of using synthetic DNA for cellular engineering seems to be at odds with the random nature of directed evolution. Traditionally, PCR-based methods have been used to create sequence diversity, inspired by the fact that mutations in nature commonly arise from errors in DNA replication. PCR-based methods are preferred when there is no prior knowledge about where mutations are likely to influence the traits of interest, but are limited in that the sequence diversity that results is restricted and biased. With single base mutations per codon – a common assumption with most protocols – only 5.7 amino acids are accessible per position on average, and in most cases, the resulting set of amino acids does not accurately represent the spectrum of physicochemical properties of naturally-occurring residues [76].

In contrast to amplification-based methods, synthetic DNA technology for library construction allows, at the very least, specifying the location where mutations are possible (e.g., saturation mutagenesis), and ultimately permits completely designing the desired sequence diversity (Fig. 4). A challenge for targeted mutagenesis is that some knowledge regarding where genotype changes are likely to affect phenotype is needed. Such knowledge can be obtained experimentally, through preliminary rounds of error-prone PCR, or computationally, based on

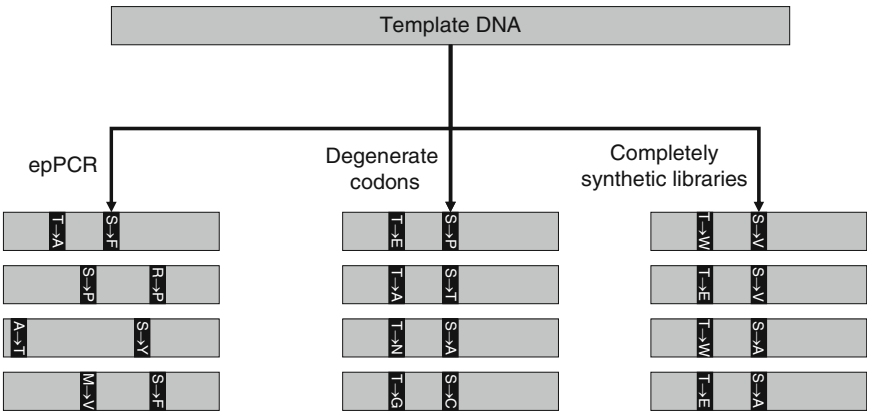


Fig. 4 The different ways in which genetic diversity can be introduced into a DNA sequence of interest. In error-prone PCR (epPCR), mutations can occur anywhere in the sequence, but the list of possible amino acid substitutions is non-exhaustive and non-random (see text). The use of synthetic oligonucleotides permits targeting mutagenesis to specific codons and, through the use of degeneracies, it allows changing any amino acid to any other. Ultimate control of the identity of the members in a library is achievable by completely synthesizing and pooling together all desired mutants; synthetic shuffling is a variation of this method in which some randomness is preserved. These three cases illustrate that the more one desires to reduce the search space, the more input one must provide for designing the library

structural information [77]. In the extreme case where the sequence diversity is entirely specified, the evolutionary effort is reduced to finding the mutant with most fitting properties. The reduction and design of the search space is therefore a key motivation for using synthetic DNA libraries for creating sequence diversity. Synthesis technologies based on sequential elongation of DNA molecules with codon-sized fragments will likely become more popular for these applications [78], but oligonucleotide-based methods have already gained sufficient attention as illustrated by the examples below.

In addition to saturation mutagenesis and related protocols, shuffling-based methods have also benefited from easy and reliable synthesis of DNA. The original gene-shuffling method that hinges upon random fragmentation followed by PCR [79] is still widely used, but synthetic versions have been proffered. The motivation is that the original method does not allow mutations that are close to each other to be recombined independently (i.e., mutations are co-inherited by the daughter molecules) and that it favours recombination in highly homologous regions [80]. By using synthetic oligonucleotides that encode all the theoretically possible variants of a library, there is no experimental limitation on the sequence diversity that can be obtained [81, 82].

The application of directed evolution has brought about significant progress in fields such as protein engineering, and these in turn have advanced cellular engineering in many ways. For instance, engineering faster enzymes, with novel substrate and co-factor specificities, product spectrum, or activity in non-native species or conditions, have a clear impact in metabolic engineering since enzymes are the basis of metabolism. Combinatorial search strategies for engineering other determinants of phenotype – such as promoters, secondary mRNA structures, and regulatory proteins – have also benefited from synthetic DNA technology. These applications shall be discussed in greater detail in the ensuing sections.

Altering substrate specificity of enzymes can serve as a tool for pathway engineering and design, and has been the focus of several protein engineering efforts. Because such specificity is dictated by the arrangement of amino acids near the catalytic site, oligonucleotide-mediated saturation mutagenesis directed to these residues can be used to impart the necessary changes. This was done, for example, to isolate an L-lactate dehydrogenase that is active on malate, phenyllactate, hydroxyisocaproate, and 4-phenyl-2-hydroxybutanoate [83]. In a similar approach, the K_m of a thymidine kinase for 3′-azido-3′-deoxythymidine was markedly lowered, making the mutant enzyme more efficient than the wild-type [84]. In another example, the fitness landscape related to the co-factor specificity of isopropylmalate dehydrogenase was investigated using a library of mutants [85]. The result of this study was the conversion of the NAD-dependent enzyme into an NADP-dependent one with the same activity as that of the wild-type. Given that redox balance is indispensable for ensuring high flux through a pathway of interest, similar approaches could be employed for metabolite overproduction.

Product specificity has also been shifted by directing mutagenesis to certain locations. The active site of a γ-humelene synthase was mutated and the pattern of how amino acid changes translate into changes in product specificity was analyzed

[86]. This allowed modelling the relationship between sequence and product preference, and opened the possibility to design novel terpene synthases. Saturation mutagenesis using synthetic oligonucleotides similarly allowed transforming a (+)-δ-cadinene synthase (which produces >98% (+)-δ-cadinene) into an enzyme that synthesizes germacrene D-4-ol with 93% purity [87]. In yet another instance, the product spectrum of farnesyldiphosphate (FPP) synthase of *Bacillus stearothermophilus* was studied by analyzing the catalytic activity of all 20 variants of a key residue in the enzyme [88]. The wild-type FPP synthase catalyzes the condensation of isopentenyl diphosphate (IPP, a C5 substrate) to FPP (a C15 product), while changes in amino acid Y81 resulted in products with 20, 25, and 30 carbons. Improving thermostability has been another focus of enzyme engineering [14] and, even though the main goal has been the isolation of enhanced proteins for cell-free biocatalysis, the resulting variants would be good candidates for metabolic engineering in thermophilic hosts.

Localized randomization and synthetic shuffling have also been applied for directed evolution of other genetic determinants that can be used to control phenotype. Promoter libraries are one example, of interest for metabolic control and pathway optimization [18, 68]. With the purpose of creating a set of promoters with different strengths, the localized bases adjacent to the consensus promoter sequences of *L. lactis* were randomized and used to express β-galactosidase [69]. Different activities of this enzyme expressed by different promoters were identified in X-gal plates, and further characterization of these elements was done in both *L. lactis* and *E. coli*. A modification of the method that alters promoter strength in a context-dependent fashion – by fusing the gene that will be expressed from the promoter library to a reporter gene – has also been developed using randomized oligonucleotides [89].

Use of synthetic oligonucleotide shuffling for pathway optimization is exemplified by the combinatorial assembly of so-called TIGRs (see the Sect. 3.1.1) [28]. A library of oligonucleotides that gave rise to different mRNA secondary structures was cloned between two fluorescent proteins to study how different combinations of stem loops at 5′- and 3′-ends of a gene affect protein levels. The TIGRs were designed to have two stem loops (one downstream of the first gene and another upstream of the second gene) separated by a RNaseE cleavage site (for a more detailed explanation, see the Sect. 3.1.1). The resulting library spans two orders of magnitude in relative fluorescence ratio, although the cited article recognizes that the library is skewed to contain more individuals with over-expression of the first gene in the sequence.

Similar approaches have also been employed in the construction of libraries of proteins that interact with DNA, which can be used for cellular engineering. Oligonucleotides are used either to target degeneracies in DNA-binding regions or to serve as individual DNA-binding determinants for synthetic shuffling. Two examples are offered. First, homing endonucleases (HEs) are sequence-specific endonucleases that effect double-strand DNA breaks, allowing for directed homologous recombination near their target sequences. To be able to direct such changes to new target sequences, randomization of residues in charge of DNA recognition resulted in a library of HEs that was screened for changes in specificity [90, 91].

Second, zinc-finger (ZF) domains have become a popular tool to direct transcription factor proteins to particular DNA sequences [92]. Briefly, a ZF domain consists of invariant tandem sequences comprising cysteine and histidine residues that collectively bind a zinc ion to form a finger-like loop within the protein. This has been exploited for construction of artificial transcription factors, fusion proteins that contain a DNA-binding ZF domain and a repressor or activator domain. Libraries of artificial transcription factors have been most commonly used to modulate expression of genes of interest [93], but since ZFs can also alter the expression of non-target genes [94], they have been used for engineering complex phenotypes [95, 96]. These phenotypes, such as tolerance to inhibitory fermentation conditions, are of interest for achieving more efficient and economically competitive biotechnological processes [97, 98].

7 The Next Frontier in Metabolic Engineering: The Road to Synthetic Organisms

Defining life has been a key and somewhat contentious philosophical question throughout human history. It is now conceivable to experiment our way out of this controversy – to synthesize chemically the components of a simple cell and see if life, broadly defined, emerges. Since DNA codes for the information needed to assemble every other cellular component, the synthetic version of this polymer is central to the task of building a cell. In fact, complete chemical synthesis of a cell will likely come much later than proof-of-concept experiments in which synthetic DNA in conjunction with cell-derived proteins and membranes are used to create a "designer cell." Synthetic DNA allows, conceptually, specifying the instructions that a designer organism will execute, and assuming that large-scale DNA synthesis is no longer a barrier [16, 99], a cardinal remaining challenge is resolving what to write in the DNA in order to have the expected outcome.

The level of difficulty for overcoming this challenge depends upon the use one has in mind for the synthetic cell (basic science or application). Many researchers have pointed out that having such a cell would permit a better understanding of how the different components give rise to observable properties [100, 101]; in other words, it would allow the construction of a genotype–phenotype map. For the purpose of assembling a simple synthetic cell, one can define the essential components bottom-up or top-down [102]. The bottom-up approach consists of enlisting the biochemical functions that are needed in a hypothetical cell and then finding the cellular components capable of carrying out these functions. The top-down approach consists in eliminating dispensable components of existing genomes either using comparative genomics or successively knocking out genes. The bottom-up approach is "cleanest" in that only user-specified parts define the cell, but since all new components will be put to the test simultaneously, troubleshooting can be a daunting task if the cell turns out to be non-operational.

After the basic cellular functions are put into a "minimal cell," synthetic DNA technology tenders the opportunity to introduce additional genes or pathways for use in more advanced applications. Of particular interest to metabolic engineers is the construction of a user-defined biocatalyst. Having such a tool would allow complete control over the metabolic functions and therefore would result in processes with high yields and purities. However, an efficient bioprocess also necessitates high titers and productivities [97], and therefore a robust microorganism is needed that continuously produces the compound of interest at an acceptable rate under harsh conditions [19]. Even if the genotype–phenotype map for basic cell functions (replication, transcription, translation) is eventually elucidated so as to allow design of the host for a production pathway, such maps for multigenic traits such as tolerance to stress are far from being conceived [17, 98]. Therefore, we will likely continue to rely on natural organisms for most metabolic engineering applications, at least in the short term.

Another aspect related to the aforementioned challenge – designing synthetic chromosomes with desirable content – is choosing the components so that they are in the apt cellular context. That is, the parts must be compatible with each other to carry out their functions, and compatible with the task we want to accomplish. The membrane system, the enzymes that carry out DNA replication, the transcription and translation machinery, the metabolic pathways, etc., must have the necessary interactions and work with the proper kinetics so that they give rise to the self-replicating, self-assembling factories that one intends to construct. This restricts the flexibility one has in designating the components to be introduced into the cell, and overcoming this limitation is anything but trivial. A modular approach in which groups of components are introduced, deleted, or exchanged conjunctively (e.g., the transcription module, the glycolysis module, etc.) could ameliorate incompatibility between elements. In addition, expecting that replication of synthetic cells obeys as strong a force as evolution in the same way as replication of natural cells does, fine-tuning of the designer genome can be carried out using mutation and selection.

The obvious way to begin is simply to emulate what nature already does and chemically synthesize an already sequenced genome and test it inside a cell devoid of its genome. This path has been partially explored by researchers at the J. Craig Venter Institute, who have reported the complete chemical synthesis of the *Mycoplasma* genome [16] and in a separate study have shown that transplantation of a genome from one species to another is possible [103]. The next step would be to perform a similar transplantation experiment, except now using a synthetic genome. Because the donor genome contained information from a single species, and because donor and recipient cells were closely related taxonomically, the aforementioned compatibility requirement was not a significant issue in these experiments. The real challenge lies in trying to design *de novo* a genome and transplant it into a cell so that the resulting organism has a specified phenotype. Since no one has tried to construct a designer cell, there is no way of anticipating the amount of work this will take. Nevertheless, given the implications that this would have, we believe this is an effort worth pursuing.

8 Conclusions

In nature, the information contained in nucleotide polymers is derived from previously-existing nucleotide polymers, the molecular version of the chicken-and-egg dilemma. To a first approximation, the information contained in such molecules gives rise to phenotypes that we, as biochemical and metabolic engineers, are interested in manipulating. Therefore, the study of cellular traits and the design of new ones depend on our ability to control the information contained in nucleic acid molecules (DNA in particular), a topic that has been the focus of many fields and most prominently of genetic engineering. The ultimate way in which this information can be controlled is *de novo* synthesis of DNA, a technology that has been quietly revolutionizing many practices in biotechnology. Its implications have remained unsung, maybe because alternatives for many of its uses have prevailed and maybe because, for it to be truly impactful, its cost must be friendlier. Nonetheless, we have almost certainly witnessed only the tip of the iceberg.

In order to probe this "iceberg," one might ponder how biotechnology will appear after easy, inexpensive, reliable, and precisely defined DNA synthesis arrives. The examples in this review portray an almost boundless landscape of possibilities if one is willing to assume that genotypes are the sum of discrete parts (promoters, operators, genes, terminators, etc.), and that metabolic masterpieces can be constructed by correctly assembling the parts into functioning, yet complex, modules. In this vision, host cells are the chassis in which the parts are encrusted, and the resulting machines have user-specified properties. It is too early to tell whether this vision will live up to its promises, or whether the intricacy of molecular interactions in biological systems will quickly constrain the number of tractable modifications that can be made to the host. Regardless of the outcome, synthetic DNA technology has already proved successful in easing the work of metabolic engineers and, through the arsenal of applications described here, will be partly or wholly responsible for crafting whatever the next paradigm of biotechnology turns out to be.

References

1. Raab RM, Tyo K, Stephanopoulos G (2005) Metabolic engineering. Adv Biochem Eng Biotechnol 100:1–17
2. Stephanopoulos G (1999) Metabolic fluxes and metabolic engineering. Metab Eng 1:1–11
3. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, Gerngross TU (2003) Production of complex human glycoproteins in yeast. Science 301:1244–1246
4. Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeast and filamentous fungi. Nat Biotechnol 22:1409–1414
5. Betengaugh M, Bentley W (2008) Metabolic engineering in the 21st century: meeting global challenges of sustainability and health. Curr Opin Biotechnol 19:411–413
6. Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. Nat Rev Genet 5:335–344

7. Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church GM (2004) Accurate multiplex gene synthesis from programmable DNA microchips. Nature 432:1050–1054
8. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. Science 304:1967–1971
9. Park JH, Lee SY, Kim TY, Kim HU (2008) Application of systems biology for bioprocess development. Trends Biotechnol 26:404–412
10. Benner SA, Sismour M (2005) Synthetic biology. Nat Rev Genet 6:533–543
11. Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438:443–448
12. Brown DM (1993) A brief history of oligonucleotide synthesis. In: Methods in molecular biology, vol 20. Humana, Totowa, NJ, pp 1–17
13. Itakura K, Rossi JJ, Wallace RB (1984) Synthesis and uses of synthetic oligonucleotides. Annu Rev Biochem 53:323–356
14. Ahmad S, Kamal MZ, Sankaranarayanan R, Rao NM (2008) Thermostable *Bacillus subtilis* lipases: in vitro evolution and structural insight. J Mol Biol 381:324–340
15. Alkan C, Karakoc E, Nadeau J, Sahinalp S, Zhang K (2006) RNA-RNA interaction prediction and antisense RNA target search. J Comput Biol 13:267–282
16. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA 3rd, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319: 1215–1220
17. Santos CN, Stephanopoulos G (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. Curr Opin Chem Biol 12:168–176
18. Alper H (2006) Development of systematic and combinatorial approaches for the metabolic engineering of microorganisms. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA
19. Keasling JD (2008) Synthetic biology for synthetic chemistry. ACS Chem Biol 3:64–76
20. Smolke CD, Keasling JD (2002) Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. Biotechnol Bioeng 80:762–776
21. Gebauer F, Hentze M (2004) Molecular mechanisms of translational control. Nat Rev Mol Cell Biol 5:827–835
22. Grunberg-Manago M (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. Annu Rev Genet 33:193–227
23. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361:13–37
24. Tijsterman M, Ketting R, Plasterk R (2002) The genetics of RNA silencing. Annu Rev Genet 36:489–519
25. Wray GA, Babbitt CC (2008) Enhancing gene regulation. Science 321:1300–1301
26. Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries: tuning of gene expression. Trends Biotechnol 24:53–55
27. Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci 102:12678–12683
28. 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
29. Munroe SH, Zhu J (2006) Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. Cell Mol Life Sci 63:2102–2118
30. Brink MF, Verbeet MP, de Boer HA (1995) Specialized ribosomes: highly specific translation in vivo of a single targetted mRNA species. Gene 156:215–222
31. Hui A, de Boer HA (1987) Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli. Proc Natl Acad Sci USA 84:4762–4766

32. Fargo D, Zhang M, Gillham N, Boynton J (1998) Shine-Dalgrano-like sequences are not required for translation of chloroplast mRNAs in *Chlamydomonas reinhardtii* chloroplasts or in *Escherichia coli*. Mol Gen Genet 257:271–282
33. Zuker M, Matthews D, Turner D (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark B (eds) RNA biochemistry and biotechnology: proceedings of the NATO advanced research workshop, Poznan, Poland, 10–17 October 1998. Springer, Berlin, pp 11–43
34. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431
35. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415
36. Both G, Furuichi Y, Muthukrishnan S, Shatkin A (1975) Ribosome binding to reovirus mRNA in protein synthesis requires 5′ terminal 7-methylguanosine. Cell 6:185–195
37. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292
38. Shabalina S, Ogurtsov A, Rogozin I, Koonin E, Lipman D (2004) Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. Nucleic Acids Res 32:1774–1782
39. Winkler W, Nahvi A, Roth A, Collins J, Breaker R (2004) Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428:281–286
40. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr Opin Biotechnol 6:494–500
41. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in Escherichia coli: a multi-gene study. Protein Expr Purif 59:94–102
42. Kleber-Janke T, Becker WM (2000) Use of modified BL21(DE3) Escherichia coli cells for high-level expression of recombinant peanut allergens affected by poor codon usage. Protein Expr Purif 19:419–424
43. Seidel HM, Pompliano DL, Knowles JR (1992) Phosphonate biosynthesis – molecular-cloning of the gene for phosphoenolpyruvate mutase from *Tetrahymena pyriformis* and overexpression of the gene-product in *Escherichia coli*. Biochemistry 31:2598–2608
44. Fuglsang A (2003) Codon optimizer: a freeware tool for codon optimization. Protein Expr Purif 31:247–249
45. Cai YJ, Sun J, Wang J, Ding YR, Tian N, Liao XR, Xu WB (2008) Optimizing the codon usage of synthetic gene with QPSO algorithm. J Theor Biol 254:123–127
46. Hebert C, Valdes J, Bentley W (2008) Beyond silencing – engineering applications of RNA interference and antisense technology for altering cellular phenotype. Curr Opin Biotechnol 19:500–505
47. Tomari Y, Zamore P (2005) Perspective: machines for RNAi. Genes Dev 19:517–529
48. Khvorova A, Reynolds A, Jayasena S (2003) Functional siRNAs and miRNAs exhibit strand bias. Cell 115:209–216
49. Schwarz D, Hutvágner G, Haley B, Zamore P (2002) Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell 10:537–548
50. Desai RP, Papoutsakis ET (1999) Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Appl Environ Microbiol 65:936–945
51. Nakayama S, Kosaka T, Hirakawa H, Matsuura K, Yoshino S, Furukawa K (2008) Metabolic engineering for solvent productivity by downregulation of the hydrogenase gene cluster hupCBA in *Clostridium saccharoperbutylacetonicum* strain N1-4. Appl Microbiol Biotechnol 78:483–493
52. Majors B, Betenbaugh M, Chiang G (2007) Links between metabolism and apoptosis in mammalian cells: applications for anti-apoptosis engineering. Metab Eng 9:317–326
53. Friedman S, Shah S (2006) Control of gene expression via light activated RNA interference. US Patent Application 11/148,891

54. Gredell J, Berger A, Walton S (2008) Impact of target mRNA structure on siRNA silencing efficiency: a large-scale study. Biotechnol Bioeng 100:744–755
55. Yiu S, Wong P, Lam T, Mui Y, Kung H, Lin M, Cheung Y (2005) Filtering of ineffective siRNAs and improved siRNA design tool. Bioinformatics 21:144–151
56. Widner B, Behr R, Von Dollen S, Tang M, Heu T, Sloma A, Sternberg D, DeAngelis PL, Weigel PH, Brown S (2005) Hyaluronic acid production in *Bacillus subtilis*. Appl Environ Microbiol 71:3747–3752
57. Yu H, Stephanopoulos G (2008) Metabolic engineering of *Escherichia coli* for biosynthesis of hyaluronic acid. Metab Eng 10:24–32
58. Chien LJ, Lee CK (2007) Hyaluronic acid production by recombinant *Lactococcus lactis*. Appl Microbiol Biotechnol 77:339–346
59. Gordon E, Horsefield R, Swarts HGP, de Pont J, Neutze R, Snijder A (2008) Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. Protein Expr Purif 62:1–8
60. Simon EM, Nanney DL, Doerder FP (2008) The "*Tetrahymena pyriformis*" complex of cryptic species. Biodivers Conserv 17:365–380
61. Kannan N, Taylor SS, Zhai YF, Venter JC, Manning G (2007) Structural and functional diversity of the microbial kinome. PLoS Biol 5:467–478
62. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu DY, 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 5:398–431
63. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu DY, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74
64. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li WZ, Jaroszewski L, Cieplak P, Miller CS, Li HY, Mashiyama ST, Joachimiak MP, van Belle C, Chandonia JM, Soergel DA, Zhai YF, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5:432–466
65. Lorenz P, Liebton K, Niehaus F, Eck J (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr Opin Biotechnol 13:572–577
66. Lorenz P, Eck J (2005) Metagenomics and industrial applications. Nat Rev Microbiol 3:510–516
67. Asara JM, Schweitzer MH, Freimark LM, Phillips M, Cantley LC (2007) Protein sequences from mastodon and Tyrannosaurus rex revealed by mass spectrometry. Science 316:280–285
68. Jensen PR, Hammer K (1998) Artificial promoters for metabolic optimization. Biotechnol Bioeng 58:191–195
69. Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64:82–87
70. Miksch G, Bettenworth F, Friehs K, Flaschel E, Saalbach A, Twellmann T, Nattkemper TW (2005) 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
71. Goler JA, Bramlett BW, Peccoud J (2008) Genetic design: rising above the sequence. Trends Biotechnol 26:538–544

72. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342
73. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338
74. Guet CC, Elowitz MB, Hsing WH, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296:1466–1470
75. Cai Y, Hartnett B, Gustafsson C, Peccoud J (2007) A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. Bioinformatics 23:2760–2767
76. Miyazaki K, Arnold FH (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J Mol Evol 49:716–720
77. Voigt CA, Mayo SL, Arnold FH, Wang ZG (2001) Computational method to reduce the search space for directed protein evolution. Proc Natl Acad Sci USA 98:3778–3783
78. Van den Brulle J, Fischer M, Langmann T, Horn G, Waldmann T, Arnold S, Fuhrmann M, Schatz O, O'Connell T, O'Connell D, Auckenthaler A, Schwer H (2008) A novel solid phase technology for high-throughput gene synthesis. Biotechniques 45:340–343
79. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391
80. Ostermeier M (2003) Synthetic gene libraries: in search of the optimal diversity. Trends Biotechnol 21:244–247
81. Coco WM, Encell LP, Levinson WE, Crist MJ, Loomis AK, Licato LL, Arensdorf JJ, Sica N, Pienkos PT, Monticello DJ (2002) Growth factor engineering by degenerate homoduplex gene family recombination. Nat Biotechnol 20:1246–1250
82. Ness JE, Kim S, Gottman A, Pak R, Krebber A, Borchert TV, Govindarajan S, Mundorff EC, Minshull J (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat Biotechnol 20:1251–1255
83. el Hawrani AS, Sessions RB, Moreton KM, Holbrook JJ (1996) Guided evolution of enzymes with new substrate specificities. J Mol Biol 264:97–110
84. Munir KM, French DC, Loeb LA (1993) Thymidine kinase mutants obtained by random sequence selection. Proc Natl Acad Sci USA 90:4012–4016
85. Lunzer M, Miller SP, Felsheim R, Dean AM (2005) The biochemical architecture of an ancient adaptive landscape. Science 310:499–501
86. Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440:1078–1082
87. Yoshikuni Y, Martin VJ, Ferrin TE, Keasling JD (2006) Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. Chem Biol 13:91–98
88. Ohnuma S, Narita K, Nakazawa T, Ishida C, Takeuchi Y, Ohto C, Nishino T (1996) A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. J Biol Chem 271:30748–30754
89. Solem C, Jensen PR (2002) Modulation of gene expression made easy. Appl Environ Microbiol 68:2397–2403
90. Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, Rolland S, Prieto J, Blanco FJ, Bravo J, Montoya G, Serrano L, Duchateau P, Paques F (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458
91. Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Paques F, Duchateau P (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34:e149
92. Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. Proc Natl Acad Sci USA 100:12271–12276

93. Beerli RR, Barbas CF 3rd (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20:135–141
94. Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973
95. Park KS, Jang YS, Lee H, Kim JS (2005) Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells. J Bacteriol 187:5496–5499
96. Park KS, Lee DK, Lee H, Lee Y, Jang YS, Kim YH, Yang HY, Lee SI, Seol W, Kim JS (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. Nat Biotechnol 21:1208–1214
97. Fischer CR, Klein-Marcuschamer D, Stephanopoulos G (2008) Selection and optimization of microbial hosts for biofuels production. Metab Eng 10:295–304
98. Klein-Marcuschamer D, Stephanopoulos G (2008) Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. Proc Natl Acad Sci USA 105:2319–2324
99. Smith HO, Hutchison CA 3rd, Pfannkoch C, Venter JC (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci USA 100:15440–15445
100. Forster AC, Church GM (2007) Synthetic biology projects in vitro. Genome Res 17:1–6
101. Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A (2005) Toward an artificial cell based on gene expression in vesicles. Phys Biol 2:P1–P8
102. Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2:45
103. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. Science 317:632–638
104. Yin E, Le YL, Pei JJ, Shao WL, Yang QY (2008) High-level expression of the xylanase from *Thermomyces lanuginosus* in *Escherichia coli*. World J Microbiol Biotechnol 24:275–280
105. Puigbo P, Guzman E, Romeu A, Garcia-Vallve S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35:W126–W131

Adv Biochem Engin/Biotechnol (2010) 120: 133–161
DOI: 10.1007/10_2009_62

Published online: 6 February 2010

Systems Biology of Recombinant Protein Production in *Bacillus megaterium*

Rebekka Biedendieck, Boyke Bunk, Tobias Fürch, Ezequiel Franco-Lara, Martina Jahn, and Dieter Jahn

Abstract Over the last two decades the Gram-positive bacterium *Bacillus megaterium* was systematically developed to a useful alternative protein production host. Multiple vector systems for high yield intra- and extracellular protein production were constructed. Strong inducible promoters were combined with DNA sequences for optimised ribosome binding sites, various leader peptides for protein export and N- as well as C-terminal affinity tags for affinity chromatographic purification of the desired protein. High cell density cultivation and recombinant protein production were successfully tested. For further system biology based control and optimisation of the production process the genomes of two *B. megaterium* strains were completely elucidated, DNA arrays designed, proteome, fluxome and metabolome analyses performed and all data integrated using the bioinformatics platform MEGABAC. Now, solid theoretical and experimental bases for primary modeling attempts of the production process are available.

Keywords Biotechnological *Bacillus megaterium*, Metabolic flux analysis, Protein production and secretion

R. Biedendieck
Protein Science Group, Department of Biosciences, University of Kent, Canterbury, Kent CT27NJ UK

B. Bunk, M. Jahn, and D. Jahn (✉)
Institute of Microbiology, Technische Universität Braunschweig, Spielmannstrasse 7, 38106 Braunschweig, Germany
e-mail: d.jahn@tu-bs.de

T. Fürch and E. Franco-Lara
Institute of Biochemical Engineering, Technische Universität Braunschweig, Gaußstrasse 17, 38106 Braunschweig, Germany

Contents

1 Introduction: What is *Bacillus megaterium*? 134
2 Tools for Recombinant Protein Production with *Bacillus megaterium* 136
2.1 Xylose-Inducible Promoter System 136
2.2 Expression Vectors for the Intracellular Production of Affinity Tagged Recombinant Proteins 137
2.3 Intracellular Recombinant Production and Affinity Chromatographic Purification of Proteins 139
2.4 High Cell Density Cultivations of GFP Producing Bacillus megaterium 141
2.5 Expression Vectors for the Export of Recombinant Proteins into the Growth Medium 143
2.6 Production, Export and Affinity Chromatographic Purification of Recombinant Proteins from the Growth Medium 147
2.7 Development of Multiple Vector Systems 148
2.8 T7 RNA Polymerase Dependent Promoter System 149
2.9 Culture Heterogeneity During Intracellular Protein Production 151
3 Towards Systems Biology 152
3.1 The Genome Sequences of Two *Bacillus megaterium* Strains 152
3.2 Metabolic Flux Analysis in *Bacillus megaterium* 153
4 Systems Biology with *Bacillus megaterium*? 158
References 159

1 Introduction: What is *Bacillus megaterium*?

The Gram-positive bacterium *Bacillus megaterium* was first described by de Bary more than a century ago [1]. With its eponymous size of up to 2 by 5 μm, this microorganism belongs to the larger bacteria (Fig. 1). Due to the dimension of its vegetative form and spores, *B. megaterium* is well suited for cell biological research, the investigation of cell-wall and cytoplasmic membrane formation, sporulation, cytoskeleton and chromosome separation [2]. In the 1960s, prior to *Bacillus subtilis*, *B. megaterium* was the "model organism" for intensive studies on sporulation.

Primarily a soil bacterium, *B. megaterium* is also found in diverse environments from rice paddies to dried food, seawater, sediments, fish and even in bee honey [3]. Taxonomically, *B. megaterium* was placed into the *B. subtilis* group of Bacilli during the 1990s [3, 4] although clear cut differences were found in the genome structure between *B. megaterium* and *B. subtilis*.

Due to its application for the production of several biotechnological relevant substances, the apathogenic *B. megaterium* is of general interest for the biotechnological industry. In contrast to Gram-negative organisms like *Escherichia coli*, it lacks endotoxins associated with the outer membrane which makes *B. megaterium* well applicable in food and even in pharmaceutical production processes. The utilisation of a wide variety of carbon sources lets this organism grow on low

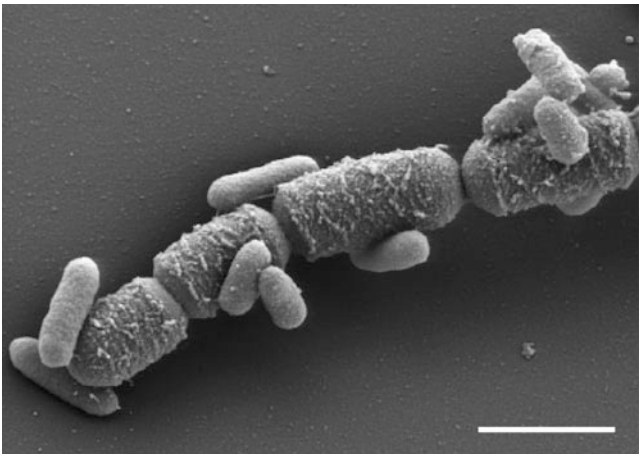


Fig. 1 Electron microscope image of vegetative cells of *B. megaterium* and *E. coli*. Aerobic cultures of *B. megaterium* and *E. coli* were separately grown in LB medium at 37°C until reaching the stationary phase. Then, both cultures were mixed in the ration of 1:1. While *B. megaterium* cells grow up to a volume of more than 60 μm^3 (2.5 × 2.5 × 10), *E. coli* cells reached a volume of 0.5 μm^3 (0.5 × 0.5 × 2), indicating an at least up to 100-fold higher volume of the *B. megaterium* cells. Aldehyde-fixed bacteria were dehydrated with a graded series of acetone, critical-point-dried with liquid CO_2, and sputter-coated with gold. Samples were examined in a field emission scanning electron microscope (FESEM) Zeiss DSM982 Gemini at an acceleration voltage of 5 kV using the Everhart-Thronley SE-detector and the SE-Inlens-detector in a 50:50 ratio. Magnification ×15,000; *white bar*: 2 μm (picture adapted from [94])

cost substances. It does not possess any obvious alkaline proteases degrading recombinant proteins. Therefore, a large amount of intact functional proteins with little or no degradation products can be obtained [5, 6]. Another advantage is its ability to secrete proteins directly into the growth medium [7]. To name only a few products of high industrial importance, *B. megaterium* is used for the production of several α- and β-amylases which are used for starch modification in the baking industry and of penicillin acylases essential for the synthesis of novel β-lactam antibiotics [8]. Moreover, *B. megaterium* is known for its ability to synthesise vitamin B_{12} aerobically and anaerobically. Various directed optimisation strategies for the enhancement of *B. megaterium* vitamin B_{12} biosynthesis were successfully employed [9].

Several *B. megaterium* strains carry significant parts of their genetic material on plasmids [3]. *B. megaterium* strain QM B1551 (ATCC12872) harbours seven indigenous plasmids of which the DNA-sequence is also known. *B. megaterium* strain PV361, a derivative of strain QM B1551 [10], lacks all seven plasmids. For biotechnological applications and research, plasmidless strains are generally used to prevent plasmid incompatibilities. A typically used strain is DSM319 of which the genome sequence was solved. In comparison to *B. subtilis*, *B. megaterium* is

known for its ability to replicate stably and maintain recombinant plasmids [3, 11]. Moreover, an efficient protoplast transformation system does exist [5, 12].

2 Tools for Recombinant Protein Production with *Bacillus megaterium*

Over several decades, multiple vector systems for recombinant gene expression in *E. coli* have been developed. Modern vectors suitable for recombinant protein production vary in the used promoter system in the presence or absence of coding sequences for affinity tags upstream or downstream of the multiple cloning site (MCS) and of sequences coding for leader peptides for the protein export. Moreover, different origins of replication (ori), antibiotic selection marker genes and MCS are used.

2.1 Xylose-Inducible Promoter System

In the 1990s, Rygus and Hillen identified the xylose-inducible promoter P_{xylA} with the according repressor protein XylR in the genome of *B. megaterium* strain DSM319 [13]. The P_{xylA} promoter is located upstream of an operon coding for the xylose isomerase XylA, the xylulokinase XylB and the xylose permease XylT, all necessary for xylose utilisation. The gene encoding the repressor protein XylR is located divergently oriented upstream of the operon with the promoter regions of *xylR* and of the *xylABT*-operon overlapping (Fig. 2a). The regulation of the *xylABT*-operon expression occurs on the transcriptional level. In the absence of xylose, XylR binds to the two tandem overlapping operator sequences located in P_{xylA} and prevents transcription of the *xylABT*-operon [14, 15]. In the presence of xylose, the sugar binds to the repressor XylR. This results in a conformational change of XylR which prevents promoter binding. In this case, the RNA-polymerase is able to recognise the promoter and initiates gene expression. An additional level of regulation is mediated by glucose [14]. A catabolite response element (*cre*) sequence is localised in the *xylA* open reading frame from bp 23 to 200 [16]. This *cis*-active *cre* element and the *trans*-active catabolite controlled protein (CcpA) are the bases for catabolite regulation in *B. megaterium*. In the presence of glucose, HPr, a phosphocarrier protein of the phosphoenolpyruvate:glucose phosphotransferase system (PTS), is phosphorylated at Ser-46 which enhances CcpA-binding to the *cre* sequence [17]. In the presence of xylose and glucose as sole energy sources, *B. megaterium* shows diauxic growth. As long as glucose is consumed, expression of the *xylABT*-operon is inhibited by CcpA binding to the *cre* sequence. When glucose becomes exhausted, the organism is able to switch from glucose to xylose consumption and enters a second logarithmic growth phase [16].

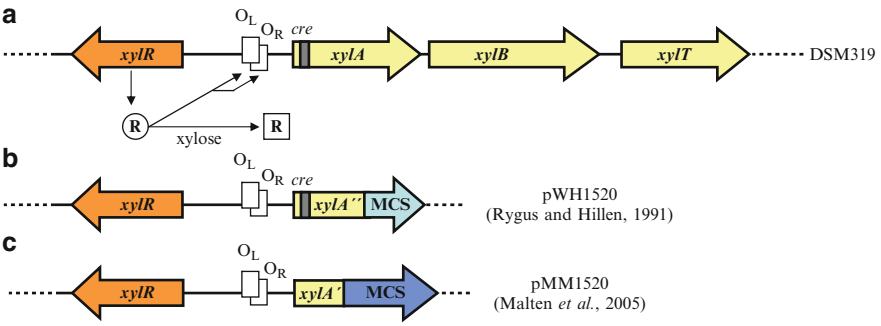


Fig. 2 Genetic organisation of the *xylABT*-operon from *B. megaterium* and the control elements of the *B. megaterium* expression vectors pWH1520 and pMM1520. The genes involved in the xylose utilisation in *B. megaterium* are indicated by *yellow arrows*. The *xylR* gene (*orange*) encodes the xylose repressor protein. *xylA, xylB* and *xylT* (*yellow*), which are co-transcribed from the xylose-inducible promoter P_{xylA}, encode the xylose isomerase, the xylulokinase and the xylose permease. The *xyl* operator region is denoted by O_LO_R (*white boxes*) indicating the two overlapping binding sites for XylR [15]. The *cre*-element (*dark grey boxes*) indicates the catabolite response element. (**a**) Genetic organisation of the *xylABT*-operon in *B. megaterium* strain DSM319. Under gene repressive conditions in the absence of xylose, XylR is an operator binding repressor protein (*circles*). In the presence of xylose the inducer interacts with XylR. This results in a conformation (*squares*) of XylR unable to bind to the operator [95, 96]. (**b**) Genetic organisation of the *xylABT*-operon elements encoded on pWH1520 [11]. The first 195 bp of *xylA*, denoted as *xylA″*, are fused to a short DNA sequence encoding the multiple cloning site (MCS, *light blue*). The operator sequences O_LO_R and the *cre*-box in *xylA″* are still present. (**c**) Genetic organisation of the *xylABT*-operon elements encoded on pMM1520 [19]. The first 15 bp of *xylA*, denoted as *xylA′*, are fused to a short DNA sequence encoding an enhanced MCS (*blue*). The operator sequences O_LO_R are still present while the *cre*-box was eliminated

2.2 *Expression Vectors for the Intracellular Production of Affinity Tagged Recombinant Proteins*

Based on the xylose-inducible promoter, Rygus and Hillen developed a xylose-dependent plasmid-borne system for the overproduction of recombinant proteins [11]. The plasmid pWH1520 encoded XylR, the promoter P_{xylA} and the first 195 bp of the *xylA* gene followed by an MCS (Fig. 2b). Further, it contains elements for replication and selection in *E. coli* which are the origin of replication *colE1* and the *β-lactamase* gene. For replication and selection in *B. megaterium* the origin of replication *oriU*, the *repU* and a tetracycline resistance gene are present. The plasmid pWH1520 was successfully used for the recombinant intracellular production of prokaryotic and eukaryotic proteins including *E. coli* β-galactosidase, *B. megaterium* glucose dehydrogenase, *Acinetobacter calcoaceticus* mutarostase, human urokinase-like plasminogen activator [11] and *Clostridium difficile* toxin A [18]. Toxin A, with a high relative molecular mass of 308,000, was produced in higher amounts compared to a recombinant production in *E. coli*.

In further studies, the promoter region of *xylA* in pWH1520 was modified (Fig. 2b, c). The *cre* sequence was eliminated and an improved MCS was inserted [19]. The resulting plasmid pMM1520 allows simple cloning of target genes by the use of 15 different DNA restriction enzyme cleavage sites located in the new designed MCS following the first 15 bp of *xylA* (Fig. 2c). To upgrade this promoter system, a new series of *B. megaterium* expression vectors for the intracellular production of recombinant proteins carrying His_6- or StrepII-tags, enabling affinity chromatography based purification, was constructed [20]. First, an additional *Bsr*GI site was introduced between ribosome binding site (RBS) and start codon of *xylA* for cloning target genes with their original start codons [20]. Further, a stop codon introduced downstream of the MCS allows for the cloning of the genes of interest without their own original stop codon (Figs. 3 and 4a). For the removal of genetic ballast, the resulting vector pSTOP1522 (Fig. 3) was liberated from a DNA fragment comprising a partial and inactive tetracycline resistance gene. This resulted in a family of plasmids all carrying the number 1622 within their names (Fig. 4). Then, two different affinity tags, the StrepII-tag [21] and the polyhistidine (His_6-) tag [22], were chosen as N- or C-terminal fusion partners for target proteins. They are encoded on the novel expression plasmids either upstream or downstream of the 63 bp open reading frame (ORF) including the MCS allowing in parallel

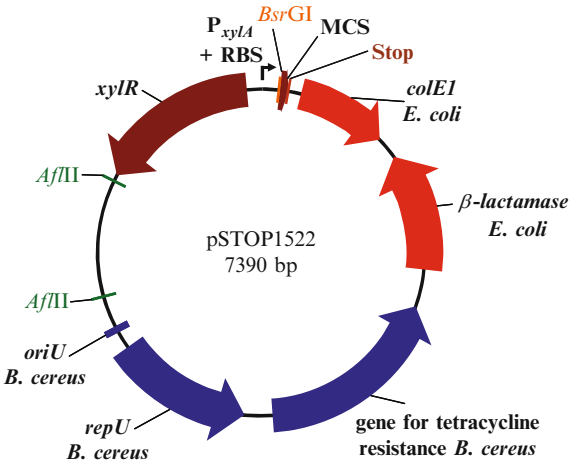


Fig. 3 Structure of the *B. megaterium* expression vector pSTOP1522. pSTOP1522 is based on the shuttle vector pMM1520 [19]. The additional *Bsr*GI site between ribosome binding site (RBS) and multiple cloning site (MCS) is marked in *orange*, the stop codon downstream of the MCS in *brown*. Elements for xylose-inducible recombinant gene expression in *B. megaterium* are the xylose-inducible promoter P_{xylA} and the gene encoding the xylose repressor (*xylR*, indicated in *brown*). Elements for plasmid replication in *Bacillus* sp. are *oriU* representing the origin of plasmid replication, *repU*, a gene essential for plasmid replication and the tetracycline resistance gene (all indicated in *blue*). Elements for plasmid replication in *E. coli* are the origin of replication *colE1* and the *β-lactamase* gene, responsible for the ampicillin resistance (both indicated in *red*). The two *Afl*II endonuclease restriction sites (in *green*) were used to create the vector series 1622 reduced in size by 855 bp

cloning of the gene of interest into all constructs. For some of the resulting fusion proteins, the corresponding N-terminal affinity tag may be removed via protease digestion, either by the tobacco etch virus (TEV) [23] or the factor Xa protease [24]. Both proteases are suited for the generation of an N-terminus with not more than one additional amino acid.

