Advances in Biochemical Engineering/Biotechnology 162
Series Editor: T. Scheper

Huimin Zhao
An-Ping Zeng *Editors*

Synthetic Biology – Metabolic Engineering

Springer

162
Advances in Biochemical Engineering/Biotechnology

Aims and Scope

This book series reviews current trends in modern biotechnology and biochemical engineering. Its aim is to cover all aspects of these interdisciplinary disciplines, where knowledge, methods and expertise are required from chemistry, biochemistry, microbiology, molecular biology, chemical engineering and computer science.

Volumes are organized topically and provide a comprehensive discussion of developments in the field over the past 3–5 years. The series also discusses new discoveries and applications. Special volumes are dedicated to selected topics which focus on new biotechnological products and new processes for their synthesis and purification.

In general, 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 cited as a journal.

More information about this series at http://www.springer.com/series/10

Huimin Zhao • An-Ping Zeng
Editors

Synthetic Biology – Metabolic Engineering

With contributions by

H.S. Alper · T. Baumann · J. Becker · N. Budisa · G.-Q. Chen · M. Deaner · M. Exner · X. Feng · E. Garcia-Ruiz · D. Gerngross · G. Gießelmann · W. Guo · M. HamediRad · S.L. Hoffmann · Y.-S. Jin · H. Kim · I.I. Kong · J.-J. Liu · C.-W. Ma · D.-C. Meng · G. Morgado · S. Panke · L. Pei · T.M. Roberts · M. Schmidt · J. Sheng · T.L. Turner · C. Wittmann · A.-P. Zeng · G.-C. Zhang · H. Zhao · L.-B. Zhou

Springer

Editors
Huimin Zhao
Department of Chemical
and Biomolecular Engineering
University of Illinois
Urbana, Illinois
USA

An-Ping Zeng
Technische Universität Hamburg-Harburg
Institut für Bioprozess- und Biosystemtechnik
Hamburg, Germany

ISSN 0724-6145 ISSN 1616-8542 (electronic)
Advances in Biochemical Engineering/Biotechnology
ISBN 978-3-319-55317-7 ISBN 978-3-319-55318-4 (eBook)
DOI 10.1007/978-3-319-55318-4

Library of Congress Control Number: 2017952559

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface: Exploring the Synergy Between Synthetic Biology and Metabolic Engineering

Synthetic biology involves the use of engineering principles to design biological parts and systems with new or improved properties, whereas metabolic engineering focuses on the engineering of microbial cell factories for the production of fuels and chemicals using recombinant DNA technologies (Zhao 2013). Both fields have been growing quickly in recent years. In particular, synthetic biology tools have increasingly been used to address scientific and technical challenges in metabolic engineering. In fact, a recent report by the National Research Council of the National Academies of Sciences, Engineering, and Medicine in the United States describes a roadmap for accelerating the development of industrial processes for production of chemicals using synthetic biology tools ("Industrialization of Biology," 2015).

This volume of *Advances in Biochemical Engineering/Biotechnology* explores the synergy between synthetic biology and metabolic engineering. It contains a total of ten reviews written by world-leading experts; roughly half of these review focus on tool development mainly in the synthetic biology area, and the other half focus on the application of synthetic biology and metabolic engineering tools for the design, engineering, and evolution of microbial cells for production of a wide variety of chemicals, materials, and fuels. In the Tool Development section, Budisa and coworkers summarize the development and application of the pyrrolysine-based system for orthogonal protein translation, a process which produces proteins containing noncanonical amino acids at specific sites. Alper and Deaner report the various strategies for the discovery and engineering of promoters and terminators with desired characteristics for controlling gene expression. Special attention is paid to the rational design of synthetic promoters and terminators. Zeng and coworkers describe the recent advances in the development of biomolecular switches or in vivo biosensors and their applications for dynamic regulation of metabolic pathways. Zhao and coworkers summarize various strategies recently developed for the design, engineering, and optimization of biochemical pathways for the microbial production of chemicals. Both computational algorithms used to design efficient metabolic routes and experimental tools to construct and improve

the efficiency of the designed pathways are discussed. Complementing the review by Zhao and coworkers, Panke and coworkers describe strategies to design novel biochemical pathways for in vitro applications such as the multi-step enzymatic synthesis of chemicals.

In the Practical Application section, Chen and Meng discuss the application of synthetic biology tools for the metabolic engineering of bacteria to produce cost-effectively polyhydroxyalkanoates, a family of biodegradable and biocompatible polyesters. Jin and coworkers provide an overview of recent advances in the engineering and evolution of *Saccharomyces cerevisiae* for the production of biofuels and chemicals. In a related review, Whittmann and coworkers highlight the application of systems biology and synthetic biology in the engineering of *Corynebacterium glutamicum* for industrial production of chemicals. Feng and coworkers report the application of ^{13}C metabolic flux analysis to identify and tackle the rate-limiting steps in metabolic pathways to improve the production of chemicals and fuels.

Synthetic biology, especially a sub-field of synthetic biology, xenobiology, which aims at changing the chemical compositions of living cells (i.e., by creating an artificial genetic code and incorporating non-conical amino acids into biosynthesis), presents many exciting potential applications and scientific challenges. At the same time, it also raises some ethical and societal issues. In the last chapter of this volume, Schmidt and coworkers review and discuss the state-of-the-art and relevant ethics and philosophical aspects of xenobiology and new-to-nature organisms.