2.3 Intracellular Recombinant Production and Affinity Chromatographic Purification of Proteins

To investigate the intracellular recombinant protein production by *B. megaterium* using the newly constructed vector series (Fig. 4), the enhanced form of the green fluorescent protein (eGFP, here referred to as GFP) from the jellyfish *Aquorea victoria* was chosen as model protein [25]. Successful cloning of *gfp* is already detectable in *E. coli* via their green colour due to a leakiness of the *xylA* promoter in the cloning host [26]. Further, the *gfp* gene has a codon adaptation index of 0.45 for *B. megaterium* which is well suited for expression in *B. megaterium* [27]. Due to a linear correlation between the fluorescence intensity and the protein amount [28], intracellular formed GFP can be quantified via fluorescence spectroscopy without cell disruption. This allows for online analysis of protein production [29]. The *gfp* gene was cloned in parallel into all new expression vectors (Fig. 4) [30]. All tested *B. megaterium* strains carrying the various plasmids showed comparable growth profiles with a maximal optical density ($OD_{578\ nm}$) of 4.0. Three distinct levels of GFP production were observed measured by luminescence spectrometer. The lowest production of GFP was observed for Strep-Xa-GFP (6.3 mg g_{CDW}^{-1}) and GFP-His (6.8 mg g_{CDW}^{-1}). An approximately 1.6-fold higher production was observed for GFP without tag (9.3 mg g_{CDW}^{-1}), Strep-TEV-GFP (10.5 mg g_{CDW}^{-1}) and GFP-Strep (11.2 mg g_{CDW}^{-1}). His-TEV-GFP (14.1 mg g_{CDW}^{-1}) was produced 2.2 times more than Strep-Xa-GFP and GFP-His. SDS-PAGE analysis of the soluble and insoluble fraction of intracellular proteins indicated that most of the recombinant GFP was found in the soluble fraction (>90% of total GFP). Only Strep-TEV-GFP (<25%) and His-TEV-GFP (<30%) were partly detected in the insoluble protein fraction (Table 1). Next step was the purification of the recombinant model protein via affinity chromatography. The amounts of purified GFP reached from 1.5 mg g_{CDW}^{-1} (GFP-His) to 6.9 mg g_{CDW}^{-1} (GFP-Strep) (Table 1) [30]. Finally, the tags were successfully cleaved off the N-terminus of the purified protein using either the TEV or the factor Xa protease (Fig. 5a). To remove the proteases, the His_6-GFP/protease solution was applied to Ni-NTA affinity chromatography. As a control, GFP still fused to its N-terminal His_6-tag was used. As expected, the still affinity tagged GFP was found bound to the affinity material. It was only released from the column material by the addition of imidazole and found in the elution fractions (Fig. 5b). In contrast, processed GFP without His_6-tag did not bind. After incubation with the tagged TEV protease, GFP was mainly found in the flow-through indicating the successful cleavage of the His_6-tag of the GFP

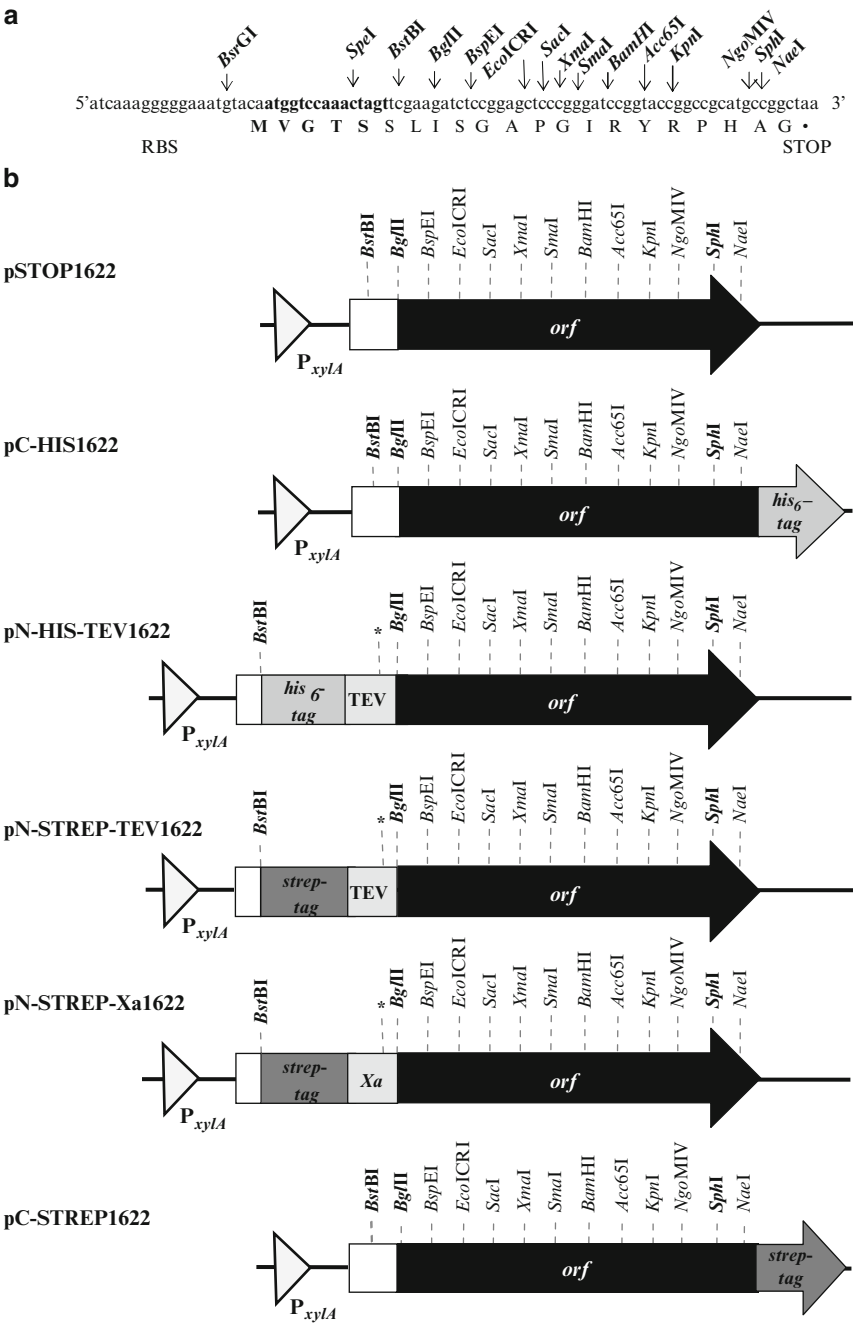


Fig. 4 Plasmids for the intracellular production of tagged proteins in *B. megaterium*. All plasmids of the series 1622 are based on the shuttle vector pSTOP1522 (Fig. 3). (**a**) DNA sequence of ribosome binding site (RBS) and multiple cloning site (MCS) of the expression plasmid pSTOP1622. The

(Fig. 5b, FT). The tagged protease as well as the cleaved off His_6-tag were found bound to the column.

2.4 *High Cell Density Cultivations of GFP Producing* Bacillus megaterium

After the establishment of GFP production at shake-flask scale, the process was transferred into a bioreactor to study GFP production in a glucose limited fed-batch cultivation. Under these conditions high cell densities of up to 80 g L^{-1} were reached [19, 31]. *B. megaterium* strain WH323 carrying pRBBm56 encoding GFP-Strep was chosen for these experiments, since it produced high amounts of GFP-Strep fusion protein. A5 medium, a minimal medium containing 0.5 g of yeast extract and 30 g L^{-1} of glucose, was used for a 1 L batch cultivation [19]. About 2 h prior to the end of the batch phase (11.2 h) recombinant gene expression was induced by the addition of 0.5% (w/v) xylose leading to the production of GFP-Strep which was measured via its fluorescence. After exhaustion of the glucose (13.3 h), further glucose was fed exponentially into the growing culture theoretically setting the growth rate (μ_{set}) to 0.12 h^{-1}. The feeding formula did not account for the maintenance metabolism [19]. Hence, an actual μ_{set} of 0.08 h^{-1} was

Table 1 Comparison of the recombinant production and affinity chromatographic purification of GFP from *B. megaterium* [30]. Different affinity tag fusion forms of GFP were produced in *B. megaterium* WH323. Purification was performed using affinity chromatography. Amounts of purified GFP-Strep were determined using a Bradford protein assay kit (Bio-Rad; Munich; Germany) and BSA (Perbio; Rockford; USA) as standard. Amounts of purified GFP-His, His-TEV-GFP, Strep-Xa-GFP and Strep-TEV-GFP were calculated via their relative fluorescence per mg protein

Protein	Protein produced		Protein purified	
	(mg L^{-1} cell culture)	(mg g_{CDW}^{-1})	(mg L^{-1} cell culture)	(mg g_{CDW}^{-1})
GFP-His	9.6 ± 0.5	6.8 ± 0.3	2.4 ± 0.1	1.5 ± 0.1
His-TEV-GFP	17.9 ± 0.9	14.0 ± 0.7	5.0 ± 0.3	3.0 ± 0.2
Strep-Xa-GFP	8.4 ± 0.4	6.3 ± 0.3	6.0 ± 0.3	4.0 ± 0.2
Strep-TEV-GFP	13.2 ± 0.7	10.5 ± 0.5	9.0 ± 0.5	6.0 ± 0.3
GFP-Strep	16.0 ± 0.8	11.2 ± 0.6	10.8 ± 0.5	6.9 ± 0.4

Fig. 4 (continued) coding sequence of an open reading frame (ORF) comprising the MCS is given. The sequence of the first five amino acids of XylA is marked in *bold*. All restriction sites of the MCS are indicated. (**b**) Scheme of expression plasmids based on pSTOP1622. All expression plasmids shown allow parallel cloning of genes of interest into the identical MCS from *Bgl*II (marked in *bold*) to *Nae*I. Restriction site *Nar*I is indicated by an *asterisk*. P_{xylA}: promoter of *xylA*; TEV: tobacco etch virus protease cleavage site; Xa: factor Xa protease cleavage site (adapted from [19])

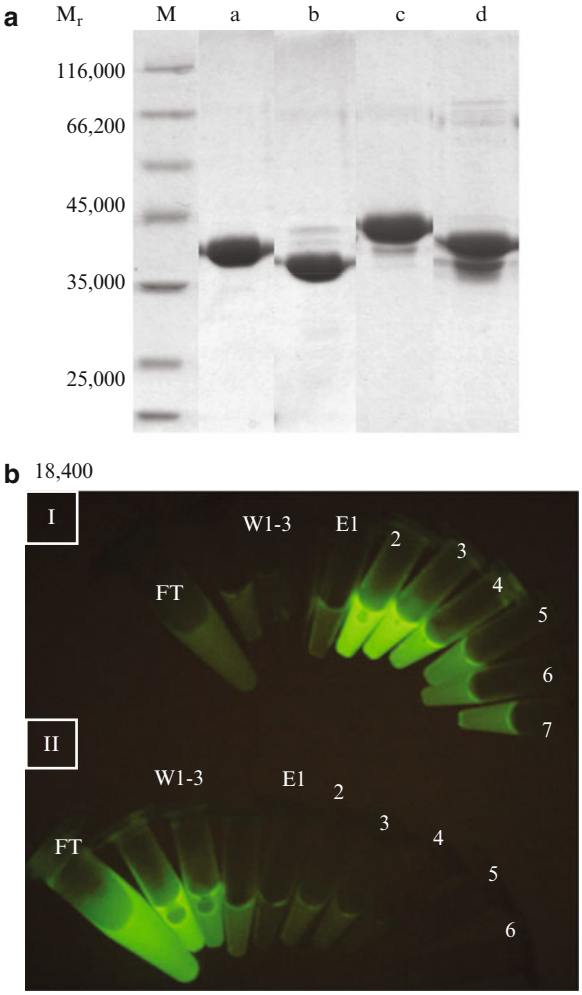


Fig. 5 Removal of the affinity tags from the GFP fusion proteins. After purification of Strep-Xa-GFP, Strep-TEV-GFP and His-TEV-GFP the affinity tags were removed. (**a**) Protease cleavage of Strep-Xa-GFP and Strep-TEV-GFP. Purified Strep-Xa-GFP and Strep-TEV-GFP were incubated with factor Xa and TEV protease, respectively. Strep-Xa-GFP before (lane 1) and after (lane 2) factor Xa protease addition and Strep-TEV-GFP before (lane 3) and after (lane 4) TEV protease addition were separated via SDS-PAGE. Lane M: Protein Molecular Weight Marker (MBI Fermentas; St. Leon-Rot; Germany). (**b**) Purification of His-TEV-GFP using a Ni-NTA affinity chromatography column. His-TEV-GFP was incubated with or without His_6-tagged TEV protease overnight. The total volume was loaded onto a Ni-NTA affinity column. (**bI**) Without protease addition, the His-TEV-GFP bound to the affinity material is shown. After washing (W1-3), the elution with imidazole released the His-TEV-GFP from the affinity material (E1-7). (**bII**) After addition of TEV protease, GFP without His_6-tag did not bind to the Ni-NTA affinity column. It was found in the flow-through (FT) and in the washing fractions (W1-3). The elution fractions showed less (E1-3) or no (E4-6) fluorescence. GFP was visualised on a light table emitting blue light and a yellow filter in front of the camera lens (adapted from [19])

obtained. During this initial stage, the amount of GFP increased from zero up to 11.2 mg of GFP per g_{CDW} and remained constant for the next 9 h. In the meantime, the biomass concentration further increased in parallel to the cell density resulting in a maximum of the volumetric GFP concentration of 226 mg L^{-1} of GFP at 25 h and 20 g L^{-1} of CDW. When specific GFP production rates were calculated, it was observed that the highest specific production rates of 3.5 mg of GFP per g_{CDW} and per hour were reached directly after xylose addition (11.2 h). In contrast, volumetric production rates stayed constant at a level of approximately 24 mg of GFP per liter and per hour for 10 h after induction.

In a second fed-batch cultivation, a different gene induction profile was tested. In order to obtain a better yield, GFP production was induced at a higher biomass concentration. Here, the batch cultivation was started with 4 g L^{-1} of glucose trying to avoid the formation of organic acids in the batch phase as observed in the case of early induction. The exponential feeding was started directly after the exhaustion of glucose with μ_{set} of 0.14 h^{-1} resulting in an actual μ_{set} of 0.11 h^{-1}. Then 5 g L^{-1} of xylose were added at 25 g L^{-1} of CDW (31.7 h) for induction of the *gfp* gene expression. The onset of GFP production quickly led to 4.8 mg of GFP per g_{CDW}. However, intracellular GFP amounts decreased in the next 2 h to 3.4 mg of GFP per g_{CDW}. A second xylose addition of 5 g L^{-1} increased the cellular GFP amount to 5.2 mg of GFP per g_{CDW} and also led to a maximal volumetric GFP amount of 274 mg L^{-1} of GFP at a biomass concentration of 52 g L^{-1} of CDW (38.2 h). The calculation of specific GFP production rates demonstrated its rapid increase after the first and second induction to maximal values of 4.7 and 1.0 mg of GFP per g_{CDW} and h, respectively.

2.5 Expression Vectors for the Export of Recombinant Proteins into the Growth Medium

Initially several naturally secreted *B. megaterium* proteins were characterised for their export efficiency. However, little research was done on the secretion of heterologous proteins. Later, *B. megaterium* gained interest as production host for polysaccharide degrading enzymes like glucanases due to its natural glucanase deficiency. Early work described the expression of an endo-β-1,4-glucanase gene from *B. subtilis* in *B. megaterium* under control of its own promoter [32]. The resulting enzyme was secreted into the growth medium. In 2003, a similar report was published on the secretion of a *Bacillus circulans* endo-β-1,3-1,4-glucanase in *B. megaterium* and for comparison in *B. subtilis*. At that time the recombinant production in *B. subtilis* yielded more protein. However, as demonstrated before, plasmid stability remained a problem for *B. subtilis* [6].

Only a few years ago, first studies were started using the xylose-inducible plasmid-borne gene expression system for the production and secretion of recombinant proteins by *B. megaterium*. The production and secretion of the high molecular mass dextransucrase DsrS (M_r = 188,000) from *Leuconostoc mesenteroides*

were investigated and systematically optimised using the original DsrS signal peptide [33]. After successful production and secretion of DsrS into the growth medium of *B. megaterium* strain DSM319, the use of a protease deficient *B. megaterium* strain MS941 (Δ*nprM*) further enhanced the secretion and extracellular stability of this enzyme [34]. High cell density cultivations (HCDCs) of *B. megaterium* resulted in up to 28,600 U L^{-1} of recombinant DsrS [19].

Based on the vectors for the intracellular production and purification of recombinant proteins, the vector system was further expanded with vectors for the extracellular recombinant protein production. Affinity tag aided protein purification from the cell-free growth medium was made possible by the addition of the corresponding His_6- and StrepII-tags [20, 35].

For the secretion of heterologous proteins, a vector encoding the functional signal peptide of the extracellular *B. megaterium* esterase LipA was constructed allowing a translational fusion of the gene of interest to this leader peptide encoding sequence. As was found earlier, the esterase LipA was found secreted in high amounts into the culture medium of *B. megaterium* [36]. The LipA signal peptide contains 28 amino acids. The resulting fusion protein is destined for the transport via the SEC-pathway into the growth medium [37]. The three typical regions of a SEC pathway-dependent signal peptide were identified in the LipA signal peptide by the SignalP algorithm [38]. Its cleavage site for the signal peptidase type I consists of an AKA motif. The vector pMM1525 incorporating all mentioned features (Fig. 6) allows the

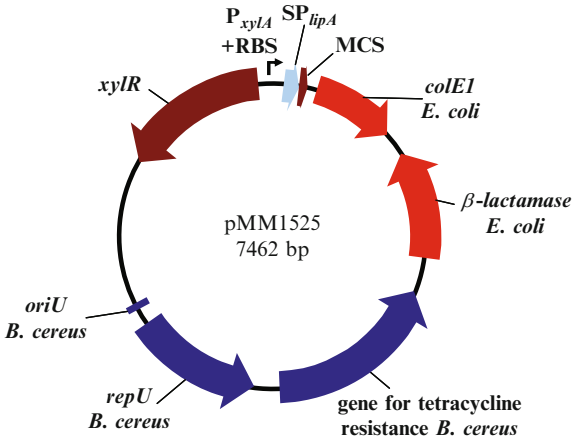


Fig. 6 Structure of the *B. megaterium* expression vector pMM1525. Vector pMM1525 is based on the shuttle vector pMM1522. The DNA sequence of the signal peptide of the extracellular *B. megaterium* esterase LipA SP_{lipA} was inserted into the *Bsr*GI/*Bst*BI digested pMM1522 using the methods of annealed synthetic oligonucleotides. The MCS is located downstream of SP_{lipA} (marked in *light blue*). Elements for xylose-inducible recombinant gene expression in *B. megaterium* are the inducible promoter (P_{xylA}) and the gene for the xylose repressor (*xylR*, marked in *brown*). Elements for plasmid replication in *Bacillus* sp. are the *oriU* representing the origin of plasmid replication, *repU*, a gene essential for plasmid replication and the tetracycline resistance gene (all marked in *blue*). Elements for plasmid replication in *E. coli* are the origin of replication *colE1* and the ampicillin resistance gene *β-lactamase* (both marked in *red*)

xylose-inducible production and secretion of proteins of interest in *B. megaterium* [20]. Under control of the promoter P_{xylA}, it encodes the LipA signal peptide followed by an MCS for the in frame insertion of genes of interest.

Further, the two affinity tags, His_6-tag and StrepII-tag, were chosen for the purification of a recombinant protein directly from the growth medium of *B. megaterium*. These new secretion vectors we named pHIS1522 and pSTREP1525 (Fig. 7). The N-terminal StrepII-tags can be removed using the also inserted factor Xa protease cleavage site. The affinity tag coding sequence of the two secretion vectors pHIS1525 and pSTREP1525 were combined leading to pSTREPHIS1525, an expression plasmid for the extracellular production of StrepII-His_6-tagged proteins (Fig. 7). Convenient parallel cloning is possible into all described vectors, when genes of interest are cloned into the identical restriction endonuclease sites of the MCS (*Bgl*II to *Nae*I) [20, 35].

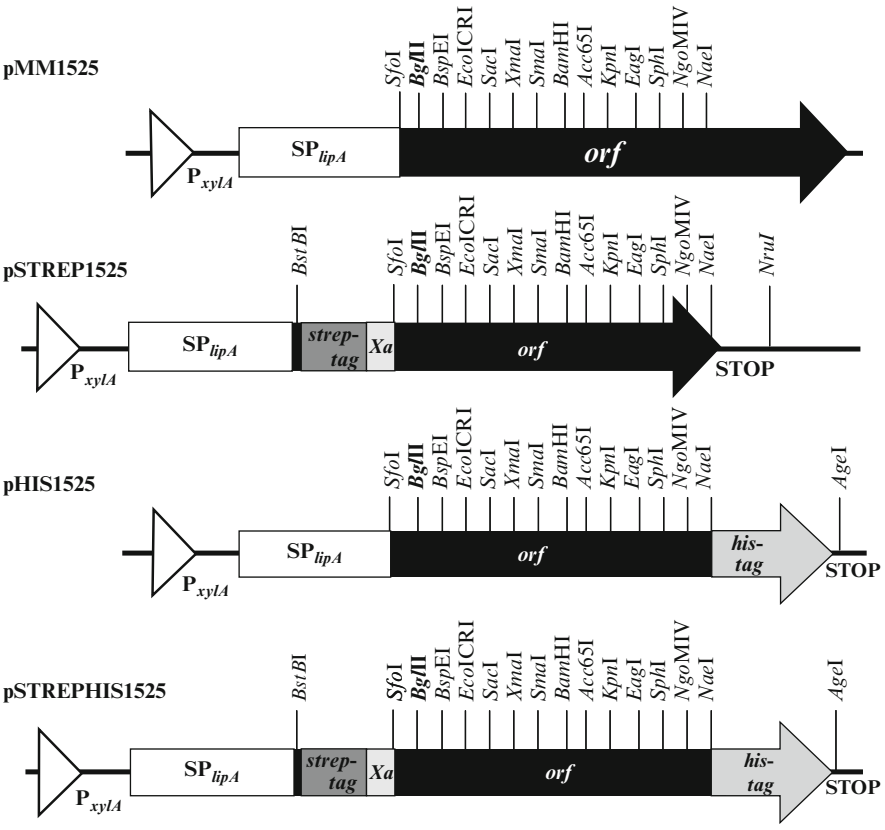


Fig. 7 Series of expression plasmids for the secretion of recombinant tagged proteins via SP_{lipA} by *B. megaterium* [20]. All plasmids are based on the shuttle vector pMM1522. The plasmids with the coding sequence for the signal peptide SP_{lipA}, their affinity tag elements and unique restriction enzyme cleavage sites are shown. All shown expression plasmids allow in parallel cloning of genes of interest into the equivalent MCS from *Bgl*II (marked in *bold*) to *Nae*I

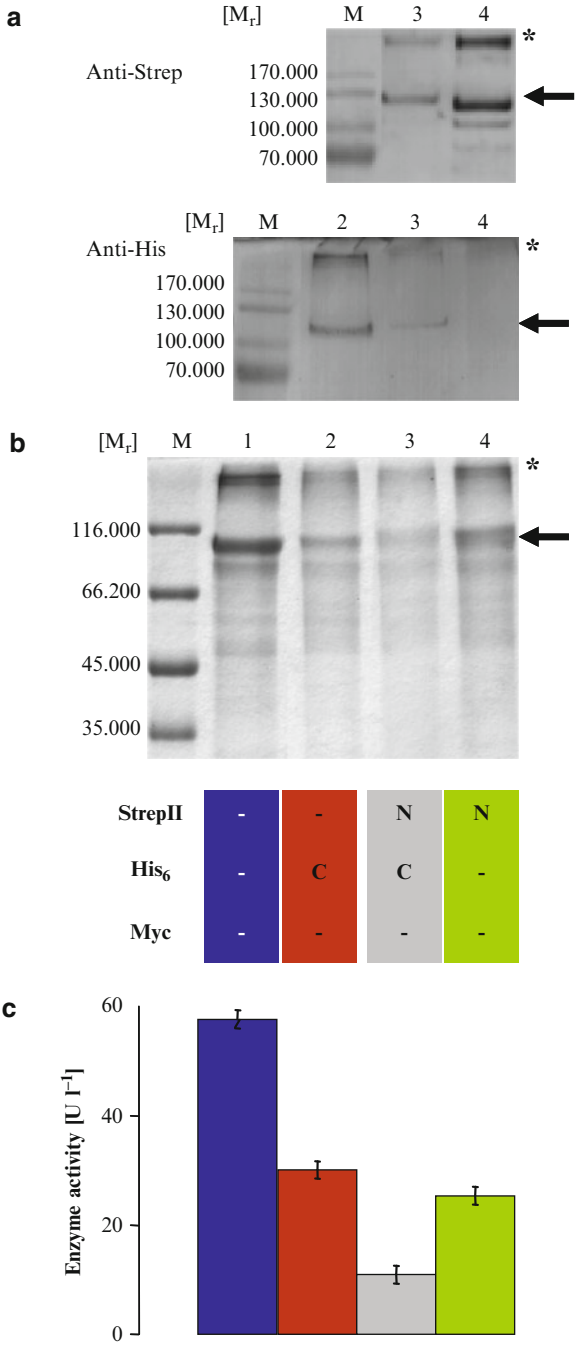


Fig. 8 Influence of affinity tags fused to levansucrase LevΔ773 on secretion by *B. megaterium* and enzyme activity. *B. megaterium* MS941 carrying the plasmids encoding LevΔ773 (lane 1),

2.6 Production, Export and Affinity Chromatographic Purification of Recombinant Proteins from the Growth Medium

For protein export studies in *B. megaterium*, the extracellular levansucrase from *Lactobacillus reuteri* strain 121, LevΔ773, lacking a C-terminal membrane spanning region, was chosen as a heterologous reporter protein [39]. The levansucrase gene expression experiments were done in shake flask experiments and the cell-free growth medium was analysed by Western blot analysis, SDS-PAGE and enzyme activity [20]. No differences were observed in the growth profiles of all *B. megaterium* plasmid carrying strains all reaching the stationary phase 3 h after induction of recombinant gene expression ($OD_{578\ nm}$ = 4.8).

The amounts of exported levansucrase by the various tested strains were visualised via SDS-PAGE (Fig. 8b). Obviously, the highest protein amount was found for the strain secreting untagged levansucrase LevΔ773. About half of the amount was observed for strains secreting LevΔ773His and StrepLevΔ773. Even lower amounts were detected for the strain producing StrepLevΔ773His. All produced enzyme variants were found active (Fig. 8). Most levansucrase activity in the growth medium was detected for the untagged form (57.6 U L^{-1}). Enzyme variants carrying one affinity tag were found exported with less efficiency. LevΔ773His and StrepLevΔ773 revealed about half of the activity of the untagged levansucrase (30.1 U L^{-1} and 25.3 U L^{-1}, respectively). The influence of small peptide tags on the secretion of heterologous proteins in bacteria has been barely studied. One report compares the secretion of the His_6-tagged with the untagged *B. licheniformis* β-lactamase in *E. coli*. Ledent et al. (1997) demonstrated that the C-terminal His_6-tag caused aberrant cleavage of the signal peptide of the investigated β-lactamase [40].

StrepLevΔ773 was found secreted in high amounts into the growth medium of recombinant *B. megaterium* MS941. Hence, the affinity chromatographic purification of StrepLevΔ773 was tested. Only small amounts of StrepLevΔ773 were purified directly from the cell free growth medium [35]. To enhance the binding to the affinity material, StrepLevΔ773 in the cell-free growth medium

Fig. 8 (continued) LevΔ773His (lane 2), StrepLevΔ773His (lane 3) and StrepLevΔ773 (lane 4) were aerobically cultivated in shaking flasks at 37°C. (**a**, **b**) Six hours after induction of recombinant gene expression proteins of cell-free growth medium were precipitated and separated by SDS-PAGE. (**a**) The precipitated proteins were blotted onto PVDF membranes and the tags fused to LevΔ773 were detected using streptavidin conjugate or anti-His-tag antibody. (**b**) The SDS-PAGE gel was stained by Coomassie Brilliant Blue G250. (**c**) The levansucrase activity in the cell-free growth medium was determined by using the D.N.S. method [97]. One unit of enzyme was defined as the release of 1 μmol glucose per min describing the transferase and hydrolase activity of LevΔ773

was concentrated by ammonium sulphate precipitation [41] prior to purification. Of the StrepII-tagged levansucrase found in the growth medium, 34% were bound to the material. After washing and elution, 0.7 mg L^{-1} StrepLevΔ773 were recovered from the growth medium [35]. Next, the C-terminal His-tagged LevΔ773 was subjected to be purified directly from the cell free growth medium. In a batch process, the chromatographic material was added to the non-concentrated cell-free growth medium containing LevΔ773His and incubated at smooth shaking to increase the contact between the tagged protein and the chromatographic affinity material. After separation of the sepharose from the growth medium and washing, proteins were eluted. This resulted in 0.9 mg L^{-1} of pure LevΔ773His in the culture medium [35]. The purified protein LevΔ773His showed a specific activity of 197.4 U mg^{-1}, which is comparable to the 138.0 U mg^{-1} obtained for LevΔ773MycHis produced in *E. coli* [39]. Here, we obtained a 3.4-fold purification with a yield of 51.5%. Hence, for the His_6-tagged levansucrase the batch application of Ni-NTA sepharose in non-concentrated cell-free growth medium was the preferred method for purification. This method was also used for the successful purification of a hydrolase from *Thermobifida fusca* which was recombinantly produced and secreted by *B. megaterium* [42].

2.7 Development of Multiple Vector Systems

In *E. coli*, multiple parallel replicating vectors are commonly used. Examples are the co-production of vector encoded tRNAs for rare codons in so called "codon plus" strains [43]. Also, the initial T7-promoter system was based on a two vector system [44]. Furthermore, *E. coli* pLysS strains, constructed for recombinant protein purification, are carrying a plasmid encoding the T7 lysozyme, a natural inhibitor of T7 RNA polymerase, which suppresses the basal expression from the T7-promoter [45]. Commercial systems for the co-production from three different plasmids are available, e.g. pACYC based vectors compatible with pET and pSC101 vectors. For stable in parallel replication of various vectors, different origins of replication (*ori*) and selection markers are required. The origin of replication *oriU* and the gene *repU* encoded by pMM1522 and all derivatives of this plasmids are derived from the plasmid pBC16 which is replicated by the rolling circle mechanism [46–48]. To utilise additional plasmids with different origins, the naturally occurring plasmids of *B. megaterium* strain QM B1551 were investigated. The seven indigenous plasmids of *B. megaterium* QM B1551 [49] belong to different compatibility classes. Two different replicons from these plasmids, *repM100* and *repM700*, were used to generate novel expression plasmids compatible with *repU* derivatives [49, 50]. The origin of replication *repM100* was first described for *B. megaterium* QM B1551 plasmid pBM100. It is replicated via a rolling circle mechanism. For the construction of pMGBm19 we used *repM100* from the plasmid pYZ5 [51, 52]. The second used replicon was *repM700* originally found in pBM700 of *B. megaterium* QM B1551.

It is a typical theta type replicon. A derivative of this plasmid named pKM700 was used to generate pMGBm21. The replicons *repU*, *repM100* and *repM700* belong to different compatibility classes. Hence, pMM1522, pMGBm19, pMGBm21 and their derivatives can be replicated in parallel by *B. megaterium*. The selection marker encoded on pMM1522 mediates tetracycline resistance while pMGBm19 and pMGBm21 are carrying a chloramphenicol resistance gene.

The *gfp* gene was inserted into pMM1522, pMGBm19 and pMGBm21, resulting in pRBBm34, pRBBm63 and pRBBm64, respectively. For all plasmid carrying strains, highest intracellular GFP amounts were measured 6 h after induction of the *gfp* gene expression. Growth medium (3.5 mg L^{-1} GFP) was produced by WH323 carrying pRBBm64 which was about four times less compared to GFP production mediated by the rolling circle plasmids pRBBm34 and pRBBm63, respectively.

2.8 T7 RNA Polymerase Dependent Promoter System

As outlined above, a two vector system was successfully developed and used for the recombinant protein production in *B. megaterium*. Based on this system, the T7 RNA polymerase (RNAP) dependant promoter system was introduced in our tool list for *B. megaterium* [51].

It was already developed in the early 1980s for *E. coli* [44] and is based on the RNA polymerase of the bacteriophage T7 (T7 RNAP). This viral RNA polymerase combines several advantageous features, such as high processivity and stringent selectivity towards its cognate promoter. Using this system for protein overproduction in *E. coli*, up to 50% or more of the total cellular proteins can consist of a desired recombinant protein [53]. Since the early 1990s, the T7 RNAP system has been adopted to different Gram-negative and -positive bacteria as well as to eukaryotic organisms [54–56].

For the recombinant *t7 rnap* expression in *B. megaterium*, *t7 rnap* was cloned under control of the xylose-inducible promoter resulting in the plasmid pT7-RNAP (Fig. 9). Further, a plasmid belonging to another compatibility class comprising the T7 RNAP promoter ϕ10 and the Tϕ transcription terminator was constructed. Additionally, this resulting plasmid pP_{T7} carries a ribosome binding site sequence optimised for use in *B. megaterium* mediating an efficient initiation of translation [57]. The introduced MCS downstream of the ϕ10 promoter with ten unique restriction enzyme cleavage sites allows for comfortable cloning of a gene of interest. The sequence of this multiple cloning site is found in all commercially available *B. megaterium* expression vectors based on the xylose-inducible promoter (MoBiTec, Göttingen, Germany) and therefore enables a simple and time-saving subcloning [51].

After confirmation of the recombinant production of the viral T7 RNA polymerase, the already introduced model protein GFP was used for the comparison of the T7 and the xylose-inducible promoter systems. Therefore, recombinant GFP

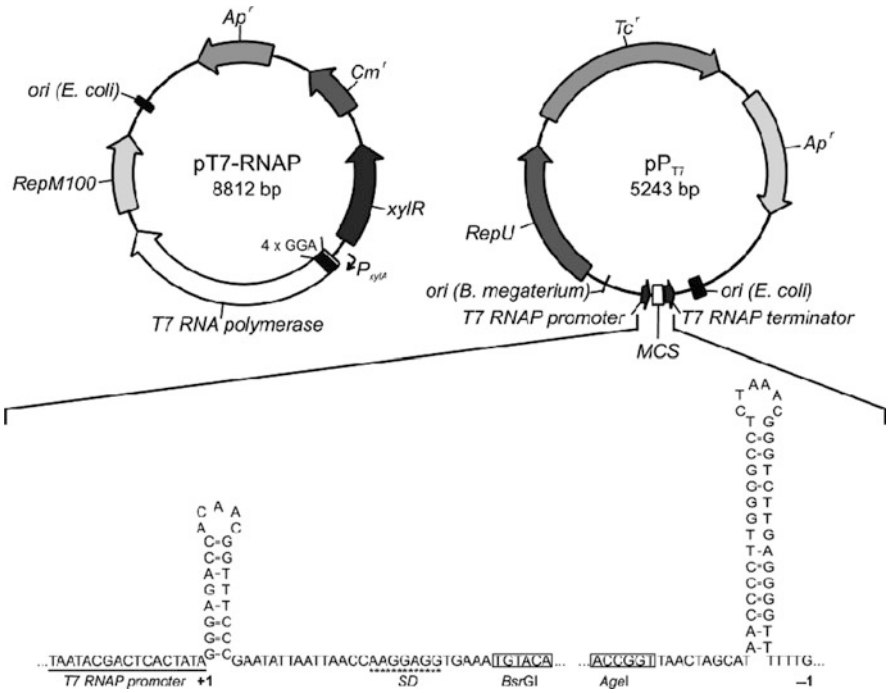


Fig. 9 The two plasmid system for T7 RNA polymerase driven protein production in *B. megaterium*. Both plasmids outlined are shuttle vectors allowing for positive selection and replication in *E. coli* and *B. megaterium*, respectively. The *Ap*^{r} gene confers resistance to ampicillin in *E. coli*. In *B. megaterium* chloramphenicol (*Cm*^{r}) can be used as selection marker for pT7-RNAP and tetracycline (*Tc*^{r}) for pP_{T7}. Due to the different nature of the replicons (*repU* and *repM100*) the plasmids can replicate in parallel in *B. megaterium*. Plasmid pT7-RNAP contains a 2,670 bp *Spe*I/*Xho*I fragment comprising *t7 rnap* and four consecutive GGA codons at the 5′-end under the transcriptional control of P_{xylA}. Plasmid pP_{T7} contains all structural elements necessary for a T7 RNAP-dependent expression of target genes. The promoter sequence of T7 RNAP is underlined and the first nucleotide (+1) of the RNA transcript is indicated [53]. The optimised Shine-Dalgarno (*SD*) sequence (marked with *stars*) is followed by the multiple cloning site (*MCS*). The first and the last unique restriction enzyme sites are shown (marked with *rectangles*). The T7 RNAP transcription terminator represented by its potential stem loop structure is located downstream of the *Age*I site and the last nucleotide of the terminated RNA transcript (−1) is indicated [51]

production with *B. megaterium* carrying pT7-RNAP and pP_{T7}-GFP was compared to those of *B. megaterium* transformed with the plasmid pRBBm34 (P_{xylA}-gfp) [58]. Using the T7 RNAP-dependent gene expression system, up to seven times more GFP (49.7 mg L^{-1}) were recombinantly produced compared to that under xylose control. Furthermore, the overall productivity was found enhanced more than six times (12.8 mg L^{-1} h^{-1} GFP) in case of using the T7 RNAP-dependent system compared to the xylose-inducible promoter (2 mg L^{-1} h^{-1} GFP) [51].

2.9 Culture Heterogeneity During Intracellular Protein Production

To study the effect of GFP production on host viability and the productivity of single *B. megaterium* cells, the fluorescence of the model protein GFP was combined with a live/dead stain using propidium iodide (PI) in a flow cytometric analysis [30]. This analysis should provide insights into the number of protein producing cells per culture and a potential population heterogeneity. *B. megaterium* cells had a slight but significant red and green auto-fluorescence visible on a green (FL1) and red (FL3) fluorescence dot plot in a flow cytometric analysis without additional staining (Fig. 10). A quadrant gating on the red-green fluorescence dot blot separated the investigated cells into four subpopulations: (1) living cells producing GFP (green), (2) living cells not producing GFP (red), (3) dead cells containing GFP (cyan), and (4) dead cells not containing GFP (black). An analysis of samples from the fed-batch HCDC with induction at 25 g L^{-1} of CDW revealed that no GFP was produced prior to induction of the *gfp* gene expression controlled by the xylose-inducible promoter P_{xylA}. Approximately 30 min after induction, 76% of all cells produced active GFP. Within the next 4 h only 3% of the GFP producing cells died. Hence, recombinant GFP production only had a minor effect on the viability of the cell population (Fig. 10). However, clear-cut culture heterogeneity

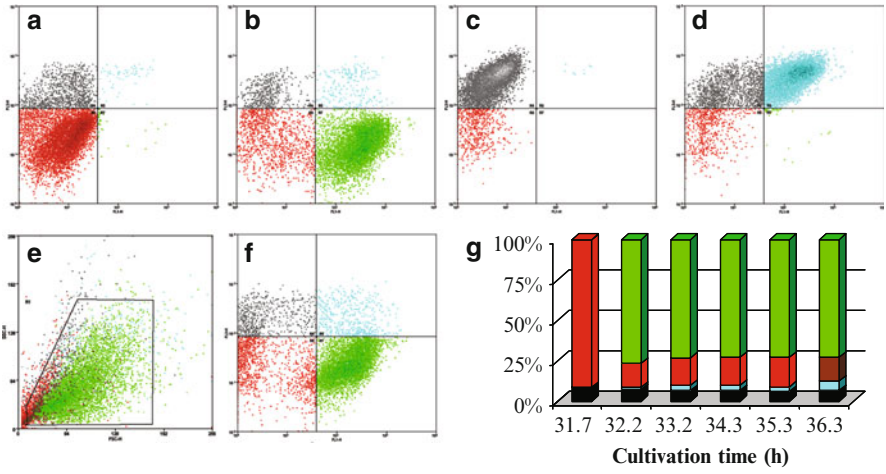


Fig. 10 Flow cytometric analysis and GFP production of cells in an high cell density fermentation. Samples taken from the fed-batch of *B. megaterium* WH323 carrying pRBBm56 (GFP-Strep) before and after induction of the *gfp* gene expression at 25 g L^{-1} of CDW were stained with propidium iodide (PI) and analysed in a FACSCalibur (Benton Dickinson; Erembodegen-Aalst; Belgium) for red (FL3) and green (FL1) fluorescence (**a–d**, **f**). As controls, fed-batch samples of living (**a**, *red*), living and GFP producing (**b**, *green*), dead (**c**, *black*) and dead and GFP producing (**d**, *cyan*) cells are shown. (**e**) The gate, used for its FL3 / FL1 dot blot. (**f**) Side (SSC) and frontal (FSC) scatter of a sample taken at 36.3 h are indicated by a box. (**g**) Accordingly coloured columns show percentage of the subpopulation compared to all cells

was observed. This might be a general phenomenon of recombinant protein production and deserves further investigation.

3 Towards Systems Biology

3.1 *The Genome Sequences of Two* Bacillus megaterium *Strains*

The complete genomic DNA sequences of the *B. megaterium* strains DSM319 and QM B1551 were determined. The chromosome of *B. megaterium* strain DSM319 consists of 5,097,447 bp and 5,299 genes, with 115 tRNAs and 11 rRNA operons. *B. megaterium* strain QM B1551 has 5,097,129 bp of genomic and 426,063 bp of plasmidic DNA. It carries 5,299 genes on its chromosome and 543 on the 7 plasmids, all including 139 tRNA genes and 12 rRNA operons. The plasmids reveal different copy numbers and comprise approximately 11% of the total cellular DNA. The size spectrum ranges from only 5.4 kb to 164.4 kb [49, 59]. One of them, plasmid pBM400 is coding for an additional rRNA operon and 18 tRNAs genes [60]. Apart from genes which products are necessary for replication and plasmid stabilisation, many of these plasmids encode genes of unknown function. Most of the known gene products are related to transposases, integrases, recombinases, glycosyltransferases and transporters. On plasmid pBM600, several genes, related to the development of additional antibiotic resistances, have been found. On the largest plasmid, pBM700, several genes involved in germination and sporulation were detected [49].

An exact and suitable genome annotation and an examination of the gene context provide a solid overview of the metabolic capabilities and regulatory strategies the organism might use to adapt to its environment. For example, experimental evidence for the oxygen-independent vitamin B_{12} biosynthesis was verified in silico due to the presence of all necessary genes. The energy metabolism can also be reconstructed from in vivo nascendi. The respiratory chain shows three types of primary dehydrogenases which are the NADH dehydrogenases Ndh, YumB and YutJ, the glycerol 3-phosphate dehydrogenase GlpD and the succinate dehydrogenase SdhCAB, respectively. Electrons are transferred from these enzymes to menaquinone, which itself passes electrons to menachinol oxidases (CydAB, YthAB), to cytochrome aa_3 (QoxABCD) as well as to the cytochrome bc_1 complex (QcrABC). Later a complex might donate electrons to cytochrome c (CccA, CccB) or to cytochrome c oxidase (CtaCDEF). Alternative anaerobic respiration systems are missing. However, *B. megaterium* was shown to grow fermentatively producing 2,3-butanediol, lactate, ethanol and acetate. The fermentation and overflow metabolism is similar to that in *B. subtilis* [61]. Compared to *B. subtilis,* a major difference has been found in the tricarboxylic acid (TCA) cycle. In *B. megaterium* the glyoxylate shunt is present evident by the presence of isocitrate lyase AceA and malate synthase GlcB.

A versatile lipid formation leading to phosphatidic acid, diglycosyl-1,2-diacylglycerol, phosphatidylethanolamine, phosphaditylglycerol and cardiolipin

(diphosphatidylglycerol) was deduced. Sporulation was found almost identical to the processes in *B. subtilis* up to the point of spore cortex and spore coat formation. The majority of *cot* and *sps* genes found in *B. subtilis* are missing in *B. megaterium*, which also explains the missing brown colour of the *B. megaterium* spores. A complete set of competence genes was detected on the *B. megaterium* genome. This might provide the basis for the establishment of natural competence needed for efficient transformation. The determined genome sequence provided the basis for the design of DNA arrays and initial transcriptional profiling. All theoretical as well as experimental data combined using the bioinformatics platform MEGABAC (http://megabac.tu-bs.de)[1].

3.2 *Metabolic Flux Analysis in* Bacillus megaterium

3.2.1 Fundamentals of Metabolic Flux Analysis: Metabolite Balancing

Metabolic flux analysis is one of the most powerful analytical and experimental tools used for physiological characterisation of cell metabolism. In its most basic form, the method is essentially based on the conservation principles used for macrochemical and biological systems applied to the internal environment of cellular systems. The fundamental equation of MFA considers the steady-state mass balances around all intracellular metabolic intermediates such that

$$\frac{\mathrm{d}\underline{X}}{\mathrm{d}t} = \underline{\underline{N}}^T \underline{v} \approx \underline{0}, \tag{1}$$

where $\underline{X} = (X_1, \ldots, X_i, \ldots, X_I)^T$ represents a vector containing all intracellular metabolite concentrations, $\underline{v} = (v_1, \ldots, v_j, \ldots, v_J)^T$ represents a vector containing all intracellular fluxes and $\underline{\underline{N}}^T$ (size = I × J) represents the stoichiometric matrix. While the coefficients of the stoichiometric matrix are usually known, the vector of fluxes can be divided for further analysis into a vector of unknown fluxes, $\underline{v}_c$ and a vector of known fluxes $\underline{v}_m$. The latter is usually made up with the substrate uptake and production transport rates which can be estimated directly from the bioreactor culture broth and from off gas analysis. Equation (1) can be transformed to

$$\underline{\underline{N}}_c{}^T \underline{v}_c + \underline{\underline{N}}_m{}^T \underline{v}_m = \underline{0} \tag{2}$$

and solved for the vector of unknown intracellular fluxes, $\underline{v}_c$ such that

$$\underline{v}_c = -(\underline{\underline{N}}_c{}^T \underline{\underline{N}}_c)^{-1} \underline{\underline{N}}_m{}^T \underline{v}_m. \tag{3}$$

[1]The platform is still password protected until the publication of the *B. megaterium* genome sequence. For accessibility requests contact b.bunk@tu-bs.de.