In summary, these ten reviews have highlighted some recently developed synthetic biology and metabolic engineering tools and their broad applications in industrial biotechnology and the future development of biology. We thank the authors for their contributions to this volume of *Advances in Biochemical Engineering/Biotechnology* and hope that the readers will enjoy their work as much as we have.

Urbana, USA — Huimin Zhao
Hamburg, Germany — An-Ping Zeng

References

Industrialization of biology: a roadmap to accelerate the advanced manufacturing of chemicals (2015) National Academies Press, Washington DC

Zhao H (2013) Synthetic biology: tools and applications. Academic Press-Elsevier, Amsterdam

Contents

Adv Biochem Eng Biotechnol (2018) 162: 1–20
DOI: 10.1007/10_2016_37

Published online: 26 October 2016

Orthogonal Protein Translation Using Pyrrolysyl-tRNA Synthetases for Single- and Multiple-Noncanonical Amino Acid Mutagenesis

Tobias Baumann, Matthias Exner, and Nediljko Budisa

Abstract To date, the two systems most extensively used for noncanonical amino acid (ncAA) incorporation via orthogonal translation are based on the *Methanococcus jannaschii* TyrRS/$tRNA^{Tyr}_{CUA}$ and the *Methanosarcina barkeri*/*Methanosarcina mazei* PylRS/$tRNA^{Pyl}_{CUA}$ pairs. Here, we summarize the development and usage of the pyrrolysine-based system for orthogonal translation, a process that allows for the recombinant production of site-specifically labeled proteins and peptides. Via stop codon suppression in *Escherichia coli* and mammalian cells, genetically encoded biomolecules can be equipped with a great diversity of chemical functionalities including click chemistry handles, post-translational modifications, and photocaged sidechains.

Keywords Expanded genetic code, Noncanonical amino acid, Orthogonal translation, Pyrrolysyl-tRNA synthetase, Stop codon suppression

Contents

T. Baumann, M. Exner, and N. Budisa (✉)
Department of Chemistry, Berlin Institute of Technology/TU Berlin, Biocatalysis Group, Müller-Breslau-Straße 10 / Sekr. L 1, D-10623 Berlin, Germany
e-mail: budisa@biocat.tu-berlin.de; http://www.biocat.tu-berlin.de

1 Introduction

Amber suppression, a widespread genetic phenomenon in bacterial species [1], can be used to reprogram coding sequences toward the incorporation of noncanonical amino acids (ncAAs) [2]. By this methodology, the gene of the target protein or peptide is mutated to a TAG amber stop codon at the desired site of ncAA incorporation. In parallel with the target, genes of an orthogonal pair, a combination of a suppressor tRNA and a suitable aminoacyl tRNA synthetase (aaRS), are expressed. Aminoacylation, namely charging the orthogonal tRNA with the ncAA, is commonly achieved by a wildtype or engineered aaRS which initially activates the ncAA via ATP. When aminoacylated tRNAs bearing a CUA anticodon are present, the translational machinery transfers the ncAAs to the growing polypeptide chain, resulting in site-specific incorporation. Being one of the three stop codons, amber sites within coding sequences naturally act as translational stop signs which trigger translation termination – a multistep process mediated by release factors [3]. One approach to improve amber suppression efficiency is thus to knock out the essential release factor 1 (RF1), either by complementation via a mutated RF2 [4] or by removal of essential amber stop codons from the *Escherichia coli* genome [5, 6].

2 Discovery and Phylogenic Distributions of PylRS as a Natural Orthogonal Pair

With natural stop (nonsense) codon suppression identified in bacteria, eukaryotes, and viruses [1], amber suppression by species of *Methanosarcina* was investigated intensively [12]. It was a study in 2002 focusing on the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB) which led to the discovery of the 22nd proteinogenic amino acid pyrrolysine, which is now commonly abbreviated as Pyl [13]. In an accompanying manuscript of the same journal, the corresponding genes encoding the tRNA and aminoacyl-tRNA synthetases, *pyl*T and *pyl*S, respectively, were described [14]. With few exceptions, natural pyrrolysine-containing proteins have so far been primarily identified in methyltransferase enzymes as part of the methanogenesis pathway in these species [15]. As part of the biocatalyst