Although (3) represents the general solution of the metabolite balancing problem, its straightforward application is restricted to several constraints of thermodynamic, physicochemical and biological nature which limit the solution space of (3) by applying upper and lower bounds to individual fluxes ($\underline{v}_{\min} \leq \underline{v} \leq \underline{v}_{\max}$) [62–64]. For example, if irreversibility of some of the reactions imposes an additional requirement to solve (3), then for every irreversible reaction considered in the metabolic network the following inequality must hold

$$v_{\text{irrev}} \geq 0 \text{ irrev} \in [1,...,i,...,I] \tag{4a}$$

or, similarly, the maximal metabolite turnover capacity of the enzymes can be expressed as

$$\underline{v} \leq \underline{v}_{\max}. \tag{4b}$$

3.2.2 Fluxome Analysis Using Isotope Labelling Methods

Isotope labelling methods have found widespread use for the in vivo characterisation of various pro- and eukaryotic cells [65–76]. They are the rigorous theoretical and experimental extensions of the metabolic balancing approach taking into account the fate of single stable isotopes atoms in a metabolic network. However, in contrast to the former approach, the isotope labelling method avoids several of the drawbacks of the metabolic balancing, for example intracellular flux dependencies that most of the systems contain in their original formulation [77–80]. The basic formulation extends the balance to all isotope atoms, which can be described in a general form as

$$\frac{\mathrm{d}\left(\underline{\underline{X}}\,\underline{x}\right)}{\mathrm{d}t} = \underline{f}\left(\underline{v}, \underline{x}^{\text{inp}}, \underline{x}\right) \approx \underline{0}, \tag{5}$$

where $\underline{x}$ is the overall isotope distribution vector, $\underline{\underline{X}}$ is a diagonal matrix corresponding to the intracellular metabolites pools size, $\underline{v}$ is the complete flux vector and $\underline{x}^{inp}$ is the known isotopomer distribution vector of labelled substrates. Considering stationary state, the intracellular metabolites pools size can be taken out of (5) and, therefore, the unknown fluxes are actually only a function of the initial known isotopomer distribution vector of labelled substrates $\underline{x}^{inp}$, and of the final measured isotopomer distribution vector of labelled metabolites. Labelling information itself can be obtained either with nuclear magnetic resonance (NMR) or mass spectroscopy (MS) techniques [81–83]. Figure 11 exemplifies the different labelling patterns that can be obtained for a C_3 metabolite, e.g. pyruvate, illustrating a theoretical distribution of the isotopes and their MS assignment.

Once the experimental information is obtained, the unknown fluxes are derived by minimisation of the difference of the modelled and measured data, i.e., based on

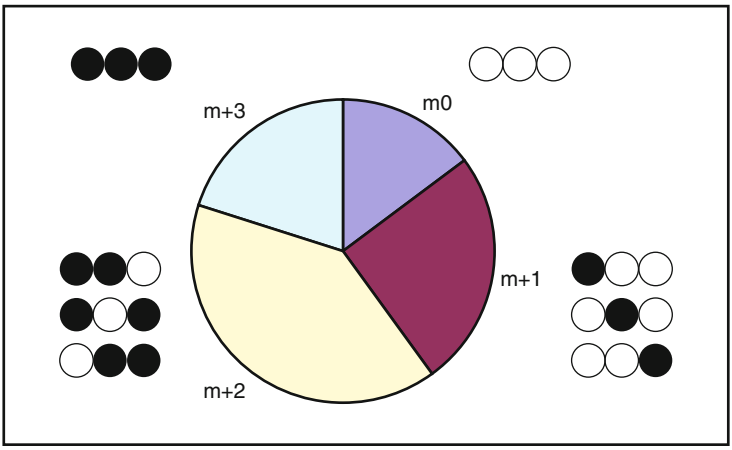


Fig. 11 Theoretical labelling patterns for a C_3 metabolite. Unlabelled molecule, m0 = m(0,0,0); molecules labelled in one carbon atom, m+1 = m(1,0,0) + m(0,1,0) + m(0,0,1); molecules labelled in two carbon atoms, m(+2) = m(1,1,0) + m(0,1,1) + m(1,0,1); full labelled molecule, m(+3) = m(1,1,1). The symbols represent: (*open circles*) ^{12}C atom, (*filled circles*) ^{13}C atom

the isotopic stoichiometric model, a simulated labelling pattern is generated and the experimental labelling pattern is used to constrain the solution by means of optimisation algorithms [80, 84–86].

3.2.3 The Fluxome of *B. megaterium*

Metabolic flux analysis using isotope labelling methods has been recently applied to the characterisation of several *B. megaterium* strains [87, 88]. The determination of the isotope labelling of the amino acids was carried out with GC-MS. Fluxes of the central carbon metabolism were estimated following the labelling pattern of the amino acids from their precursors.

The metabolic flux distributions around the intermediate pyruvate for different strains and environmental conditions are summarised in Fig. 12. This part of the metabolism has been shown to be an important node for the interconversion between glycolytic C_3 metabolites and C_4 metabolites of the tricarboxylic acid (TCA) cycle. The different anaplerotic reactions are of special importance for the production of recombinant proteins as they provide precursors, such as oxaloacetate, for amino acid biosynthesis. Due to that, the flux distribution is noticeably affected by both the cultivation conditions and the carbon source used which indicates flexible adaptation to the environmental situation. The flux from pyruvate to oxaloacetate through the reaction catalysed by pyruvate carboxylase was found to be the main anaplerotic pathway in *B. megaterium*.

In connection with the decarboxylation of malate by malic enzyme yielding pyruvate, a metabolic cycle occurs that is additionally involved in the energy and

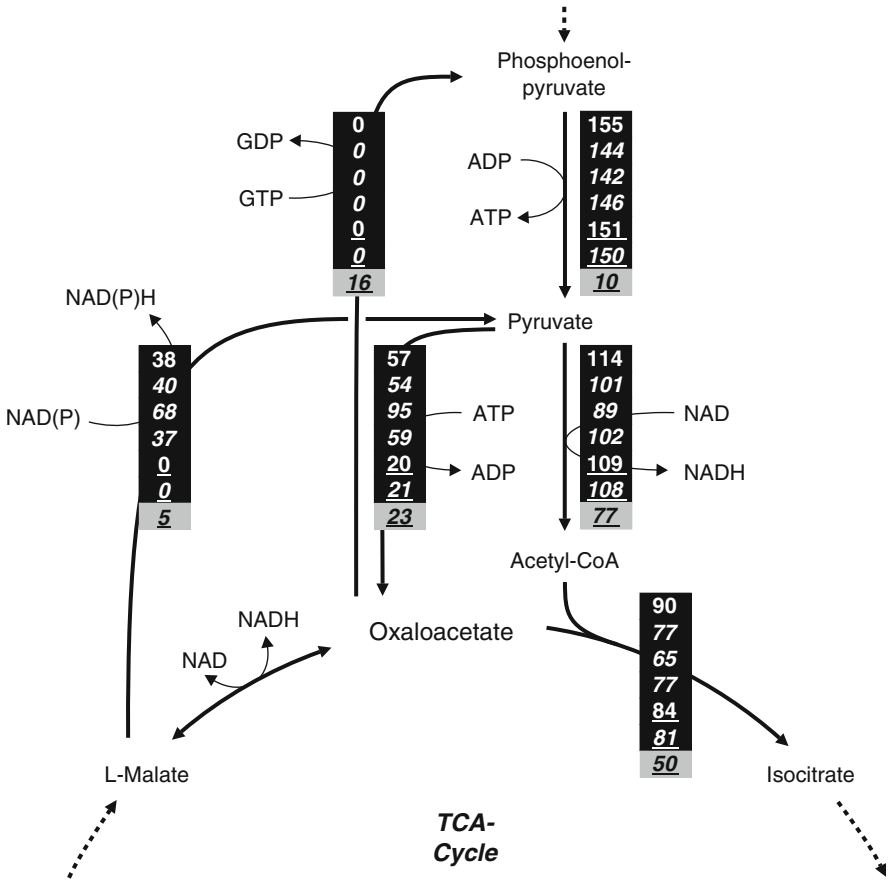


Fig. 12 Metabolic flux distributions at the interface between glycolysis and tricarboxylic acid (TCA) cycle for different *B. megaterium* strains and cultivation conditions. Values represent molar fluxes normalised to the substrate uptake rate (100). From top to bottom, black background = cultivation on glucose, grey background = cultivation on pyruvate: DSM319, $D = 0.106\ h^{-1}$, C-limited; MS941, $D = 0.11\ h^{-1}$, C-limited; MS941, $D = 0.426\ h^{-1}$, C-limited; MS941, $D = 0.108\ h^{-1}$, N-limited; WH320, $D = 0.096\ h^{-1}$, C-limited; WH323, $D = 0.107\ h^{-1}$, C-limited; (in all cases with glucose as carbon source); WH323, $D = 0.103\ h^{-1}$, C-limited, pyruvate as carbon source. Data adapted from [87, 88]

reducing equivalent metabolism. Depending on the cofactor specificity of malic enzyme which requires either NAD or NADP, the series of reactions in the cycle consumes just one ATP per turn but could possibly also generate one NADPH required for anabolic reactions. Hence, in the latter case the pyruvate (PYR)/oxalacetate (OAA)/malate (MAL)/pyruvate (PYR) cycle could be regarded as a replacement of the transhydrogenase catalysed reaction: NADH + ATP → NADPH + ADP. This assumption correlates with the observed higher flux through that metabolic cycle at higher growth rates to meet the increased demand of

reduction equivalents for biosynthetic processes. Another aspect concerning the necessity of metabolic cycles has been discussed for *B. subtilis* [89], namely, that the cycle may act as an additional pathway to keep the intermediate pool sizes at levels required for optimal growth.

3.2.4 The Influence of Recombinant Protein Production on the *B. megaterium* Fluxome

The metabolic link between glycolysis and TCA cycle is centred around the metabolic intermediate pyruvate including metabolites such as phosphoenolpyruvate, oxaloacetate and malate [90]. It has been shown that this metabolic node represents an important switch for the distribution of carbon fluxes in dependence of the environmental conditions. Consequently, the pyruvate node became a major target for metabolic engineering in order to optimise recombinant protein production.

In order to gain information about the importance of this metabolic node with regard to recombinant protein production in *B. megaterium* WH323, different environmental conditions were applied. Their effect on the metabolic flux distribution and the production of the chosen model protein *T. fusca* hydrolase TFH [91] were investigated. The noticeable influence of the pyruvate node on the metabolic network and, thus, on biosynthetic processes in general originates from reactions involved in the energy and reducing equivalent metabolism. In this context, Fig. 13 illustrates the relative changes of the metabolic fluxes for some key reactions after the induction of recombinant protein production. In the case of glucose (Fig. 13a), induction of recombinant protein production leads to an increased flux (+46%) through the pentosephosphate pathway, which was a source for NADPH required for biosynthetic processes. Furthermore, activation as well as a higher flux through the reaction cycle from PYR to OAA, and subsequently back to PYR via MAL catalysed by pyruvate carboxylase, malate dehydrogenase and malic enzyme was determined representing a potential alternative source for NADPH. For pyruvate as carbon source, similar results were obtained for the metabolic cycle around the intermediate pyruvate. Here, the fluxes through pyruvate carboxylase, phosphoenolpyruvate (PEP) carboxykinase and malic enzyme catalysed reactions were found increased by 4, 13 and 20% after induction of recombinant protein production (Fig. 13b). In contrast to growth on glucose, the TCA cycle flux was found to be increased as well (+17%) whereas the activity of gluconeogenic reactions was lowered by up to 50%. These findings indicate that the TCA cycle is used as the main source for reducing equivalents whereas the pentose phosphate pathway (PPP) is used to a low extent only.

More detailed metabolic flux analyses on both nutritional states revealed that the energetic parameters, namely Y_{ATP} and $Y_{NADPH/X}$, are changed noticeably after induction when using pyruvate as carbon source compared to growth on glucose. For pyruvate, a decrease of about 50% of the Y_{ATP} value from 13.2 to 7.0 g_{BM} mol^{-1} combined with a twofold increase of the $Y_{NADPH/X}$ value from

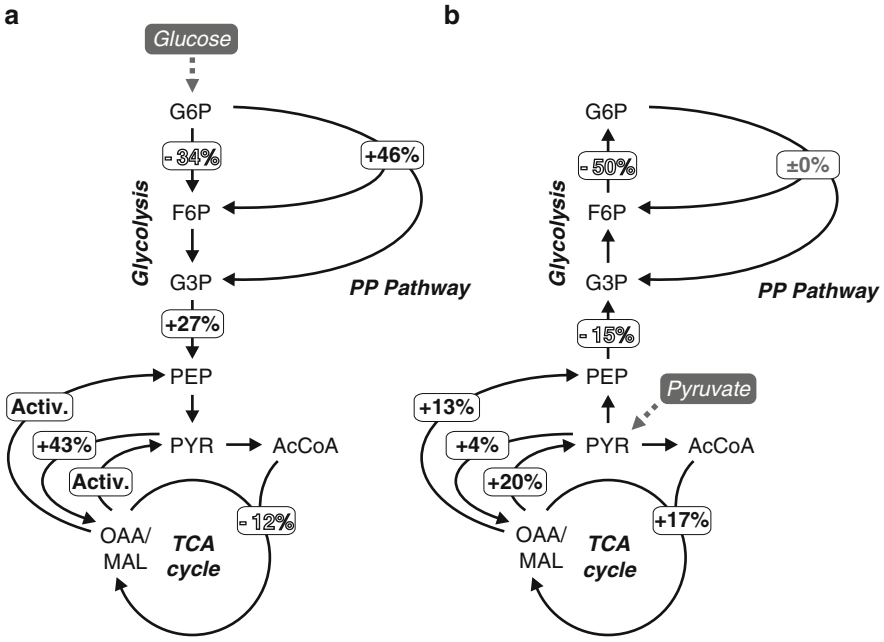


Fig. 13 Relative changes in metabolic fluxes. Relative changes in the metabolic fluxes of *B. megaterium* WH323 carrying TFH encoding pYYBm9 after induction of recombinant protein production with either glucose (**a**) or pyruvate (**b**) as the sole carbon source are given. *AcCoA* acetyl-coenzyme A; *Activ* activation; *F6P* fructose 6-phosphate; *G6P* glucose 6-phosphate; *MAL* malate; *OAA* oxaloacetate; *PEP* phosphoenolpyruvate; *PPP* pentose phosphate pathway; *PYR* pyruvate; *TCA* tricarboxylic acid. Data adapted from [88]

17.5 to 30.6 mmol g_{BM}^{-1} was determined [88]. Both the lower amount of biomass generated per mol ATP (Y_{ATP}) and the increased specific production of NADPH indicate a beneficial energetic situation for the biosynthesis of the recombinant TFH which has been experimentally confirmed by an up to fivefold increase of productivity of recombinant protein [88].

4 Systems Biology with *Bacillus megaterium*?

We define systems biology as the establishment of bioinformatic models for the prediction of cellular processes. With *B. megaterium* we focus on processes related to intra- and extracellular protein production. For this purpose, we already employ the genome data to deduce metabolic networks, model metabolic fluxes and confront our models with the results from metabolome and fluxome analyses [87, 88]. The second level of model building focuses on gene regulation. Here we deduce gene regulating networks using the PRODORIC database and the VIRTUAL

FOOTPRINT software tool (www.prodoric.de, [92]). Visualisation is achieved via ProdoNet [93]. Consequently, DNA microarray and proteomic experiments are the experimental validation for our proposals. Furthermore, integration of regulatory and metabolic networks is desired. The identification of limiting steps in protein production and the deduction of molecular optimisation strategies are the major goals in the future.

References

1. De Bary A (1884) Vergleichende Morphologie und Biologie der Pilze, Mycetozoen und Bakterien. Wilhelm Engelmann, Leipzig, Germany
2. Vary PS (1992) In: Doi R, McGloughlin M (eds) Biology of bacilli: application to industry. Butterworths-Heinemann, Boston, pp 251–310
3. Vary PS (1994) Microbiology 140:1001–1013
4. Priest FG (1993) In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other Gram-positive bacteria. American Society for Microbiology, Washington, DC, pp 3–16
5. von Tersch MA, Robbins HL (1990) FEMS Microbiol Lett 70:305–310
6. Kim JY (2003) Biotechnol Lett 25:1445–1449
7. Priest FG (1977) Bacteriol Rev 41:711–753
8. Panbangred W, Weeradechapon K, Udomvaraphant S, Fujiyama K, Meevootisom V (2000) J Appl Microbiol 89:152–157
9. Biedendieck R, Malten M, Barg H, Bunk B, Martens J-H, et al (2009) Microb Biotechnol 3:24–37
10. Sussman MD, Vary PS, Hartman C, Setlow P (1988) J Bacteriol 170:4942–4945
11. Rygus T, Hillen W (1991) Appl Microbiol Biotechnol 35:594–599
12. Barg H, Malten M, Jahn M, Jahn D (2005) In: Barredo JL (ed) Microbial processes and products. Humana Press, Totowa, NJ, USA, pp 165–184
13. Rygus T, Scheler A, Allmansberger R, Hillen W (1991) Arch Microbiol 155:535–542
14. Gärtner D, Geissendorfer M, Hillen W (1988) J Bacteriol 170:3102–3109
15. Dahl MK, Degenkolb J, Hillen W (1994) J Mol Biol 243:413–424
16. Rygus T, Hillen W (1992) J Bacteriol 174:3049–3055
17. Deutscher J, Kuster E, Bergstedt U, Charrier V, Hillen W (1995) Mol Microbiol 15:1049–1053
18. Burger S, Tatge H, Hofmann F, Genth H, Just I, Gerhard R (2003) Biochem Biophys Res Commun 307:584–588
19. Malten M, Hollmann R, Deckwer WD, Jahn D (2005) Biotechnol Bioeng 89:206–218
20. Malten M, Biedendieck R, Gamer M, Drews AC, Stammen S et al (2006) Appl Environ Microbiol 72:1677–1679
21. Schmidt TG, Koepke J, Frank R, Skerra A (1996) J Mol Biol 255:753–766
22. Porath J, Carlsson J, Olsson I, Belfrage G (1975) Nature 258:598–599
23. Kapust RB, Waugh DS (2000) Protein Expr Purif 19:312–318
24. Nagai K, Thogersen HC (1984) Nature 309:810–812
25. Cormack BP, Valdivia RH, Falkow S (1996) Gene 173:33–38
26. Jordan E, Hust M, Roth A, Biedendieck R, Schirrmann T et al (2007) Microb Cell Fact 6:2
27. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B et al (2005) Nucleic Acids Res 33: W526–W531
28. Scholz O, Thiel A, Hillen W, Niederweis M (2000) Eur J Biochem 267:1565–1570
29. Reischer H, Schotola I, Striedner G, Potschacher F, Bayer K (2004) J Biotechnol 108:115–125
30. Biedendieck R, Yang Y, Deckwer WD, Malten M, Jahn D (2007) Biotechnol Bioeng 96:525–537

31. Hollmann R, Deckwer WD (2004) J Biotechnol 111:89–96
32. Lee DS, Pack MY (1987) Enzyme Microb Technol 9:594–597
33. Malten M (2005) Protein production and secretion in *Bacillus megaterium*. Dissertation thesis. Technical University of Braunschweig, Braunschweig, Germany
34. Wittchen KD, Meinhardt F (1995) Appl Microbiol Biotechnol 42:871–877
35. Biedendieck R, Beine R, Gamer M, Jordan E, Buchholz K et al (2007) Appl Microbiol Biotechnol 74:1062–1073
36. Ruiz C, Blanco A, Pastor FI, Diaz P (2002) FEMS Microbiol Lett 217:263–267
37. Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E et al (2004) Microbiol Mol Biol Rev 68:207–233
38. Nielsen H, Krogh A (1998) Proc Int Conf Intell Syst Mol Biol 6:122–130
39. van Hijum SA, Szalowska E, van der Maarel MJ, Dijkhuizen L (2004) Microbiology 150:621–630
40. Ledent P, Duez C, Vanhove M, Lejeune A, Fonze E et al (1997) FEBS Lett 413:194–196
41. Englard S, Seifter S (1990) In: Deutscher MP (ed) Methods in enzymology. Academic Press, San Diego, pp 285–301
42. Yang Y, Biedendieck R, Wang W, Gamer M, Malten M et al (2006) Microb Cell Fact 5:36
43. Makrides SC (1996) Microbiol Rev 60:512–538
44. Tabor S, Richardson CC (1985) Proc Natl Acad Sci USA 82:1074–1078
45. Stano NM, Patel SS (2004) J Biol Chem 279:16136–16143
46. Bernhard K, Schrempf H, Goebel W (1978) J Bacteriol 133:897–903
47. Khan SA (2005) Plasmid 53:126–136
48. Malten M, Nahrstedt H, Meinhardt F, Jahn D (2005) Biotechnol Bioeng 91:616–621
49. Kieselburg MK, Weickert M, Vary PS (1984) Bio/Technology 2:254–259
50. Stevenson DM, Kunnimalaiyaan M, Muller K, Vary PS (1998) Plasmid 40:175–189
51. Gamer M, Frode D, Biedendieck R, Stammen S, Jahn D (2009) Appl Microbiol Biotechnol 82:1195–1203
52. Kunnimalaiyaan M, Zhou Y, Scholle M, Baisa G, Vary PS (2001) Plasmid 45:142–143
53. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Methods Enzymol 185:60–89
54. Brunschwig E, Darzins A (1992) Gene 111:35–41
55. Conrad B, Savchenko RS, Breves R, Hofemeister J (1996) Mol Gen Genet 250:230–236
56. Elroy-Stein O, Moss B (1990) Proc Natl Acad Sci USA 87:6743–6747
57. Vellanoweth RL (1993) In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other Gram-positive bacteria. ASM Press, Washington, pp 699–711
58. Homann A, Biedendieck R, Gotze S, Jahn D, Seibel J (2007) Biochem J 407:189–198
59. Vary PS, Biedendieck R, Fuerch T, Meinhardt F, Rohde M et al (2007) Appl Microbiol Biotechnol 76:957–967
60. Scholle MD, White CA, Kunnimalaiyaan M, Vary PS (2003) Appl Environ Microbiol 69:6888–6898
61. Nicholson WL (2008) Appl Environ Microbiol 74:6832–6838
62. Palsson B (2002) Nat Biotechnol 20:649–650
63. Palsson BO (2006) Cambridge University Press
64. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Academic Press
65. Blank LM, Lehmbeck F, Sauer U (2005) FEMS Yeast Res 5:545–558
66. Christensen B, Nielsen J (1999) Metab Eng 1:282–290
67. Dauner M, Sauer U (2000) Biotechnol Prog 16:642–649
68. Fischer E, Zamboni N, Sauer U (2004) Anal Biochem 325:308–316
69. Hans MA, Heinzle E, Wittmann C (2001) Appl Microbiol Biotechnol 56:776–779
70. Kleijn RJ, van Winden WA, Ras C, van Gulik WM, Schipper D, Heijnen JJ (2006) Appl Environ Microbiol 72:4743–4754
71. Noh K, Gronke K, Luo B, Takors R, Oldiges M, Wiechert W (2007) J Biotechnol 129:249–267
72. Noh K, Wahl A, Wiechert W (2006) Metab Eng 8:554–577

73. Schmidt K, Carlsen M, Nielsen J, Villadsen J (1997) Biotechnol Bioeng 55:831–840
74. Shastri AA, Morgan JA (2007) Phytochemistry 68:2302–2312
75. van Winden WA, Heijnen JJ, Verheijen PJ, Grievink J (2001) Biotechnol Bioeng 74:505–516
76. van Winden WA, van Dam JC, Ras C, Kleijn RJ, Vinke JL et al (2005) FEMS Yeast Res 5:559–568
77. Mollney M, Wiechert W, Kownatzki D, de Graaf AA (1999) Biotechnol Bioeng 66:86–103
78. Wiechert W, Mollney M, Isermann N, Wurzel M, de Graaf AA (1999) Biotechnol Bioeng 66:69–85
79. Wiechert W, Mollney M, Petersen S, de Graaf AA (2001) Metab Eng 3:265–283
80. Wiechert W, Siefke C, de Graaf AA, Marx A (1997) Biotechnol Bioeng 55:118–135
81. Wittmann C (2007) Microb Cell Fact 6:6
82. Wittmann C, Hans M, Heinzle E (2002) Anal Biochem 307:379–382
83. Zupke C, Stephanopoulos G (1995) Biotechnol Bioeng 45:292–303
84. Rantanen A, Mielikainen T, Rousu J, Maaheimo H, Ukkonen E (2006) Bioinformatics 22:1198–1206
85. Rantanen A, Rousu J, Jouhten P, Zamboni N, Maaheimo H, Ukkonen E (2008) BMC Bioinformatics 9:266
86. Zhao J, Shimizu K (2003) J Biotechnol 101:101–117
87. Fürch T, Hollmann R, Wittmann C, Wang W, Deckwer WD (2007) Bioprocess Biosyst Eng 30:47–59
88. Fürch T, Wittmann C, Wang W, Franco-Lara E, Jahn D, Deckwer WD (2007) J Biotechnol 132:385–394
89. Diesterhaft MD, Freese E (1973) J Biol Chem 248:6062–6070
90. Sauer U, Eikmanns BJ (2005) FEMS Microbiol Rev 29:765–794
91. Yang Y, Malten M, Grote A, Jahn D, Deckwer WD (2007) Biotechnol Bioeng 96:780–794
92. Grote A, Klein J, Retter I, Haddad I, Behling S et al (2009) Nucleic Acids Res 37:D61–D65
93. Klein J, Munch R, Biegler I, Haddad I, Retter I, Jahn D (2009) Nucleic Acids Res 37: D494–D498
94. Biedendieck R, Gamer M, Jaensch L, Meyer S, Rohde M et al (2007) J Biotechnol 132:426–430
95. Schmiedel D, Kintrup M, Kuster E, Hillen W (1997) Mol Microbiol 23:1053–1062
96. Dahl MK, Schmiedel D, Hillen W (1995) J Bacteriol 177:5467–5472
97. Sumer JB, Howell SF (1935) J Biol Chem 108:51–54

Adv Biochem Engin/Biotechnol (2010) 120: 163–193
DOI: 10.1007/10_2009_54

Published online: 17 February 2010

Extending Synthetic Routes for Oligosaccharides by Enzyme, Substrate and Reaction Engineering

Jürgen Seibel, Hans-Joachim Jördening, and Klaus Buchholz

Abstract The integration of all relevant tools for bioreaction engineering has been a recent challenge. This approach should notably favor the production of oligo- and polysaccharides, which is highly complex due to the requirements of regio- and stereoselectivity. Oligosaccharides (OS) and polysaccharides (PS) have found many interests in the fields of food, pharmaceuticals, and cosmetics due to different specific properties. Food, sweeteners, and food ingredients represent important sectors where OS are used in major amounts. Increasing attention has been devoted to the sophisticated roles of OS and glycosylated compounds, at cell or membrane surfaces, and their function, e.g., in infection and cancer proliferation. The challenge for synthesis is obvious, and convenient approaches using cheap and readily available substrates and enzymes will be discussed. We report on new routes for the synthesis of oligosaccharides (OS), with emphasis on enzymatic reactions, since they offer unique properties, proceeding highly regio- and stereoselective in water solution, and providing for high yields in general.

Keywords Biocatalysis, Enzyme engineering, Fructosyltransferases, Glucosyltransferases, Immobilization, Oligosaccharides, Reaction engineering

J. Seibel (✉)
Institute of Organic Chemistry, University of Würzburg, Am Hubland, 97074 Würzburg, Germany
e-mail: seibel@chemie.uni-wuerzburg.de

H.-J. Jördening and K. Buchholz
Department of Technical Chemistry, Technical University Braunschweig, Hans-Sommer-Str. 10, 38106 Braunschweig, Germany

Contents

1 Introduction 164
2 Enzymatic Oligosaccharide Synthesis 166
2.1 Acceptor Reactions of Dextransucrase, Products, Kinetics 166
2.2 Engineering of Glycosyltransferases (GTF, FTF) 172
3 From Gene to Product 179
3.1 Biocatalyst Development (Immobilization, Two Enzyme System) 180
3.2 Engineering of GTFR 183
3.3 Application of Engineered Dextransucrase (GTFR) 183
3.4 Integrated Reaction Engineering 185
4 Conclusions 187
References 187

Abbreviations

I	Inulosucrase
L	Liter(s)
LS	Levansucrase
min	Minute(s)
mol	Mole(s)
OS	Oligosaccharide
PS	Polysaccharide
rt	Room temperature
Ts	Tosyl, 4-toluenesulfonyl

1 Introduction

The integration of all relevant tools for bioreaction engineering has been a recent challenge. It has also been in the focus of a common research project on "Integration of genetic and biochemical engineering methods for the development of biotechnological processes" ("From gene to product"). It aims at combining and integrating methods available both from the basic sciences, notably genetics and molecular biology, and from the chemical engineering sciences. It includes the optimization of the biological system, or the biocatalyst, respectively, bioreactor performance, and product isolation to be integrated by appropriate methods [1, 2].

This approach should notably favor the production of oligo- and polysaccharides, which is highly complex due to the requirements of regio- and stereoselectivity. Enzymes are the tools that are best suited to meet these requirements. A considerable number of industrial processes have been established for the production of oligosaccharides. They are, however, limited to few naturally occurring substrates and enzymes. That means that only products based on

sucrose and starch, with D-glucose and D-fructose as the constituents, have become available at reasonable prices. However, many other products, including other less common monosaccharides, such as galactose, xylose, rhamnose, etc., as well as other regio-specific combinations of sugars, are of current interest for different purposes in the food and pharmaceuticals sectors. Oligosaccharides (OS) have been applied worldwide in major amounts in the sectors of food, notably sweeteners and food ingredients [3]. A wide variety of different OS are produced and sold in Japan.

Increasing attention has been devoted to the sophisticated roles of OS and glycosylated compounds at cell or membrane surfaces. Oligosaccharides are found on cell surfaces as glycoprotein or glycolipid conjugates and play important structural and functional roles in numerous biological recognition processes. These processes include viral and bacterial infection, cancer metastasis, inflammatory response, innate, and adaptive immunity, and many other receptor-mediated signaling processes [4, 5]. Inhibition of cell or bacterial adhesion and vaccine formulation have therefore become relevant topics of pharmaceutical research.

The challenge for developing synthetic methods for making OS, notably complex OS, that can be scaled up to an industrial level is obvious, using substrates, as well as enzymes that are available at reasonable prices and purity.

Strategies for OS synthesis include:

- Chemical synthesis, limited by complex synthetic approaches using protection of the sugar hydroxyl groups in order to achieve regioselectivity, and more so stereoselective reactions
- Highly selective enzymatic synthesis using Leloir glycosyl transferases (GTF), but restricted due to limited availability of enzymes, and nucleotide activated substrates at high prices
- Sucrase type (non-Leloir GTF) enzymes that operate both regio- and stereoselective, using sucrose as a cheap substrate, or, in some cases, such as cyclodextrin transferases, starch; they are however limited to the transfer of glucose or fructose in synthetic routes

(For an overview on activated sugars in synthesis see Seibel et al. [6].
Convenient routes for new OS however, are rarely available [3].
We report here on new routes for the synthesis of oligosaccharides (OS), with emphasis on enzymatic reactions, since they offer unique properties, proceeding highly regio- and stereoselective in water solution, and providing for high yields in general. We used new or modified substrates and both wild type and genetically modified glycosyltransferases (sucrase type GTF of the non-Leloir type), both glucosyl and fructosyl transferases that synthesize OS and polysaccharides (PS). We extended the range of monosaccharides to be transferred, e.g., D-galactose, D-mannose, D-xylose, L-rhamnose, L-fucose, by using sucrose analogs and genetically engineered GTF, in order to design new classes of OS. The different glycopyranosyl-residues in such new OS may be recognized by carbohydrate binding cell receptors and function as inhibitors of bacterial adhesion [7].

We selected approaches that may be scaled up and transferred to industrial manufacturing. Reactions using nucleotide activated sugars are therefore not included.

Enzyme engineering will be demonstrated to be an interesting tool for tailoring oligosaccharide products and the glycosidic linkage specificity of carbohydrate active enzymes. The combination of site-directed/random mutagenesis and substrate engineering further results in high product yields and novel products. The described methods may be subjected to similar enzyme and substrate engineering approaches.

As an example, the reaction engineering for the isomaltose formation from sucrose is discussed. We started the studies with an immobilized wildtype dextransucrase (DSR-S). By modeling the kinetics of the system, a first optimization by using suitable substrate and acceptor concentrations could be achieved. Further improvements were possible by using appropriate zeolites for reaction integrated product separation. The introduction of a coimmobilized bi-enzymatic biocatalyst (DSR-S and dextranase) and especially the utilization of a genetically engineered enzyme variant led to the optimal design for the whole process.

2 Enzymatic Oligosaccharide Synthesis

2.1 Acceptor Reactions of Dextransucrase, Products, Kinetics

Glucansucrases (non-Leloir GTFs) naturally act on substrates, sucrose in most cases, with the transfer of one glycosyl unit, either D-glucose or D-fructose, to growing polysaccharide chains, or to acceptors, producing oligosaccharides elongated by one or more glycosyl units [8, 9]. The reactions proceed with high efficiency, with high yields, regio- and stereoselectivity, and in water as the solvent. The background of this convenient synthetic pathway is the high energy of the glyosidic bond of sucrose which is similar to that of nucleotide activated sugars [6]. GTFs are bacterial enzymes expressed extracellularly by various bacterial species of *Leuconostoc*, *Lactobacillus* and *Streptococcus* (the structure of GTFs will be discussed in Sect. 2.2).

Among GTF, dextransucrase (EC 2.4.1.5) has attracted much attention, since it elaborates two kinds of reactions: (1) the primary reaction (substrate reaction), the synthesis of dextran from sucrose; and (2) the secondary reaction (acceptor reaction), the transfer of D-glucose from sucrose to carbohydrate acceptors that are added to the reaction solution, thus synthesizing a range of oligosaccharides. Dextran is a D-glucopyranosyl polysaccharide containing mostly α-1,6 glycosidic bonds, with branches linked with α-1,2 or α-1,3 bonds to the main chain, and with a molecular weight ranging from 0.5 to 6.10^6 kDa [8, 9]. Dextran has many industrial

applications due to its nonionic character, high molecular weight, and good stability under normal operating conditions. It has been widely used in the pharmaceutical industry and as a chromatographic medium [10].

Isomaltooligosaccharides (as mixtures of different α-glucooligosaccharides) are produced in part by dextransucrases using sucrose as the substrate and maltose as an acceptor. They have been used as sweeteners in Europe and Japan for years, mainly in the field of prebiotics, and dermocosmetics [11–13]. Several different methods can be used to produce them by using such carbohydrates as sucrose, maltose, starch, and dextran. For the industrial production of α-glucooligosaccharides with sucrose as the substrate and maltose as the acceptor by one specific dextransucrase (of *L. mesenteroides* NRRL B-1299), the kinetic behavior of the enzyme has been characterized in order to optimize the synthesis [14].

Recently isomaltose, as a singular product, has gained interest, so its formation by dextransucrase will be discussed in more detail in this chapter. Optimal reaction conditions were elaborated in this project for isomaltose production in a pilot reactor, including the following steps [15]:

- The formation rate of isomaltose (mmol L^{-1} min^{-1}) was maximized
- The isomaltose yield and absolute concentration in the product mixture were maximized
- The formation of dextran and other by-products was minimized
- The amount of sucrose in the product solution was minimized
- The thermal and operational stability of the immobilized enzyme was investigated

The kinetics of all major dextransucrase reactions, including polysaccharide and acceptor reactions, will be discussed in some detail later. The kinetics of dextran synthesis have been investigated early by Ebert, Patat, and Schenk [16, 17]. With respect to the acceptor reaction, the glucosyl moiety is transferred from sucrose to the acceptor molecule instead of the growing dextran chain to give the acceptor molecule elongated by one single glucosyl unit as the primary product. Usually the new bond formed is an α-1,6-glucosidic bond. In most cases, the product can itself serve as an acceptor, so that a homologous series of glucosylated oligosaccharides is formed [8, 9]. In the presence of sucrose only, the single substrate, the initial reaction rate can be calculated by Michaelis-Menten kinetics up to 200 mM sucrose concentration, but at higher sucrose concentrations substrate inhibition becomes significant and has to be taken into account. The inhibition constant for sucrose is 730 mM [18, 19]. The inhibition can be overcome by the addition of acceptors. The enzyme activity is significantly enhanced, and stabilized by the presence of dextran, and by calcium ions.

It must be pointed out that Michaelis kinetics does not apply in the presence of acceptors. Constants like apparent V_{max}, K_M, and K_I are dependent on acceptor concentrations. In order to understand the reactions occurring in the presence of different types of acceptors, and to provide an appropriate background for reaction engineering in industrial processes, the kinetics of the different reaction pathways

have been analyzed in detail, and a mathematical model has been established (using the dextransucrase DSR-S from *L. mesenteroides* B-512F). It aims at predicting the optimal reaction conditions and reactor configurations for different acceptors, and notably for the production of leucrose and isomaltooligosaccharides (isomalto-OS) [20, 21]. The model takes into account most of the different reactions occurring – PS and OS formation in parallel or in sequence, enzyme complexes with substrate, a covalent glucosyl-enzyme intermediate, binding sites for the acceptor and the growing dextran chain, respectively, the relevant parameters, notably concentrations of enzyme, substrate and all intermediates or acceptors, respectively, oligosaccharides formed and the growing chain of dextran (Fig. 1) [20, 21]. Kinetics and modeling have been based on a major range of experimental investigations [20–23]. For all species (intermediates, products) shown in Fig. 1, differential equations were established. The parameters of the model were calculated by numerical integration of experimental data with the fourth-order Runge-Kutta method and optimized with the simplex algorithm.

With good acceptors, e.g., maltose, apparent V_{max} values increase with increasing acceptor concentration by a factor of about three at 600 mM maltose concentration [22]. They decrease, as do initial reaction rates, with increasing concentrations of weak acceptors like glucose or fructose, again by a factor of three, at 2.75 M

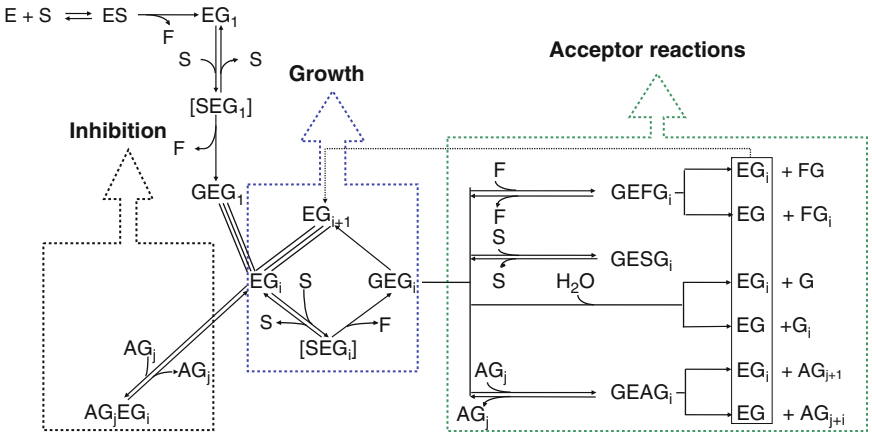


Fig. 1 Scheme of reactions catalyzed by dextransucrase. The scheme summarizes all relevant reactions with: enzyme complexes (E, ES, Gi, GE Gi etc.); sucrose (S); acceptors (A, AGj); fructose (F); acceptor products (AGj). The enzyme is assumed to have one active center, where a glucosyl group is bound covalently, and binding sites for the growing dextran chain (Gi) and acceptors (A, AGj). Types of reactions are indicated: (dextran) chain growth, inhibition by acceptor products AGj, and acceptor reactions. Fructose may be bound to the acceptor site and react with G to form leucrose FG. Reaction with sucrose (S) corresponds to (reversible) substrate inhibition. Reactions with water (very slow) yield glucose or dextran (Gi). Acceptor reactions (reactions with acceptors A and acceptor products (AGj) give acceptor products elongated by one glucose (G); these may again react as acceptors

fructose concentration [21]. Apparent K_M values also vary with acceptor type and concentration. Dextran formation may be nearly suppressed at high acceptor concentration, which is important for oligosaccharide production.

The final product concentration depends significantly on the ratio of substrate and acceptor. The first acceptor product (panose in the presence of maltose) is favored at high maltose excess, and further tetra- and pentasaccharides are formed in significant amounts. Thus, isomalto-oligosaccharides can be produced in high yield under appropriate reaction conditions.

In the presence of weak acceptors, such as glucose or fructose, initial overall reaction rates decrease significantly with increasing acceptor concentration, whereas they increase with the substrate concentration. Initial rates of acceptor product formation increase with both substrate and acceptor concentration, the dextran formation being suppressed to a major extent only at very high acceptor concentration (at about 3 M glucose or fructose concentration) [20, 21]. The acceptor product of fructose, leucrose, does not act as an acceptor, so that high yields (up to about 70%) can be obtained, which is of interest for industrial application. Leucrose has been developed up to the pilot scale (500 kg leucrose produced) for application as an alternative sweetener [24, 25].

The model was then extended to two different dextransucrases from *L. mesenteroides* 1299, a soluble and an immobilized form. The parameter sets for the soluble form fitted well with those obtained from the B-512F enzyme. The immobilized form showed a good agreement only for parameters linked to the sucrose concentration (turnover number, K_M, and substrate inhibition constant), but significant differences occurred with other parameters [26].

Summarizing the results of many investigations, monosaccharides and their derivatives (e.g., sorbitol) are rather weak acceptors. Disaccharides, including acceptor products like isomaltose, are much better acceptors [9, 27, 28]. The decrease of enzyme activity with time has been described by a first order reaction, and inactivation parameters have been calculated.

The immobilization of the enzyme has been achieved efficiently by inclusion in alginate beads ($d_p \sim 2$ mm) [29]. The productivity of the immobilized enzyme was found to be optimal in a continuous tubular reactor [23]. Silica flour was added to the alginate, to increase the bead density and to allow operation in a fluidized bed reactor with a high-density fluid phase like highly concentrated sugar solutions. Because of the strong complex between the enzyme and dextran and the resulting high dextran content in the enzyme preparation, the immobilized enzyme seems to be diffusion-limited, depending on the enzyme concentration.

The special case treated in this chapter – the reactions with glucose as an acceptor – will be discussed in more detail later since they are essential for the project aiming at production of isomaltose as a single product (in contrast to mixtures of OS). When glucose is present, the first acceptor product formed is isomaltose. The next acceptor products are higher isomalto-OS like isomaltotriose, -tetraose, -pentaose, and so on. However, only oligosaccharides up to dp 4 were observed. Higher initial glucose concentrations lead to higher final isomaltose concentrations after sucrose exhaustion. Furthermore, the formation of higher

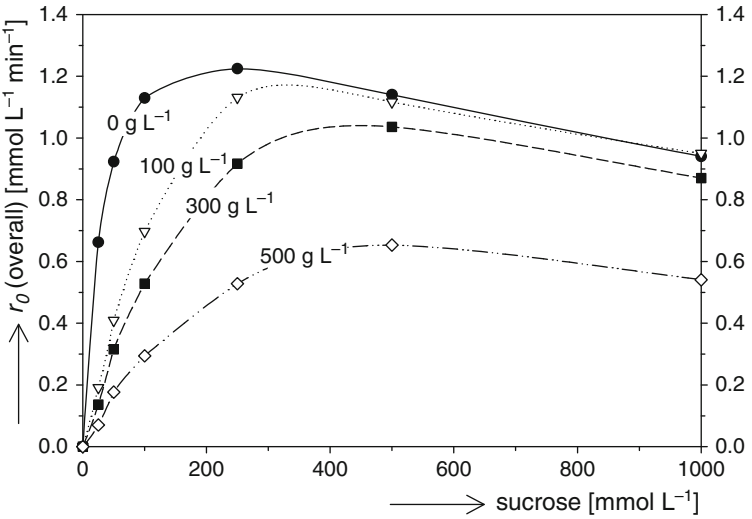


Fig. 2 Overall reaction rates (sum of all oligosaccharides formed, *y*-axis) at different initial sucrose and glucose concentrations. (glucose: 0 (*dotted lines with filled circle*), 100 (*dotted lines with inverted open triangle*), 300 (*dotted lines with filled square*), 500 (*dotted lines with open diamond*) g L^{-1}; soluble dextransucrase; 1.32 U mL^{-1}; 30°C)

oligosaccharides is reduced, whereas the sucrose consumption rate decreases. The high yield of acceptor products correlates with a repression of dextran formation, which is important for effective use of immobilized enzyme. Simulated data fit the experimental data very well.