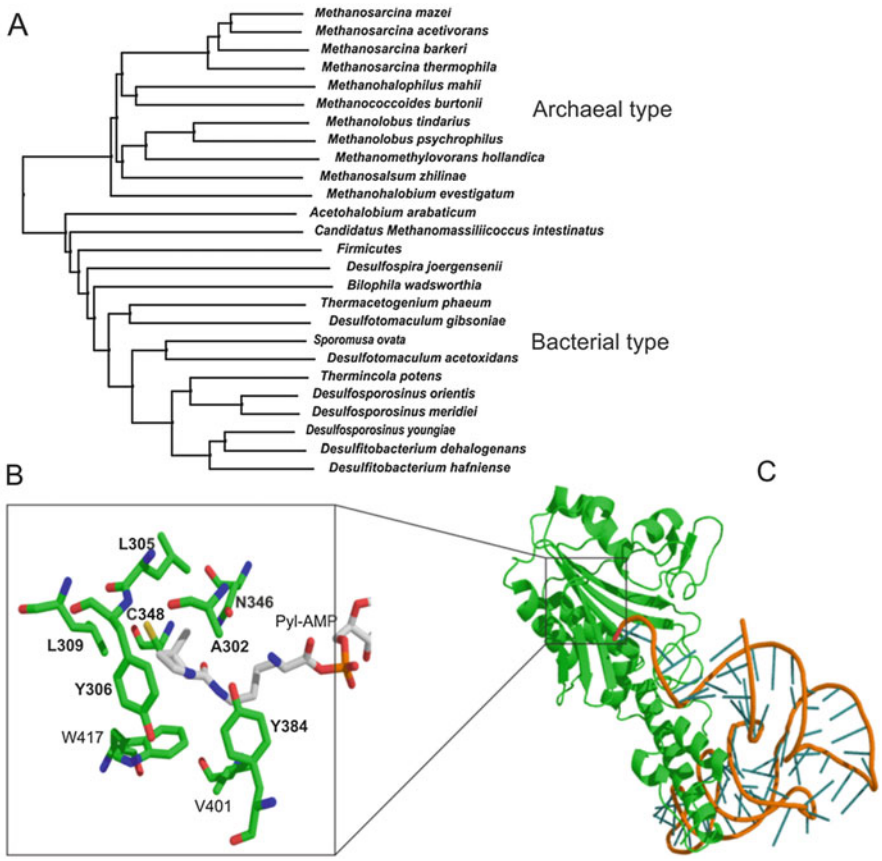

Fig. 1 Phylogenic distribution and structure of pyrrolysyl-tRNA synthetase. (**a**) Phylogenic tree of PylRS distribution reveals a separation of archaeal and bacterial forms. Sequences were retrieved using protein BLAST [7] with *Methanosarcina mazei* PylRS (accession number **Q8PWY1**) as query. Retrieved sequences (excluding duplicate entries from different strains and engineered variants, hypothetical proteins, bacterial N-terminal domains and non-pyrrolysyl-tRNA synthetases) were aligned with clustalW2 web server [8]. The phylogenetic tree was visualized using SeaView [9]. (**b**) Amino acid binding pocket of *Mm*PylRS with pyrrolysyl-AMP, taken from PDB ID 2ZIM [10]. (**c**): *Desulfitobacterium hafniense* PylRS bound to $tRNA^{Pyl}$, taken from PDB ID 2ZNI [11]

structure, the function of this lysine derivate is to enable the use of methylamines as energy sources for the host cell. In contrast to the archaeal counterpart, the bacterial enzymes discovered so far (see Fig. 1a for a phylogenetic tree based on protein sequence alignments) are encoded by two separate genes, with a structurally different aaRS as expression product [11]. To date, their natural function remains unclear [15].

3 Basic Features of the Natural PylRS:tRNA^{Pyl} System

Structurally, L-pyrrolysine presents a large lysine derivative with a methyl-pyrroline ring at the ε-amino group (see Fig. **2**). The pyrrolysine-tRNA synthetase (PylRS) belongs to the aaRS class II (subclass IIc) and bears the corresponding conserved fold of the catalytic domain. To bind and accommodate the pyrrolysine moiety followed by activation via ATP-hydrolysis, the PylRS biocatalyst structure bears an unusually large substrate binding pocket (see Fig. **1**b) [10, 11]. Key functional parameters of the synthetase were identified in 2004, where in vitro experiments showed that PylRS catalyzes the formation of pyrrolysyl-tRNA$^{Pyl}_{CUA}$ in an ATP-dependent manner [16]. Structure-function studies revealed several PylRS residues important for substrate recognition and discrimination against other metabolites. In *Methanosarcina mazei* PylRS, for example, Asn346 functions as a so-called gatekeeper residue and significantly restricts the substrate range toward defined sets of structure [17]. Accordingly, mutation of this residue can result in an altered substrate spectrum that allows the charging of tRNA$^{Pyl}_{CUA}$ with ncAAs which are rejected in the case of wild-type enzymes [18]. The target of aminoacylation, tRNA$^{Pyl}_{CUA}$, also exhibits several unique features. These shape it toward a more compact but still L-shaped and structurally similar molecule in comparison to conventional bacterial tRNAs. In conjunction with several recognition elements, discrimination against other tRNAs is achieved [15].

4 First Engineering Reports: Substrate Range and Design

Back in 1980, Kwok and Wong proposed that transferring a tRNA/aaRS pair from one organism to another could provide a route toward an expanded genetic code [19]. This concept was picked up in a study by Furter in 1998, where a yeast tRNA/phenylalanyl-tRNA synthetase pair was shown to work in *E. coli*. With efficiencies exceeding 60%, this strategy allowed the incorporation of *p*-fluoro-phenylalanine (as naturally occurring in yeast cells exposed to the ncAA) at amber stop codon sites [20]. It was in the pioneering work of Peter Schultz and coworkers where the door toward human-made orthogonal pairs was eventually opened. Using a tyrosyl-tRNA/synthetase pair of the archaeal hyperthermophilic organism *Methanococcus jannaschii* combined with mutations selected from an amber suppressor tRNA library, high-fidelity orthogonal translation was accomplished [2].