The initial overall reaction rate (sucrose consumption rate) decreases considerably with higher initial glucose concentration (Fig. 2), whereas the isomaltose formation rate increases up to 300 g L^{-1} glucose, which reflects the shift from dextran formation (substrate reaction) to the acceptor reaction (Fig. 3). However, with increasing glucose concentration, the maximal reaction rate shifts towards higher initial sucrose concentrations. However sucrose concentrations >0.5 M inhibit the reaction (Fig. 3).

With respect to the final isomaltose concentration, higher glucose concentrations are favorable, although the maximal isomaltose formation rate occurs at 300 g L^{-1}. Up to one-third higher yields can be obtained if the glucose concentration is increased up to 500 g L^{-1}. Experimental data of the selectivity for isomaltose formation are a function of the glucose concentration. The selectivity (isomaltose formation vs higher OS formation) increases with increasing glucose concentration and reaches a maximal value around 0.5. Furthermore, increasing final isomaltose concentrations correlate with increasing initial sucrose concentrations, but at the expense of decreasing yield. These results show that the glucose concentrations should be varied between 300 g L^{-1} and 500 g L^{-1} to achieve a high isomaltose formation rate and isomaltose end concentration. The sucrose concentration should

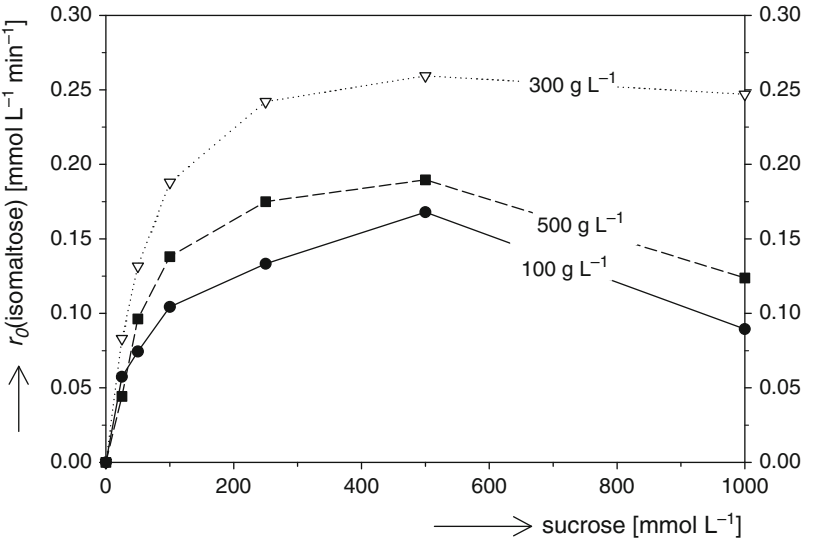


Fig. 3 Initial isomaltose formation rates with soluble dextransucrase (1.32 U mL^{-1}) at different glucose and sucrose concentrations. *Filled circles*: 100 g L^{-1} glucose; *open triangles*: 300 g L^{-1} glucose; *filled squares*: 500 g L^{-1} glucose

be <500 mM, to prevent inhibition, but also not too low, so that high reaction rates and product end concentrations can be achieved.

However, for long-term operation sucrose should be kept at rather low concentration in order to prevent dextran formation, whereas the glucose concentration should be high.

In order to evaluate dextransucrase operational stability during isomaltose formation, enzyme half-lives were investigated over extended time of operation with the given parameters of the reaction over 6 h. The decrease of enzyme activity E as a function of time may be described by a first order reaction:

$$E = E_0 \cdot e^{(-k_d \cdot t)}, \quad t_{1/2} = \frac{\ln 2}{k_d}$$

Immobilization increases the stability more than twofold over that of the free enzyme between 30°C and 40°C. However, an even higher stabilizing effect at 30°C results from high glucose concentrations (here 500 g L^{-1}) in the reaction mixture, when the immobilized enzyme is almost stable at 30°C over 1 month [30]. This stabilization effect is important for continuous long-term operation at 30°C. (For other cases the inactivation constants k_d of the immobilized enzyme were 0.0135 (1/d) when maltose was the acceptor (stabilizing), and 0.029 (1/d) when fructose was the acceptor [20]).

2.2 Engineering of Glycosyltransferases (GTF, FTF)

2.2.1 Engineering of Glucansucrases (GTF)

Glucansucrases (non-Leloir glycosyltransferases, GTFs) are expressed by lactic acid bacteria. They are structurally related to glycosidases, but their dominant reaction is the glucosyltransfer from the substrate sucrose for the high efficient synthesis of glucan polymers [31]. Glucansucrases are classified as glycoside hydrolase (GH) enzymes of family 70 according to the carbohydrate active enzyme database CAZy [32]. In terms of the EC-classification system, glucansucrases are found as EC 2.4.1.5 (dextransucrase, but mutansucrase and reuteransucrase are still classified here at the moment) and EC 2.4.1.140 (alternansucrase). More than 40 glucansucrase primary structures are known at the moment, all showing a high similarity in organization. In all glucansucrases, an *N*-terminal, variable region of different length can be found, followed by a conserved catalytic domain and a *C*-terminal, so-called "glucan binding"-domain [33]. However, so far no crystal structure of a lactic acid glucansucrase has been published. Structural comparison with GH-13 (α-amylase) enzymes predicts that the catalytic domain is organized as a $(\beta/\alpha)_8$-barrel, but circularly permutated [34]. This prediction was confirmed by Dijkstra et al. presenting a first, not published crystal structure of GTF180, a glucansucrase of *L. mesenteroides 180* [35]. Interestingly, the enzyme not only showed a circular permutation, but also a huge U-shaped folding of the whole enzyme, forming distinct domains out of severed sequence-sections. (personal communication, (see Fig. 8 by Vujičić et al. [36]). Glucansucrases are relevant enzymes for industrial applications. The production of the polysaccharide dextran (α-1,6-glucan) by *L. mesenteroides* is one example. Such polymers are used in the food sector, as additives for dyes and in health care [37, 38]. Variants and differences in glycosidic linkage type, degree and type of branching, and molecular mass of glucans yield different structural and functional properties.

Thus a major scientific goal is to understand how the enzymes control the regiospecificity of the polymer synthesis. This knowledge should be used for the design of enzymes for production of tailor-made saccharides, including different purposes: more selective production of either di- or oligosaccharides or polysaccharides, changing regioselectivity towards favorable glycosidic bonds, or even introducing other types of OS, such as thioglycosid bonds.

One target enzyme is the glycosyltransferase R (GTFR) from *Streptococcus oralis*, a dextransucrase, which catalyzes glucan synthesis. Its high stability without significant loss of activity at 30°C for weeks makes GTFR an interesting candidate for industrial applications. Structure-functions studies of glucansucrases identified three catalytic residues crucial for their activity previously [33]. Also the catalytic triad of this GH 70 enzymes has already been investigated, including the nucleophile D516 (GTFR numbering) for the formation of the covalent glucosyl-enzyme complex [34, 39], the acid/base catalyst E554, and a so-called transition state stabilizer D627 by site-directed mutagenesis experiments for similar enzymes

Table 1 Polymer linkage type and alignment of amino acid sequences of various (mutant) glucansucrase enzymes with GTFR wild-type and mutant variants derived [52]

Wild-type and mutant enzymes	Main α-linkage in glucan products	Amino acid sequence around transition state stabilizer D627 in GTFR
DSRS	(1–6) PS	655 YSFVRAHDSEVQTVI
GTFI	(1–3) PS	557 YSFIRAHDSEVQDLI
GTFA	(1–4) PS	1126 YSFVRAHDNNSQDQI
GTFA'	(1–6) PS	1126 YSFVRAHDSEVQDQI
GTFR	(1–6) PS	620 YIFVRAHDSEVQTVI 713 SPYHDAIDA
S628D	OS	620 YIFVRAHDDEVQTVI
S628R	OS, PS	620 YIFVRAHDREVQTVI
R624G/V630I/D717A	(1–3, 1–6) PS	620 YIFVGAHDSELQTVI 713 SPYHAAIDA
R624G		620 YIFVGAHDSEVQTVI
V630I	(1–6) PS	620 YIFVRAHDSELQTVI
R624G/V630I	(1–3, 1–6) PS	620 FVGAHDSELQTVI

DSRS: *Leuconostoc mesenteroides* NRRL B-1355 [41], GTFI: *Streptococcus downei* Mfe28 [53], GTFA: *Lactobacillus reuteri* 121 [54], GTFA': mutant GTFA [50], GTFR: *Streptococcus oralis* [55], GTFR variants: mutants constructed in this study. Mutations are shown with *black underlay*, the putative transition state stabilizer is shown with *gray underlay*. *PS* polysaccharide; *OS* oligosaccharide

[40–43] and sequence alignments with the corresponding residues [44]. Based on structural modeling, further amino acid residues putatively forming the enzyme active center could be identified [45, 46]. In particular, a motif around the transition state stabilizer D627 seemed to be crucial regarding its transglycosylation activity [47–51]. As a consequence the "RAHDSEV" segment was chosen to be randomized by PCR mutagenesis. This promising approach should lead to variants which synthesize polysaccharides with novel properties. Indeed, a variant was selected which produced from sucrose a polymer which exhibited a more textured nature when compared to the wildtype polysaccharide dextran. This interesting polysaccharide was analyzed as a polymer with dramatically increased α-1,3-glucosidic linkages. The GTFR variant carried mutations at R624G:V630I:D717A (Table 1). Interestingly all mutations contributed to the dramatic switch in regioselectivity from a dextran type with mainly α-1,6- to a mutant type polymer with predominantly α-1,3-glucosidic linkages (Fig. 4) [52].

2.2.2 Donor and Acceptor Substrate Engineering

A second novel aspects approach is substrate directed synthesis. The point of this aspect is that wild-type enzymes could be used to generate more diversity without engineering the enzyme. Therefore acceptor substrates were designed which allow the control of linkage specificity by the enzyme and further chemical reactions and

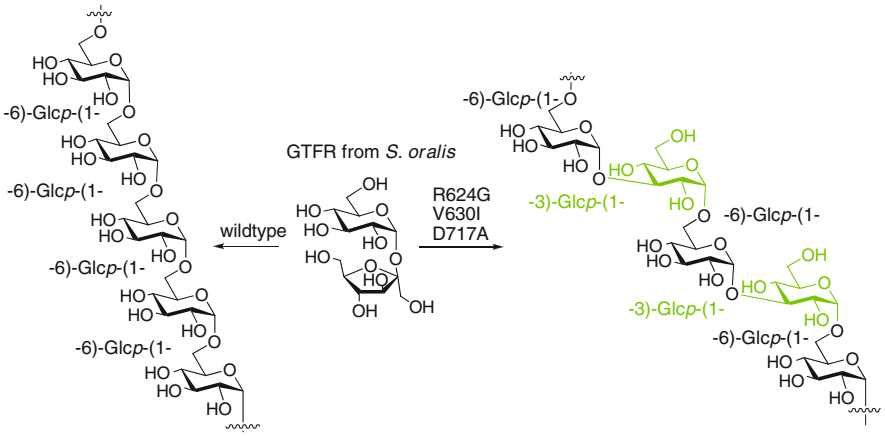


Fig. 4 GTFR variant (R624G/V630I/D717A) forming a polyglucan with increased α-(1,3) linkages instead of α-(1,6) linkages in polysaccharide dextran formed by the wildtype enzyme [52]

thus enhancing glycodiversity. The blockage of the preferred transglycosylation-site by 6-*O*-para-tolouenesulfonyl-group blocked and switched GTFR glycosylation preference from the major α-1,6- to the minor α-1,3-activity (Fig. 5) [56]. These findings were used into a two step chemoenzymatic synthesis where the chemoselectivity of the used glucansucrases can be variably guided from α-1,6 to α-1,2, α-1,3, or α-1,4 linked glucose to the acceptor (blocked in 6-position) followed – in a second step – by a nucleophilic substitution of the tosyl-group by thiosugar-residues (galactose, glucose, neuraminic acid) (Fig. 5). This reaction sequence offers the short and efficient synthesis of a wide repertory of glycostructures [56].

2.2.3 Engineering of Fructansucrase Enzymes

A second interesting industrial class of carbohydrate processing enzymes are fructansucrases (FSs). They are mostly found in lactic acid bacteria and, like GTFs, they also use the cheap substrate sucrose [57]. There are two major types of fructansucrases [58], one producing the homopolymer levan with β-2,6-bound fructose (Fru) units thus called levansucrases (LSs) [59, 60] and the second the inulosucrases (ISs) synthesizing inulin (β-2,1-bound Fru units) [61, 62]. The first detailed kinetic measurements of FTF reactions by Chambert et al. [60, 63] and by Ouarne and Guibert [64] allowed the proposal of enzymatic reaction mechanisms for these enzymes. LSs and ISs are grouped in the glycoside hydrolase family 68 (GH68) and together with GH32 (e.g., invertase, inulinase) they comprise clan GH-J, according to CAZy [65] (www.cazy.org).

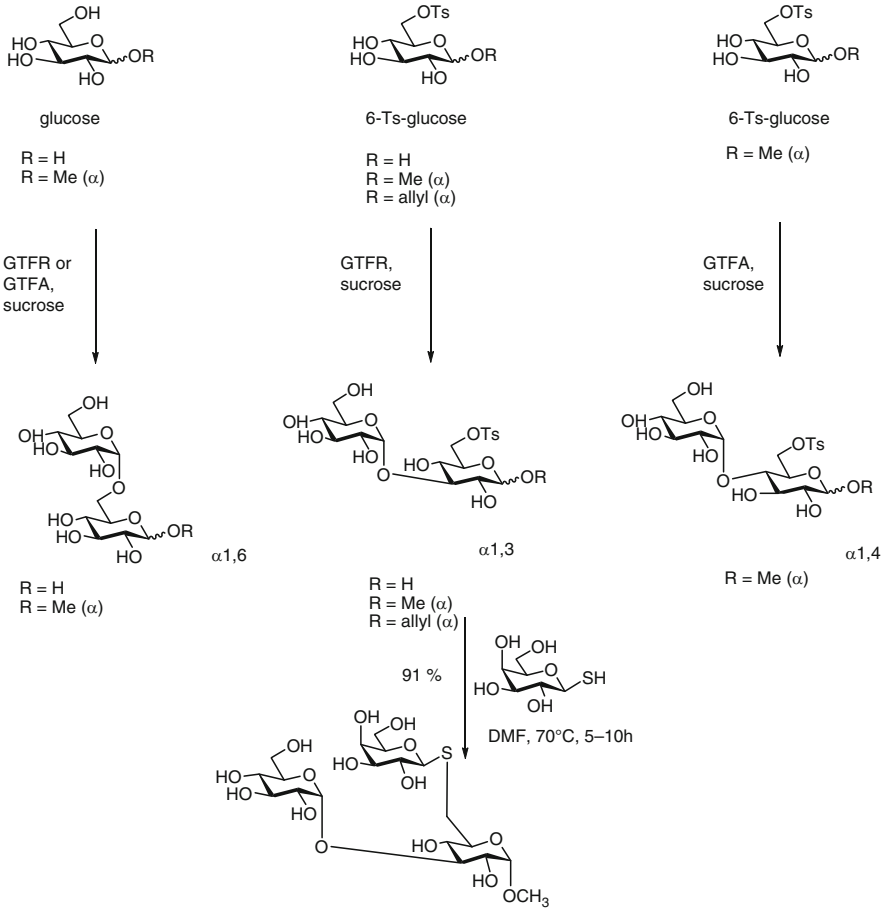


Fig. 5 Acceptor-substrate-directed synthesis by GTFR and GTFA enzymes followed by chemical substitution with thiosugars [56]

So far, two LS X-ray structures have been published, the first from the Gram-positive bacterium *Bacillus subtilis* (SacB, also with bound sucrose and raffinose) [66, 67] and one from the Gram-negative bacterium *Gluconobacter diazotrophicus* [68]. Furthermore, GH32 3D-structures of invertase from *Thermotoga maritima* (also with the trisaccharide raffinose in the active site), fructan 1-exohydrolase from *Cichorum intybus*, exo-inulinase from *Aspergillus niger*, and an E203Q *Arabidopsis* invertase mutant [69–73] have been solved. All protein-structures contain a five-bladed β-propeller fold with a deep, negatively-charged, central cavity with the active site at the end of this cavity. The substrate sucrose is fixed in subsites –1 (fructosyl residue) and +1 (glucosyl residue) [66].

Amino acid sequence alignments of LSs and ISs, site-directed mutagenesis experiments and X-ray structures of such enzymes enabled the functional analysis

of some amino acid residues [66, 68, 74, 75]. Levansucrases constitute a catalytic triad Asp^{95}, Glu^{352}, and Asp^{257} (*Bacillus megaterium* numbering) [75, 76], where Glu^{352} as acid/base catalyst protonates the glycosidic bond of sucrose and Asp^{95} attacks the anomeric carbon to form a covalent enzyme–substrate intermediate with fructose [75]. The role of Asp^{257} is most likely the stabilization of the fructofuranoside at positions 3-OH and 4-OH. While the reaction mechanism has been proposed and experimentally supported in an acceptable way, it is, in contrast, still not known why and how these enzymes discriminate between polysaccharide (PS), oligosaccharide (OS), and hydrolysis reactions. The LSs of *B. subtilis*, *Lactobacillus reuteri* 121 [77] and *B. megaterium* [75] synthesize high-molecular-mass levan [7, 60, 78], but the LS of *G. diazotrophicus*, *Zymomonas mobilis*, and *Lactobacillus sanfranciscensis*, and the IS of *L. reuteri* 121, synthesize short FOSs (kestose and nystose) from sucrose [77–81]. Substitutions of key amino acid residues in LSs like Arg^{360} from *B. subtilis* SacB [82] at subsite –1, and the similar amino acid residue Arg^{370} of *B. megaterium* SacB effected accumulation of neokestose (β-Fru*f*-2,6-α-Glc*p*-1,2-β-Fru*f*) and blastose (β-Fru*f*-2,6-α-Glc*p*) [75]. Also *B. megaterium* LS residue Asn^{252} (subsite +2). Substitution of Asn^{252} to Ala or Gly stopped PS formation without affecting K_m and k_{cat} values and switched from mainly PS synthesis to hydrolysis [75]. From the X-ray structure of the *B. subtilis* SacB (74% amino acid identity with *B. megaterium* LS) it can be seen that Asn^{252} residue may stabilize at subsite +2 the third fructosyl unit of the growing OS chain and direct it as an acceptor substrate into the optimal position for further transfructosylations [75].

Lessons from these results are that carbohydrate processing enzymes can be optimized by site-directed and random mutagenesis in a more rational way for the production of tailored oligo- and polysaccharides under various aspects, also including up- and downstream processing for industrial purposes.

2.2.4 Oligo- and Polysaccharide Synthesis with Sucrose Analogues

A further approach – the synthesis of sucrose analogues – should be realized in order to expand the substrate repertoire of sucrose converting enzymes. Sucrose analogues are structurally similar compounds (isomers) to sucrose, and act as highly activated substrates for the synthesis of novel oligo- and polysaccharides [83]. Docking experiments of sucrose analogues with the levansucrase from *B. subtilis* suggested that sucrose analogues could be synthesized and transformed by LSs. Indeed, the synthesis of sucrose analogues has succeeded (EC 2.4.1.162) with Ls from *B. subtilis* NCIMB 11871 by transferring the fructosyl residue of the substrate sucrose to the monosaccharide D-glycopyranosyl acceptors (D-mannose, D-galactose, 2-deox-D-glucose, D-fucose, D-xylose) to yield the β-D-fructofuranosyl-α-D-glycopyranosides (D-Man-Fru, D-Gal-Fru, D-2-Deoxy-Glc-Fru, D-Fuc-Fru, D-Xyl-Fru) (Fig. 6) [84, 85]. A range of these have been shown to be formed in good, or high yields, with respect to the quasi-equilibrium, under kinetic control and optimum conditions. Docking experiments were performed in order to get an

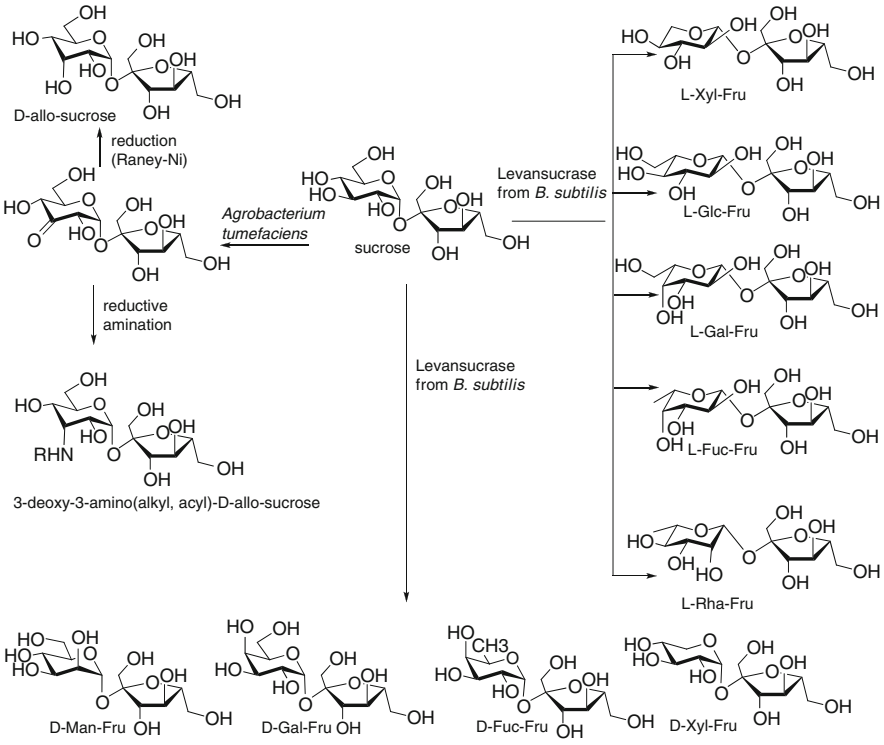


Fig. 6 Different synthetic routes to sucrose analogues

idea of the binding mode of sucrose analogues in the active site of the enzyme. As a result the orientation of D-Gal-Fru was almost identical with the binding mode of sucrose. Thus it was assumed that D-Gal-Fru may function as donor substrate with high efficiency. In contrast the orientation of D-Man-Fru in the active site was not as productive as sucrose. Using L-glycopyranosides as acceptors led to the formation of β-D-fructofuranosyl-β-L-glycopyranoside (L-Glc-Fru, L-Gal-Fru, L-Fuc-Fru L-Xyl-Fru, L-Rha-Fru), sucrose analogues with a β-(1, 2)-glycosidic linkage (Figs. 6 and 7) [84].

Sucrose analogues (β-D-fructofuranosyl-α-D-glycopyranosides) open the gate for the transfer of a wide repertoire of monosaccharides and may be used for OS and PS synthesis (1) using fructansucrases for the synthesis of new glycopyranosyloligofructosides or (2) alternatively using glucansucrases for the transfer of the glycopyranoside [6, 86].

However, the question arose whether such analogues are efficient substrates for useful products. In principal these alternative donor substrates should be highly activated as is sucrose which has a $\Delta G_A^0 = -26.5$ kJ mol^{-1} [87, 88] and the difference in the Gibbs energy change should be usable for the synthesis of

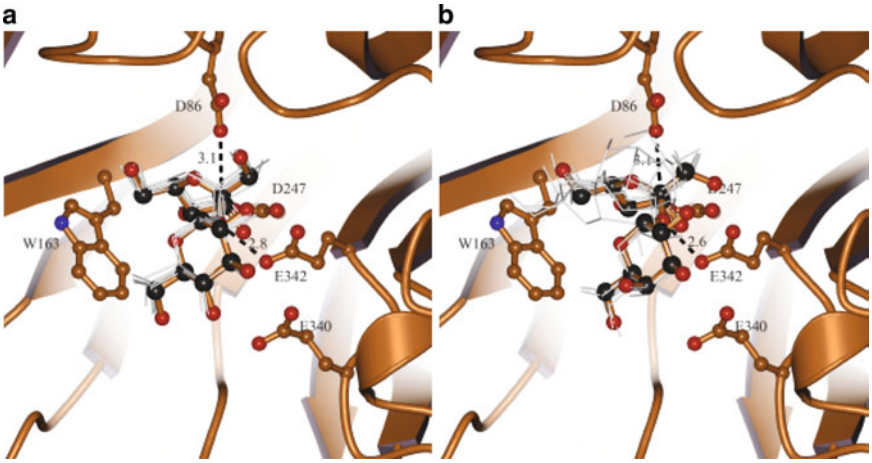


Fig. 7 Lowest energy dockings of the substrates sucrose (*left*) and D-Gal-Fru (*right*) with levansucrase from *B. subtilis* show identical orientation in the active site of the enzyme. Further conformations of D-Gal-Fru docking experiments are also superimposed (*gray*)

oligo- and polysaccharides by sucrase type enzymes for the transfer of the glycopyranoside or fructose. The transfer of different glycopyranosides to other sugars or natural products as acceptors is a challenging perspective. Advantages of this system apart from the currently employed enzymes such as glycosynthases and glycosyltransferases [89, 90] are that industrially established glucansucrases and fructansucrases, or genetically modified variants, may be usable for extended substrate and product spectra.

In initial studies fructosyltransferases were examined, which can use sucrose analogues for the synthesis of a considerable variety of new oligo- and polysaccharides. With the levansucrases from *B. subtilis* and *B. megaterium* the main results were the synthesis of polyfructans (Fig. 7). The products were believed to be not much different from levan since another glycopyranosyl-residue like xylose will not have a major effect for the structural and biochemical properties of the new molecule. As a consequence the levansucrase N242H variant was generated by random mutagenesis that does not produce polysaccharides but instead short-chain FOS (Fig. 8) [7]. The engineered enzyme in combination with the substrate analogue Xyl-Fru facilitated the synthesis of the 6-kestose analogue (αXyl-(1,2)-β-Fru-(2,6)-β-Fru) [7].

Industrially, 1-kestose is produced by β-fructofuranosidase from *A. niger* with sucrose as the substrate. The novel substrate analogues were thus tested with a genetically optimized strain of *A. niger* which overexpressed β-fructofuranosidase (Fig. 8). From the sucrose analogues a highly efficient synthesis of 1-kestose and 1-nystose analogues headed with different monosaccharides of potential interest (Fig. 9) were realized [91]. Kestose analogues will be tested if they provide novel prebiotic activities.

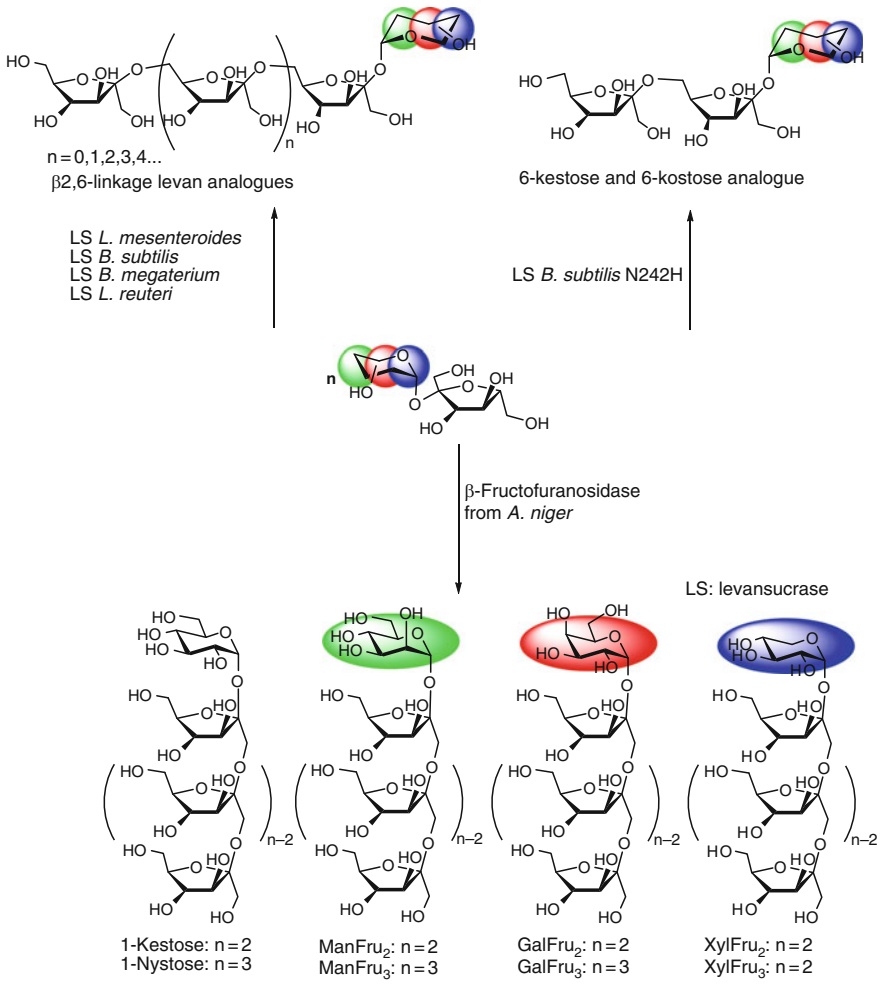


Fig. 8 Different products with sucrose analogues as substrates [91]. Enzymatic synthesis of levans headed with different glycopyranosides by levansucrases of different sources. 6-kestose by a Ls variant N252A, and 1-kestose, 1-nystose analogues by β-fructofuranosidase of *A. niger*

3 From Gene to Product

Process integration with combining downstream operation and biocatalytic reaction and further enzyme engineering to improve the selectivity has been a challenge in recent years [2, 92, 93]. As an example for an advanced integrated process, the enzymatic production of isomaltose from sucrose has been developed. This includes genetic engineering of the enzyme, the choice of biocatalyst design, and optimized reaction conditions. Because of the special characteristics of this reaction

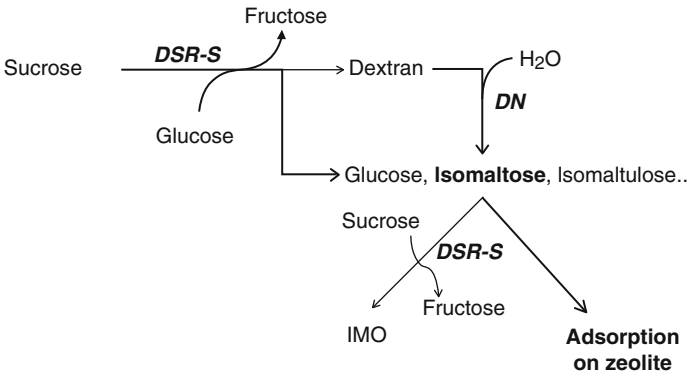


Fig. 9 Scheme of coimmobilizate catalyzed reaction of sucrose. *DSR-S* wildtype dextransucrase; *DN* dextranase; *IMO* isomaltooligosaccharides

(the reaction product isomaltose is a better acceptor than glucose, and thus undergoes subsequent reactions (8, 9, 41)), and a reaction integrated product isolation has also been included.

3.1 Biocatalyst Development (Immobilization, Two Enzyme System)

In principle, several routes exist for enzymatic isomaltose synthesis. With respect to cost-effectiveness it is obvious that one should use substrates like sucrose or starch that can be exploited by dextransucrase (EC 2.4.1.5) or glucoamylase (EC: 3.2.1.3), respectively. However, in both cases, isomaltose represents a side product which is released only in small proportions next to dextran or glucose as main products. Realization of higher yields requires extensive time and effort with respect to engineering of reaction and catalyst design. Based on kinetic investigations dextransucrase has been chosen for the production of isomaltose with sucrose as the substrate, and glucose as an acceptor (see Sect. 2.1).

Bearing in mind technical application with scale-up, enzyme immobilization is a prerequisite. Since the large enzyme (ca. 180 kDa) is additionally linked with a dextran chain, it can be easily and economically immobilized by entrapment in calcium alginate [29, 94].

Coimmobilization as an advanced biocatalytic tool has been investigated since the mid-1980s, when one of the first two enzyme systems, coimmobilized glucose oxidase and catalase, was successfully designed by the Boehringer Manheim company (todays Roche Diagnostics). It was investigated experimentally and theoretically with respect to the dual function, recycling half of the oxygen

consumed in glucose oxidation by degradation of hydrogen peroxide formed as a byproduct, and reducing its deactivation effect on both enzymes [95, 96].

Hartmeyer [97] coined the term second generation catalysts when he undertook attempts to coimmobilize cell/enzyme systems in hydrogels (e.g., agarose, alginate), especially yeasts with coimmobilized glucoamylase, β-galactosidase, or β-glucosidase for production of ethanol from alternative resources like starch, whey, and cellobiose [98–100]. The coimmobilization of two enzymes within alginate beads is less common because only a few enzymes are retained in this matrix. Instead of that, polymers like polystyrene, polyurethane foam, or polyacrylamide have been used [101–103]. The first approach to optimize the isomaltose formation is based on detailed kinetic investigations of acceptor reactions of dextransucrase (see before) by the optimization of substrate and acceptor concentrations, as it can be calculated from the model (Fig. 1). Thus the follow-up reaction of isomaltose glycosylation is reduced significantly. However, this optimization also causes high glucose concentrations in the reaction and product solution, requiring tedious separation-operations. Hence further attempts have to be made for finding the most efficient production system for isomaltose at high concentration and purity.

Therefore a two enzyme system, a combination of dextransucrase (DSR-S, from *L. mesenteroides* B-511F) and dextranase (DN) (EC 3.2.1.11), was studied. DN catalyses the hydrolysis of α-1,6-glycosidic linked oligoglucans and dextran to yield isomaltose as the final product. As is shown in Fig. 9, dextran synthesized by DSR-S can be hydrolyzed and in this way it contributes to an increase in product concentration. A mixture of DSR-S and DN has been applied as soluble enzymes for production of isomaltooligosaccharides [104, 105]. However, application of native soluble enzymes is not an option for technical purposes. Reactions in batch-mode as well as in a membrane reaction lead to oligomers with degrees of polymerization of five or more. Furthermore, Fig. 9 explains the advantage of using reaction integrated product separation with zeolites. By doing so the concentration of isomaltose in solution decreases and cannot therefore be utilized as an acceptor, resulting in production of isomaltooligosaccharides (IMO).

Coimmobilization of both enzymes offers the most efficient approach for this type of reaction and its potential scale-up. Basically, action of DN on the DSR-S: dextran aggregates must be considered if combined operation of both enzymes is intended. Degradation of dextran which is tightly associated with DSR-S occurs by DN and causes complete inactivation of DSR-S [27].

Owing to the low molecular weight of DN its immobilization on a particulate carrier, before inclusion in alginate, was necessary, in order to prevent enzyme bleeding out of the alginate matrix [106, 107].

Several boundary conditions must be taken into account for successful coimmobilization: conservation of the DN activity upon adsorption, relationship of particle size of preimmobilized DN and resulting coimmobilizate matrix, interactions of adsorbent and DSR-S, as well as the activity ratio of both enzymes [108]. The literature-known adsorbent for DN adsorption (bentonite) could not be applied, because it completely inactivates DSR-S activity upon coimmobilization even at

low admixtures (50% decrease of enzyme activity at 0.1% w/v admixture (111). By contrast, hydroxyapatite is almost inert and DSR-S exhibits 50% activity even at 9% w/v hydroxyapatite [106]. Hydroxyapatite exists as a fine-disperse solid and thus allows manufacture of stable coimmobilizates. Hence, this material has been used for further experiments.

Effects of the above-mentioned parameters upon DSR-S activity have been outlined for coimmobilizates based on hydroxyapatite (Fig. 10). With increasing ratio of DN:DSR-S first the DSRS-activity increases up to a certain limit, due to the growing coverage of hydroxyapatite with DN, which hinders inactivation of DSR-S by enzyme adsorbent interaction. Further rise of the activity ratio DN:DSR-S causes degradation of the enzyme associated dextran which leads to inactivation of DS. The optimum activity ratio of both enzymes exhibits a maximum at 1.2 U_{DN}:$U_{DS}{}^{-1}$.

The progress through application of coimmobilizates can be well represented by the increase in yield. Figure 11 clearly shows the positive impact of coimmobilizates, especially at a low ratio of acceptor:substrate. At a molar ratio of glucose to sucrose of 5:1 coimmobilizates produce twice as much product (Y = 35%) as compared to the reference with immobilized DSR-S only (Y = 15%). At low glucose concentration much dextran is produced which serves as an efficient substrate for DN. At high glucose concentration the DSR-S immobilizate not only synthesizes isomaltose, but also concomitantly increased amounts of oligosaccharides with low degree of polymerization are formed. The latter are converted by DN only at low rates. Thus at these high glucose:sucrose ratios molar yield coefficients of both systems approach each other. However, isolation of isomaltose from the product mixture with high glucose concentration is laborious.

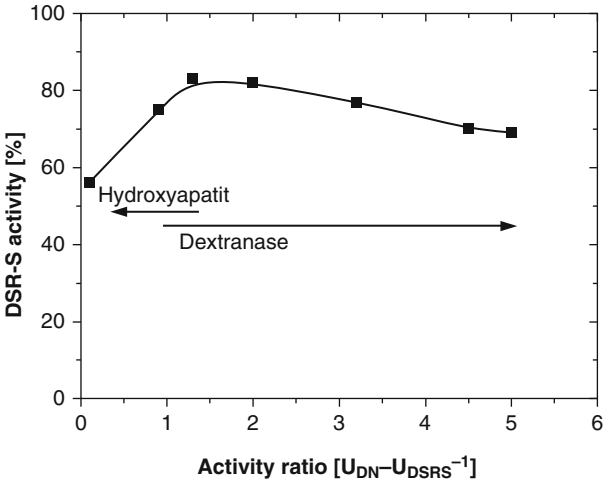


Fig. 10 Effect of enzyme activity ratio between dextranase (DN) and dextransucrase (DSR-S) on DSR-S activity

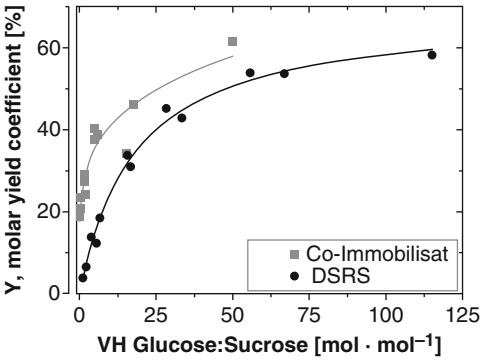


Fig. 11 Molar yield coefficient Y of immobilizate at different acceptor/substrate ratios for the enzyme-coimmobilizate (DSR-S and DN) and DSR-S

3.2 Engineering of GTFR

For the purpose of further improved isomaltose production, the GTFR enzyme was tailored more specifically by random mutagenesis of the gene around the transition state stabilizer in the "RAHDSEV" segment (see Sect. 2.2.1). First, the resulting GTFR mutant library was screened for clones exhibiting high activity with the donor substrate sucrose. Second, positive clones were screened for enzymes which do not produce polysaccharides but instead oligosaccharides using thin layer chromatography (TLC). As a result the modifications at position S628 achieved by saturation mutagenesis guided the reaction towards the synthesis of short chain oligosaccharides with a drastically increased yield of oligosaccharides. [52] No polymer formation was detected. With sucrose only, the S628D-variant synthesized isomaltose in 12% yield, also leucrose synthesis (22%) increased. Additional products, such as palatinose (14%) and several unknown compounds, were formed as well. In initial studies glucose as acceptor substrate, the isomaltose yield (47%) was increased 25-fold as compared to the wild-type enzyme. Addition of fructose as acceptor substrate enhanced the leucrose yield (64%) and palatinose yield (21%) impressively (Fig. 12).

Kinetic studies revealed that the apparent K_m value for sucrose of the S628D variant was comparable to that of the wild-type GTFR. While substitution of serine for arginine in S628R resulted in a 14-fold reduction of activity, the aspartate variant S628D was not significantly affected either in K_m or k_{cat} value. This engineered enzyme showed a favorable potential for technical applications, particularly for production of isomaltose, leucrose or palatinose (isomaltulose) oligosaccharides.

3.3 Application of Engineered Dextransucrase (GTFR)

The genetically modified glucosyltransferase GTFR 72p5611 (GTFR S628D from *S. oralis*, see before) was transferred to the pilot reactor scale and studied with the

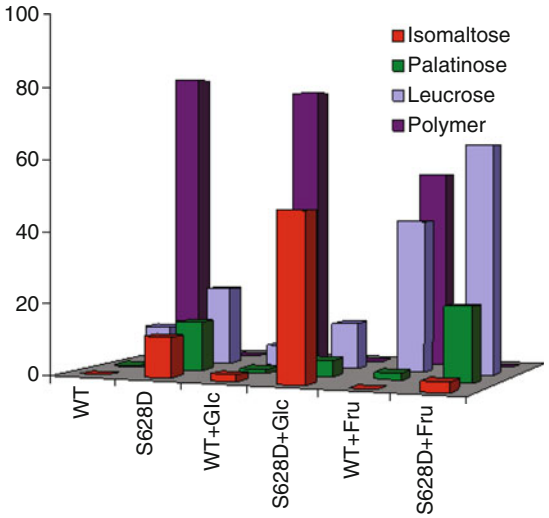


Fig. 12 Product spectra of the wild-type (WT) GTFR and mutant S628D enzymes (200 U/L) incubated (7 days at 30°C) with sucrose (146 mM) and different acceptor substrates (292 mM, Glc: glucose, Fru: fructose). Yields are given in % (mol/mol Glc, Fru). Yields of higher oligosaccharides (DP > 5) and hydrolysis products are not shown [52]

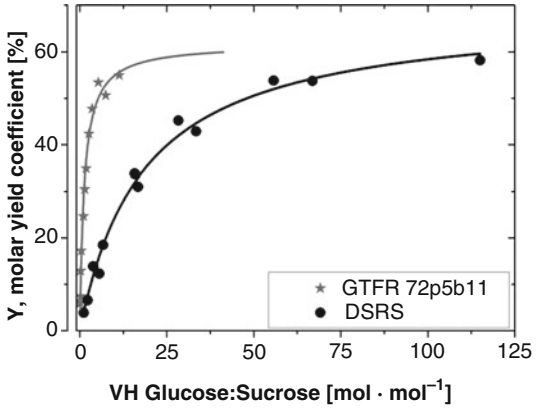


Fig. 13 Molar yield coefficient Y of immobilizate at different acceptor/substrate ratios for the GTFR variant (S628D) and DSR-S

aim to optimize further the reaction yield. The modified GTFR produces with practically no dextran at favorable low acceptor concentration. Without addition of glucose even significant isomaltose formation occurs and excess addition of glucose leads to a further increase in isomaltose formation. Figure 13 outlines that this enzyme provides for a much enhanced affinity for the transglucosylation reaction with glucose, so that high isomaltose yields are obtained at a low ratio of glucose:sucrose. As a characteristic parameter y_{50} can be inferred from the ration of glucose and sucrose, at which half the maximal yield is obtained: commercial DSR-S exhibits a y_{50} of 18.3 while the GTFR enzyme variant exhibits a favorable y_{50} of only 1.2 [3].

Application of this enzyme variant in the bioreactor allows far easier product isolation. Protein engineering turns out to be the most efficient method and also

particularly beneficial for downstream processing – evidence for the necessity of networking of all disciplines from molecular biology up to chemical and process engineering.

3.4 Integrated Reaction Engineering

Previous studies of our work group demonstrated that isomaltose exhibits a distinct higher affinity towards certain dealuminated β-zeolites as opposed to other carbohydrates like fructose or glucose [94, 109]. Sucrose is not adsorbed at all. As a consequence, a process could be developed which directly removes the isomaltose from the reaction solution by adsorption onto zeolite. For this purpose a fluidized bed reactor has been utilized with a special focus on the separation of the two solid phases (Fig. 14). The biocatalyst containing entrapped dextransucrase is produced by the jet-cutter method [110]; the alginate beads have a mean particle size of 0.5 mm. To accomplish an adequate high density of biocatalyst, silica flour (30% w/v) is included. The particle diameter of the second solid phase (zeolite) is adjusted to 10 μm. As a consequence, zeolite is loaded with isomaltose inside the reactor and can then freely exit the reactor together with the product solution, whereas the biocatalyst is retained inside the fluidized bed reactor [92, 94].

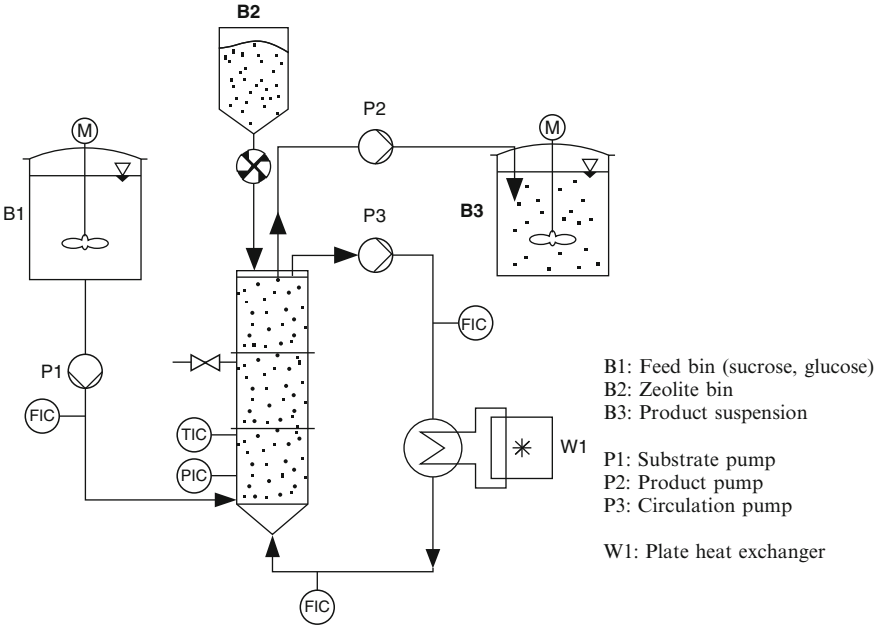


Fig. 14 Process flow sheet of fluidized multiphase bed reactor [92]. (FIC, TIC, PIC: Flow, Temperature, and Pressure Indicator Control)

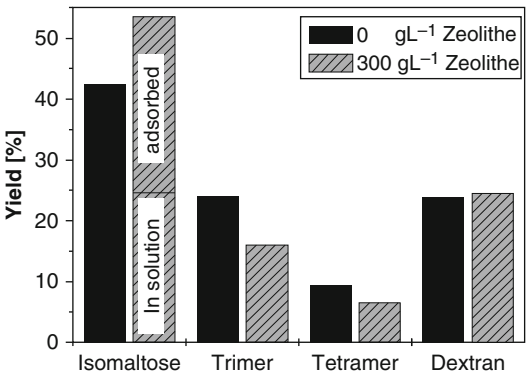


Fig. 15 Increase of yield by application of β-zeolite for in situ product adsorption (c_{suc}: 146 mM, c_{glc}: 2523 mM, according to Ergezinger [92, 111])

A comparison of reaction with and without addition of zeolite (300 g L^{-1}) under otherwise identical conditions in a stirred tank reactor resulted in considerably higher overall isomaltose yields for the system containing zeolite (Fig. 15). Since a major part of isomaltose in the reaction system is withdrawn by adsorption, the lower concentration in solution diminishes formation of glucosylated follow-up products such as isomaltotriose or isomaltotetrose. In general DSR-S exhibits a bimodal product distribution. This can be well explained by the enzyme mechanism; either product is released due to the termination of acceptor reaction (oligomers formation), or glucose is continuously polymerized until the displacement the growing chain occurs [112]. Because of these two mutual exclusive mechanisms, dextran formation is not affected in the presence of zeolite. Analogously, leucrose synthesis remains unaffected as it solely depends on sucrose and fructose concentration.

Glucose acts as acceptor and is consumed during the reaction corresponding to the number of moles of converted sucrose. For the experiment presented in Fig. 15, that implicates a glucose concentration above 2 M even in case of complete sucrose turnover. This high concentration comes along with several adverse effects: obstruction of isomaltose separation from the viscous suspension and difficulties in the downstream processing of zeolite (viscous and highly concentrated reaction solution in void volume).

With a reduction in the glucose content, as is possible with the coimmobilizates (Fig. 11) and/or the genetically engineered GTFR (Fig. 13), the viscosity and concentration of the solution in the void volume also decreases and hence the downstream processing becomes more convenient.

For the separation of the loaded zeolite from the product suspension, centrifugation can be used. The interparticle void volume of zeolite consists of a solution that contains the substrates and fructose as a byproduct. Therefore several washing steps are required. The more washing effort applied, the higher the resulting purity of isomaltose (Fig. 16). However, a rise in isomaltose purity comes along with lower product yields.

With the new enzyme variant the product purity is, even without washing, much higher than with the original utilized DSR-S (35% compared to only 10%). Because

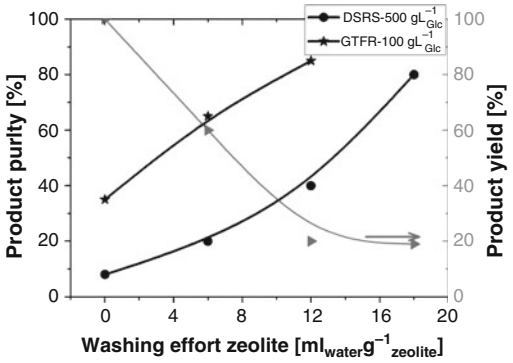


Fig. 16 Downstream processing of zeolite

the purification procedure reduces yield, this also means that higher yields can be obtained in the case of using the GTFR variant. The washing/separation procedure can be optimized furthermore by introducing a countercurrent washing procedure as suggested by Holtkamp et al. [113].

4 Conclusions

In the framework of the collaborative research centre SFB 578 integrated production and isolation of isomaltose in a multiphase fluidized bed reactor have been achieved in a case where the product to be synthesized has been a byproduct of wild type enzymes formed at low concentration. Focal points of studies were optimization of reaction conditions, the biocatalytic system, as well as downstream processing. Due to the interdisciplinary orientation of this project it became possible to establish an optimal reactor configuration. This involves a specific reactor setup (fluidized bed reactor with a second solid phase within the suspension) as well as process and biocatalyst optimization starting first with a wild type enzyme, then developing a coimmobilized biocatalyst with two enzymes, and finally designing and applying a genetically engineered enzyme. In further studies the developed platform will be extended to other processes.

Acknowledgement This project is supported by the German Research Foundation via SFB 578 "From Gene to Product".

References

1. Deckwer W-D, Jahn D, Hempel D, Zeng A-D (2006) Systems biology approaches to bioprocess development. Eng Life Sci 6:455–469
2. Hempel DC (2006) Development of biotechnological processes by integrating genetic and engineering methods. Eng Life Sci 5:443–447

3. Jördening H-J, Erhardt FA, Holtkamp M, Buchholz K, Scholl S (2008) Verfahrens- und Katalysatordesign als Aufarbeitungsstrategie für die enzymatische Darstellung von Isomaltose. Chem Ing Tech 80:867–874
4. Varki A (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3:97–130
5. Wong CH (2005) Protein glycosylation: new challenges and opportunities. J Org Chem 70:4219–4225
6. Seibel J, Buchholz K, Jördening HJ (2006) Glycosylation with activated sugars using glycosyltransferases and transglycosidases. Biocatal Biotrans 24:311–342
7. Beine R, Moraru R, Nimtz M, Na'amnieh S, Pawlowski A, Buchholz K, Seibel J (2008) Synthesis of novel fructooligosaccharides by substrate and enzyme engineering. J Biotechnol 138:33–41
8. Robyt JF, Walseth TF (1978) The mechanism of acceptor reactions of *Leuconostoc mesenteroides* B-512F dextransucrase. Carbohydr Res 61:433–445
9. Robyt JF, Eklund SH (1982) Stereochemistry involved in the mechanism of action of dextransucrase in the synthesis of dextran and the formation of acceptor products. Bioorg Chem 11:115–132
10. Shamala TR, Prasad MS (1995) Preliminary studies on the production of high and low viscosity dextran by *Leuconostoc spp*. Proc Biochem 30:237–241
11. Quirasco M, Lopez-Munguia A, Pelenc V, Remaud M, Paul F, Monsan P (1995) Enzymatic production of glucooligosaccharides containing alpha-(1,2) glycosidic bonds. Potential application in nutrition. Ann N Y Acad Sci 750:317–320
12. Remaud-Simeon M, Lopez-Munguia A, Pelenc V, Paul F, Monsan P (1994) Production and use of glucosyltransferases from *Leuconostoc mesenteroides* NRRL B-1299 for the synthesis of oligosaccharides containing alpha-(1,2) linkages. Appl Biochem Biotech 44:101–117
13. Olano-Martin E, Mountzouris KC, Gibson GR, Rastall RA (2000) In vitro fermentability of dextran, oligodextran and maltodextrin by human gut bacteria. Brit J Nutr 83:247–255
14. Dols M, Remaud-Siméon M, Willemot RM, Vignon M, Monsan P (1998) Structural characterization of the maltose acceptor products synthesized by *Leuconostoc mesenteroides* NRRL B-1299 dextransucrase. Carbohydr Res 305:549–559
15. Berensmeier S, Jördening H-J, Buchholz K (2006) Isomaltose formation by free and immobilised dextransucrase. Biocata Biotrans 24:280–290
16. Ebert KH, Schenk G, Stricker H (1964) Ueber den Mechanismus des Aufbauschrtittes der enzymatischen Dextran- und Laevanbildung. Ber Bunsenges 68:765–767
17. Ebert KH, Patat F (1962) Kinetische Betrachtungen ueber die enzymatische Dextransynthese. Z Naturforschung 17b:738–748
18. Hehre EJ, Hamilton DM (1946) Bacterial synthesis of an amylopectin-like polysacchride from sucrose. J Biol Chem 166:77–78
19. Mayer RM, Matthews MM, Futerman CL, Parnaik VK, Jung SM (1981) Dextransucrase: acceptor substrate reactions. Arch Biochem Biophys 208:278–287
20. Demuth B, Jordening HJ, Buchholz K (1999) Modelling of oligosaccharide synthesis by dextransucrase. Biotechnol Bioeng 62:583–592
21. Boker M, Jordening HJ, Buchholz K (1994) Kinetics of leucrose formation from sucrose by dextransucrase. Biotechnol Bioeng 43:856–864
22. Heincke K, Demuth B, Joerdening H-J, Buchholz K (1999) Kinetics of the dextransucrase acceptor reaction with maltose: experimental results and modeling. Enz Microb Technol 24:523–534
23. Reh K-D, Noll-Borchers M, Buchholz K (1996) Productivity of immobilized dextransucrase for leucrose formation. Enz Microb Technol 19:518–524
24. Schwengers D (1991) Leucrose, a ketodisaccharide of industrial design. In: Lichtenthaler FW (ed) Carbohydrates as organic raw materials. VCH, Weinheim, pp 183–195
25. Buchholz K, Noll-Borchers M, Schwengers D (1998) Production of leucrose by dextransucrase. Starch/Stärke 50:164–172

26. Dols M, Remaud-Simeon M, Willemot RM, Demuth B, Jördening HJ, Buchholz K, Monsan P (1999) Kinetic modeling of oligosaccharide synthesis catalyzed by *Leuconostoc mesenteroides* NRRL B-1299 dextransucrase. Biotechnol Bioeng 63:308–315
27. Robyt JF (1995) Mechanism in the glucansucrase synthesis of polysaccharides and oligosaccharides from sucrose. Adv Carbohydr Chem Biochem 51:133–168
28. Demuth K, Jördening H-J, Buchholz K (2000) Oligosaccharide synthesis with dextransucrase. Bielecki S, Tramper J, Polak J (eds) Food Biotechnology. Elsevier Science, Amsterdam, pp 123–135
29. Reischwitz A, Reh K-D, Buchholz K (1995) Unconventional immobilization of dextransucrase with alginate. Enzyme Microb Technol 17:457–461
30. Berensmeier S, Ergezinger M, Bohnet M, Buchholz K (2004) Design of immobilised dextransucrase for fluidised bed application. J Biotechnol 114:255–267
31. Monchois V, Willemot RM, Monsan P (1999) Glucansucrases: mechanism of action and structure-function relationships. FEMS Microbiol Rev 23:131–151
32. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280(Pt 2):309–316
33. van Hijum SA, Kralj S, Ozimek LK, Dijkhuizen L, van Geel-Schutten IG (2006) Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. Microbiol Mol Biol Rev 70:157–176
34. MacGregor EA, Jespersen HM, Svensson B (1996) A circularly permuted alpha-amylase-type alpha/beta-barrel structure in glucan-synthesizing glucosyltransferases. FEBS Lett 378:263–266
35. Pijning T, Vujicic-Zagar A, Kralj S, Eeuwema W, Dijkhuizen L, Dijkstra BW (2008) Biochemical and crystallographic characterization of a glucansucrase from *Lactobacillus reuteri* 180. Biocatal Biotrans 26:12–17
36. Buchholz K (2008) Introduction: the spirit of carbohydrates – carbohydrate bioengineering. Biocatal Biotransform 26:3–11
37. Lacaze G, Wick M, Cappelle S (2007) Emerging fermentation technologies: development of novel sourdoughs. Food Microbiol 24:155–160
38. Coviello T, Matricardi P, Alhaique F (2006) Drug delivery strategies using polysaccharidic gels. Expert opinion on drug delivery 3:395–404
39. Mooser G, Hefta SA, Paxton RJ, Shively JE, Lee TD (1991) Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two *Streptococcus sobrinus* alpha-glucosyltransferases. J Biol Chem 266:8916–8922
40. Kato C, Nakano Y, Lis M, Kuramitsu HK (1992) Molecular genetic analysis of the catalytic site of *Streptococcus mutans* glucosyltransferases. Biochem Biophys Res Commun 189:1184–1188
41. Monchois V, Remaud-Simeon M, Russell RR, Monsan P, Willemot RM (1997) Characterization of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase (DSRS) and identification of amino-acid residues playing a key role in enzyme activity. Appl Microbiol Biotechnol 48:465–472
42. Kralj S, van Geel-Schutten GH, van der Maarel MJ, Dijkhuizen L (2004) Biochemical and molecular characterization of Lactobacillus reuteri 121 reuteransucrase. Microbiology (Reading, England) 150:2099–2112
43. Devulapalle KS, Goodman SD, Gao Q, Hemsley A, Mooser G (1997) Knowledge-based model of a glucosyltransferase from the oral bacterial group of mutans streptococci. Protein Sci 6:2489–2493
44. Seibel J, Hellmuth H, Hofer B, Kicinska AM, Schmalbruch B (2006) Identification of new acceptor specificities of glycosyltransferase R with the aid of substrate microarrays. Chembiochem 7:310–320
45. Swistowska AM, Gronert S, Wittrock S, Collisi W, Hecht HJ, Hofer B (2007) Identification of structural determinants for substrate binding and turnover by glucosyltransferase R supports the permutation hypothesis. FEBS Lett 581:4036–4042

46. Seibel J, Moraru R, Gotze S (2005) Biocatalytic and chemical investigations in the synthesis of sucrose analogues. Tetrahedron 61:7081–7086
47. Monchois V, Vignon M, Russell RR (2000) Mutagenesis of asp-569 of glucosyltransferase I glucansucrase modulates glucan and oligosaccharide synthesis. Appl Environ Microbiol 66:1923–1927
48. Shimamura A, Nakano YJ, Mukasa H, Kuramitsu HK (1994) Identification of amino acid residues in *Streptococcus mutans* glucosyltransferases influencing the structure of the glucan product. J Bacteriol 176:4845–4850
49. Remaud-Simeon M, Willemot R, Sarcabal P, Potocki de Montalk G, Monsan P (2000) Glucansucrases: molecular engineering and oligosaccharide synthesis. J Mol Catal B Enzym 10:117–128
50. Kralj S, van Geel-Schutten IG, Faber EJ, van der Maarel MJ, Dijkhuizen L (2005) Rational transformation of Lactobacillus reuteri 121 reuteransucrase into a dextransucrase. Biochemistry 44:9206–9216
51. Kralj S, Eeuwema W, Eckhardt TH, Dijkhuizen L (2006) Role of asparagine 1134 in glucosidic bond and transglycosylation specificity of reuteransucrase from Lactobacillus reuteri 121. FEBS J 273:3735–3742
52. Hellmuth H, Wittrock S, Kralj S, Dijkhuizen L, Hofer B, Seibel J (2008) Engineering the glucansucrase GTFR enzyme reaction and glycosidic bond specificity: toward tailor-made polymer and oligosaccharide products. Biochemistry 47:6678–6684
53. Ferretti JJ, Gilpin ML, Russell RR (1987) Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. J Bacteriol 169:4271–4278
54. Kralj S, van Geel-Schutten GH, Rahaoui H, Leer RJ, Faber EJ, van der Maarel MJ, Dijkhuizen L (2002) Molecular characterization of a novel glucosyltransferase from Lactobacillus reuteri strain 121 synthesizing a unique, highly branched glucan with alpha-(1,4) and alpha-(1,6) glucosidic bonds. Appl Environ Microbiol 68:4283–4291
55. Fujiwara T, Hoshino T, Ooshima T, Sobue S, Hamada S (2000) Purification, characterization, and molecular analysis of the gene encoding glucosyltransferase from *Streptococcus oralis*. Infec Immun 68:2475–2483
56. Hellmuth H, Hillringhaus L, Hobbel S, Kralj S, Dijkhuizen L, Seibel J (2007) Highly efficient chemoenzymatic synthesis of novel branched thiooligosaccharides by substrate direction with glucansucrases. Chembiochem 8:273–276
57. Cerning J (1990) Exocellular polysaccharides produced by lactic acid bacteria. FEMS Microbiol Rev 7:113–130
58. Yun JW (1996) Fructooligosaccharides – occurence, preparation and application. Enz Microb Technol 19:107–117
59. van Hijum SA, Bonting K, van der Maarel MJ, Dijkhuizen L (2001) Purification of a novel fructosyltransferase from *Lactobacillus reuteri* strain 121 and characterization of the levan produced. FEMS Microbiol Lett 205:323–328
60. Chambert R, Treboule G, Dedonder R (1974) Kinetic studies of levansucrase of *Bacillus subtilis*. Eur J Biochem 41:285–300
61. van Hijum SAFT, Kralj S, Ozimek LK, Dijkhuizen L, van Geel-Schutten IGH (2006) Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. Microbiol Mol Biol Rev 70:157–176
62. van Hijum SA, van Geel-Schutten GH, Rahaoui H, van der Maarel MJ, Dijkhuizen L (2002) Characterization of a novel fructosyltransferase from *Lactobacillus reuteri* that synthesizes high-molecular-weight inulin and inulin oligosaccharides. Appl Environ Microbiol 68:4390–4398
63. Chambert R, Gonzy-Treboul G (1976) Levansucrase of *Bacillus subtilis*: kinetic and thermodynamic aspects of transfructosylation processes. FEBS 62:55–64
64. Ouarne F, Guibert A (1995) Fructo-oligosaccharides: enzymatic synthesis from sucrose. Zuckerindustrie 120:793–798
65. CAZy (2009) Carbohydrate-active enzymes, online database, http://www.cazy.org. 7 Sep

66. Meng G, Fütterer K (2003) Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. Nat Struc Biol 10:935–941
67. Meng G, Futterer K (2008) Donor substrate recognition in the raffinose-bound E342A mutant of fructosyltransferase *Bacillus subtilis* levansucrase. BMC Struct Biol 8:16
68. Martinez-Fleites C, Ortiz-Lombardia M, Pons T, Tarbouriech N, Taylor EJ, Arrieta JG, Hernandez L, Davies GJ (2005) Crystal structure of levansucrase from the Gram-negative bacterium Gluconacetobacter diazotrophicus. Biochem J 390:19–27
69. Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J Biol Chem 279:18903–18910
70. Verhaest M, Ende WV, Roy KL, De Ranter CJ, Laere AV, Rabijns A (2005) X-ray diffraction structure of a plant glycosyl hydrolase family 32 protein: fructan 1-exohydrolase IIa of Cichorium intybus. Plant J 41:400–411
71. Nagem RA, Rojas AL, Golubev AM, Korneeva OS, Eneyskaya EV, Kulminskaya AA, Neustroev KN, Polikarpov I (2004) Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. J Mol Biol 344:471–480
72. Alberto F, Jordi E, Henrissat B, Czjzek M (2006) Crystal structure of inactivated Thermotoga maritima invertase in complex with the trisaccharide substrate raffinose. Biochem J 395:457–462
73. Matrai J, Lammens W, Jonckheer A, Le Roy K, Rabijns A, Van den Ende W, De Maeyer M (2008) An alternate sucrose binding mode in the E203Q Arabidopsis invertase mutant: an X-ray crystallography and docking study. Proteins 71:552–564
74. Ozimek LK, Kralj S, Kaper T, van der Maarel MJ, Dijkhuizen L (2006) Single amino acid residue changes in subsite−1 of inulosucrase from Lactobacillus reuteri 121 strongly influence the size of products synthesized. FEBS J 273:4104–4113
75. Homann A, Biedendieck R, Gotze S, Jahn D, Seibel J (2007) Insights into polymer versus oligosaccharide synthesis: mutagenesis and mechanistic studies of a novel levansucrase from Bacillus megaterium. Biochem J 407:189–198
76. Ozimek LK, van Hijum SA, van Koningsveld GA, van Der Maarel MJ, van Geel-Schutten GH, Dijkhuizen L (2004) Site-directed mutagenesis study of the three catalytic residues of the fructosyltransferases of *Lactobacillus reuteri* 121. FEBS Lett 560:131–133
77. Ozimek LK, Kralj S, van der Maarel MJ, Dijkhuizen L (2006) The levansucrase and inulosucrase enzymes of Lactobacillus reuteri 121 catalyse processive and non-processive transglycosylation reactions. Microbiology 152:1187–1196
78. Hernandez L, Arrieta J, Menendez C, Vazquez R, Coego A, Suarez V, Selman G, Petit-Glatron MF, Chambert R (1995) Isolation and enzymatic properties of levansucrase secreted by *Acetobacter diazotrophicus* SRT4, a bacterium associated with sugar cane. Biochem J 309 (Pt 1):113–118
79. Korakli M, Rossmann A, Ganzle MG, Vogel RF (2001) Sucrose metabolism and exopolysaccharide production in wheat and rye sourdoughs by *Lactobacillus sanfranciscensis*. J Agric Food Chem 49:5194–5200
80. Korakli M, Pavlovic M, Ganzle MG, Vogel RF (2003) Exopolysaccharide and kestose production by Lactobacillus sanfranciscensis LTH2590. Appl Environ Microbiol 69:2073–2079
81. Doelle HW, Kirk L, Crittenden R, Toh H, Doelle MB (1993) Zymomonas mobilis – science and industrial application. Crit Rev Biotechnol 13:57–98
82. Chambert R, Petit-Glatron MF (1991) Polymerase and hydrolase activities of *Bacillus subtilis* levansucrase can be separately modulated by site-directed mutagenesis. Biochem J 279:35–41
83. Kralj S, Buchholz K, Dijkhuizen L, Seibel J (2008) Fructansucrase enzymes and sucrose analogues: new approach for the synthesis of unique fructo-oligosaccharides. Biocatal Biotransform 26:32–41

84. Seibel J, Moraru R, Gotze S, Buchholz K, Na'amnieh S, Pawlowski A, Hecht HJ (2006) Synthesis of sucrose analogues and the mechanism of action of *Bacillus subtilis* fructosyltransferase (levansucrase). Carbohydr Res 341:2335–2349
85. Cheetam PSJ, Hacking AJ, Vlitos M (1989) Synthesis of novel disaccharides by a newly isolated fructosyltransferase from *Bacillus subtilis*. Enz Microb Technol 11:212–219
86. Seibel J, Beine R, Moraru R, Behringer C, Buchholz K (2006) A new pathway for the synthesis of oligosaccharides by the use of non-Leloir glycosyltransferases. Biocatal Biotrans 24:157–165
87. Goldberg R, Tewari J (1989) Thermodynamic and transport properties of carbohydrates and their monophosphates: the pentoses and hexoses. J Phys Chem Ref Data 18:809–822
88. Goldberg R, Tewari J, Ahluwalia J (1989) Thermodynamics of the hydrolysis of sucrose. J Biol Chem 264:9901–9904
89. Hancock SM, Vaughan MD, Withers SG (2006) Engineering of glycosidases and glycosyltransferases. Curr Opin Chem Biol 10:509–519
90. Trincone A, Giordano A (2006) Glycosyl hydrolases and glycosyltransferases in the synthesis of oligosaccharides. Curr Org Chem 10:1163
91. Zuccaro A, Götze S, Kneip S, Dersch P, Seibel J (2008) Tailor-made fructooligosaccharides by a combination of substrate and genetic engineering. Chembiochem 9:143–149
92. Ergezinger M, Bohnet M, Berensmeier S, Bucholz K (2006) Integrated enzymatic synthesis and adsorption of Isomaltose in a multiphase fluidized bed reactor. Eng Life Sci 5:1–8
93. Buchholz K, Kasche V, Bornscheuer U (2005) Biocatalysts and enzyme technology. Wiley-VCH, Weinheim
94. Berensmeier S, Buchholz K (2004) Separation of isomaltose from high sugar concentrated enzyme reaction mixture by dealuminated β-zeolite. Sep Purif Technol 38:129
95. Buchholz K, Gödelmann B (1978) Macrokinetics and operational stability of immobilized glucose oxidase and catalase. Biotechnol Bioeng 20:1201–1220
96. Reuss M, Buchholz K (1979) Analysis of the coupled transport, reaction and deactivation phenomena in the immobilized system glucose oxidase and catalase. Biotechnol Bioeng 21:2061–2081
97. Hartmeyer W (1985) Immobilized biocatalysts – from simple to complex systems. Trends Biotechnol 3:149–153
98. Förster U (1986) Co-Immobilisierung von Saccharomyces cerevisiae und beta-Galactosidase zur Ethanolherstellung aus Molke. PhD-Thesis, Mathematisch-Naturwissenschaftliche Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen
99. Hahn-Hägerdal B (1984) An enzyme coimmoilized with a microorganism: the conversion of cellobiose to ethanol using [beta]-glucosidase and Saccaromyces cerevisiae in calcium alginate gels. Biotechnol Bioeng 26:365–385
100. Giordano RL, Trovati J, Schmidell W (2008) Continuous production of ethanol from starch using glucoamylase and yeast co-immobilized in pectin gel. Appl Biochem Biotechnol 147:47–61
101. Ge Y, Wang Y, Zhou H, Wang S, Tong Y, Li W (1999) Coimmobilization of glucoamylase and glucose isomerase by molecular deposition technique for one-step conversion of dextrin to fructose. J Biotechnol 67:33–40
102. van de Velde F, Lourenco ND, Bakker M, van Rantwijk F, Sheldon RA (2000) Improved operational stability of peroxidases by coimmobilization with glucose oxidase. Biotechnol Bioeng 69:286–291
103. Atia KS, Ismail SA, El-Arnaouty MB, Dessouki AM (2003) Use of coimmobilized beta-amylase and pullulanase in reduction of saccharification time of starch and increase in maltose yield. Biotechnol Progr 19:853–857
104. Goulas AK, Cooper JM, Grandison AS, Rastall RA (2004) Synthesis of isomaltooligosaccharides and oligodextrans in a recycle membrane bioreactor by the combined use of dextransucrase and dextranase. Biotechnol Bioeng 88:778–787

105. Goulas AK, Fisher DA, Grimble GK, Grandison AS, Rastall RA (2004) Synthesis of isomaltooligosaccharides and oligodextrans by the combined use of dextransucrase and dextranase. Enzyme Microb Technol 35:327–338
106. Erhardt FA, Stammen S, Jordening HJ (2008) Production, characterization and (co-)immobilization of dextranase from Penicillium aculeatum. Biotechnol Lett 30:1069–1073
107. Erhardt FA, Jordening HJ (2007) Immobilization of dextranase from Chaetomium erraticum. J Biotechnol 131:440–447
108. Erhardt FA, Kugler J, Chakravarthula RR, Jordening HJ (2008) Co-immobilization of dextransucrase and dextranase for the facilitated synthesis of isomalto-oligosaccharides: preparation, characterization and modeling. Biotechnol Bioeng 100:673–683
109. Buttersack C, Fornefett I, Mahrholz J, Buchholz K (1997) Specific adsorption from aqueous phase on apolar zeolites in surface science and calatalysis. In: Chon H, Ihm SK, Uh YS (eds) Studies in surface science and calatalysis. Elsevier, Amsterdam, pp 1723–1730
110. Prüße U, Fox B, Kirchhoff M, Bruske F, Breford J, Vorlop K-D (1998) The jet cutting method as a new immobilization technique. Biotechnol Tech 12:105
111. Ergezinger M (2006) Integrierte enzymatische Produktion und Adsorption von Isomaltose in einem Mehrphasenreaktor. Phd-thesis, Technische Universität Braunschweig
112. Kim D, Robyt JF, Lee SY, Lee JH, Kim YM (2003) Dextran molecular size and degree of branching as a function of sucrose concentration, pH, and temperature of reaction of *Leuconostoc mesenteroides* B-512FMCM dextransucrase. Carbohydr Res 338:1183–1189
113. Holtkamp M, Erhardt FA, Jördening H-J, Scholl S (2009) Reaction-integrated separation of isomaltose by ad- and desorption on zeolite. Chem Eng Process 48:852–858

Adv Biochem Engin/Biotechnol (2010) 120: 195–242
DOI: 10.1007/10_2009_55

Published online: 17 February 2010

Regeneration of Nicotinamide Coenzymes: Principles and Applications for the Synthesis of Chiral Compounds

Andrea Weckbecker, Harald Gröger, and Werner Hummel

Abstract Dehydrogenases which depend on nicotinamide coenzymes are of increasing interest for the preparation of chiral compounds, either by reduction of a prochiral precursor or by oxidative resolution of their racemate. The regeneration of oxidized and reduced nicotinamide cofactors is a very crucial step because the use of these cofactors in stoichiometric amounts is too expensive for application. There are several possibilities to regenerate nicotinamide cofactors: established methods such as formate/formate dehydrogenase (FDH) for the regeneration of NADH, recently developed electrochemical methods based on new mediator structures, or the application of gene cloning methods for the construction of "designed" cells by heterologous expression of appropriate genes.

A very promising approach is enzymatic cofactor regeneration. Only a few enzymes are suitable for the regeneration of oxidized nicotinamide cofactors. Glutamate dehydrogenase can be used for the oxidation of NADH as well as NADPH while L-lactate dehydrogenase is able to oxidize NADH only. The reduction of NAD^+ is carried out by formate and FDH. Glucose-6-phosphate dehydrogenase and glucose dehydrogenase are able to reduce both NAD^+ and $NADP^+$. Alcohol dehydrogenases (ADHs) are either NAD^+- or $NADP^+$-specific. ADH from horse liver, for example, reduces NAD^+ while ADHs from *Lactobacillus* strains catalyze the reduction of $NADP^+$. These enzymes can be applied by their

A. Weckbecker and W. Hummel (✉)
Institute of Molecular Enzyme Technology, Heinrich-Heine-University of Düsseldorf, Research Centre Jülich, Stetternicher Forst, 52426, Jülich, Germany
e-mail: w.hummel@fz-juelich.de

H. Gröger
Department of Chemistry and Pharmacy, University of Erlangen-Nuremberg, Henkestr. 42, 91054, Erlangen, Germany

A. Weckbecker
evocatal GmbH, Merowingerplatz 1a, 40225, Düsseldorf, Germany

inclusion in whole cell biotransformations with an $NAD(P)^+$-dependent primary reaction to achieve in situ the regeneration of the consumed cofactor.

Another efficient method for the regeneration of nicotinamide cofactors is the electrochemical approach. Cofactors can be regenerated directly, for example at a carbon anode, or indirectly involving mediators such as redox catalysts based on transition-metal complexes.

An increasing number of examples in technical scale applications are known where nicotinamide dependent enzymes were used together with cofactor regenerating enzymes.

Keywords Chiral compounds, Dehydrogenases, Enzymatic reduction, Nicotinamide coenzymes, Regeneration of coenzymes

Contents

1 Introduction 197
2 Enzymatic Methods to Regenerate Reduced Nicotinamide Cofactors 199
2.1 Formate Dehydrogenase 199
2.2 Glucose-6-Phosphate Dehydrogenase 201
2.3 Glucose Dehydrogenase 202
2.4 Alcohol Dehydrogenase 203
2.5 Hydrogenase 204
2.6 Phosphite Dehydrogenase 205
2.7 Further Enzyme Systems for the Regeneration of Reduced Nicotinamide Cofactors 205
3 Enzymatic Methods to Regenerate Oxidized Nicotinamide Cofactors 206
3.1 Glutamate Dehydrogenase 207
3.2 L-Lactate Dehydrogenase 208
3.3 Alcohol Dehydrogenase 208
3.4 NADH Oxidase 209
4 Electrochemical Methods 210
4.1 Electroenzymatic Oxidations 211
4.2 Electroenzymatic Reductions 213
5 Further Methods 220
5.1 Photoelectrochemical Cofactor Regeneration 220
5.2 Chemical Cofactor Regeneration 220
6 Whole Cell Applications Based on Recombinant Enzymes ("Designer Cells") 222
7 Coenzyme-Dependent Reactions in Organic Solvents 223
8 The Nicotinamide Dinucleotide Binding Motif and Its Reversal by Protein Engineering 227
9 Examples of Small-Scale and Industrial Applications 228
9.1 L-*tert*-Leucine as an Example of Ton-Scale Synthesis 228
9.2 L-6-Hydroxynorleucine 228
9.3 Synthesis of Allysine Ethylene Acetal in 80 kg-Scale 230
9.4 Synthesis of Chiral Alcohols 231
9.5 Alcohol Dehydrogenase Catalyzed Synthesis of the Chiral Side Chain of Statins 232
9.6 Synthesis of Optically Active L-Amino Acids with Amino Acid Dehydrogenases 234
9.7 Optically Active α-Hydroxy Acids or α-Hydroxy Esters 234

9.8 Multistep "One-Pot" Processes: Direct Coupling of a Chemical and an Enzymatic Step 235
10 Conclusions: Capability and Limitations 235
References 238

Abbreviations

ADH	Alcohol dehydrogenase
FDH	Formate dehydrogenase
GDH	Glucose dehydrogenase
GluDH	Glutamate dehydrogenase
HLADH	Horse liver alcohol dehydrogenase
LDH	Lactate dehydrogenase
LeuDH	Leucine dehydrogenase
PTDH	Phosphite dehydrogenase
TBADH	Alcohol dehydrogenase from *Thermoanaerobacter brockii*

1 Introduction

The number of enzymes for industrial synthetic applications is growing fast. Enzymatic synthesis can be performed under mild reaction conditions so that many problems of chemical synthesis like isomerization or racemization can be prevented. Furthermore, enzymes are highly specific and selective, especially for enantio- or regio-selective introduction of functional groups. For the preparation of chiral enantiopure compounds, the resolution of racemic mixtures by hydrolases is a well-established route, which has the advantage to be able to use enzymes free of coenzymes. Otherwise, only a maximum yield of 50% can be reached by the primary reaction and further steps of reracemization must follow to avoid loss of the undesired enantiomer.

A lot of these feasible enzymes are cofactor dependent. Cofactors are low-molecular compounds which are responsible for example the transfer of hydrogen, electrons or functional groups in enzyme-catalyzed reactions. Examples for cofactors are ATP, FAD, coenzyme A, or the nicotinamide cofactors NAD^+ and $NADP^+$. In contrast to FAD and FMN, neither NAD^+ nor $NADP^+$ are covalently bound to the enzyme. Coenzymes are modified during the reaction, and they are converted stoichiometrically. Particularly the $NAD(P)^+$-dependent oxidoreductases which catalyze redox reactions are very interesting for technical processes. By means of these enzymes the production of building blocks for fine chemicals and important pharmaceuticals is possible. The dehydrogenases are very useful catalysts for chiral synthesis. These enzymes which belong to the class of oxidoreductases catalyze the transfer of hydrogen from one compound to another, mainly to $NAD(P)^+$. Due to the high costs of the nicotinamide cofactors, their stoichiometric use is not acceptable from an economical point of view. Besides, by coupling suitable cofactor regenerating reactions with thermodynamically unfavorable reactions, these can be driven towards the desired product. For these reasons industrial applications require

efficient in situ regeneration of the consumed cofactors. During the last years the development of capable regenerating processes has been subject to intensive studies. Regeneration of nicotinamide cofactors can be accomplished through several ways, namely enzymatic, chemical, photochemical, or electrochemical.

Several requirements must be considered in order to choose an appropriate regeneration system. The method should be practical and inexpensive. The regeneration system must be stable over a long period of time. Products need to be separated without much effort. Enzymes as well as reagents should be stocked or commercially available. The required equipment needs to be present. There should be no cross reactions between educts and products of the producing reaction or the compounds needed for cofactor regeneration. The formation of product should be thermodynamically as well as kinetically preferred. Byproduct formation should be negligible concerning the cofactor. The reduction of the cofactor $NAD(P)^+$ has to be regioselective, otherwise it would result in a partly inactive form of the cofactor. Reduction of $NAD(P)^+$ can lead to the 1,4-dihydropyridine product and to the 1,6-dihydropyridine product. The 1,4-dihydropyridine product is the only active form (Fig. 1).

Enzymatic ways for regeneration of nicotinamide cofactors are the best investigated and most convenient strategies. They show high selectivity for the formation of the active form of the cofactor. Further considerable advantages of enzymatic regeneration systems are the high compatibility with other reagents and the fact that enzymatic assays can be monitored easily [1]. There are two main principles to regenerate nicotinamide cofactors enzymatically (Fig. 2) [2].

The first approach requires a second enzyme as well as a second substrate for regeneration. In the second method, the substrate-coupled regeneration, one single enzyme is responsible for the formation of the desired product as well as for cofactor recycling.

Since the approach of enzymatic cofactor regeneration is well-investigated and favorable, the main part of this review deals with this strategy. Additionally, a minor part of this chapter describes ways of electrochemical regeneration of reduced and oxidized nicotinamide cofactors. This method has not yet been studied so extensively but is showing some considerable benefits such as low costs and the utilization of mass-free electrons as regenerating agents. Finally, chemical and photochemical regeneration of nicotinamide cofactors are dealt with briefly.

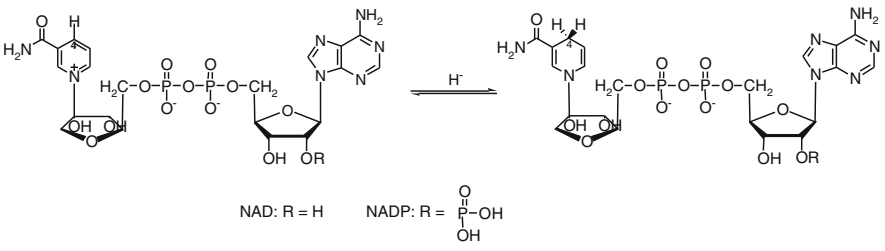


Fig. 1 Reduction of $NAD(P)^+$ to the active form 1,4-dihydropyridine

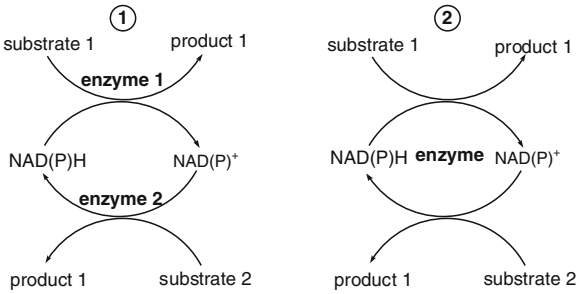


Fig. 2 Different principles for the regeneration of nicotinamide cofactors. Method 1 describes the regeneration using a second enzyme, method 2 shows the substrate-coupled approach utilizing one enzyme for the main reaction, the reduction of the substrate as well as for the regeneration of NAD(P)H

The introduction of organic–water two-phase systems permits the performance of efficient conversions of less water-soluble compounds. For this reason organic–water two-phase systems play an important role and thus are discussed extensively in this review.

A further important aspect is the feasibility of whole cell biotransformations. Whole cell biotransformations show a lot of advantages as compared to isolated enzymes, such as the improved stability of enzymes. If both, producing and regenerating enzymes, are available in one single strain, no addition of expensive cofactor is necessary because the intracellular cofactor pool can be utilized. Whole cell biotransformations are therefore very promising for technical applications, and making these conversions an intensively studied subject in the last years. The use of recombinant DNA techniques offers many possibilities to create capable systems. This chapter describes the most important whole cell biotransformations developed in the past as well as relevant processes with small-scale and technical application.

2 Enzymatic Methods to Regenerate Reduced Nicotinamide Cofactors

Several enzymes are known to be in use for the regeneration of reduced nicotinamide cofactors. The most important ones are described in Table 1 and some applications are summarized in the following Sections.

2.1 Formate Dehydrogenase

Formate dehydrogenase (FDH, EC 1.2.1.2) catalyzes the oxidation of formate to carbon dioxide while NAD^+ is reduced to NADH. A major advantage is the

Table 1 Enzymes useful for the regeneration of reduced nicotinamide coenzymes

Enzyme EC number	Organism	Activity (U mg^{-1})	References
Formate DH EC 1.2.1.2	*Candida boidinii*	4–6	[3–6]
	Candida methylica	4–6	[4, 5, 7]
	Candida methanolica	4–6	[8, 9]
	Pseudomonas sp.	4–6	[10–12]
	Thiobacillus sp.	7.6	[13]
	Mycobacterium vaccae		[14, 15]
Glucose-6-phosphate DH EC 1.1.1.49	*Leuconostoc mesenteroides*	290	[16, 17]
	Bacillus stearothermophilus		[18]
Glucose DH EC 1.1.1.47	*Thermoplasma acidophilum*		[19, 20]
	Bacillus megaterium	550	[21–23]
	Bacillus subtilis	375	[24]
Alcohol dehydrogenase (NAD^+) EC 1.1.1.1	*Saccharomyces cerevisiae*		[25]
	Horse liver		[26]
Alcohol dehydrogenase ($NADP^+$) EC 1.1.1.2	*Saccharomyces cerevisiae*		[27]
	Thermoanaerobacter brockii		[28, 29]
	Lactobacillus brevis		[30–32]
	Lactobacillus kefir		[30, 33–35]
Hydrogenase EC 1.12.1.2	*Alcaligenes eutrophus*	54	[36–38]
	Hydrogenomonas H 16		[39]
Hydrogenases I EC 1.12.7.2, formerly EC 1.18.99.1	*Pyrococcus furiosus*	360	[40]
Phosphite DH EC 1.20.1.1	*Pseudomonas stutzeri*	16	

irreversibility of the reaction. The carbon dioxide formed is chemically inert and can be easily removed. Furthermore, formate is a very cheap substrate as well as being innocuous towards most enzymes and the enzyme is commercially available.

In particular, FDH from *Candida boidinii* is often used for the regeneration of NADH. However, the low specific activity of this enzyme (4–6 U mg^{-1}) is a considerable disadvantage. The K_M values for NAD^+ and formate are 0.09 and 13 mM, respectively. The enzyme has a broad pH optimum of 7.5–8.5; while 55°C is the optimal temperature [3].

There are many applications involving FDH from *C. boidinii*. In 1980, Shaked and Whitesides published the first example to demonstrate the regeneration of a coenzyme in a preparative synthesis by coupling this enzyme with a simple NADH-consuming reaction, the production of D-lactate from pyruvate (Fig. 3) [41].

For the production of chiral hydrophobic alcohols, FDH from *C. boidinii* was combined with an NAD^+-dependent alcohol dehydrogenase (ADH) from *Rhodococcus erythropolis*. As a first example for the production of hydrophobic alcohols, an enzyme membrane reactor (EMR) was used for the synthesis enantiomerically pure (*S*)-1-phenylpropan-2-ol and some related structures out of their corresponding ketones [42].

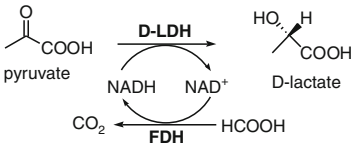


Fig. 3 Preparation of D-lactate from pyruvate catalyzed by D-LDH (D-lactate dehydrogenase) using formate/FDH for the regeneration of NADH

Nanba et al. described an FDH isolated from a *Thiobacillus* strain, which, in contrast to the *C. boidinii* enzyme, shows a high stability against haloketone compounds [13]. This property is important for the conversion of halogene substituted ketones, in particular ethyl 4-chloro-3-oxobutanoate esters, serving as building blocks of lipid regulating drugs (see section 9.5). The enzyme is also resistant to SH reagents.

Another approach to stabilize FDH against haloketones was followed by Yamamoto et al. They studied a set of cysteine mutants of FDH from *Mycobacterium vaccae* [14]. A triple cysteine mutant was successfully used for the synthesis of (*S*)-4-chloro-3-hydroxybutanoate ester by recombinant *Escherichia coli* cells coexpressing the FDH mutant and a carbonyl reductase from *Kluyveromyces aestuarii* [15].

2.2 Glucose-6-Phosphate Dehydrogenase

Glucose-6-phosphate dehydrogenase (G-6-P-DH) catalyzes the oxidation of D-glucose-6-phosphate to D-glucono-1,5-lactone-6-phosphate with a simultaneous reduction of $NAD(P)^+$. So far, the enzyme has been isolated from a large number of animal, plant, and microbial sources. One group of enzymes is $NADP^+$-specific, another is NAD^+-specific, and furthermore there are enzymes which prefer either $NADP^+$ or NAD^+. Finally, there are also glucose-6-phophate dehydrogenases with dual nucleotide specificity with comparable NAD^+- and $NADP^+$-linked activities. Because of the rapid hydrolysis of the formed glucono-1,5-lactone-6-phosphate to 6-phosphogluconate, the catalyzed reaction runs definitely irreversible.