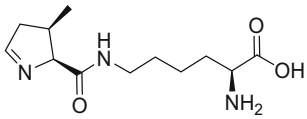

Fig. 2 Structure of pyrrolysine

As in the case of many follow-up studies based on the pyrrolysine system reviewed here, iterative rounds of negative and positive selection were used to isolate tRNA/synthetase pairs specific for the target ncAA. For negative selection, a toxic barnase protein gene harboring one or more amber stop codons is commonly used. Aminoacylation of the amber suppressor tRNA by endogenous synthetases or expression of an unspecific synthetase variant which charges canonical amino acids results in cell death. During positive selection for specific ncAA incorporation at amber sites, libraries of tRNA/synthetase combinations are screened for clones dependent on the presence of the ncAA in the growth medium. For this scenario, the chloramphenicol acetyltransferase gene (CAT) is frequently employed, whose functional full-length gene product confers resistance to the antibiotic chloramphenicol [2].

Besides pyrrolysine, wild-type PylRS enzymes (including the variant produced by the Gram-positive bacterium *Desulfitobacterium hafniense* depicted in Fig. **1**c) recognize several alternative substrate molecules which can be activated and loaded onto $tRNA^{Pyl}$ [21]. Combining random or rationally chosen mutations with stringent positive and negative selection systems, the substrate spectrum of the enzyme can be broadened or reshaped significantly toward a variety of ncAAs. These are grouped and summarized in the next sections.

5 Simple Chemical Handles and Hydroxy Amino Acids

Several studies concerning the functional characterization of PylRS revealed a surprisingly broad substrate tolerance, presumed to result from pyrrolysine recognition via hydrophobicity and, for example, not via its α-group. Figure **3** summarizes diverse chemical handles including hydroxy amino acids which can be incorporated into proteins via *Methanosarcina mazei* (*Mm*), *Methanosarcina barkeri* (*Mb*), and *Desulfitobacterium hafniense* (*Dh*) pyrrolysine tRNA/synthetase combinations. In addition to the natural substrate and with high efficiency, *N*-ε-*tert*-butoxycarbonyl-L-lysine (BocK) is transferred to $tRNA^{Pyl}$ both in vitro and in vivo [31]. Within the same study, the α-hydroxyacid *N*-*tert*-Boc-6-amino-2-hydroxy-L-hexanoic acid (**15**, Boc-LysOH) was also incorporated in a site-specific manner in *E. coli*. With this non-natural bond in the polypeptide chain, more alkali-liable compared to the natural amide bond, α-hydroxyacids allow hydrolysis under mild conditions. Using an *Mb*PylRS evolved for incorporation of the azide-bearing cyclic Pyl analogue *N*-ε-(((1*R*,2*R*)-2-azidocyclo-pentyloxy)carbonyl)-L-lysine (**5**, ACPK) enabled expressed protein ligation via hydrazinolysis. First, an oxoester was cotranslationally incorporated into the protein backbone via a noncanonical α-hydroxyacid. Second, addition of hydrazine led to site-selective cleavage in vitro, which allowed for the subsequent ligation of a chemically synthesized cysteine-bearing peptide. In vitro refolding of the non-natural protein fusion reconstituted a folded active protein [32].

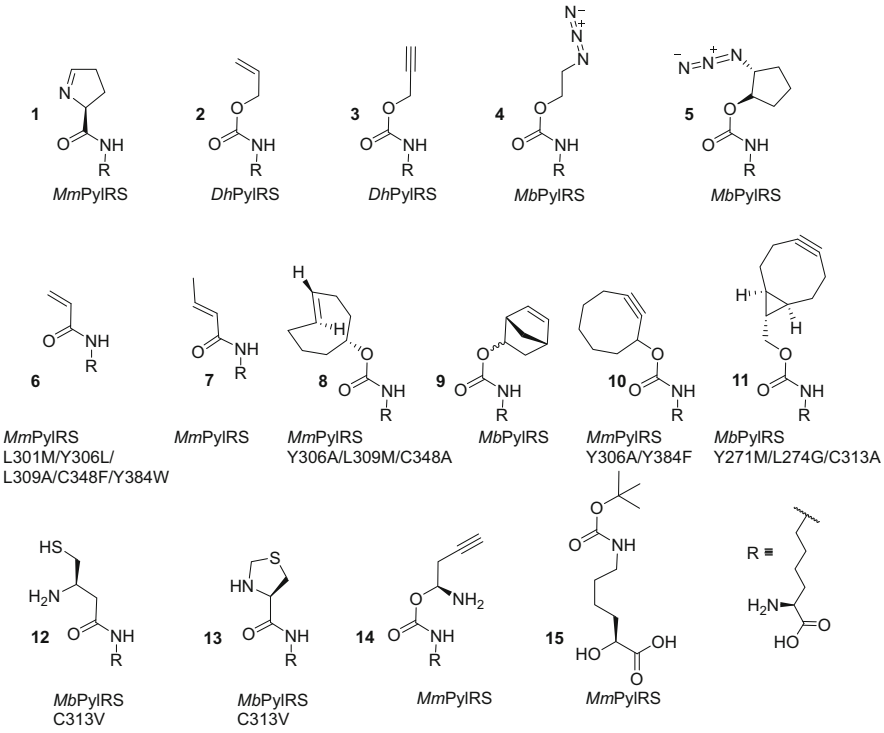

Fig. 3 Chemical handles. *Top row*: In vivo synthesized pyrroline-carboxylysine, simple alkenes, alkynes, and azido-amino acids for ligation chemistry. *Center row*: Amino acids with highly reactive double and triple bonds. *Bottom row*: Amino acids with multiple functional groups for complex ligation chemistry and backbone analogs. References: **1**: [22]; **2, 3**: [21]; **4, 9**: [23]; **5**: [24]; **6**: [25]; **7**: [26]; **8, 11**: [27]; **10**: [28]; **12, 13**: [29]; **14**: [30]; **15**: [31]

6 Post-translational Modifications

In earlier PylRS studies, the Lys moiety of the substrate ncAA as well as the *N*-ε-carbonyl group remained unchanged as they represent substrate identity elements recognized by the wild-type enzyme [18].

Despite the high interest in their study, efficient protein- and site-selective post-translational modifications are difficult to achieve in *E. coli* and mammalian cells. Solid-phase peptide synthesis, on the other hand, suffers from limitations in the maximum polypeptide chain length. Consequently, orthogonal translation using appropriate ncAAs and compatible orthogonal pairs presents an excellent methodology because of its high selectivity. With methylation as a key post-translational modification (PTM) in eukaryotic organisms, methyllysine residues could be site-specifically created in histone proteins [34]. To reach this goal, the ncAA *N*-ε-allyloxycarbonyl-*N*-ε-methyl-L-lysine (**19** as ncAA scaffold) was incorporated using the *Mb*PylRS system. Histone proteins recombinantly produced as *E. coli*

inclusion bodies were refolded in vitro followed by conversion into the methyllysine-modified variants via a ruthenium catalyst.

Ubiquitination, that is an isopeptide linkage of a substrate protein to ubiquitin, was achieved via a *Mb*PylRS system and directed evolution toward several ncAA substrates. To achieve discrimination against lysine (which differs from the desired PTM-residue only by an inserted sulfur atom), Boc protection groups were employed. These were removed in vitro subsequent to protein production and purification. Within the same study, another lysine-PTM, δ-hydroxy-L-lysine, was also incorporated [35].

Protein acetylation as another PTM was achieved via cotranslational incorporation of *N*-ε-acetyl-L-lysine (**16**, AcK). To incorporate this ncAA, an *Mm*PylRS variant was created by directed evolution. Careful inspection of the enzyme's catalytic parameters revealed that K_M values for AcK remained high but still enabled relatively good production yields of modified CAT [36]. After optimization of tRNA processing and *Mb*PylRS expression, the same posttranslational lysine modification could be artificially created in human superoxide dismutase (hSOD) using *Saccharomyces cerevisiae* as expression host [37].

Aiming for the study of chromatin modifications, three PTMs, namely *N*-ε-propionyl- (**17**, Kpr), *N*-ε-butyryl- (**18**, Kbu), and *N*-ε-crotonyl-lysine (Kcr) were successfully incorporated into histone H3 lysine at position 9. Modified target proteins were produced in *E. coli*, refolded in vitro, and obtained in milligram quantities. Not naturally occurring in *E. coli*, some of these histone modifications were found to be partially deacylated by unknown mechanisms. Supplementation of a deacylase inhibitor significantly reduced this target protein fraction [26]. Chemical structures of amino acid analogs enabling the creation of posttranslational protein modifications via orthogonal translation are shown in Fig. **4**.

7 Complex Chemical Handles: Crosslinkers and Photocages

Highly selective crosslinking reactions present an important tool for protein interaction studies. Inevitably exposed to non-specific interactions inside the highly crowded environment of the cellular host, the artificially introduced chemical handles have to remain as stable and inert as possible. Once purified and/or exposed to their interaction partner, however, highly specific and fast reactions are desirable. At the same time, temperatures need to remain low so that protein denaturation is avoided or at least kept at a minimal level. Despite the availability of several chemical methods, those compatible with physiologic conditions, namely ambient temperature and close to neutral pH in aqueous solution, are consequently most promising. Bearing the potential to introduce new-to-nature chemistries site-specifically into proteins, several PylRS variants with altered ncAA substrate