It is highly disadvantageous that many products are difficult to separate from 6-phosphogluconate and the substrate glucose-6-phosphate is rather expensive. Again, glucose-6-sulfate is a profitable alternative because it is easier to prepare than glucose-6-phosphate. Finally, glucose-6-phosphate as well as its product can cause severe problems because both catalyze the decomposition of NAD(P)H in solution [43].

G-6-P-DH from *Leuconostoc mesenteroides* is used in organic synthesis for example to synthesize D-lactic acid or (*S*)-benzyl alcohol [43]. Commercially available G-6-P-DH was coupled with ADH from *Lactobacillus kefir* for the regeneration of NADPH to produce optically pure (*R*)-phenylethanol [33].

2.3 Glucose Dehydrogenase

Glucose dehydrogenase (GDH) is a useful catalyst for the conversion of β-D-glucose to D-glucono-1,5-lactone. The formed lactone is quickly converted into the corresponding acid. GDHs from both *Bacillus subtilis* and *Bacillus megaterium* are well characterized. Their biochemical properties are very similar and the alignment of the primary structures of both enzymes shows 85% conservations [44]. *B. megaterium* possesses several genes that code for isozymes of GDH. A simultaneous synthesis of enantiomerically pure (*R*)-1-phenylethanol and (*R*)-α-methylbenzylamine from racemic α-methylbenzylamine was realized using ω-transaminase from *Vibrio fluvialis* SH17 and ADH from *L. kefir* in combination with GDH from *B. subtilis*, which is used as an efficient coenzyme regenerating tool (Fig. 4) [45].

GDHs are also applied for the regeneration of NADH. For the production of L-carnitine starting from 3-dehydrocarnitine by means of an L-carnitine dehydrogenase NADH regeneration was achieved by GDH from *Bacillus* sp. [46]. In a two-step enzymatic asymmetric reduction system 2,6,6-trimethyl-2-cyclohexen-1,4-dione (ketoisophorone) was reduced via (6*R*)-2,2,6-trimethyl-cyclohexane-1,4-dione ((6*R*)-levodione) to (4*R*,6*R*)-4-hydroxy-2,2,6-trimethylcyclohexanone (actinol) using "old yellow enzyme 2" from *Saccharomyces cerevisiae* and levodione reductase from *Corynebacterium aquaticum* M-13 as biocatalysts, both expressed in *E. coli* (Fig. 5).

In both steps, commercially available GDH was used for NADH regeneration [47].

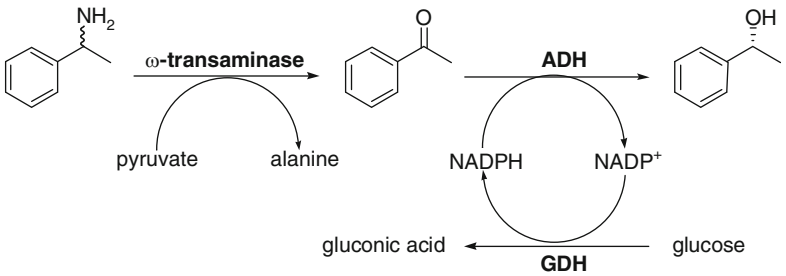


Fig. 4 Synthesis of (*R*)-phenylethanol and (*R*)-α-methylbenzylamine using ω-transaminase and ADH in combination with GDH for recycling of cofactor

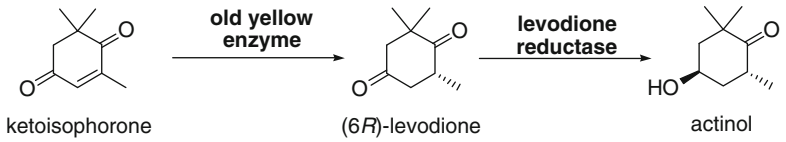


Fig. 5 Two-step preparation of actinol starting from ketoisophorone using "old yellow enzyme" and levodione reductase

(6*S*)-Tetrahydrofolate was used as starting material for the synthesis of L-leucovorin. Applying an NADPH-dependent dihydrofolate reductase in combination with GDH from *Gluconobacter scleroides* KY3613 for recycling of the cofactor, (6*S*)-tetrahydrofolate was synthesized from dihydrofolate [48].

2.4 Alcohol Dehydrogenase

ADHs catalyze the oxidation of alcohols to ketones with simultaneous reduction of $NAD(P)^+$. Due to the reversibility of this reaction, ADH-catalyzed reactions can either be used for the synthesis of (chiral) compounds or for the regeneration of the coenzyme. The latter holds true, for example, in the case of substrate-coupled ADH-catalyzed reduction reactions using isopropanol or ethanol as the hydrogen donor. Several kinds of ADHs have already been described. ADHs of the EC 1.1.1.1 group are dependent on NAD^+. They act on primary or secondary alcohols or on hemiacetals. In contrast, ADHs of the group EC 1.1.1.2 depend on $NADP^+$. Some enzymes of this group oxidize only primary alcohols; others act on secondary alcohols as well.

2.4.1 Alcohol Dehydrogenase (NAD^+) (EC 1.1.1.1)

The best-investigated enzymes from this NAD^+-dependent group are the ADHs from *Drosophila melanogaster* [49–51], *S. cerevisiae* [25], horse liver [26], and *R. erythropolis* [52–54]. Horse liver ADH (HLADH) is a commercially available biocatalyst. HLADH accepts a variety of substrates; it generally catalyzes the reversible oxidation of primary and secondary alcohols [26]. By means of HLADH, 2-phenylpropionaldehyde was stereoselectively reduced to 2-phenyl-1-propanol while ethanol was simultaneously oxidized to acetaldehyde in order to regenerate NADH [55]. Another important source of ADH is yeast. The enzyme was used for NADH regeneration in the course of the regioselective reduction of androstandione to androsterone by 3α-hydroxysteroid dehydrogenase from *Pseudomonas testosteroni* [56].

2.4.2 Alcohol Dehydrogenase ($NADP^+$) (EC 1.1.1.2)

Well characterized $NADP^+$-dependent ADHs from microbial sources have been isolated from several sources (Table 1). ADHs from *Lactobacillus kefir* and *L. brevis* convert prochiral ketones into the corresponding (*R*)-alcohols in a highly stereospecifically manner. They show a broad substrate spectrum, high specific activities (up to 100–500 U mg^{-1}), and they both need the presence of Mg^{2+} ions to maintain their activity [30]. ADH from *Thermoanaerobacter brockii* (TBADH)

is frequently used for NADPH regeneration. The enzyme is remarkably stable at a temperature up to 85°C, and it is very robust in organic solvents [28]. Furthermore, it is able to catalyze both oxidation and reduction. NADPH is conveniently regenerated by simultaneous oxidation of 2-propanol to acetone. Advantages of this system are the low cost of the substrate 2-propanol and the high regeneration rate. Cyclohexanone monooxygenase (CHMO) from *Acinetobacter calcoaceticus* was combined with TBADH to perform efficient Baeyer-Villiger and sulfide oxidations [57].

2.5 Hydrogenase

The enzyme hydrogenase (hydrogen dehydrogenase; EC 1.12.1.2) is able to reduce electron acceptors by molecular hydrogen. When it is used in cofactor regenerating systems, consumed NADH can be regenerated directly by molecular hydrogen.

Hydrogenases were isolated from microbial sources (Table 1). Due to their high reduction potential, the cheap substrate H_2, and the facilitated product isolation which is not disturbed by interfering products of the regenerating reaction, hydrogenases are convenient enzymes for regeneration purposes. However, the enzymes are highly sensitive against oxygen and they require the addition of electrontransferring enzymes and dyes. Hydrogenase from *Alcaligenes eutrophus* has a specific activity of 54 U mg^{-1} [58]. It has a very poor stability but this problem is less prominent when used in an immobilized form [36].

Another type of hydrogenase (hydrogenases I, EC 1.12.7.2, formerly EC 1.18.99.1) catalyzes the oxidation of molecular hydrogen by transferring the electrons to ferredoxin as the primary electron acceptor. One prominent candidate from this class is the soluble hydrogenase I from *Pyrococcus furiosus*. This enzyme is remarkably thermostable [40]. In contrast to other hydrogenases, it preferentially catalyzes the production of molecular hydrogen rather than its oxidation. The specific activity of hydrogenase from *P. furiosus* is 360 U/mg. Reduced ferredoxin appears to be the physiological electron donor. Although the production of molecular hydrogen represents the preferred reaction, this enzyme can also be used for the reduction of nicotinamide coenzymes by H_2. A partially purified sample of the soluble hydrogenase from *P. furiosus* generates NADPH directly from oxidized $NADP^+$ without a generation of any byproducts, except for protons. For this reason, the enzyme was applied for the regeneration of NADPH utilizing cheap molecular hydrogen coupled with TBADH to reduce prochiral ketones. Two prochiral model substrates, acetophenone and (2*S*)-hydroxy-1-phenyl-propanone (HPP), were quantitatively reduced to the corresponding (*S*)-alcohol and (1R,2S)-diol. An *ee* >99.5% and *de* >98%, respectively, with total turnover numbers (ttn: mol product/mol consumed cofactor $NADP^+$) of 100 and 160, could be reached (Fig. 6) [59].

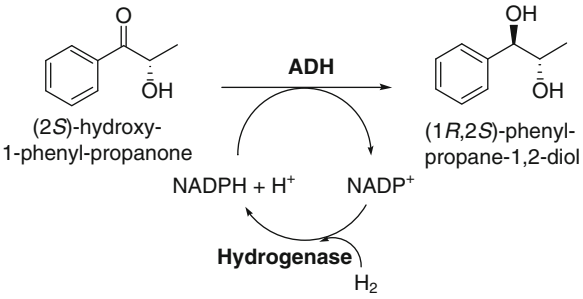


Fig. 6 Use of ferredoxin-hydrogenase for NADPH regeneration here in a system for the stereoselective reduction of (2*S*)-hydroxy-1-phenyl-propanone to (1*R*,2*S*)-phenyl-propane-1,2-diol catalyzed by ADH

2.6 Phosphite Dehydrogenase

Phosphite dehydrogenase (PTDH) represents a further attractive method for regeneration of the reduced nicotinamide coenzyme. This enzyme was isolated from *Pseudomonas stutzeri* [60, 61]. It catalyzes the nearly irreversible oxidation of phosphite to phosphate with the concomitant reduction of NAD^+ to NADH. Advantages of this method are the use of the inexpensive phosphite substrate, formation of the benign phosphate as the product, and the favorable equilibrium constant. Since the wild type PTDH accepts $NADP^+$ as the coenzyme very poorly, a mutant enzyme was created that reduces both NAD^+ and $NADP^+$ with high efficiency [62]. Additionally, the thermostability was significantly improved by directed evolution [63]. This highly stable $NADP^+$-accepting PTDH mutant was used in an EMR to regenerate the coenzyme during the xylose reductase catalyzed xylitol synthesis [64] (Fig. 7).

2.7 Further Enzyme Systems for the Regeneration of Reduced Nicotinamide Cofactors

Lactate dehydrogenases (LDHs) can also be used for the regeneration of nicotinamide coenzymes. Although the equilibrium of the reaction favors the formation of lactate or NAD^+, respectively, the regeneration of NADH can also be carried out utilizing a surplus of lactate. LDHs are well described from several sources such as bacteria from the genera *Lactobacillus*, *Leuconostoc*, or *Bacillus* or from eucaryotic cells. Applying the commercially available L-lactate dehydrogenase (L-LDH) from rabbit muscle, Gu and Chang developed a multienzyme system consisting of leucine dehydrogenase (LeuDH) (EC 1.4.1.9), L-lactic dehydrogenase (EC 1.1.1.27), urease (EC 3.5.1.5), and dextran-NAD^+ microencapsulated within artificial cells. Several

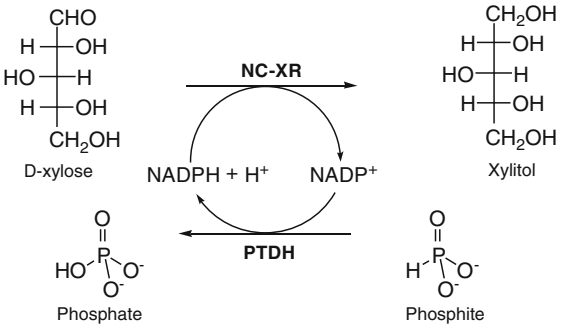


Fig. 7 Xylose reductase catalyzed xylitol synthesis. The coenzyme NADPH was regenerated by a $NADP^+$-accepting mutant of phosphite dehydrogenase

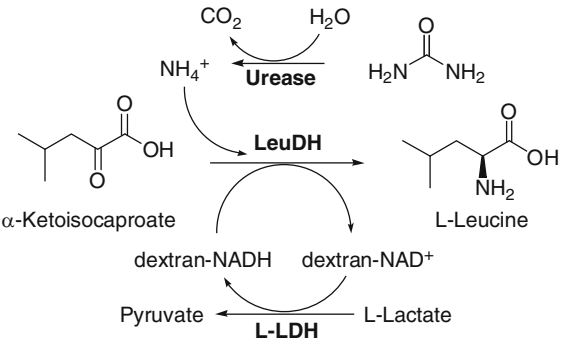


Fig. 8 Synthesis of amino acids by a multienzyme system consisting of leucine dehydrogenase (LeuDH) catalyzing the reductive amination of the corresponding keto acid, L-lactate dehydrogenase (L-LDH), and lactate for the regeneration of NADH and urease for the in situ generation of ammonia. The coenzyme NAD^+ was covalently bond to dextran, enzymes and dextran-coupled NAD^+ were microencapsulated within artificial cells

essential amino acids like L-leucine, L-valine, and L-isoleucine were synthesized using these cells as the catalyst. Urease catalyzes the in situ generation of ammonia and lactate/LDH are responsible for the regeneration of NADH, coupled onto high molecular-weight dextran (Fig. 8). More than 90% of the original activity was retained when artificial cells were kept at 4°C for 6 weeks [65].

3 Enzymatic Methods to Regenerate Oxidized Nicotinamide Cofactors

So far not many enzymatic methods have been described for the regeneration of oxidized nicotinamide cofactors. These are presented in the following section.

3.1 Glutamate Dehydrogenase

The synthesis of L-glutamate from ammonium and α-ketoglutarate is catalyzed by glutamate dehydrogenase (GluDH). Several GluDHs exist with different coenzyme specificities (Table 2). Lee and Whitesides used the system 2-ketoglutarate/GluDH to regenerate NAD^+ during the oxidative resolution of 1,2-butanediol with glycerol dehydrogenase (Fig. 9) [78].

Commercially available GluDH was applied for $NADP^+$ regeneration during the preparation of 12-ketochenodeoxycholic acid from cholic acid by means of 12α-hydroxysteroid dehydrogenase from *Clostridium* group P (Fig. 10) [79].

Table 2 Enzymes useful for the regeneration of oxidized nicotinamide coenzymes

Enzyme EC number	Organism	Activity (U mg^{-1})	References
Glutamate DH (NAD^+) EC 1.4.1.2	*Neurospora crassa*		[66]
Glutamate DH ($NADP^+$) EC 1.4.1.4	*Escherichia coli*	130 (reductive amination of ketoglutarate)	[67, 68]
Glutamate DH; dual specificity EC 1.4.1.3	*Bos taurus*	167 (oxidative deamination of L-glu)	[69, 70]
L-Lactate DH EC 1.1.1.27	*Lactobacillus casei*	2,290 (pyruvate reduction)	[71–73]
	Lactobacillus lactis		[71]
	Lactobacillus curvatus	2,030 (pyruvate reduction)	[71]
	Streptococcus epiderminis		[71]
NADH oxidase (H_2O-forming) EC 1.6.3.1	*Lactobacillus brevis*	350	[74]
	Lactobacillus sanfranciscensis		[75]
NADH oxidase (H_2O_2-forming) EC 1.6.99.3	*Escherichia coli*	5.2 (NADH oxidation)	[76, 77]

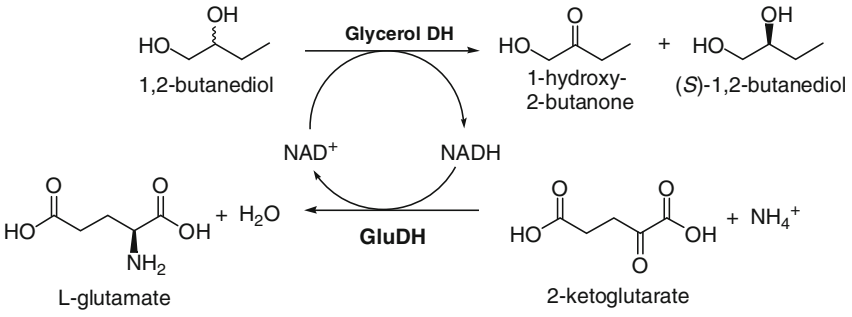


Fig. 9 Stereoselective oxidation of 1,2-butanediol to (*S*)-1,2-butanediol catalyzed by glycerol dehydrogenase with simultaneous cofactor regeneration using the system 2-ketoglutarate/glutamate dehydrogenase (GluDH)

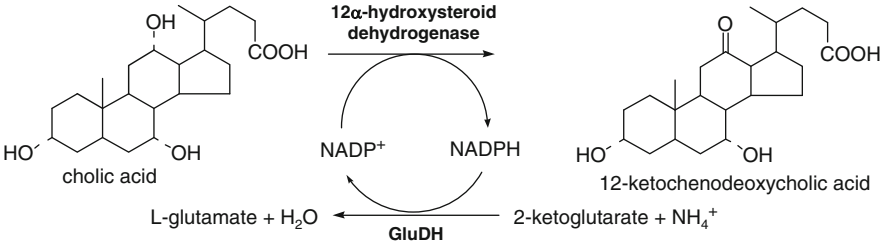


Fig. 10 Production of 12-ketochenodeoxycholic acid from cholic acid catalyzed by 12α-hydroxysteroid dehydrogenase with coupled cofactor regeneration using GluDH

3.2 *L-Lactate Dehydrogenase*

L-Lactate dehydrogenase (L-LDH, EC 1.1.1.27) catalyzes the reduction of pyruvate to (*S*)-lactate with a simultaneous oxidation of NADH. L-LDH is found in all higher organisms. There are two kinds of L-LDHs: enzymes from one group are activated by fructose 1,6-diphosphate while the other group stays independent [71]. L-LDH is highly selective for pyruvate, short-chain 2-keto acids and phenylpyruvic acid [80]. All bacterial NAD^+-dependent LDHs form lactate from pyruvate in vivo, and there is no evidence at all that they catalyze the other direction as well. The equilibrium constant lies far on the direction of lactate formation, and thus the reaction catalyzed by bacterial LDHs can be considered almost irreversible. LDHs from some lactobacilli like *Lactobacillus fermentum* or *L. cellobiosus* show no or just poor reaction with lactate [71], whereas mammalian LDHs can be considered as reversible [71]. Well characterized L-LDHs are summarized in Table 2.

All these enzymes are commercially available and inexpensive. They have extraordinarly high specific activities. The specific activity of L-LDH from *L. casei* is 2,290 U mg^{-1} and that from *L. curvatus* lays by 2,030 U mg^{-1}, but both enzymes catalyze only the reduction of pyruvate, not the reverse reaction [72]. L-LDH from rabbit muscle was used to regenerate NAD^+ [81]. The system pyruvate/LDH is convenient in use, because enzymes and substrates are stable and inexpensive. It was coupled with pyruvic and glyoxylic acids as oxidants to regenerate NAD^+ in oxidations of glucose-6-phosphate to 6-phosphogluconate catalyzed by G-6-P-DH. The system pyruvate/LDH has also been used to regenerate NAD^+ in the HLADH catalyzed oxidation of *cis*-1,2-bis(hydroxymethyl)cyclohexane to (+)-(1*R*,6*S*)-*cis*-8-oxabicyclofnonan-7-one (Fig. 11). This regeneration proceeds efficiently and is economically acceptable.

3.3 *Alcohol Dehydrogenase*

The system acetaldehyde/ADH is often employed for the oxidation of NADH. In general, ADH from horse liver is used for this purpose. As mentioned earlier,

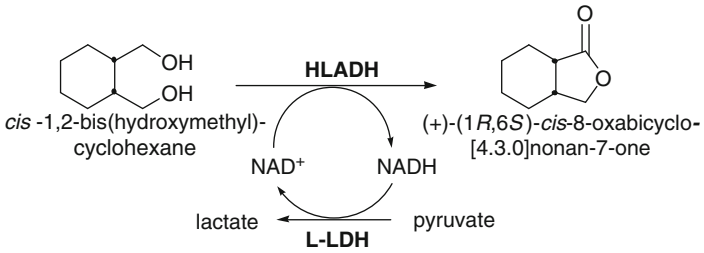


Fig. 11 Synthesis of (+)-(1*R*,6*S*)-*cis*-8-oxabicyclo[4.3.0]nonan-7-one from *cis*-1,2-bis(hydroxymethyl)cyclohexane catalyzed by HLADH. The cofactor NAD^+ is regenerated by the system pyruvate/L-LDH

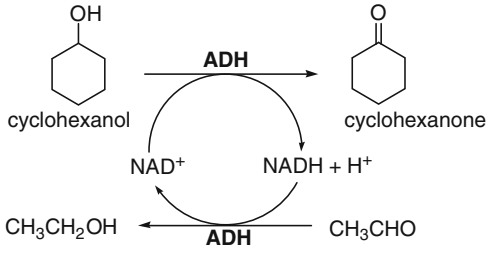


Fig. 12 Oxidation of cyclohexanol catalyzed by HLADH with acetaldehyde as a coenzyme regenerating substrate

this enzyme catalyzes the reversible oxidation of primary and secondary alcohols. Lemière et al. examined HLADH catalyzed oxidations of nine structurally different alcohols in combination with acetaldehyde as a coenzyme recycling substrate. Amongst them, cyclohexanol was oxidized nearly quantitatively within 1 h (Fig. 12) and 2-cyclohexenol within 6 h [82].

3.4 NADH Oxidase

NADH oxidases catalyze either the two-electron reduction of molecular oxygen to peroxide (Nox-1) or the four-electron reduction of molecular oxygen to water (Nox-2) [83]. Both enzymes are flavoproteins, but, as confirmed by sequence analyses, the enzymes are only distantly related [84]. They have been isolated from organisms like *Streptococcus mutans* [83–85], *Streptococcus faecalis* [86], *Archaeoglobus fulgidus* [87], *L. brevis* [74, 88], *L. sanfranciscensis*, and *Borrelia burgdorferi* [75]. Water-forming NADH oxidase from *L. brevis* has a specific activity of 350 U mg^{-1}, and the K_M value for NADH is 24 μM. This enzyme was coupled with an NAD^+-dependent (*R*)-specific ADH in order to regenerate NAD^+. Starting from *rac*-1-phenylethanol, the (R)-enantiomer could be oxidized completely resulting in pure (*S*)-1-phenylethanol (Fig. 13) [88].

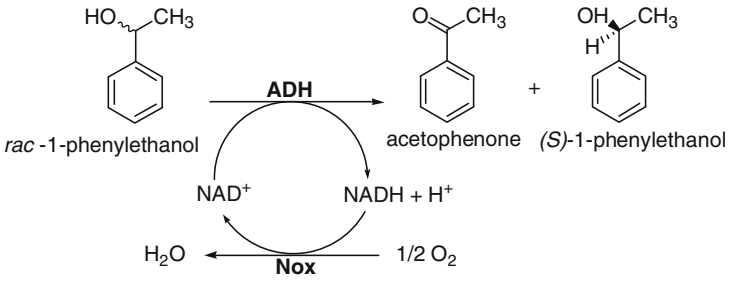


Fig. 13 Production of (*S*)-1-phenylethanol from *rac*-1-phenylethanol catalyzed by (*R*)-specific ADH with coupled cofactor regeneration using NADH oxidase from *Lactobacillus brevis*

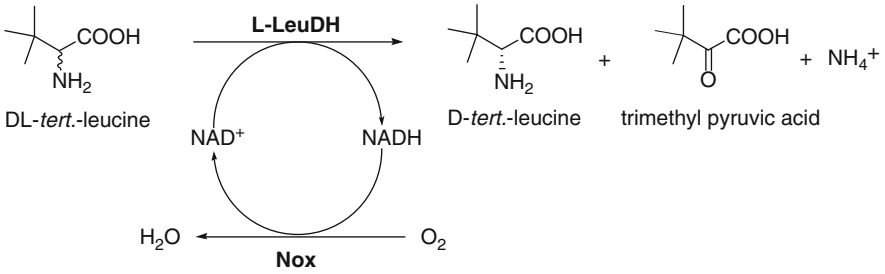


Fig. 14 Kinetic resolution of DL-*tert*-leucine catalyzed by L-leucine dehydrogenase (L-LeuDH) for the preparation of D-*tert*-leucine with simultaneous NAD^+ regeneration using NADH oxidase (Nox) from *Lactobacillus brevis*

Another example for the regeneration of NAD^+ is represented by the LeuDH catalyzed preparation of D-*tert*-leucine [89]. Starting from DL-*tert*-leucine, the L-enantiomer was oxidized to trimethylpyruvate while D-*tert*-leucine remained as the desired product. The consumed cofactor NAD^+ can be regenerated by use of NADH oxidase (Nox) from *L. brevis* (Fig. 14).

The driving force in the coupled enzyme process is independent of the equilibrium of the LeuDH catalyzed reaction because of the irreversibility of the NOX reaction. This is most favorable because the oxidation reaction of many $NAD(P)^+$-dependent dehydrogenases is hampered by their equilibrium, which prefers the reduction reaction. By applying this system D-*tert*-leucine was obtained with an excellent *ee* >99%.

4 Electrochemical Methods

Apart from enzymatic cofactor regeneration, both chemical and electrochemical regeneration methods have attracted attention due to their greater flexibility. The electrochemical regeneration of cofactors has a few advantages, especially the use

of the mass-free electrons as regenerating agents. There are no co-substrates necessary and no couple products are formed [90].

4.1 Electroenzymatic Oxidations

Direct electrochemical reduction of oxidized nicotinamide cofactors is not useful because of the formation of dimers via intermediate radicals. On the other hand, direct electrochemical oxidation of NAD(P)H to $NAD(P)^+$ can be performed successfully [90]. However, it requires relatively high oxidation potentials and may result in electrode passivation.

To accelerate the otherwise slow NADH oxidation, carbon electrodes with high surface area can be used. Anode potentials need to be at least + 900 mV vs NHE for the direct electrochemical oxidation of NAD(P)H. Under these conditions only substrates with rather high stability against oxidation can be transformed without any loss of selectivity. The introduction of a mediator leads to a faster transfer of electrons between NADH and the anode. A further advantage of this method is the prevention of a possible anodic oxidation of NADH [90].

This approach is used in the oxidation of 2-hexene-1-ol and 2-butanol in the presence of either yeast ADH (YADH) or ADH from *T. brockii* with in situ regeneration of $NAD(P)^+$ by indirect electrolysis with tris(3,4,7,8-tetramethyl-1,10-phenanthroline) iron (II/III) complex at an anode potential of 630 mV vs NHE (Fig. 15) [91].

If mediators, which act as one-electron transfer agent towards NADH, are used, they must possess relatively positive potentials. Substrates containing further oxidation sensitive groups cannot be used. This is why mediator systems with very low oxidation potentials could be considered for obtaining high chemoselectivities. Nevertheless, mediators with very low oxidation potentials remove hydride ions instead of reacting via electron transfer [90].

Phenanthrolindiones are often used as mediators because of their high stability. To reduce the oxidation potential for the electrochemical regeneration and to enhance the hydride transfer from NAD(P)H, the electron density of the *o*-chinoid structure should be decreased [92]. This can be done by complexation with a

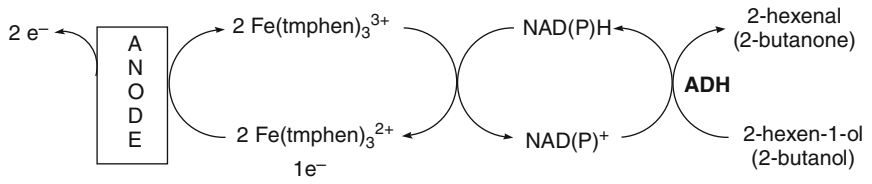


Fig. 15 Electrochemically driven oxidation of 2-hexen-1-ol and 2-butanol to 2-hexenal and 2-butanone, respectively, catalyzed by ADH from *Thermoanaerobacter brockii*. *tmphen* 3,4,7,8-tetramethyl-1,10-phenanthroline

transition-metal ion. Simultaneously, the solubility of the mediators in the aqueous buffer system can be enhanced. Co^{2+}, Ni^{2+}, and Cu^{2+} complexes with 1,10-phenanthroline-5,6-dione as a chelating ligand react under fast hydride ion transfer from NAD(P)H to the ligand. Unfortunately, insoluble precipitates of telomers or oligomers of the reduced forms are formed, so that the application of these systems as mediators for NAD^+ regeneration is not accomplishable [90]. To overcome this problem, Steckhan et al. developed efficient and stable systems. They blocked coordination sites of 1,10-phenanthroline-5,6-dione by various ligands. In the first system tris(1,10-phenanthroline-5,6-dione)ruthenium(II) perchlorate is used as redox catalyst while in the second and third system telomerization of the reduced Co^{2+} or Ru^{2+} complex is prevented through the use of a mixed ligand system. This consists of one 1,10-phenanthroline-5,6-dione ligand as the catalytically active unit and one *N*,*N*,*N*-tris(aminoethyl)amine or *N*,*N*,*N*-tris(2-pyridylmethyl)amine ligand to block the other sites of the metal center (Fig. 16) [93, 94].

These regeneration systems were combined with the HLADH catalyzed reaction of cyclohexanol to cyclohexanone as a model system (Fig. 17). The catalytic

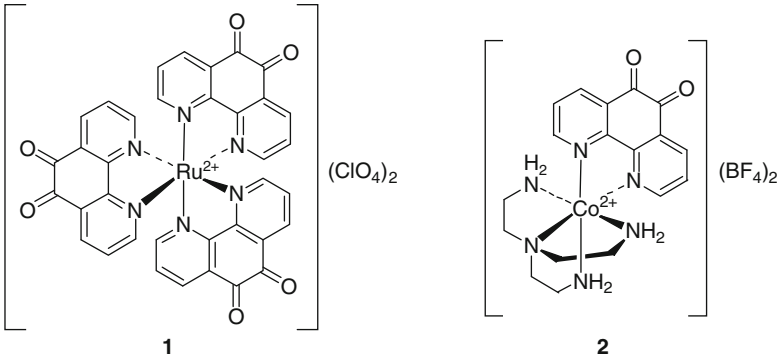


Fig. 16 Structures of the redox catalyst tris(1,10-phenanthroline-5,6-dione) ruthenium(II) perchlorate 1 and the mixed ligand system consisting of 1,10-phenanthroline-5,6-dione and one *N*,*N*,*N*-tris(aminoethyl)amine ligand 2

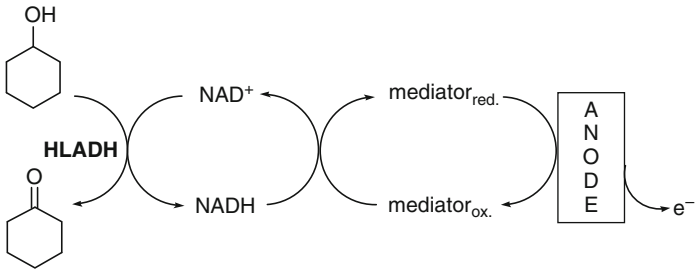


Fig. 17 Conversion of cyclohexanol to cyclohexanone catalyzed by HLADH with simultaneous regeneration of NAD^+ by the phenanthrolinedione complexes described above

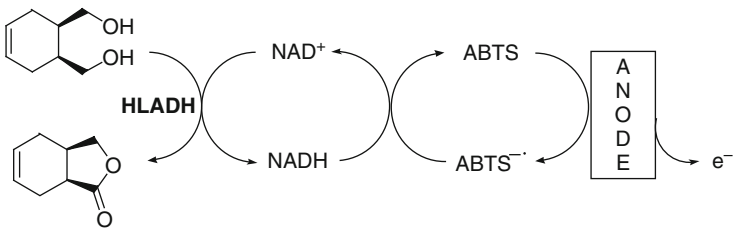


Fig. 18 Conversion of *meso*-3,4-dihydroxymethylcyclohex-1-ene to the lactone (3a*R*,7a*S*)-3a,4,7,7a-tetrahydro-3*H*-isobenzofurane-1-one catalyzed by HLADH

turnover frequency could be increased by a factor of 14 in comparison with the uncomplexed phenanthrolinediones.

For NADH oxidation in oxidoreductase-catalyzed synthesis three surface-modified electrodes were investigated [95], namely 3,4-dihydroxybenzaldehyde electropolymerized on glassy carbon, hexacyanoferrate modified nickel electrodes and Meldola Blue, adsorbed on graphite. The required high overpotential was reached with all electrodes and especially with the Meldola Blue electrode. Furthermore, this electrode ensured a good storage stability, though under operating synthesis conditions the stability was decreased.

Another efficient method is the electrochemical oxidation of NADH at 0.585 V vs Ag/AgCl by means of $ABTS^{2-}$ (2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonate)) as an electron transfer mediator [96]. Due to the unusual stability of the radical cation ABTS•$^{-}$, the pair $ABTS^{2-}$/ABTS•$^{-}$ is a useful mediator for application in large-scale synthesis even under basic conditions. Basic conditions are favorable for dehydrogenase catalyzed reactions. This electrochemical system for the oxidation of NADH using $ABTS^{2-}$ as mediator was successfully coupled with HLADH to catalyze the oxidation of a *meso*-diol (*meso*-3,4-dihydroxymethylcyclohex-1-ene) to a chiral lactone ((3a*R*,7a*S*)-3a,4,7,7a-tetrahydro-3*H*-isobenzofurane-1-one) with a yield of 93.5% and *ee* >99.5% (Fig. 18).

4.2 Electroenzymatic Reductions

The direct electrochemical reduction of $NAD(P)^+$ requires high overpotentials and leads to enzymatically inactive $NAD(P)^+$-dimers originating from the one-electron transfer reaction. Therefore the electrochemical regeneration of NAD(P)H has always to be performed by indirect electrochemical methods. The formation of dimers can be prevented by using an additional regenerating enzyme. This second enzyme must be able to accept two electrons in two steps from the mediator and then transfer one electron pair to $NAD(P)^+$ [90]. Examples for such regenerating enzymes are lipoamide dehydrogenase (diaphorase) [97], ferredoxin-$NADP^+$-reductase (FNR) [98], enoate reductase [99–101] and the "VAPOR" enzymes (Viologen accepting pyridine-nucleotide oxidoreductases) [102]. The latter enzymes

are applied in combination with methyl viologen *N*,*N*′-dimethyl-4-4′-bipyridinium dichloride as one-electron mediator. Ferredoxin instead of viologens can also be used as a mediator system for FNR.

By coimmobilization of viologen and the VAPOR enzymes at the surface of a carbon cathode, NADH can be regenerated at a rate of 9 nmol h^{-1} cm^{-2} [102, 103]. The VAPOR enzymes are flavoenzymes occurring for example in thermophilic bacilli. They are able to regenerate all four species of the pyridine nucleotides NADH, NAD^+, NADPH, and $NADP^+$ according to the equations in Fig. 19.

Clostridium thermoaceticum contains the so-called AMAPOR (artificial-mediator-accepting pyridine-nucleotide oxidoreductases), which are useful for electromicrobial regeneration of all four forms of pyridine nucleotides, too. An NADP(H) dependent AMAPOR from *C. thermoaceticum* has been purified and characterized [104]. It is able to react with rather different artificial mediators such as viologens or quinones, for example 1,4-benzoquinone, anthraquinone-2,6-disulfonate, or 2,6-dichloro-indophenol.

AMAPOR can be coupled with commercially available pyridine-nucleotide-dependent oxidoreductases to synthesize compounds like (*S*)-glutamate, (2*R*,3*S*)-isocitrate or 6-phosphogluconate (Fig. 20) [105]. The required redox equivalents are provided by electrochemically regenerated artificial mediators like methylviologen or cobalt sepulchrate for NAD(P)H regeneration and carboxamidomethyl-viologen or anthraquinone sulfonates for $NAD(P)^+$ regeneration, respectively.

This system operates well to reduce nonactivated carboxylic acids to the corresponding alcohols and to convert primary alcohols into the corresponding aldehydes and carboxylates using an aldehyde and an alcohol oxidoreductase.

In a quite different approach, NAD^+ can be reduced in a process involving as initial step the electrochemical reduction of the disulfide groups of lipoamide and

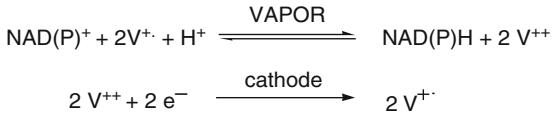


Fig. 19 Regeneration of all four species of the pyridine nucleotides NADH, NAD^+, NADPH, $NADP^+$ by the VAPOR enzymes. V^{++} viologen

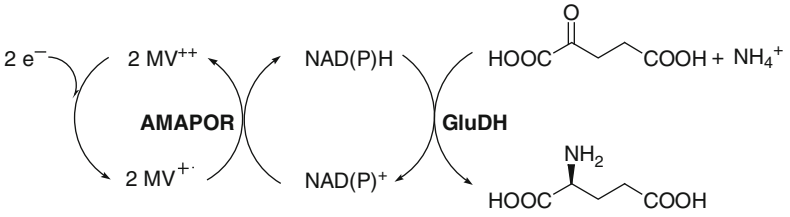


Fig. 20 Regeneration of NAD(P)H by means of AMAPOR coupled with the reductive amination of 2-oxoglutarate to (*S*)-glutamate catalyzed by glutamate dehydrogenase (GluDH). *MV* methyl-viologen

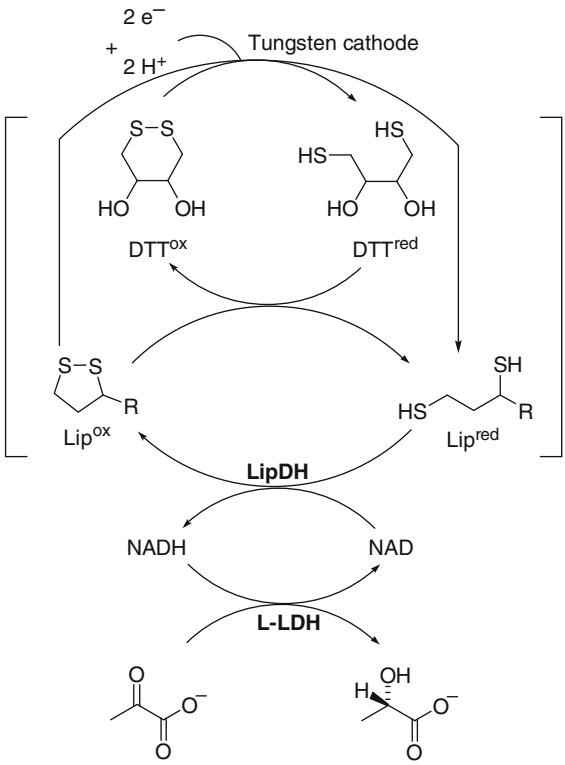


Fig. 21 L-LDH catalyzed reduction of pyruvate coupled with regeneration of NADH based on electrochemical reduction of disulfides. *LipDH* Lipoamide Dehydrogenase; *DTT* Dithiothreitol; *Lip* Lipoamide (R = $(CH_2)_4CONH_2$)

oxidized dithiothreitol to the corresponding dithiols. The formed dihydrolipoamide reduces NAD^+ to NADH in a reaction catalyzed by lipoamide dehydrogenase [106]. This method is based on the ability of a tungsten electrode to reduce selectively stable cyclic disulfides to the corresponding dithiols. After the electrochemical cleavage of the disulfide bond the reduction of NAD^+ to NADH occurs in a regiospecific enzymatic step. By this process L-lactate can be synthesized from pyruvate with 85% yield and 96% *ee* (Fig. 21).

Limitations of the system are the low turnover numbers as well as the failure to regenerate NADPH. However, if ferredoxin reductase is used for coenzyme regeneration instead of lipoamide dehydrogenase, this method can also be applied to $NADP^+$-dependent systems [97].

Further research for useful mediators led to the anthracycline antibiotic adriamycin which serves as a novel mediator for the FNR or diaphorase (DP) catalyzed electrochemical reduction of $NAD(P)^+$ [107]. This regeneration system has been satisfactory combined with NAD(P)H-dependent enzymatic reactions. The $NADP^+$-dependent FNR/adriamycin system was coupled with GluDH. GluDH was entrapped together with FNR on the electrode surface. NADPH was efficiently

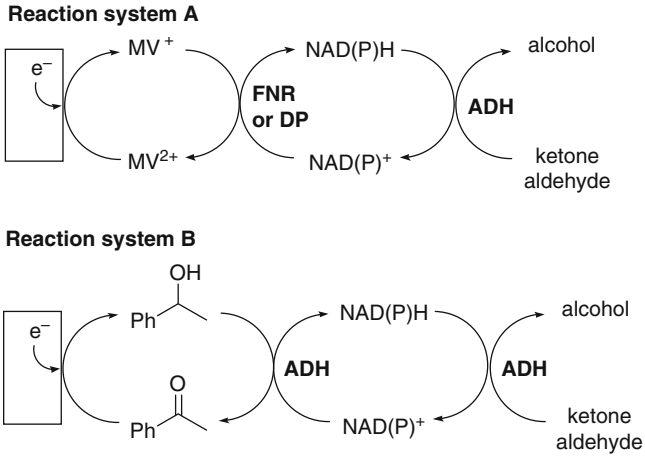


Fig. 22 Electroreduction of ketones or aldehydes using ADH as catalyst. Reduction system A shows the ADH-catalyzed reduction coupled with regeneration of NADPH or NADH by ferredoxin-$NADP^+$ reductase (FNR) or diaphorase (DP), respectively with assistance of methyl viologen as an electron mediator. In system B, ADH is used as sole enzyme which catalyzes both reduction of substrates and regeneration of cofactors

recycled by the coupled conversion of α-ketoglutarate into L-glutamate. The NAD^+-dependent system DP/adriamycin on the other hand was combined with LDH transforming pyruvate effectively into L-lactate.

For the asymmetric reduction of ketone and aldehyde derivates, two electrochemical reduction systems using ADH as catalyst were examined (Fig. 22) [108]. In system A, the reduced coenzymes are regenerated using either FNR for NADPH or DP for NADH. Methyl viologen serves as electron mediator between the electrode and FNR/DP. System B contains ADH as sole enzyme, which catalyzes both reduction of substrates and regeneration of cofactors. Phenylethanol is oxidized by ADH accompanied by reduction of $NADP^+$ to NADPH and its oxidation product acetophenone is reduced electrochemically at a glassy carbon cathode.

Using these systems, stereoselective reduction of some compounds like propiophenone or pyruvic acid was obtained with high *ee* values. Both systems, however, show very slow reaction rates, the rate in system B being even much lower than that in system A (Fig. 22).

Several aspects have to be considered in order to regenerate NAD(P)H by an indirect electrochemical procedure without the application of a second regeneration enzyme: The active redox catalyst should transfer two electrons in one step or a hydride ion. At potentials more negative than −0.9 V vs SCE, NAD^+ dimers will be formed, so the electrochemical activation of the catalyst should be possible at potentials less negative than −0.9 V. To prevent low chemoselectivity and low enantioselectivity, the active form of the catalyst should not transfer the electrons or the hydride ion directly to the substrate but to $NAD(P)^+$. Furthermore, only active 1,4-NAD(P)H should be formed [90].

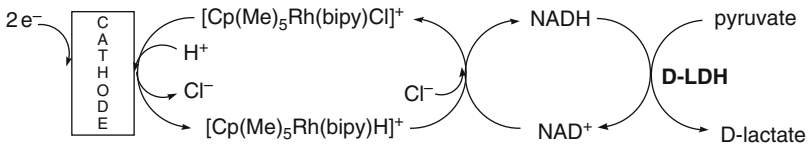


Fig. 23 Electrochemically driven D-LDH catalyzed enzymatic reduction of pyruvate to D-lactate using pentamethylcyclopentadienyl-2-2′-bipyridinechloro-rhodium(III) ([Cp*Rh(bpy)Cl]Cl) as redox catalyst

There are two systems so far which fulfill these requirements: tris(2,2′-bipyridyl) rhodium complexes [109, 110] and substituted or nonsubstituted (2,2′-bipyridyl) (pentamethylcyclopentadienyl)-rhodium complexes [111]. At potentials between −680 and −840 mV vs SCE, the electrochemical reduction of these complexes leads to the formation of rhodium hydride complexes. Hydride ions are transferred from the complex to $NAD(P)^+$ under specific formation of 1,4-NAD(P)H and the initial complex.

This system has been efficiently applied in the in situ electroenzymatic reduction of pyruvate to D-lactate by means of the NADH-dependent D-lactate dehydrogenase (Fig. 23). Using pentamethylcyclopentadienyl-2-2′-bipyridinechloro-rhodium(III) ([Cp*Rh(bpy)Cl]Cl) as redox catalyst, D-lactate was formed with an *ee* value of 93.5% after 3 h at a rate of five turnovers per hour [112].