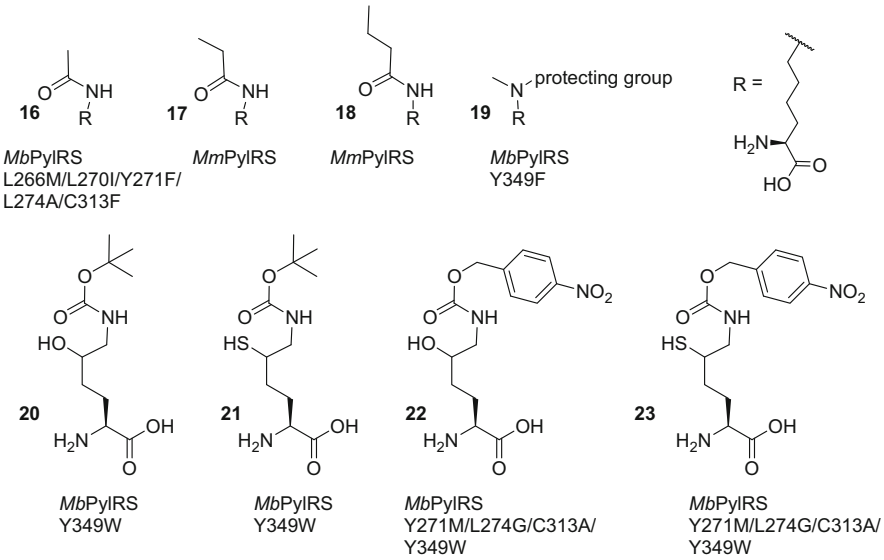

Fig. 4 Post-translational modifications. *Top row*: *N*-Acetylated lysines and *N*-methyl lysine. *N*-Methyl lysine has not directly been incorporated but can be introduced by chemical deprotection of *N*-ε-protected and methylated lysine derivatives. *Bottom row*: δ-Branched lysine derivatives used for traceless ubiquitination. References: **16**: [33]; **17, 18**: [26]; **19**: See [34] for an example. The original publication falsely states the PylRS mutation as Y384F, which is the corresponding position in *Mm*PylRS. **20**–**23**: [35]

spectra were created. Structures of the following types of chemical handles suitable for crosslinking reactions and photocaging are depicted in Fig. **5**.

7.1 Crosslinkers

With the selective joining of small functional groups under simple reaction conditions, the field of click chemistry meets several demands of protein crosslinking. Using pyrrolysine tRNA/synthetase pairs, several alkyne-containing ncAAs were successfully introduced into proteins via orthogonal translation. Using a *Mb*PylRS/$tRNA^{Pyl}$ pair, the aliphatic azide (*S*)-2-amino-6-((2-azidoethoxy)carbonylamino) hexanoic acid and the alkyne (*S*)-2-amino-6-((prop-2-ynyloxy)carbonylamino) hexanoic acid were successfully incorporated into model proteins using the *E. coli* translation apparatus. Via copper-catalyzed Huisgen [3 + 2] cycloaddition, the latter protein modification allowed subsequent biotin-labeling in vitro [43]. A follow-up study managed to functionalize yeast cells with this orthogonal translation system [37].

Charging $tRNA^{Pyl}$ with ncAAs bulkier than pyrrolysine requires alteration and enlargement of the PylRS substrate binding pocket. Both in *E. coli* and mammalian

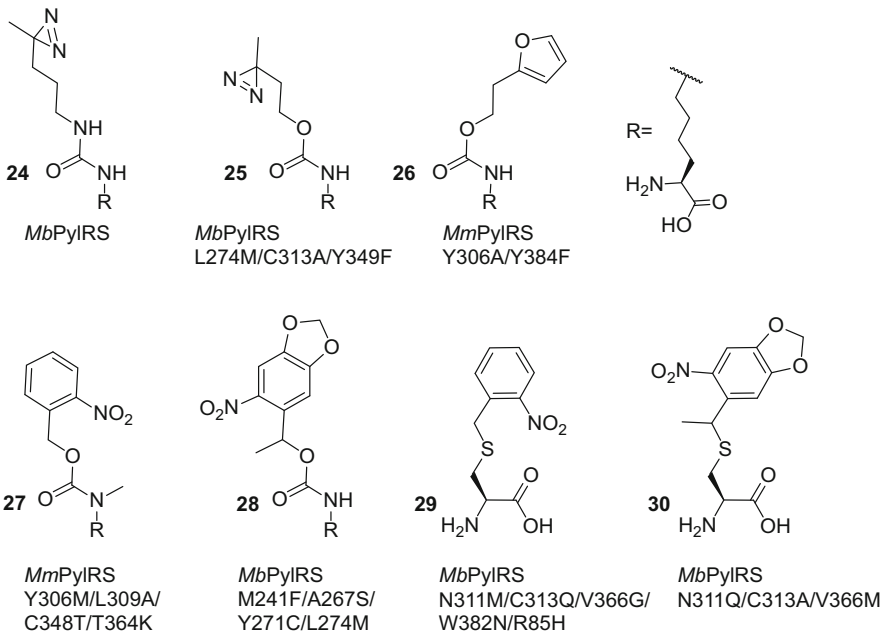

Fig. 5 Crosslinkers and photocaged amino acids. *Top row*: Photo-crosslinkers. *Bottom row*: Photocaged (methyl) lysine and cysteine. References: **24**: [24]; **25**: [38]; **26**: [39]; **27**: [40]; **28**: [37], [41]; **29, 30**: [42]

cells, an azide-bearing cyclic pyrrolysine analogue *N*-ε-(((1*R*,2*R*)-2-azidocyclopentyloxy)carbonyl)-L-lysine (**5**, ACPK) could be incorporated via an expanded genetic code. With this artificial chemical functionality introduced, the biocompatible Cu(I) ligand BTTES (2-[4-{(bis[(1-*tert*-butyl-1*H*-1,2,3-triazol-4-yl)methyl] amino)-methyl}-1*H*-1,2,3-triazol-1-yl]ethyl hydrogen sulfate) enabled copper-induced azide-alkyne cycloaddition (CuAAC) [44]. Rational design of an *Mm*PylRS double-alanine mutant (positions 346 and 348), for example, freed space which is filled by side chain moieties in the wild-type enzyme. This enabled the efficient incorporation of large para-substituents such as *p*-propargyloxy phenylalanine (**35**) among six additional ncAAs [45].

Multiple reactivities were enabled via the electron-deficient olefin *N*-ε-acryloyl-L-lysine (**6**, AcrK) and *N*-ε-crotonyl-L-lysine (**7**, CrtK) incorporated into proteins via an evolved *Mm*PylRS. Reactions tested with these ncAAs span a 1,4-addition for protein PEGylation, radical polymerization toward a copolymer hydrogel, and 1,3-dipolar cycloaddition. Although wild-type PylRS only afforded the incorporation of CrtK at low efficiency, the evolved synthetase variant enabled target protein production levels of 25 mg/L of *E. coli* culture. AcrK-modified superfolder GFP (sfGFP) efficiently reacted with thiol-containing nucleophiles, which resulted in turn-on fluorescence. The acrylamide moiety was further employed for labeling the outer membrane of *E. coli* via incorporation into OmpX [25].

Several ncAA crosslinking approaches suffer from relatively low reaction rate constants. To tackle this limitation, the Chin group managed to incorporate three ncAAs, namely *N*-ε-L-thiaprolyl-L-lysine (**14**), as well as *N*-ε-D-cysteinyl-L-lysine and *N*-ε-L-cysteinyl-L-lysine (**12**), for cyanobenzo-thiazole condensation. Once introduced, the 1,2-aminothiol moiety meets key demands of bioorthogonal reactions: Although not occurring naturally in proteins, it allows for their efficient, rapid, and specific labeling. Multiple rounds of mutagenesis and selection afforded the creation of an *Mb*PylRS variant that charges $tRNA^{Pyl}$ with *N*-ε-L-thiaprolyl-L-lysine, an ncAA which can be efficiently deprotected via *O*-methylhydroxyamine to form *N*-ε-L-cysteinyl-L-lysine. At high rates, this modification can be reacted with 2-cyanobenzothiazole (CBT) at physiological temperature and pH 7 [29].

To avoid copper-based catalysis of the crosslinking reaction, norbornene amino acids have proven useful for orthogonal translation and tetrazine click chemistry [23]. PylRS from *Methanosarcina mazei* was evolved via iterative saturation mutagenesis to incorporate such ncAAs. Subsequently, the modified protein was reacted with nitrile imines created from hydrazonoyl chloride or with tetrazines in an inverse electron demand Diels–Alder reaction [46]. By employing tetrazines for this type of reaction, rapid fluorogenic protein labeling could also be achieved with bicyclo[6.1.0]non-4-yn-9-ylmethanol (BCN). Because of its high specificity, the reaction between these two moieties proceeded in *E. coli* with low background. Incorporation of a BCN-containing ncAA in mammalian cell culture via an evolved *Mb*PylRS variant allowed TAMRA fluorescence labeling via tetrazine-conjugated fluorophores supplied to the growth medium [27].

To enrich interaction partners for mass spectrometry-based identification, a "click-and-release" strategy was developed. Human small ubiquitin-related modifier (SUMO) was C-terminally labeled with an alkyne-containing pyrrolysine analog further bearing an ester bond. This dual functionality enabled both copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) to bind SUMOylated proteins to a resin for enrichment and subsequent release via mild alkaline treatment followed by vacuum-assisted removal of the base [30].

Instead of supplying the final ncAA for tRNA-charging via PylRS, cells genetically further equipped with an appropriate modification pathway can be used to synthesize the desired ncAA in vivo from precursor molecules. Pyrroline-carboxy-lysine (Pcl, **1**) as a multi-purpose crosslinker was intracellularly produced from D-ornithine by two Pyl biosynthetic genes (*pyl*C and *pyl*D). At neutral pH, modified proteins were shown to react efficiently with 2-amino-benzaldehyde or 2-amino-acetophenone. Consequently, the incorporated ncAA allowed target protein PEGylation as well as labeling with diverse substrates including peptides, oligosaccharides, oligonucleotides, fluorescence, and biotin. Surprisingly, the demethylated analog Pcl even proved superior over pyrrolysine during orthogonal translation [22].

To overcome limitations from the above-mentioned CuAAC, Plass et al. introduced two previously characterized mutations to shape the *Mm*PylRS substrate binding pocket toward bulky pyrrolysine derivatives. This enabled the genetic incorporation of strained alkynes with relatively large side-chain sizes such

as *N*-ε-(cyclooct-2-yn-1-yloxy)carbonyl)-L-lysine (**10**). With applications for single molecule FRET studies and bacterial cell labeling, incorporation into the model protein GFP was efficient. Yields exceeded 10 mg/L of *E. coli* culture and copper-free click reactions with commercially available azide-functionalized dyes were shown [28].