The same system was employed in the reduction of 4-phenyl-2-butanone to (*S*)-4-phenyl-2-butanol using HLADH as well as *S*-ADH from *Rhodococcus* sp. with high enantioselectivity [113]. With pentamethylcyclopentadienyl-4-ethoxymethyl-2,2′-bipyridinechloro-rhodium(III) as mediator and HLADH as catalyst, after 5 h 70% of 4-phenyl-2-butanone was reduced to (*S*)-4-phenyl-2-butanol with 65% *ee*. Using *S*-ADH, 76% of the ketone was converted to the (*S*)-alcohol after 5 h with 77% *ee*. Furthermore, this system has been applied in an electrochemical EMR with a polymer bound rhodium complex as mediator.

Shimizu et al. used simple rhodium-aqua ions (Rh^{3+}) immobilized onto polymer-modified electrodes to perform the electrochemical reduction of NAD^+ [114]. Rh^{3+} was loaded onto polymeric anion doped-polypyrrole membranes coated on the surface. Electrochemical reduction of NAD^+ with immobilized Rh^{3+} was performed at −0.85 V, where Rh^{3+} was reduced to Rh^+. NADH was produced without detectable formation of NAD-dimers.

Another system for electroenzymatic regeneration of NADH includes lipoamide dehydrogenase covalently modified with *N*-ethyl-*N*′-(γ-aminopropyl)-viologen bromide and dip-coated onto a carbon electrode under a Nafion® film [115]. Nafion® is a perfluorinated sulfonated polymer. The chemical bond between lipoamide dehydrogenase and the viologen species prevents the leakage of the toxic methyl viologen from the electrode and therefore the contamination of the final product. This system can be used to carry out syntheses employing NADH-dependent enzymes such as LDH and malate dehydrogenase. A similar enzyme electrode was developed by electrochemical copolymerization of a redox enzyme

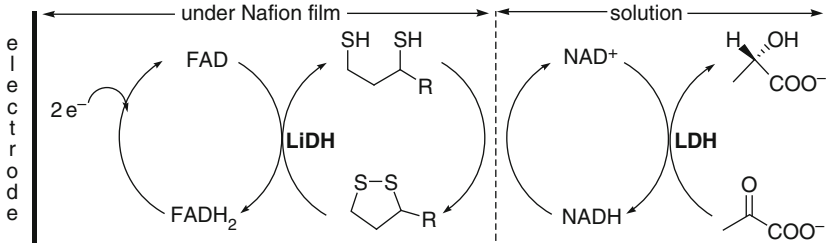


Fig. 24 Electroenzymatic synthesis of L-lactate coupled with cofactor regeneration using a FAD-modified electrode

with an oligomeric organic redox mediator on the surface of a glassy carbon electrode [116]. This electrode serves as an efficient catalyst for the regeneration of cofactors in enzymatic syntheses requiring NADH. An alternative method to avoid completely the use of the toxic redox mediator viologen is represented by the application of the natural redox cofactor FAD for this purpose [117]. An excess of FAD entrapped together with lipoamide dehydrogenase allows self-mediated electron transfer. In order to test the capacity of FAD as a mediator, it was used in monomeric as well as in polymeric form in L-lactate synthesis using LDH coupled with coenzyme regeneration (Fig. 24).

In both cases, pyruvate was completely consumed and L-lactate was produced. The polymerized FAD film shows remarkable long-time stability.

To form a highly stable modified electrode, L-histidine was attached on the surface of a silver electrode by covalent bonds. At the surface of this histidine-modified electrode, NAD^+ can be reduced to NADH. High reduction rates were observed so that it is possible to integrate this regeneration system into bioreactors and biosensors [118].

Yun et al. used a cholesterol-modified gold amalgam electrode for the direct electrochemical regeneration of NADH [119]. This electrode was prepared by immersing gold plate in mercury and casting few drops of cholesteryl oleate solution over the gold amalgam. The reduction of NAD^+ occurs directly at the electrode, which is supposed to hinder the dimerization of the NAD radicals on its membrane surface. The NAD^+ reduction process was combined with the enzymatic oxidation of NADH in the presence of pyruvate and D-lactate dehydrogenase. In this system, the turnover number of NAD^+ was estimated as 1,400.

In another system, the $[Cp^*Rh^{III}(bpy)(H_2O)]^{2+}$ complex was used for the regeneration of NADH in combination with an isolated flavine-dependent monooxygenase (2-hydroxybiphenyl-3-monooxygenase) from *P. azelaica* [120]. This enzyme catalyzes the specific *ortho*-hydroxylation of several α-substituted phenol derivatives with simultaneous oxidation of NADH (Fig. 25).

In the nonoptimized process, productivity rates of 204 mg L^{-1} h^{-1} were achieved which can be considered to be attributed for about 50% of the in vivo process and 50% of the in vitro process with enzymatic regeneration of the coenzyme. The complex $[Cp^*Rh(bpy)(H_2O)]^{2+}$ shows high stability and activity

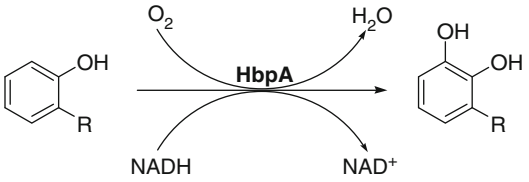


Fig. 25 Specific *ortho*-hydroxylation of α-substituted phenols catalyzed by NAD^+-dependent 2-hydroxybiphenyl-3-monooxygenase (HbpA). R = alkyl (Et, Pr, iPr), aryl (Ph, 2-HOC6H4), Hal (F, Cl, Br)

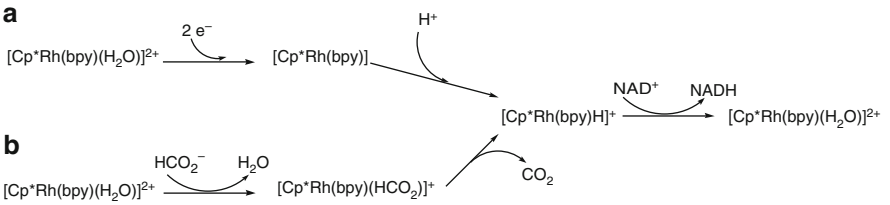


Fig. 26 Regeneration of NADH by the $[Cp^*Rh(bpy)(H_2O)]^{2+}$ complex. The corresponding hydridorhodium complex can be generated either electrochemically by cathodic reduction (method **a**) or chemically with formate (method **b**). ($Cp^* = C_5Me_5$; bpy = 2,2-bipyridine)

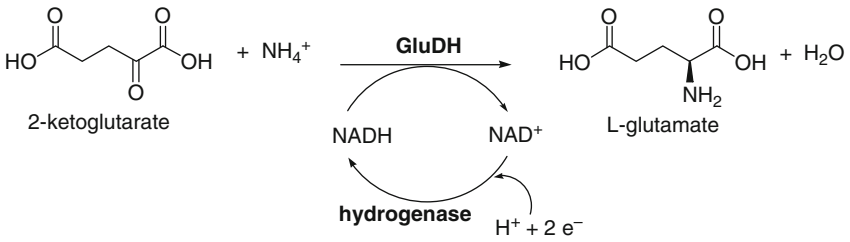


Fig. 27 Conversion of α-ketoglutarate to L-glutamate catalyzed by GluDH with coupled electrochemical regeneration of NADH using hydrogenase from *Alcaligenes eutrophus*

over a very broad range of pH and temperature. Electrical power as well as chemical compounds like formate or alcohols serves as electron donor (Fig. 26) [121].

The electroenzymatic reduction of NAD^+ was successfully coupled with a synthesis reaction [122]. Hydrogenase from *A. eutrophus* was applied to regenerate NADH electrochemically during the transformation of α-ketoglutarate into L-glutamate catalyzed by an L-glutamate dehydrogenase (Fig. 27).

For this purpose an analytical thin-layer cell and a laboratory-scale batch reactor were successfully used. The coupled system was controlled by the NADH concentration level. The major disadvantage of the process is the poor stability of the hydrogenase.

5 Further Methods

5.1 Photoelectrochemical Cofactor Regeneration

To overcome the low current densities obtained in direct electrochemical oxidation of NAD(P)H, photochemical regeneration methods have been investigated. Photoelectrochemical regeneration of $NAD(P)^+$ was developed and applied to the oxidation of cyclohexanol and butan-2-ol using HLADH as well as ADH from *T. brockii* [123]. The photosensitizer tris(2,2′-bipyridyl)ruthenium(II) is photochemically excited by visible light (Fig. 28). The excited Ru(II)* complex acts as electron donor for *N*,*N*′-dimethyl-4,4′-bipyridinium sulfate (MV^{2+}) forming tris(2,2′-bipyridyl)ruthenium(III) and the MV-cation radical. The Ru(III) complex oxidizes NAD(P)H effectively, thus regenerating the cofactor for the oxidizing enzyme. The MV-cation radical can be oxidized again at an appropriate anode. Cofactor as well as enzyme remained stable during the reaction process.

5.2 Chemical Cofactor Regeneration

Chemical cofactor regeneration systems often consist of *o*-chinoid mediators acting as hydride ion removing agents. The final electron acceptor oxygen forms hydrogen peroxide, which is destroyed by added catalase. Several conditions must be fulfilled for the process to be efficient [92]. To obtain high yields in acceptable periods, the turnover frequencies of the quinoid mediators need to be very high. The mediators have to be stable in reduced as well as oxidized states. Furthermore, they should be readily soluble in aqueous solutions and they should react selectively only with the

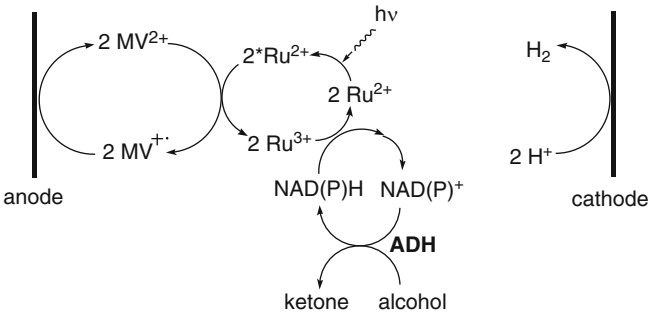


Fig. 28 Photoelectrochemical cofactor regeneration with simultaneous alcohol oxidation using ADHs from horse liver (HLADH) or *Thermoanaerobacter brockii*. Ru^{2+} serves as photosensitizer while MV^{2+} (methyl viologen) functions as primary electron acceptor which can be reoxidized at the anode

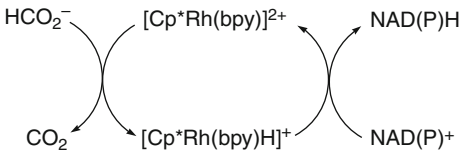


Fig. 29 Chemical cofactor regeneration with a homogeneous reduction system using cationic rhodium complexes. The reducing equivalents are supplied by formate

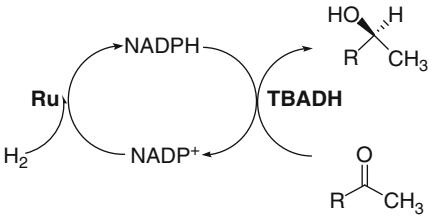


Fig. 30 Reduction of ketones catalyzed by TBADH (ADH from *Thermoanaerobacter brockii*) with simultaneous regeneration of NADPH using H_2 and the Ru-catalyst $[RuCl_2(TPPTS)_2]_2$. *TPPTS* tris(*m*-sulfonatophenyl)-phosphine

reduced cofactor. Phenanthroline transition-metal complexes like tris(1,10-phenanthroline-5,6-dione)ruthenium(II) fulfill these requirements. They are effective catalysts for the chemical aerobic regeneration of NAD^+ [124]. $NAD(P)^+$ can be reduced chemically via an efficient, purely homogeneous reduction system in which the reducing equivalents, two electrons and one proton, are supplied by formate and transferred to pentamethylcyclopentadienyl-2,2′-bipyridine-rhodium (III) (Fig. 29). This mediator is able to reduce $NAD(P)^+$ [124].

A further method for chemical cofactor regeneration is the NAD(P)H regeneration using dihydrogen (H_2). This method is preferred because H_2 is the lowest cost reducing agent and it yields no considerable byproducts. A simple transition-metal catalyst using ruthenium(II) complexes was developed for the direct reaction between H_2 and $NADP^+$ under conditions suitable for coupling of NADPH regeneration to an enzymatic reduction [125]. Mild reaction conditions are sufficient; the reduction of the coenzyme shows high selectivity vs the enzymatic substrate as well as high regioselectivity yielding 1,4-dihydropyridine. The ruthenium(II) complex $[RuCl_2(TPPTS)_2]_2$ (TPPTS = tris(*m*-sulfonatophenyl)-phosphine) is a water-soluble catalyst for the reaction of H_2 with carbohydrates under relatively mild conditions. This complex is able to reduce $NADP^+$ in quantities corresponding to three to four turnovers. The reduction of 2-heptanone to (*S*)-2-heptanol using ADH from *T. brockii* with a catalytic amount of $NADP^+$ and $[RuCl_2(TPPTS)_2]_2$ was performed in order to determine whether this system is effective in conjunction with NADPH-dependent enzymatic reductions (Fig. 30).

The sample was stirred at 60°C under 5 atm of H_2. After 4 h 18% of 2-heptanone was reduced, corresponding to 10 turnovers of both $NADP^+$ and ruthenium. The 2-heptanol formed after 18 turnovers showed an *ee* of 40%. This value, although low, demonstrates the involvement of the enzyme in the reduction. The investigated coenzyme recycling system does operate, but the enantioselectivity of the ketone reduction is quite poor.

6 Whole Cell Applications Based on Recombinant Enzymes ("Designer Cells")

The use of isolated purified enzymes obviously avoids the formation of unwanted byproducts which may be generated from otherwise present and usually unknown enzymes. Isolated enzymes may be more active and selective than whole cells, but their recovery and purification is very expensive. Whole cell biotransformations show a lot of advantages such as the increased stability of enzymes which exist in whole cells. If these enzymes depend on cofactors, whole cell biotransformations become even more favorable because the use of whole cells allows the intracellular cofactor pool to be exploited, and any addition of exogenous cofactors becomes unnecessary. Otherwise the costs for the required cofactors often exceed that for the enzyme or the value of the product. Therefore, whole cell biotransformations are very interesting for industrial applications.

Wild type cells often contain several enzymes which carry out the same reaction. Unfortunately, in many cases these enzymes produce compounds with opposite stereoselectivity. Therefore, whole cells in which such enzymes are present cannot be applied for the synthesis of enantiomerically pure products. To increase the stereoselectivity of a whole cell reaction, recombinant DNA techniques need to be applied. Very common is the overexpression of the enzyme, which catalyzes the particular reaction in a suitable heterologous host such as *E. coli*. The simultaneous overexpression of an enzyme which catalyzes the regeneration of the consumed cofactor is highly efficient. Ideally, growing cells should provide simultaneously the enzyme for the desired reaction as well as the cofactor regenerating enzyme. Such so-called "designer cells" seem to be very promising for technical applications.

There are different possibilities for performing whole cell biotransformations with several enzymes. Each enzyme can be produced from another strain. To carry out the desired reaction, these strains are combined in different amounts which depend on the particular enzyme activities. Disadvantages of this method are diffusion problems and the membrane barrier [126]. Alternatively, all enzymes can be expressed in one single host strain. This can be reached by means of recombinant DNA techniques in several ways: The genes can be expressed under one single promoter or under different promoters. Furthermore, it is possible to express all genes from the same promoter but with different copy numbers. Using these methods it is possible to create tailor-made catalysts for manifold purposes.

For the synthesis of (*R*)-alcohols by reduction of the corresponding ketones, an *E. coli* whole-cell catalyst was constructed based on an ADH from *L. kefir*. Some preparative examples using these "designer cells" are summarized in section 9.4.

For Baeyer-Villiger oxidations of ketones into lactones, recombinant whole-cell catalysts were constructed using the CHMO from *A. calcoaceticus* and cells of *S. cerevisiae* [127–129] or *E. coli* [130–132] as host system. These "designer cells" were used for example for the biotransformation of bicyclo[3.2.0]-hept-2-ene-6-one [133–136]. In order to avoid the substrate and product inhibition, an adsorber

resin was used for in situ substrate feeding and product removal [137]. Another example refers to the oxidation of 2-benzyloxycyclopentanone. Out of the racemic mixture of this substrate, the (*R*)-enantiomer is oxidized to the corresponding lactone (*ee* 97%), whereas the remaining (*S*)-substrate is racemized in situ at basic pH or in the presence of an anionic exchange resin [133].

7 Coenzyme-Dependent Reactions in Organic Solvents

Even though enzyme-catalyzed conversions provide plenty of advantages, there are also some drawbacks. Especially unsatisfactory stability of the enzymes and elaborate isolation and purification processes may cause problems [138]. Because of the limited solubility of many organic substrates in aqueous solution, concentrations of reagents are generally particularly lower in biocatalytic than in chemical processes. This circumstance requires larger reaction volumes and complicated product recovery. Furthermore, the highly nucleophilic property of water often causes undesirable side reactions such as hydrolysis and decomposition of educts as well as products. To extend the field of application of biotransformations the idea of carrying out enzymatic processes in organic solvents was investigated. The use of organic solvents in enzymatic systems can overcome some of the problems. The catalysis of reactions that are unfavorable in water can be accomplished. Water-induced side reactions can be suppressed and hydrophobic substrates can be solubilized. Furthermore, the recovery of some products and biocatalysts is facilitated [138]. However, all these advantages can only take effect if the problem of coenzyme dependence is also solved. The influence of organic solvents on coenzyme-dependent reactions has also been studied.

A general method has been developed for utilization of cofactor-requiring enzymes in organic media [139]. ADH from horse liver as well as NADH were attached onto the surface of glass beads and afterwards suspended in a water-immiscible organic solvent containing the substrate. This method can be applied to other NAD^+-dependent enzymes as well. Both NADH and NAD^+ are efficiently regenerated with ADH-catalyzed oxidation of ethanol and reduction of isobutyraldehyde, respectively (Fig. 31).

Asymmetric oxidoreductions performed in isopropyl ether allow syntheses of optically active alcohols with *ee* >95% on a 1–10 mmol scale. Nakamura et al. investigated the effect of organic solvents on the reduction of α-keto esters mediated by bakers' yeast [140]. The reduction of ethyl 2-oxoheptanoate was tested in various solvents. Best results were achieved with benzene so further experiments were performed with benzene. Conversion only takes place in the presence of small amounts of water. The reduction of six α-keto esters was examined regarding the stereoselectivity of the corresponding α-hydroxy esters. The reactions were performed in aqueous systems as well as in benzene. In aqueous systems, the formed hydroxy esters show (*S*)-stereoselectivity while the stereochemistry of the reaction shifts markedly towards the production of (*R*)-α-hydroxy esters in benzene.

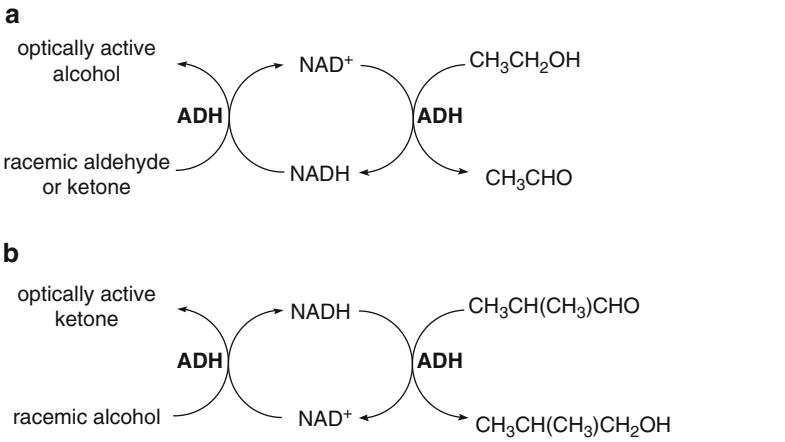


Fig. 31 Preparation of optically active compounds employing HLADH and NADH, which are codeposited onto glass beads in a monophasic organic solvent. (**a**) Reduction reaction to produce chiral alcohols in the presence of ethanol for NADH regeneration. (**b**) Oxidation reaction to produce enantiomerically pure alcohol or a ketone out of the racemic mixture coupled with the reduction of isobutyraldehyde to regenerate NAD^+

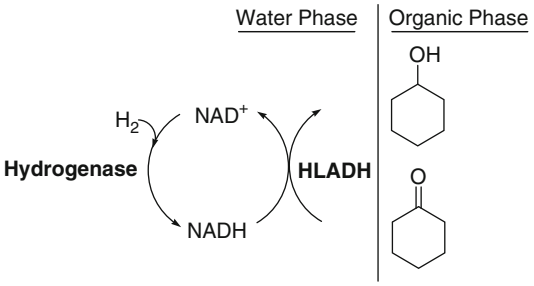


Fig. 32 Reduction of cyclohexanone catalyzed by HLADH with simultaneous hydrogenase-driven regeneration of NADH in an organic–aqueous two-phase system

To catalyze the reduction of cyclohexanone to cyclohexanol in an organic–aqueous two-phase system, soluble hydrogenase from *A. eutrophus* as an NADH regenerating catalyst was coupled with HLADH (Fig. 32) [37].

Several organic solvents were investigated with regard to stability and activity of HLADH as well as their influence on the hydrogenase-driven reaction. Hydrophobic solvents such as heptane or toluene proved to be the most suitable solvents for the coupled enzyme-system. Furthermore, it became apparent that nonimmobilized cells, permeabilized with cetyl-trimethylammonium bromide, showed the best results for NADH regeneration. After optimization the conversion in heptane with 10% water yields 99% cyclohexanol by reduction of cyclohexanone.

The bioconversion of hydrocortisone to prednisolone by free, immobilized, and reused immobilized cells of *Bacillus sphaericus* ATCC 13805, *B. sphaericus* SRP III, and *Arthrobacter simplex* 6946 was investigated in an aqueous as well as a two-liquid-phase system using various organic solvents [141]. The immobilized cells

lead to higher product yield than the free cells, both applied in an aqueous system. The best solvent for a two-phase system for immobilized cells as the catalyst was *n*-decane with a surplus of water in a ratio of 1:6. In this system a high yield of product was formed within a short incubation period. The system butyl acetate/aqueous buffer (1:30) lead to the highest product yield, but the reaction to reach this high yield took a very long time. While immobilized *B. sphaericus* ATCC 13805 and *B. sphaericus* SRP III cells gave about 5–30% higher prednisolone yields in the aqueous phase than in the two-phase system, the cells of *A. simplex* 6946 gave higher activity in the two-phase system.

The reduction of 2-pentanone was carried out with ADH from *T. brockii* and $NADP^+$ immobilized on celite, and 2-butanol was added as co-substrate for coenzyme regeneration [142]. The reactions were performed in hexane at controlled water activity because the water content in the organic solvent is crucial for enzyme catalysis. An enzyme is generally more active and flexible at higher hydration degree. In this system it was found that the initial activity was strongly influenced by the water activity. By increasing the water activity from 0.32 to 0.96 at 40°C the initial activity increased by a factor of 130. The stereoselectivity of the enzyme was also affected by the water activity. For optimum enantioselectivity high water activities are required.

NAD^+-dependent 2-hydroxybiphenyl-3-monooxygenase (HbpA) was applied for an effective in vitro biotransformation system with enzymatic cofactor regeneration in organic–aqueous reaction media [143]. The stability and activity of HbpA was investigated in the presence of different apolar and polar organic solvents. Good results were achieved in 80 vol.% decanol and 10 vol.% methanol. As YADH was used as cofactor regenerating enzyme, conversions were investigated in the presence of 10 vol.% methanol which serves both as cosolvent and substrate for YADH (Fig. 33).

This monophasic biotransformation system was used for the hydroxylation of 2,2′-dihydroxybiphenyl to 2,2′,3-trihydroxybiphenyl within a reaction time of 2.5 h and 2,2′,3,3′-tetrahydroxybiphenyl after 23 h. A mean volume productivity of 0.43 and 0.05 g l^{-1} h^{-1}, respectively, was achieved for these biotransformations. Furthermore, HbpA and FDH were applied in 80 vol.% decanol to hydroxylate 2-hydroxybiphenyl to 2,3-dihydroxybiphenyl with a productivity of 0.46 g l^{-1} h^{-1} and a total turnover number of 503. In 10 vol.% aqueous methanol the total turnover number of 30 and enzyme stability for at least 60 h were observed. Because of the

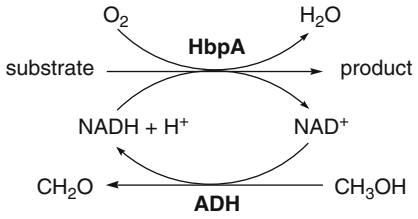


Fig. 33 Reaction scheme of conversions using 2-hydroxybiphenyl-3-monooxygenase (HbpA) as well as ADH for cofactor regeneration performed in the presence of methanol which acts as both cosolvent and substrate

good operational stability of HbpA in decanol and methanol this enzyme can be used for hydroxylations of phenolic substrates in methanol/YADH and decanol/FDH systems with high productivities.

Biphasic systems proved to be advantageous as well in the biocatalytic synthesis of (–)-1-trimethylsilylethanol which was performed by asymmetric reduction of acetyltrimethylsilane with an isolate from *Rhodotorula* sp. AS2.2241 [144]. Immobilized cells were employed due to the easy separation of the product as well as the improved tolerance against unfavorable factors. In an aqueous/organic solvent biphasic system higher product yield and enantiomeric excess were achieved as compared to an aqueous monophasic system. Several organic solvents were examined, and isooctane was found to be the most suitable organic phase for the reaction.

For the asymmetric biocatalytic reduction of ketones with in situ cofactor regeneration an enzyme-compatible biphasic reaction medium has been developed (Fig. 34) [53].

Reactions with these compounds suffer from very low substrate concentrations due to the low solubility of hydrophobic ketone substrates in aqueous media, which leads to unsatisfactory volumetric productivities. To achieve higher substrate concentrations, a biphasic reaction medium was introduced. The system water/*n*-heptane (4:1) proved to be the most suitable system with regard to stability of the examined enzymes. The large-scale available (*S*)-specific ADH from *R. erythropolis* as well as FDH from *C. boidinii* are stable for long periods of time in this aqueous–organic solvent system. Preparative conversions with a variety of aromatic ketone substrates were carried out with this reaction medium. For example, *p*-chloroacetophenone was converted into the corresponding (*S*)-alcohol with >99% *ee* and 69% conversion. The obvious increase in volumetric productivity is due to the higher substrate concentrations. The reduction of *p*-chloroacetophenone

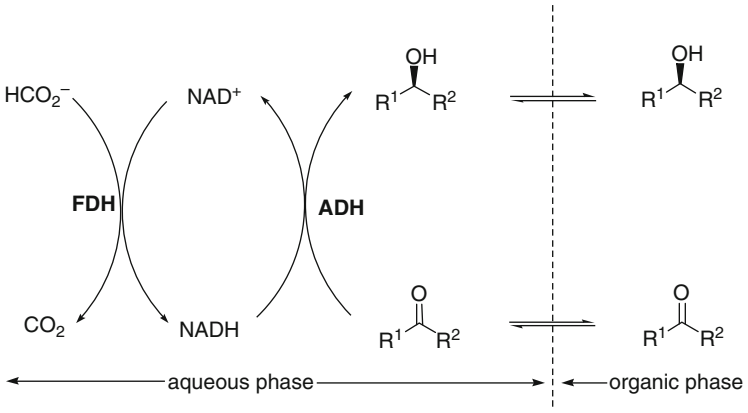


Fig. 34 Enzymatic reduction in biphasic media. Ketones are reduced enantioselective to the corresponding (*S*)-alcohols by ADH from *Rhodococcus erythropolis*. Regeneration of the cofactor NADH is carried out by FDH from *C. boidinii*. The introduction of a biphasic system allows higher substrate concentrations

was investigated as a model reaction. At 20 and 40 mM substrate concentrations conversions were enhanced to yield 77 and 75%, respectively. Even at 200 mM concentration, a still satisfactory conversion of 63% was achieved. Thus, with this biphasic system, ketones can be reduced efficiently via enzymatic in situ cofactor regeneration at high substrate concentrations in the direct presence of an organic solvent.

8 The Nicotinamide Dinucleotide Binding Motif and Its Reversal by Protein Engineering

Coenzyme specificity is an important property of $NAD(P)^+$-dependent oxidoreductases which is associated with different metabolic functions. Most of the NAD(H)-dependent enzymes are involved in oxidative catabolic reactions, whereas NADP(H) is needed for reductive biosynthetic processes [145]. As both coenzymes differ in the phosphate group esterified at the 2′-position of adenosine ribose only [146], dehydrogenases require highly efficient structural features determined by a limited number of amino acid residues to discriminate between the two coenzyme types. For preparative or industrial applications, use of NAD^+ as the coenzyme is significantly preferred over $NADP^+$ due to the more favorable price, the increased stability, and the better feasibility of cofactor regeneration. NADPH decomposes 3–4 times faster at pH 5 and 10–20 times faster at pH 7 [1]). Therefore, it is of interest to investigate ways to create mutants of $NADP^+$-dependent enzymes that prefers NAD^+. Amino acid residues to discriminate between the two cofactor types are located in most cases 20–25 amino acids C-terminal of the glycine-rich consensus motif (GxGxxG) of dehydrogenases, which bind the cofactor in a so-called Rossman-fold. In the last few years, different strategies and methods have been published to alter the cofactor specificity of enzymes, each depending on indications from mutagenesis studies or 3D-structures focusing on the cofactor interacting amino acids within the Rossmann fold [147, 148]. In NAD^+-dependent dehydrogenases, typically aspartic or glutamic acid is responsible for the discrimination by rejecting the negatively charged 2′-phosphate group of $NADP^+$. Otherwise, in $NADP^+$-dependent enzymes, charged arginine or lysine residues can support electrostatic interactions with this 2′-phosphate group. These structural insights are the basis for structure guided engineering of cofactor binding sites [146, 148–152]. Rosell and colleagues (2003) altered the cofactor specificity of an $NADP^+$ dependent medium-chain ADH towards NAD^+ by creating a triple mutant G223D/T224I/H225N. The incorporation of a negatively charged residue and a bigger amino acid led to $NADP^+$ rejection. Additional discharge of this region stabilizes the binding of NAD^+. Another example is represented by the *Candida tenuis* xylose reductase, whose coenzyme preference was changed from NADPH to NADH by structure guided site-directed mutagenesis of two amino acids ($Lys^{274} \rightarrow$ Arg and $Asn^{276} \rightarrow$ Asp) [150].

9 Examples of Small-Scale and Industrial Applications

9.1 L-tert-*Leucine as an Example of Ton-Scale Synthesis*

Because of its bulky, inflexible, and hydrophobic side chain, *tert*-leucine (2-amino-3,3-dimethylbutanoic acid, Tle) is an important amino acid used as template or catalyst compound in asymmetric synthesis and in peptidic medicinal compounds. L-Tle has attracted much attention as a key component of newly emerged drugs or as building block of ligands, catalysts, and auxiliaries for asymmetric synthesis. It is synthesized in ton-scale by reductive amination of trimethylpyruvic acid by means of LeuDH from *Bacillus stearothermophilus* with very high yield and excellent optical purity [153]. NADH, which is consumed during the reaction, can be regenerated by FDH from *C. boidinii* (Fig. 35).

Further improvement of this process can be reached by immobilization of the enzymes in an EMR. The coenzyme can be enlarged chemically by binding it to polyethyleneglycol (PEG) in order to prevent leakage of the coenzyme through the membrane. This LeuDH catalyzed synthesis of enantiomerically pure L-Tle runs now routinely on a multiton scale at Degussa [154].

9.2 L-6-Hydroxynorleucine

L-6-Hydroxynorleucine is a key intermediate used for the synthesis of a vasopeptidase inhibitor. It was synthesized from 2-keto-6-hydroxyhexanoic acid by reductive amination using beef liver GluDH and GDH from *Bacillus* sp. for regeneration of NADH (Fig. 36) [155]. The educt of the reaction, 2-keto-6-hydroxyhexanoic acid is in equilibrium with 2-hydroxytetrahydropyran-2-carboxylic acid.

The reaction was completed within 3 h with reaction yields of 89–92% and *ee* >99%.

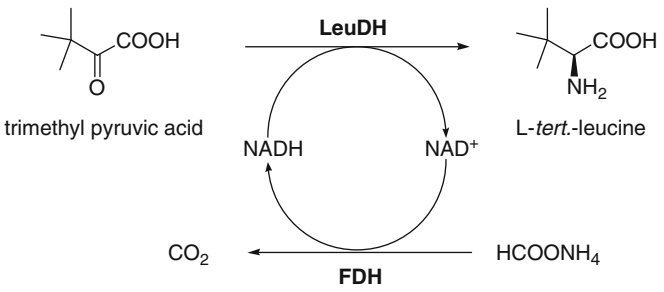


Fig. 35 Production of L-*tert*-leucine by reductive amination of trimethyl pyruvic acid catalyzed by leucine dehydrogenase (LeuDH) and formate dehydrogenase (FDH) for cofactor regeneration

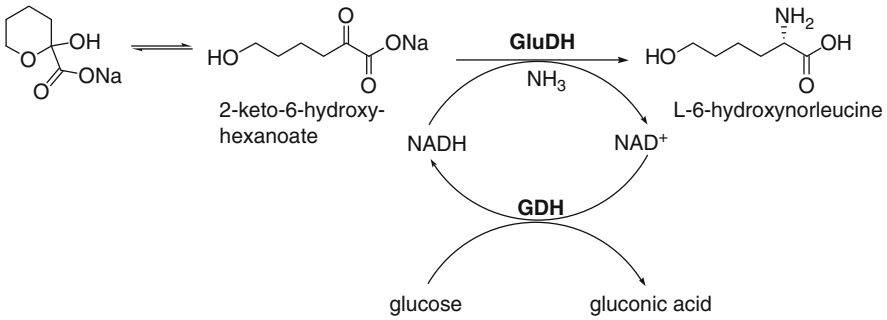


Fig. 36 Conversion of 2-keto-6-hydroxyhexanoic acid to L-6-hydroxynorleucine. Reductive amination is carried out by glutamate dehydrogenase (GluDH), coenzyme regeneration by glucose dehydrogenase (GDH)

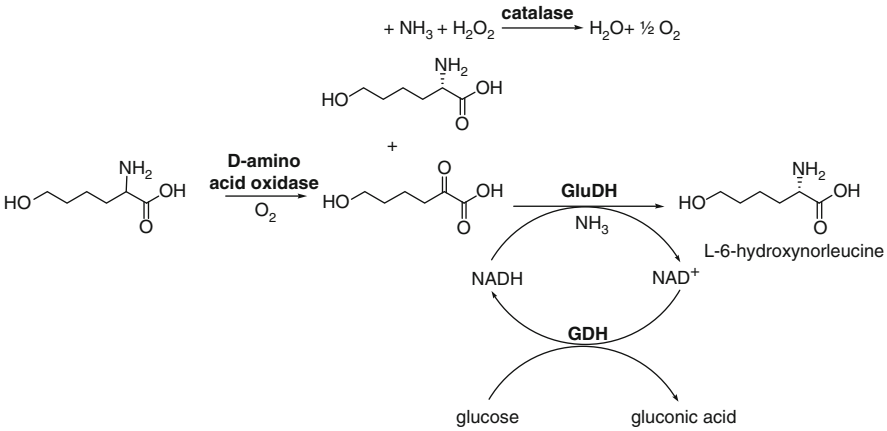


Fig. 37 Conversion of racemic 6-hydroxynorleucine acid to L-6-hydroxynorleucine. The keto acid is formed in situ by D-amino acid oxidase, H_2O_2 is destroyed by catalase

A modification of this reaction concerns the availability of the keto acid substrate. To circumvent its complicated lengthy chemical synthesis, 2-keto-6-hydroxyhexanoic acid was synthesized by treatment of racemic 6-hydroxynorleucine with D-amino acid oxidase and catalase (Fig. 37). The production of racemic 6-hydroxynorleucine occurs by hydrolysis from 5-(4-hydroxybutyl)hydantoin. D-Amino acid oxidase converts the D-enantiomer of racemic 6-hydroxynorleucine to the corresponding ketoacid which is reductively aminated to L-6-hydroxynorleucine by GluDH.

6-Hydroxynorleucine was formed in this modified process with yields of 91–97% and optical purities of >99%.

9.3 Synthesis of Allysine Ethylene Acetal in 80 kg-Scale

(*S*)-2-Amino-5-(1,3-dioxolan-2-yl)-pentanoic acid (allysine ethylene acetal) is an important building block used for an alternative production of VANLEV (omapatrilat, Fig. 38) which is a vasopeptidase inhibitor.

Allysine ethylene acetal is synthesized from the corresponding keto acid by reductive amination using phenylalanine dehydrogenase (PDH) from *Thermoactinomyces intermedius* ATCC 33205 combined with FDH from *C. boidinii* SC13822 for the regeneration of NADH (Fig. 39).

With heat-dried cells of *T. intermedius* as PDH source and heat-dried *C. boidinii* as FDH source, reaction yields of 84% and *ee* >98% were achieved. As a major drawback of this system the production of *T. intermedius* could not be scaled up. To solve this problem heat-dried *E. coli* containing *T. intermedius* PDH and heat-dried *C. boidinii* as source of FDH were used. In this procedure, recovery was not a problem and 197 kg allysine ethylene acetal were synthesized in three batches with an average yield of 91% and *ee* >8%. In a third approach, recombinant *Thermoactinomyces* PDH as well as heat-dried methanol-grown *Pichia pastoris* expressing endogenous FDH were used to produce 15 kg allysine ethylene acetal with a yield of 97% and *ee* >98%. This process allowed both enzymes to be produced during a

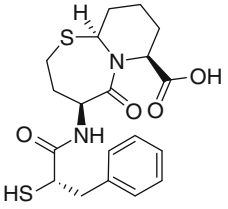


Fig. 38 Structure of the vasopeptidase inhibitor VANLEV (omapatrilat)

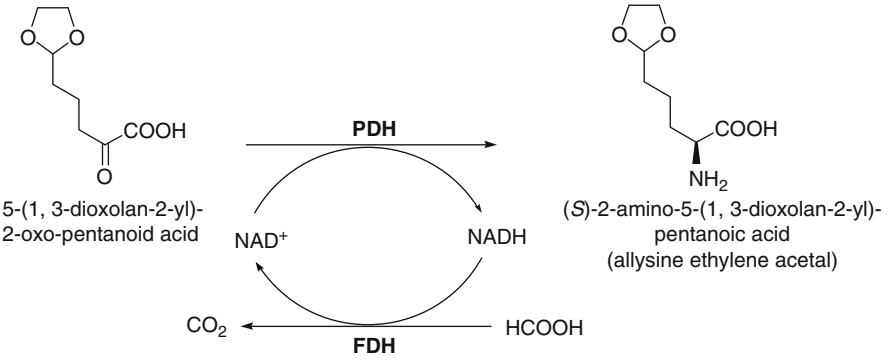


Fig. 39 Conversion of 5-(1,3-dioxolan-2-yl)-2-oxo-pentanoid acid to allysine ethylene acetal by reductive amination using phenylalanine dehydrogenase (PDH) and formate dehydrogenase (FDH) for cofactor recycling

single fermentation batch. Furthermore, both enzymes were conveniently expressed in about the right ratio that was used for the reaction.

9.4 Synthesis of Chiral Alcohols

9.4.1 (S)-Alcohols

Several chiral alcohols were prepared out of their corresponding ketones at Degussa AG utilizing recombinant (*S*)-specific ADH from *R. erythropolis* [2, 52] as the catalyst and FDH of *C. boidinii* for regeneration of NADH. Due to the low solubility of hydrophobic ketones in water, an enzyme-compatible biphasic reaction medium consisting of water and *n*-heptane (4:1) has been developed (see Fig. 34). The broad substrate specificity of the *S*-ADH allows the formation of a broad range of aliphatic or aromatic *sec*-alcohols, or hydroxy esters. As an example, 200 mM (*S*)-*p*-chloroacetophenone resulted in a yield of 63% of product, purified by chromatography. A conversion of 65%, and a slightly lower enantioselectivity of 97% *ee* was observed when using *p*-bromoacetophenone, and the asymmetric reduction of phenoxyacetone proceeded quantitatively with an enantioselectivity of 99% *ee* [53, 54, 156].

9.4.2 (R)-Alcohols

Based on the (*R*)-specific ADH from *L. kefir*, a recombinant *E. coli* strain was constructed as a whole-cell biocatalyst, and co-expressed GDH was used for regeneration of NADPH [157]. These "designer cells" were applied for the reduction of 4-fluoroacetophenone to the corresponding optically active (*R*)-4-fluorophenylethan-1-ol at 0.5 M educt concentration [158]. After a reaction time of 23 h, a conversion of >95% has been achieved, and the purified isolated chiral alcohol showed an *ee* value of >99% (87% yield). (*S*)-β-Halohydrins were obtained with this whole-cell catalyst by means of an enantioselective reduction of the corresponding ketones with both high conversions of >95% and enantioselectivities of >99% (Fig. 40). Base-induced cyclization of the β-halohydrin led to enantiomerically pure (*S*)-epoxides in high yield and enantiomeric purity (>99% *ee*) [159].

9.4.3 Glyceraldehyde

Optically active L-glyceraldehyde can be obtained biocatalytically by use of a D-specific ADH. This enzyme reduces the D-enantiomer yielding glycerol and enantiopure L-glyceraldehyde (Fig. 41). NADPH as the coenzyme for the reduction was regenerated by GDH.

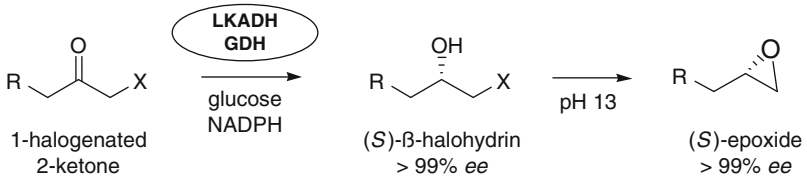


Fig. 40 Concept for the two-step synthesis of enantiomerically pure (*S*)-epoxides out of aliphatic 1-halogenated 2-ketones. The ketone was reduced by a recombinant whole-cell catalyst bearing alcohol dehydrogenase from *Lactobacillus kefir* (LKADH) and glucose dehydrogenase (GDH) for regeneration of NADPH. Base-induced cyclization of the enantiomerically pure (*S*)-β-halohydrin intermediate gave the desired (*S*)-epoxides in high yield and enantiomeric purity (>99% *ee*)

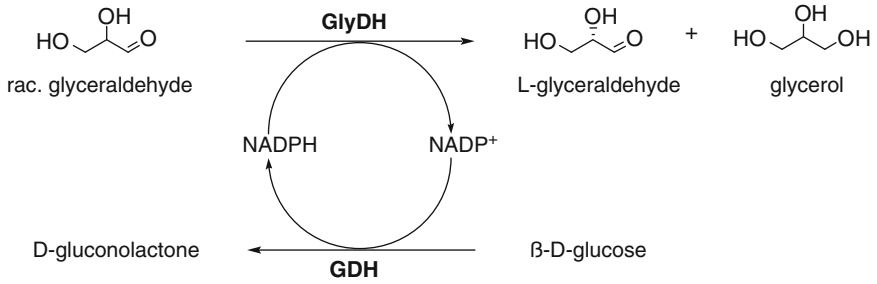


Fig. 41 Scheme of the synthesis of L-glyceraldehyde by reduction of D-glyceraldehyde out of the racemic mixture using NADP-dependent glycerol dehydrogenase from *Gluconobacter oxydans*

This reaction was carried out in a 250 mL-scale with 150 mM racemic glyceraldehyde. The kinetic resolution at a conversion of 50% resulted in L-glyceraldehyde with an *ee* >99% [160].

9.5 *Alcohol Dehydrogenase Catalyzed Synthesis of the Chiral Side Chain of Statins*

Synthetic statins are important lipid regulating drugs for the treatment of atherosclerosis and other diseases related to hyperlipidaemia, especially coronary heart disease. As the pharmacophore, all synthetic statins contain a saturated or partially unsaturated syn-3,5-dihydroxy C_7-carboxylate. An important building block for the synthesis of the side chain is represented by 4-chloro-3-hydroxybutanoate esters (CHBE). Both enantiomers can be obtained by enzyme-catalyzed reduction of the β-keto ester. The group of Kataoka and Shimizu found that an aldehyde reductase of *Sporobolomyces salmonicolor* [161] and a carbonyl reductase of *Candida*

magnoliae [162] can be used to catalyze NADPH-dependent stereospecific reduction of 4-chloro-3-oxobutanoate esters (COBE) to (*R*)-CHBE and (*S*)-CHBE, respectively [163] (Fig. 42). Aldehyde reductase from *S. salmonicolor* was cloned and overproduced in *E. coli* [164]. To use the advantages of a whole-cell biocatalyst, *E. coli* cells were constructed expressing both the reductase gene from *S. salmonicolor* and the GDH gene from *B. megaterium*. The asymmetric production of (*R*)-CHBE was investigated with this system [165]. The reaction was carried out in an organic solvent–water two-phase system. (*R*)-CHBE was formed in the organic phase with a molar yield of 94.1% and an *ee* of 91.7%.

Recombinant *E. coli* cells coexpressing GDH from *B. megaterium* for regeneration of NADPH were applied in industrial scale to produce (*R*)-CHBE or (*S*)-CHBE in high optical purity in 300–350 g L^{-1} scale [163, 166, 167]. Besides COBE, only glucose and a catalytic amount of $NADP^+$ were fed as the substrates, the turnover numbers for NADPH were calculated to be 13,500 and 35,000 mol CHBE/mol, respectively [168].

Furthermore, much work has been devoted to the development of stereoselective methods for preparing terminally functionalized 3,5-dihydroxy carboxylates. Müller et al. developed a new chemoenzymatic synthesis of the chlorinated 5-hydroxy-3-keto ester as the precursor for the dihydroxy hexanoate [169, 170]. The diketo ester is reduced exclusively in the C_5 position by $NADP^+$-dependent ADH from *L. kefir* or *L. brevis*. NADPH is regenerated by an excess of isopropanol by the same enzyme [171, 172] (Fig. 43). High concentrations of isopropanol are

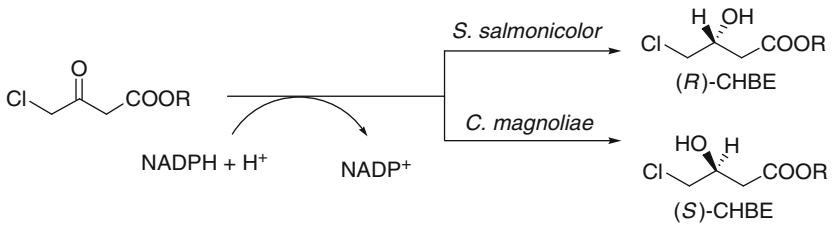


Fig. 42 Stereospecific reduction of 4-chloro-3-oxobutanoate ester (COBE) to (*R*)- and (*S*)-4-chloro-3-hydroxybutanoate ester (CHBE) by aldehyde reductase from *Sporobolomyces salmonicolor* and carbonyl reductase from *Candida magnoliae*, respectively

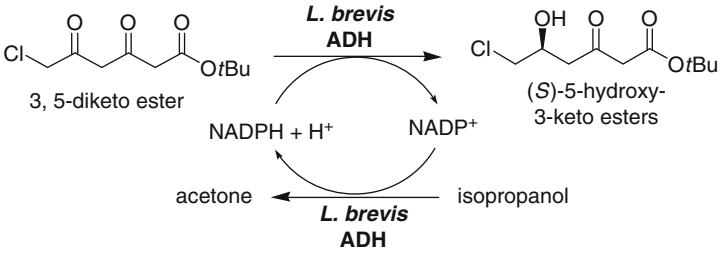


Fig. 43 Regio- and enantio-selective reduction of 3,5-diketoester

the driving force of this reaction, further on, the high concentrations of isopropanol enhance the solubility of the lipophilic diketo ester. A conversion of more than 90% is attained with an enantiomeric purity of the product of >99.5%. Reduction of the chlorinated compound is routinely performed on a 75 g scale (8 L fed batch).

9.6 Synthesis of Optically Active L-Amino Acids with Amino Acid Dehydrogenases

Amino acids are required worldwide on a large scale. For this reason their synthesis has been studied extensively. L-Amino acids were produced with various L-amino acid dehydrogenases coupled with FDH (Fig. 44) [173].

L-Leucine, L-valine, L-norvaline, L-methionine, L-phenylalanine, and L-tyrosine were produced by means of the recombinant *E. coli* cells with high yields (>80%) and high optical purities (up to 100% *ee*).

9.7 Optically Active α-Hydroxy Acids or α-Hydroxy Esters

Commercially available *C. boidinii*-FDH was used to recycle the NADH cofactor in stereospecific reductions by (*R*)-2-hydroxyisocaproate dehydrogenase from *L. casei* [174]. Enantiomerically pure (*R*)-2-hydroxy-4-methylpentanoic acid was obtained with 88% yield. The broad substrate specificity of this enzyme enables the synthesis of a broad range of enantiomerically pure α-hydroxy acids with aliphatic or aromatic side chains.

Methyl (*R*)-*o*-chloromandelate represents the key building block for the synthesis of clopidogrel, a platelet aggregation inhibitor widely administered to atherosclerotic patients with the risk of a heart attack or stroke that are caused by the formation of a clot in the blood. The worldwide sales of this pharmaceutical

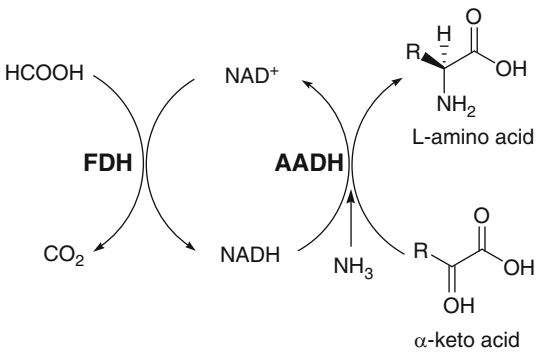


Fig. 44 Production of L-amino acids by coupling of FDH and amino acid dehydrogenase

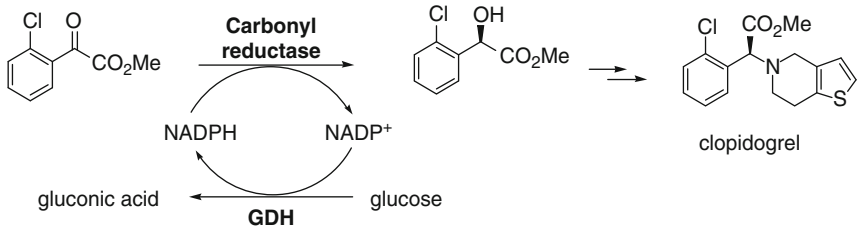


Fig. 45 Synthesis of methyl (*R*)-*o*-chloromandelate as the main precursor of the pharmaceutical compound clopidogrel

compound amounted to $6.4 billion per year. Ema et al. describe the stereospecific synthesis of this intermediate using a recombinant *E. coli* strain overproducing carbonyl reductase from *S. cerevisiae* (bakers' yeast) as the catalyst and GDH for regeneration of NADPH [175–177]. A product level as high as 178 g L^{-1} was attained on a 2-g scale (1.0 M) [178] (Fig. 45).

9.8 Multistep "One-Pot" Processes: Direct Coupling of a Chemical and an Enzymatic Step

The chemoenzymatic synthesis of chiral compounds usually implies one or several chemical steps in organic solvents, the separation and purification of the prochiral intermediate followed by an enzyme-catalyzed step in aqueous environment. Up to now there has only been a few examples to combine these processes as a "one-pot" synthesis in order to avoid the time-, effort-, and solvent-intensive steps of isolation of the prochiral compound. Such direct coupling of multistep syntheses contributes to a significantly improved process economy as well as to more sustainable synthetic routes. Quite recently, Gröger et al. published two examples to demonstrate the feasibility of a direct combination of chemical and enzymatic steps: a Wittig reaction, carried out in aqueous buffer has been coupled with an ADH catalyzed step yielding enantiomerically pure allylic alcohols [179] (Fig. 46) and a Suzuki cross-coupling has been directly linked with an ADH step yielding optically active biaryl alcohols or biaryl diols, respectively, with high conversion yield and excellent *ee* values (>99%) [180] (Fig. 47).

10 Conclusions: Capability and Limitations

This review describes the use of different methods for the regeneration of nicotinamide coenzymes, particularly enzymatic and electrochemical strategies.

An advantage of electrochemical cofactor regeneration is the use of mass-free electrons as regenerating agents. The low cost of electricity can be exploited.

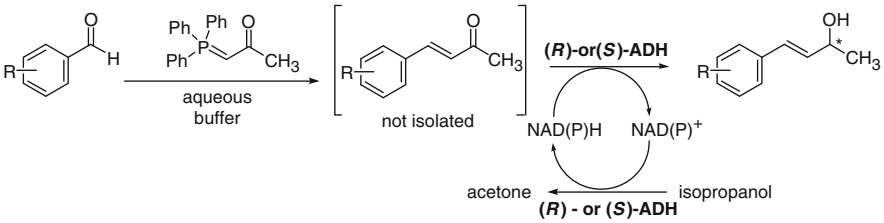


Fig. 46 One-pot two-step synthesis of allylic alcohols. The reductive step is catalyzed by (*R*)-ADH from *Lactobacillus kefir* or by (*S*)-ADH from *Rhodococcus* sp., respectively. The coenzyme is regenerated by an excess of isopropanol

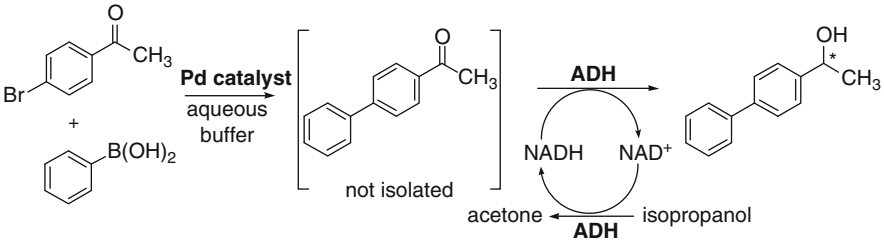


Fig. 47 One-pot two-step synthesis of biaryl alcohols. The reductive step is catalyzed by (*S*)-ADH from *Rhodococcus* sp., the coenzyme is regenerated by an excess of isopropanol

Furthermore, there are no co-substrates necessary and no couple products are formed. Hence, products can easily be isolated.

The direct electrochemical oxidation of NAD(P)H is possible but for this purpose relatively high oxidation potentials are necessary and potential passivation of the electrodes can occur. Because of the low speed of NADH oxidation and to avoid fouling of the electrodes, mediators like phenanthroline derivates are often used for advanced electron transfer. Those systems can be coupled efficiently with enzyme-catalyzed reactions which require NAD(P)H oxidation.

A major disadvantage of electrochemical reduction systems is the one-electron reduction of $NAD(P)^+$ followed by radical coupling which results in an enzymatically inactive dimer. Therefore, indirect electrochemical methods become necessary like the introduction of an additional regenerating enzyme which transfers two electrons from the mediator to $NAD(P)^+$. Enzymes like ferredoxin-$NADP^+$ reductase or DP can be employed as regenerating enzymes while methyl viologen is often used as mediator. In order to prevent the need for a second regenerating enzyme, the redox catalyst should transfer electrons simultaneously or one hydride ion at potentials less negative than -0.9 V. Substituted bipyridyl rhodium complexes, for example, are well suited for this task. A further limitation for the application of electrochemical regenerating methods concerns the technical realization of production plants. Until now, neither large-scale apparatus suitable for the batch-wise

production nor continuously working plants have been described which include an electrochemical regeneration of the nicotinamide coenzyme.

Though electrochemical regeneration of nicotinamide cofactors has some drawbacks, it is a beneficial technology for ecological reasons. Because of the use of electrodes as clean reagents, this method is very interesting for synthetic applications.

Enzymatic regeneration of nicotinamide coenzymes shows many advantages. In general, the selectivity of enzymes for nicotinamide coenzymes is very high. This strategy shows great compatibility with enzyme-catalyzed synthesis. Furthermore, by the use of enzymatic assays, reaction control is very easy to accomplish. However, the properties of enzymes to be used for regeneration of coenzymes have to be well considered, especially for recombinant whole-cell catalysts. The specific activity and the expression level of the selected enzyme has to fit the activity of the coenzyme-consuming primary enzyme. Some enzymes catalyze a nearly irreversible reaction which can be utilized advantageously. The regeneration system has to be stable over a long period of time and under the chosen reaction conditions. Products of the regenerating reaction must not interfere with the primary enzymatic reaction or hamper severely the downstream process. Furthermore, the chosen enzyme and the regeneration substrate must be cheap.

The enzymatic regeneration of reduced nicotinamide cofactors is possible in several ways. The best-investigated enzyme for regeneration of NADH is FDH and this method has found a wide area of application. A drawback of this enzyme is its inability to reduce $NADP^+$ and the low specific activity. Other enzymes like G-6-P-DH or GDH are able to regenerate NADH as well as NADPH, which makes them more promising for application. NAD^+ as well as $NADP^+$ dependent ADHs themselves are important enzymes for enzymatic regeneration of reduced or oxidized nicotinamide cofactors. The standard method for the regeneration of $NAD(P)^+$ is the use of GluDH. In addition, L-LDH is able to regenerate NAD^+. A recently published and very promising approach is the application of NADH oxidase for the regeneration of oxidized nicotinamide cofactors.

Most of these enzymatic ways to regenerate nicotinamide cofactors have been successfully established in a multitude of synthetic processes. The introduction of organic–water two-phase systems seems to be a useful method to overcome problems with the low solubility of many organic substrates. Bioconversions using whole cells have been well investigated because they show some decisive advantages like increased enzyme stability. The use of recombinant DNA techniques offers a wide field of application. Tailor-made cells can be created to perform whole cell biotransformations efficiently.

Presently, quite a lot of examples of enzymatic redox systems are known which include an efficient regenerating step and which are applied for large-scale manufacturing of important chiral compounds like fine chemicals or building blocks for pharmaceuticals in a very efficient way.

The developments of recombinant DNA techniques will later provide promising approaches of using enzymatic cofactor regeneration in the years to come. The preparation of suitable “designer cells” will become simpler and new bioconversion processes will be designed.

References

1. Chenault HK, Whitesides GM (1987) Appl Biochem Biotechnol 14:147–197
2. Hummel W, Kula MR (1989) Eur J Biochem 184:1–13
3. Schütte H, Flossdorf J, Sahm H, Kula M-R (1976) Eur J Biochem 62:151–160
4. Allen SJ, Holbrook JJ (1995) Gene 162:99–104
5. Avilova TV, Egorova OA, Ioanesyan LS, Egorov AM (1985) Eur J Biochem 152:657–662
6. Kato N, Sahm H, Wagner F (1979) Biochim Biophys Acta 566:12–20
7. Gul-Karaguler N, Sessions RB, Clarke AR, Holbrook JJ (2001) Biotechnol Lett 23:283–287
8. Izumi Y, Kanzaki H, Morita S, Futazuka H, Yamada H (1989) Eur J Biochem 182:333–341
9. Izumi Y, Kanzaki H, Morita S, Yamada H (1987) FEMS Microbiol Lett 48:139–142
10. Karzanov VV, Bogatsky YA, Tishkov VI, Egorov AM (1989) FEMS Microbiol Lett 51: 197–200
11. Müller U, Willnow P, Ruschig U, Hopner T (1978) Eur J Biochem 83:485–498
12. Tishkov VI, Galkin AG, Marchenko GN, Tsygankov YD, Egorov AM (1993) Biotechnol Appl Biochem 18:201–207
13. Nanba H, Takaoka Y, Hasegawa J (2003) Biosci Biotechnol Biochem 67:2145–2153
14. Yamamoto H, Mitsuhashi K, Kimoto N, Kobayashi Y, Esaki N (2005) Appl Microbiol Biotechnol 67:33–39
15. Yamamoto H, Mitsuhashi K, Kimoto N, Matsuyama A, Esaki N, Kobayashi Y (2004) Biosci Biotechnol Biochem 68:638–649
16. Cosgrove MS, Naylor C, Paludan S, Adams MJ, Levy HR (1998) Biochemistry 37:2759–2767
17. Wong CH, Gordon J, Cooney CL, Whitesides GM (1981) J Org Chem 46:4676–4679
18. Okuno H, Nagata K, Nakajima H (1985) J Appl Biochem 7:192–201
19. Bright JR, Byrom D, Danson MJ, Hough DW, Towner P (1993) Eur J Biochem 211:549–554
20. Smith LD, Budgen N, Bungard SJ, Danson MJ, Hough DW (1989) Biochem J 261:973–977
21. Mitamura T, Urabe I, Okada H (1989) Eur J Biochem 186:389–393
22. Nagao T, Mitamura T, Wang XH, Negoro S, Yomo T, Urabe I, Okada H (1992) J Bacteriol 174:5013–5020
23. Pauly HE, Pfleiderer G (1975) Hoppe Seylers Z Physiol Chem 356:1613–1623
24. Hilt W, Pfleiderer G, Fortnagel P (1991) Biochim Biophys Acta 1076:298–304
25. Drewke C, Ciriacy M (1988) Biochim Biophys Acta 950:54–60
26. Adolph HW, Maurer P, Schneiderbernlohr H, Sartorius C, Zeppezauer M (1991) Eur J Biochem 201:615–625
27. van Iersel M, Eppink M, van Berkel W, Rombouts F, Abee T (1997) Appl Environ Microbiol 63:4079–4082
28. Keinan E, Hafeli EK, Seth KK, Lamed R (1986) J Am Chem Soc 108:162–169
29. Korkhin Y, Kalb Gilboa AJ, Peretz M, Bogin O, Burstein Y, Frolow F (1998) J Mol Biol 278:967–981
30. Hummel W (1997) Adv Biochem Eng Biotechnol 58:145–184
31. Niefind K, Muller J, Riebel B, Hummel W, Schomburg D (2003) J Mol Biol 327:317–328
32. Niefind K, Riebel B, Mueller J, Hummel W, Schomburg D (2000) Acta Crystallogr D Biol Crystallogr 56:1696–1698
33. Hummel W (1990) Biotechnol Lett 12:403–408
34. Bradshaw CW, Hummel W, Wong C-H (1992) J Org Chem 57:1532–1536
35. Hummel W (1990) Appl Microbiol Biotechnol 34:15–19
36. Danielsson B, Winqvist F, Malpote JY, Mosbach K (1982) Biotechnol Lett 4:673–678
37. Andersson M, Holmberg H, Adlercreutz P (1998) Biotechnol Bioeng 57:79–86
38. Hasumi F, Nakamura S, Nakada N (1991) Mem Numazu Coll Technol 25:63–65
39. Pfitzner J, Linke HAB, Schlegel HG (1970) Arch Mikrobiol 71:67–70
40. Bryant F, Adams M (1989) J Biol Chem 264:5070–5079

41. Shaked Z, Whitesides GM (1980) J Am Chem Soc 102:7104–7105
42. Kruse W, Hummel W, Kragl U (1996) Recl Trav Chim Pays Bas 115:239–243
43. Wong C-H, Whitesides GM (1981) J Am Chem Soc 103:4890–4899
44. Lampel KA, Uratani B, Chaudhry GR, Ramaley RF, Rudikoff S (1986) J Bacteriol 166: 238–243
45. Yun H, Yang YH, Cho BK, Hwang BY, Kim BG (2003) Biotechnol Lett 25:809–814
46. Lin S-S, Miyawaki O, Nakamura K (1999) J Biosci Bioeng 87:361–364
47. Wada M, Yoshizumi A, Noda Y, Kataoka M, Shimizu S, Takagi H, Nakamori S (2003) Appl Environ Microbiol 69:933–937
48. Eguchi T, Kuge Y, Inoue K, Yoshikawa N, Mochida K, Uwajima T (1992) Biosci Biotechnol Biochem 56:701–703
49. Benach J, Atrian S, Gonzalez-Duarte R, Ladenstein R (1999) J Mol Biol 289:335–355
50. Smilda T, Kamminga AH, Reinders P, Baron W, van Hylckama Vlieg JE, Beintema JJ (2001) J Mol Evol 52:457–466
51. Winberg JO, McKinley-McKee JS (1998) Biochem J 329:561–570
52. Abokitse K, Hummel W (2003) Appl Microbiol Biotechnol 62:380–386
53. Gröger H, Hummel W, Buchholz S, Drauz K, Van Nguyen T, Rollmann C, Husken H, Abokitse K (2003) Org Lett 5:173–176
54. Hummel W, Abokitse K, Drauz K, Rollmann C, Gröger H (2003) Adv Synth Catal 345:A34
55. Kelemen-Horvath I, Nemestothy N, Belafi-Bako K, Gubicza L (2002) Chem Pap Chem Zvesti 56:52–56
56. Hirakawa H, Kamiya N, Yata T, Nagamune T (2003) Biochem Eng J 16:35–40
57. Zambianchi F, Pasta P, Carrea G, Colonna S, Gaggero N, Woodley JM (2002) Biotechnol Bioeng 78:489–496
58. Schneider K, Schlegel HG (1976) Biochim Biophys Acta 452:66–80
59. Mertens R, Greiner L, van den Ban ECD, Haaker HBCM, Liese A (2003) J Mol Catal B Enzym 24–25:39–52
60. Vrtis JM, White AK, Metcalf WW, van der Donk WA (2001) J Am Chem Soc 123:2672–2673
61. Vrtis JM, White AK, Metcalf WW, van der Donk WA (2002) Angew Chem Int Ed 41:3257–3259
62. Woodyer R, van der Donk WA, Zhao HM (2003) Biochemistry 42:11604–11614
63. Johannes TW, Woodyer RD, Zhao HM (2005) Appl Environ Microbiol 71:5728–5734
64. Johannes TW, Woodyer RD, Zhao HM (2007) Biotechnol Bioeng 96:18–26
65. Gu KF, Chang TM (1990) Appl Biochem Biotechnol 26:115–124
66. Veronese FM, Nyc JF, Degani Y, Brown DM, Smith EL (1974) J Biol Chem 249:7922–7928
67. Sakamoto N, Kotre AM, Savageau MA (1975) J Bacteriol 124:775–783
68. Veronese FM, Boccu E, Conventi L (1975) Biochim Biophys Acta 377:217–228
69. Graham LD, Griffin TO, Beatty RE, McCarthy AD, Tipton KF (1985) Biochim Biophys Acta 828:266–269
70. Peterson PE, Pierce J, Smith TJ (1997) J Struct Biol 120:73–77
71. Garvie EI (1980) Microbiol Rev 44:106–139
72. Hensel R, Mayr U, Fujiki H, Kandler O (1977) Eur J Biochem 80:83–92
73. Gordon GL, Doelle HW (1976) Eur J Biochem 67:543–555
74. Hummel W, Riebel B (2003) Biotechnol Lett 25:51–54
75. Riebel BR, Gibbs PR, Wellborn WB, Bommarius AS (2002) Adv Synth Catal 344:1156–1168
76. Park HJ, Kreutzer R, Reiser COA, Sprinzl M (1992) Eur J Biochem 205:875–879
77. Park HJ, Reiser COA, Kondruweit S, Erdmann H, Schmid RD, Sprinzl M (1992) Eur J Biochem 205:887–893
78. Lee LG, Whitesides GM (1986) J Org Chem 51:25–36
79. Carrea G, Bovara R, Cremonesi P, Lodi R (1984) Biotechnol Bioeng 26:560–563
80. Kim M-J, Whitesides GM (1988) J Am Chem Soc 110:2959–2964

81. Chenault HK, Whitesides GM (1989) Bioorg Chem 17:400–409
82. Lemiere GL, Lepoivre JA, Alderweireldt FC (1985) Tetrahedron Lett 26:4527–4528
83. Higuchi M, Yamamoto Y, Poole LB, Shimada M, Sato Y, Takahashi N, Kamio Y (1999) J Bacteriol 181:5940–5947
84. Poole LB, Higuchi M, Shimada M, Li Calzi M, Kamio Y (2000) Free Radic Biol Med 28:108–120
85. Matsumoto J, Higuchi M, Shimada M, Yamamoto Y, Kamio Y (1996) Biosci Biotechnol Biochem 60:39–43
86. Ross RP, Claiborne A (1992) J Mol Biol 227:658–671
87. Kengen SWM, van der Oost J, de Vos WM (2003) Eur J Biochem 270:2885–2894
88. Geueke B, Riebel B, Hummel W (2003) Enzyme Microb Technol 32:205–211
89. Hummel W, Kuzu M, Geueke B (2003) Org Lett 5:3649–3650
90. Steckhan E (1994) Electrochem V 170:83–111
91. Komoschinski J, Steckhan E (1988) Tetrahedron Lett 29:3299–3300
92. Hilt G, Jarbawi T, Heineman WR, Steckhan E (1997) Chemistry 3:79–88
93. Hilt G, Steckhan E (1993) J Chem Soc Chem Commun:1706–1707
94. Steckhan E, Brielbeck B, Frede M, Hilt G (1996) Power Electrochem:230–253
95. Delecouls-Servat K, Bergel A, Basseguy R (2001) J Appl Electrochem 31:1095–1101
96. Schröder I, Steckhan E, Liese A (2003) J Electroanal Chem 541:109–115
97. DiCosimo R, Wong C-H, Daniels L, Whitesides GM (1981) J Org Chem 46:4622–4623
98. Ito M, Kuwana T (1971) J Electroanal Chem 32:415
99. Simon H (1992) Pure Appl Chem 64:1181–1186
100. Simon H, Bader J, Gunther H, Neumann S, Thanos J (1985) Angew Chem Int Ed 24:539–553
101. Thanos ICG, Simon H (1987) J Biotechnol 6:13–29
102. Günther H, Paxinos AS, Schulz M, Vandijk C, Simon H (1990) Angew Chem Int Ed 29:1053–1055
103. Nagata S, Gunther H, Bader J, Simon H (1987) FEBS Lett 210:66–70
104. Günther H, Walter K, Kohler P, Simon H (2000) J Biotechnol 83:253–267
105. Schulz M, Leichmann H, Gunther H, Simon H (1995) Appl Microbiol Biotechnol 42:916–922
106. Shaked Z, Barber JJ, Whitesides GM (1981) J Org Chem 46:4100–4101
107. Kano K, Takagi K, Ogino Y, Ikeda T (1995) Chem Lett:589–590
108. Yuan R, Watanabe S, Kuwabata S, Yoneyama H (1997) J Org Chem 62:2494–2499
109. Franke M, Steckhan E (1988) Angew Chem Int Ed 27:265–267
110. Wienkamp R, Steckhan E (1982) Angew Chem Int Ed 21:782–783
111. Steckhan E, Herrmann S, Ruppert R, Dietz E, Frede M, Spika E (1991) Organometallics 10:1568–1577
112. Ruppert R, Herrmann S, Steckhan E (1987) Tetrahedron Lett 28:6583–6586
113. Westerhausen D, Herrmann S, Hummel W, Steckhan E (1992) Angew Chem Int Ed 31:1592–1531
114. Shimizu Y, Kitani A, Ito S, Sasaki K (1993) Denki Kagaku 61:872–873
115. Voivodov KI, Sobolov SB, Leonida MD, Fry AJ (1995) Bioorg Med Chem Lett 5:681–686
116. Leonida MD, Fry AJ, Sobolov SB, Voivodov KI (1996) Bioorg Med Chem Lett 6:1663–1666
117. Leonida MD, Sobolov SB, Fry AJ (1998) Bioorg Med Chem Lett 8:2819–2824
118. Long Y-T, Chen H-Y (1997) J Electroanal Chem 440:239–242
119. Baik SH, Kang C, Jeon IC, Yun S-E (1999) Biotechnol Tech 13:1–5
120. Hollmann F, Schmid A, Steckhan E (2001) Angew Chem Int Ed 40:169–171
121. Hollmann F, Witholt B, Schmid A (2003) J Mol Catal B Enzym 19:167–176
122. Cantet J, Bergel A, Comtat M (1996) Enzyme Microb Technol 18:72–79
123. Ruppert R, Steckhan E (1989) J Chem Soc Perkin Trans 2:811–814
124. Ruppert R, Herrmann S, Steckhan E (1988) J Chem Soc Chem Commun:1150–1151
125. Wagenknecht PS, Penney JM, Hembre RT (2003) Organometallics 22:1180–1182

126. Wilms B, Wiese A, Syldatk C, Mattes R, Altenbuchner J (2001) J Biotechnol 86:19–30
127. Cheesman MJ, Kneller MB, Kelly EJ, Thompson SJ, Yeung CK, Eaton DL, Rettie AE (2001) Protein Expr Purif 21:81–86
128. Stewart JD (1998) Curr Org Chem 2:195–216
129. Stewart JD, Reed KW, Kayser MM (1996) J Chem Soc Perkin Trans 1:755–757
130. Chen YC, Peoples OP, Walsh CT (1988) J Bacteriol 170:781–789
131. Doig SD, O'Sullivan LM, Patel S, Ward JM, Woodley JM (2001) Enzyme Microb Technol 28:265–274
132. Iwaki H, Hasegawa Y, Teraoka M, Tokuyama T, Bergeron H, Lau PCK (1999) Appl Environ Microbiol 65:5158–5162
133. Gutierrez MC, Furstoss R, Alphand V (2005) Adv Synth Catal 347:1051–1059
134. Hilker I, Alphand V, Wohlgemuth R, Furstoss R (2004) Adv Synth Catal 346:203–214
135. Hilker I, Gutierrez MC, Furstoss R, Ward J, Wohlgemuth R, Alphand V (2008) Nat Protoc 3:546–554
136. Law HEM, Baldwin CVF, Chen BH, Woodley JM (2006) Chem Eng Sci 61:6646–6652
137. Hilker I, Baldwin C, Alphand W, Furstoss R, Woodley J, Wohlgemuth R (2006) Biotechnol Bioeng 93:1138–1144
138. Carrea G, Riva S (2000) Angew Chem Int Ed 39:2226–2254
139. Grunwald J, Wirz B, Scollar MP, Klibanov AM (1986) J Am Chem Soc 108:6732–6734
140. Nakamura K, Kondo S, Kawai Y, Ohno A (1993) Bull Chem Soc Jpn 66:2738–2743
141. Manosroi J, Sripalakit P, Manosroi A (1998) J Chem Technol Biotechnol 73:203–210
142. Jonsson A, van Breukelen W, Wehtje E, Adlercreutz P, Mattiasson B (1998) J Mol Catal B Enzym 5:273–276
143. Lutz J, Mozhaev VV, Khmelnitsky YL, Witholt B, Schmid A (2002) J Mol Catal B Enzym 19–20:177–187
144. Luo DH, Zong MH, Xu JH (2003) J Mol Catal B Enzym 24–25:83–88
145. Mittl PR, Perham RN, Schulz GE, Berry A, Scrutton NS (1995) Protein Sci 3:1504–1514
146. Rosell A, Valencia E, Ochoa WF, Fita I, Pares X, Farres J (2003) J Biol Chem 278:40573–40580
147. Bellamacina CR (1996) FASEB J 10:1257–1269
148. Bubner P, Klimacek M, Nidetzky B (2008) FEBS Lett 582:233–237
149. Leitgeb S, Petschacher B, Wilson DK, Nidetzky B (2005) FEBS Lett 579:763–767
150. Petschacher B, Leitgeb S, Kavanagh KL, Wilson DK, Nidetzky B (2005) Biochem J 385:75–83
151. Petschacher B, Nidetzky B (2005) Appl Environ Microbiol 71:6390–6393
152. Watanabe S, Makino K, Kodaki T (2005) J Biol Chem 280:10340–10349
153. Drauz K, Eils S, Schwarm M (2002) Chim Oggi Chem Today 20:15–21
154. Bommarius AS, Schwarm M, Stingl K, Kottenhahn M, Huthmacher K, Drauz K (1995) Tetrahedron Asymmetry 6:2851–2888
155. Hanson RL, Schwinden MD, Banerjee A, Brzozowski DB, Chen BC, Patel BP, McNamee CG, Kodersha GA, Kronenthal DR, Patel RN, Szarka LJ (1999) Bioorg Med Chem 7:2247–2252
156. Gröger H, Hummel W, Rollmann C, Chamouleau F, Huesken H, Werner H, Wunderlich C, Abokitse K, Drauz K, Buchholz S (2004) Tetrahedron 60:633–640
157. Gröger H, Chamouleau F, Orologas N, Rollmann C, Drauz K, Hummel W, Weckbecker A, May O (2006) Angew Chem Int Ed 45:5677–5681
158. Gröger H, Rollmann C, Chamouleau F, Sebastien I, May O, Wienand W, Drauz K (2007) Adv Synth Catal 349:709–712
159. Berkessel A, Rollmann C, Chamouleau F, Labs S, May O, Gröger H (2007) Adv Synth Catal 349:2697–2704
160. Richter N, Neumann M, Liese A, Wohlgemuth R, Eggert T, Hummel W (2009) Chembiochem 10:1888–1896

161. Kataoka M, Sakai H, Morikawa T, Katoh M, Miyoshi T, Shimizu S, Yamada H (1992) Biochim Biophys Acta 1122:57–62
162. Wada M, Kataoka M, Kawabata H, Yasohara Y, Kizaki N, Hasegawa J, Shimizu S (1998) Biotechnol Biochem 62:280–285
163. Kataoka M, Rohani LP, Yamamoto K, Wada M, Kawabata H, Kita K, Yanase H, Shimizu S (1997) Appl Microbiol Biotechnol 48:699–703
164. Kita K, Matsuzaki K, Hashimoto T, Yanase H, Kato N, Chung MC, Kataoka M, Shimizu S (1996) Appl Environ Microbiol 62:2303–2310
165. Shimizu S, Kataoka M, Katoh M, Morikawa T, Miyoshi T, Yamada H (1990) Appl Environ Microbiol 56:2374–2377
166. Kataoka M, Rohani LPS, Wada M, Kita K, Yanase H, Urabe I, Shimizu S (1998) Biosci Biotechnol Biochem 62:167–169
167. Kataoka M, Yamamoto K, Kawabata H, Wada M, Kita K, Yanase H, Shimizu S (1999) Appl Microbiol Biotechnol 51:486–490
168. Ogawa J, Shimizu S (2002) Curr Opin Biotechnol 13:367–375
169. Müller M, Wolberg M, Schubert T, Hummel W (2005) Adv Biochem Eng Biotechnol 92:261–287
170. Wolberg M, Villela M, Bode S, Geilenkirchen P, Feldmann R, Liese A, Hummel W, Müller M (2008) Bioprocess Biosyst Eng 31:183–191
171. Wolberg M, Hummel W, Wandrey C, Müller M (2000) Angew Chem 112:4476–4478
172. Wolberg M, Ji AG, Hummel W, Müller M (2001) Synthesis 2001:937–942
173. Krix G, Bommarius AS, Drauz K, Kottenhahn M, Schwarm M, Kula M-R (1997) J Biotechnol 53:29–39
174. Kallwass HK (1992) Enzyme Microb Technol 14:28–35
175. Ema T, Okita N, Ide S, Sakai T (2007) Org Biomol Chem 5:1175–1176
176. Ema T, Sugiyama Y, Fukumoto M, Moriya H, Cui JN, Sakai T, Utaka M (1998) J Org Chem 63:4996–5000
177. Ema T, Yagasaki H, Okita N, Nishikawa K, Korenaga T, Sakai T (2005) Tetrahedron Asymmetry 16:1075–1078
178. Ema T, Ide S, Okita N, Sakai T (2008) Adv Synth Catal 350:2039–2044
179. Krausser M, Hummel W, Gröger H (2007) Eur J Org Chem:5175–5179
180. Burda E, Hummel W, Gröger H (2008) Angew Chem Int Ed 47:9551–9554

Index

A
Acinetobacter calcoaceticus, 137
Acrylamide, 79
Actinol, 202
Adriamycin, 215
Affinity chromatographic purification, 147
Affinity tagged recombinant proteins, 137
Alcohol dehydrogenase, 201, 203, 208
Alcohols, chiral, 231
Allysine ethylene acetal, 230
Amino acid dehydrogenases, 234
Amino acids, aromatic, 40
  optically active, 234
7-Aminodeacetoxycephalosporanic acid (7-ADCA), 54
Aminotransferases, 3
Amylases, 79, 135
Amyloglucosidase (AMG), 79
Anaplerosis, 21
Asparaginase, 80
Aspergillus, 57
  cell factories, 76
  chemicals production, 78
  examples, 87
Aspergillus niger, 60, 77
Aspergillus oryzae, 60, 77
Aspergillus terreus, 60
Assimilatory pathways, 23

B
Bacillus megaterium, 133
  fluxome, 155
  genome sequences, 152
  metabolic flux analysis, 153
  recombinant proteins, 136
  systems biology, 158
Bacillus stearothermophilus, 123
Bacillus subtilis, 60
Baker's yeast, 68
Benzyl alcohol, 201
Biocatalysis, 163
Bioethanol, 84
Bioprocess engineering, 1, 2
Biosynthetic pathways, 25
Biotechnology products, 25
Brevibacterium lactofermentum, 9
Brewer's yeast, 68

C
Cadaverine, 44
Calmodulin, 61
Candida antarctica (CALB), 82
Candida utilis, 74
Carbon metabolism, 23
Cardiolipin, 152
Carnitine, 202
β-Carotene desaturase/methyltransferase, 3
Carotenoids, 73
CE-ESI-MS, 62
Cell factories, 57
Cellulases, 77
Chiral compounds, 195
Cichorum intybus, 175
Citric acid, 79

Clostridium difficile, 137
Clostridium thermoaceticum, AMAPOR, 214
Coenzyme-dependent reactions, organic solvents, 223
Coenzymes, regeneration, 195
Cofactor regeneration, chemical, 220
photoelectrochemical, 220
Coimmobilization, 180
Compactin, 81
Corynebacterium glutamicum, metabolic fluxes, 34
metabolism, 23
Cryoresistance, 71
Culture heterogeneity, intracellular protein production, 151
Cyclohexanol, 212

D
Dehydrogenases, 195
Designer cells, 124, 222
Dextransucrase, 143, 166
acceptor reactions, 166
Diglycosyl-1,2-diacylglycerol, 152
DNA microarrays, 4, 60

E
Electrochemical methods, 210
Electrofusion, 69
Enzymatic reduction, 195
Enzyme engineering, 163
Ergosterol, 75
Error-prone PCR, 121
Erwinia uredovora, 74
Escherichia coli, 1
Ethanol, 44, 70

F
Farnesyldiphosphate (FPP), 74
synthase, 123
Filamentous fungi, cell factories, 57
Flux balance analysis (FBA), 64
Fluxomics, 63
Formate dehydrogenase, 199
Fructansucrases, 174
Fructosyltransferases (FTF), 163, 177
Fungi, cell factories, 76
filamentous, 57

G
Gene circuits, 101
engineering, synthetic DNA, 117
Gene expression, molecular engineering, 107
Genome sequencing, 27
Genome-scale metabolic models (GSMMs), 21, 67
Genomics, 58, 82
Geodin, 88
GFP, *Bacillus megaterium*, 141
Global transcription machinery engineering (gTME), 5
Glucanases, 143
Glucansucrases, 166
Gluconobacter diazotrophicus, 175
Glucose dehydrogenase, 202
Glucose-6-phosphate dehydrogenase, 201
Glucosyltransferases (GTF), 163, 172
Glucuronidases (GUS), 90
Glutamate, 21, 22
metabolic engineering, 41
Glutamate dehydrogenase, 207
L-Glyceraldehyde, 231
Glycerol, 84
Glycosyltransferase R (GTFR), 172, 183
Green fluorescent protein (GFP), 139, 141

H
Halohydrins, 231
Homing endonucleases (HEs), 123
Homolactic bacteria, 72
Homologous recombination (HR), 68
Human growth hormone (hGH), 52
Human urokinase-like plasminogen activator, 137
Humelene synthase, 122
Hydrocortisone, 75
Hydrogenase, 204
Hydroxy acids, optically active, 234
Hydroxy esters, optically active, 234
Hydroxyisocaproate, 122
L-6-Hydroxynorleucine, 228
β-Hydroxynorvaline, 9

I
Immobilization, 163
Industrial biotechnology, 51, 54
  systems biology, 52
Industrial systems biology, 54
Insulin, 52
Integrated reaction engineering, 185
Internal ribosome entry sequences (IRESs), 108
Inulin, 174
Inulosucrases, 174
Invertase, 175
Isocitrate dehydrogenase (ICDH), 34
Isocitrate lyase (ICL), 13
Isomaltooligosaccharides, 167

K
Kestoses, 179
α-Ketoglutarate dehydrogenase complex (ODHC), 35
Ketoisophorone, 202
Kluyveromyces lactis, 73
Koji processes, 79

L
Lactamase, 146
D-Lactate, 201
L-Lactate dehydrogenases (LDHs), 122, 205, 208
L-Lactic acid, 71
Lactobacillus spp., 72
Lactoferrin, 82
Lactonitrile, 71
Leucine dehydrogenase (LeuDH), 206
Leuconostoc mesenteroides, 143, 201
L-Leucovorin, 203
Levansucrases, 146, 174
Levodione reductase, 202
Lipase, 82
Lipoamide dehydrogenase, 217
Lovastatin, 6, 88
Lycopene, 7, 55, 73, 103
L-Lysine, 6, 21
  metabolic engineering, 38

M
Mannheimia succiniciproducens, 3, 5
Melibiose, 70
Metabolic control, 101
Metabolic engineering, 21, 51, 53
  targeted, 1
Metabolic fluxes, 32
  analysis, 21, 133
Metabolic footprinting, 62
Metabolic modeling, 64
Metabolic network, analysis, in silico, 28
  genome-scale reconstruction, 28
  reconstruction, 27
Metabolic pathway engineering, 38
Metabolomics, 62
L-Methionine, 3, 9, 30
MIAPE-MS, 61
Modeling, dynamic, 32
Molasses, 70
Molecular breeding, 4
Multiple vector systems, 148
Mutarostase, 137

N
NADH dehydrogenases, 152
NADH oxidase, 209
NADPH, 21
  metabolism, 37
  supply, 40
Neokestose, 176
Nicotinamide coenzymes, 195, 199
Nicotinamide dinucleotide binding motif, 227
p-Nitrophenol butyrate, 82
Non-homologous end joining (NHEJ), 77
Novel products, 43
Nutritional requirements, 23

O
Oligosaccharides, 163
  synthesis, 101, 104
    enzymatic, 166
Omics, 2, 58
One-pot processes, 235
Oxidations, electroenzymatic, 211

P
Pantothenate, 7
Penicillin acylases, 135

Penicillium chrysogenum, 57
Phaffia rhodozyma, 74
Phage shock protein A, 5
4-Phenyl-2-hydroxybutanoate, 122
Phenyllactate, 122
Phenylpropanol, 200
Phosphaditylglycerol, 152
Phosphatidic acid, 152
Phosphatidylethanolamine, 152
Phosphite dehydrogenase, 205
Phosphodiester chemistry, 106
Phosphoenolpyruvate
  carboxylase (PPC), 13
Phospholipid interacting proteins, 61
Phosphotriester synthesis, 106
Pichia pastoris, 68
Poly(hydroxyalkanoates), 103
Poly(hydroxybutyrate), 103
Polymerase chain reaction (PCR), 106
Precursor supply, 39
Progesterone, 75
Promoter, xylose-inducible, 136
Promoter system, T7 RNA
  polymerase-dependent, 149
Propanediol, 54, 103
Proteins, affinity chromatographic
  purification, 139
  affinity tagged recombinant, 137
  export into growth medium, 143
  production, 133
    heterologous, 81
  secretion, 81, 133
Proteomics, 61
Protoplast mediated
  transformation (PMT), 78
Pyruvate, 157

R

Random mutation, 1
Reaction engineering, 163
Recombinant enzymes, 222
Recombinant proteins, *Bacillus*
  megaterium 136
Reductions, electroenzymatic, 213
Regulatory engineering, 101
Renierapurpurin, 3
Rhizopus oryzae, 77
Rhodococcus erythropolis, 200
Riboflavin, 54
RNA silencing, 108, 114

S

Saccharomyces cerevisiae, 57, 68, 83
  cell factory, 67
  chemical production, 71
  lactic acid, 72
L-Serine, 41
Serratia marcescens, 9
Shine-Dalgrano (SD) sequences, 109
Statins, alcohol dehydrogenase
  catalyzed synthesis, 232
Stoichiometric modeling, 29
Strain engineering, systems biology, 42
Strain improvements, 68, 77
Streptococcus mutants, 86
Streptococcus oralis, 172
Streptomyces coelicolor, 60
Substrates, 41
Succinic acid, 6
Sucrose analogs, 176
Synechoxanthin, 3
Synthetic biology, 101
Synthetic libraries, 121
Synthetic organisms, 124
Systems biology, 1, 21, 51, 53, 152
  toolbox, 58
Systems metabolic engineering, 1, 21

T

T7 RNA polymerase-dependent
  promoter system, 149
Taka-diastase, 79
L-*tert*-Leucine, 228
Tetracycline resistance, 137
Tetrahydrofolate, 203
Thermoactinomyces intermedius, 230
Thermotoga maritima, 175
L-Threonine, 1, 7
  systems biology/bioprocess
    development, 12
Toxin A, 137
Transcription, modulating, 115
Transcriptomics, 60
Translation modulators, 108
Translational control, re-engineering, 108

Trehalose, 71
Triacylglycerol lipase, 82
Trichoderma reesei, 77
Tryptophan, 21
Tunable intergenic regions (TIGRs), 108
Two enzyme system, 180

U
Upstream open reading frames (uORFs), 108

U
Vector systems, 148

W
Whole cell applications, 222

X
Xanthophyllomyces dendrorhous, 74
Xylanases, 77
Xylitol, 4, 206
Xylose reductase, 205
Xylose-inducible promoter system, 136

Y
Yeasts, 83
  cell factories, 57, 67