7.2 Photocages

To reveal chemical functionalities in a spatiotemporal manner, recombinant proteins can be functionalized with photocaged ncAAs via orthogonal translation. Key prerequisites for this technique include non-toxic light wavelengths and intensities for rapid deprotection as well as efficient genetic incorporation of the ncAAs. Although enzymatic target protein labeling could in principle prove to be an in vitro labeling alternative, it frequently does not reach completeness and appropriately modified products often remain difficult to isolate.

As described in the previous section, lysine residues present important targets for PTM. Nuclear localization of eukaryotic proteins is frequently dependent on the presence of these residues. In yeast cells, successful usage of a photocaged lysine derivative *N*-ε-[(1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethoxy)-carbonyl]-L-lysine (**28**) for genetic code expansion extension was reported [37]. As part of another study, this ncAA was incorporated into nuclear localization sequences (NLS) of nucleoplasmin and the tumor suppressor p53 in human cells. This afforded directed evolution of *Mb*PylRS toward the new substrate. Using ncAA-modified proteins, Gautier et al. managed to change their cellular localization via controlled photolysis of the caged residue [41].

Photocaged *N*-ε-methyl-L-lysine was used to facilitate *Mm*PylRS discrimination against lysine. Positive and negative selection led to the isolation of a PylRS variant which used the ncAA for $tRNA^{Pyl}$ charging. Photolytic deprotection was reported to proceed efficiently during exposure to 365 nm UV light for 1 h under physiological pH [40].

7.3 Photo-Crosslinkers

Compared to copper-catalyzed or chemically-induced reactions which frequently suffer from limited biocompatibility, photo-induced crosslinking reactions present a promising alternative and feature spatiotemporal control. Consequently, several lysine derivatives have been synthesized and tested with pyrrolysine tRNA/synthetase pairs for incorporation into proteins.

Orthogonal translation was reported successful in yeast cells with *N*-ε-[(2-(3-methyl-3*H*-diazirin-3-yl)ethoxy)carbonyl]-L-lysine (**25**) as a photocrosslinking ncAA and human superoxide dismutase (hSOD) as the target protein

[37]. Using an evolved variant instead of wild-type *Mb*PylRS, the same ncAA (alternatively called 3'-azibutyl-*N*-carbamoyl-lysine, AbK) was incorporated into cyclin-dependent kinase 5 (Cdk5). Photoactivation via UV light (360 nm) led to crosslinking to its substrate, p21-activated kinase 1 (Pak1) in mammalian HEK 293 T cells [38]. In a similar fashion, the acid chaperone HdeA of enteric bacterial *Shigella* pathogens was used to study host cell infection mechanisms and in vivo protein-protein interactions via a photoaffinity group. ((3-(3-Methyl-3*H*-diazirin-3-yl)propamino)carbonyl)-*N*-ε-L-lysine (DiZPK, **24**) proved acid-stable and resulted in covalent protein coupling superior to *p*-benzoylphenylalanine (Bpa) [24].

Nitrile imines created from a tetrazole moiety via UV-irradiation can be used for photo-crosslinking cycloaddition reactions with norbornene-modified proteins. However, Kaya et al. also showed that harmful effects can arise from the UV irradiation required for crosslinking [46]. Compared to the latter ncAA, cyclopropene presents a less bulky moiety and should thus be incorporated into proteins more efficiently. Its inherently high reactivity stems from ring strain and enables rapid photoinduced cycloaddition reactions. These have been demonstrated with two tetrazoles in *E. coli* and mammalian HEK 293 cells. Biocompatibility with the cellular environment was further assessed by exposure to glutathione as an abundant biological nucleophile [47].

Liberated by light, radicals can be used to drive copper-free crosslinking reactions. Via anti-Markovnikov thiol-ene and thiol-yne coupling (TEC and TYC), regioselective reactions were induced with low-energy, near-UV light (365–400 nm). Robust product formation was achieved in aqueous buffer using VA-044 or 2,2-dimethoxy-2-phenylaceto-phenone (DPAP, 10%) as photo-induced catalysts. Able to link thiols to an alkyne, TYC enabled fluorescent protein labeling with *N*,*N*′-bis(dansyl)cystamine [48].

Less harmful for biomolecules than UV light and further benefitting from higher penetration depths in biological samples and tissues, red light was successfully employed for protein crosslinking. Instead of protein–protein interactions, a new method to study the binding of proteins to nucleotides was developed using orthogonal translation. Schmidt and Summerer managed to genetically incorporate *N*-ε-[2-(furan-2-yl)ethoxy]carbonyllysine (**26**) via amber suppression. Because the wild-type enzyme did not result in detectable amounts of modified target protein, the substrate binding pocket of *Mm*PylRS was evolved toward incorporation of the ncAA. Complex formation of an HIV-1 protein (trans-activator of transcription, TAT) with a hairpin RNA as its natural interaction partner could be detected via the new chemical functionality. The required reactive singlet oxygen was induced via photosensitizers such as *N*-ε-[2-(furan-2-yl)ethoxy]carbonyllysine [39].

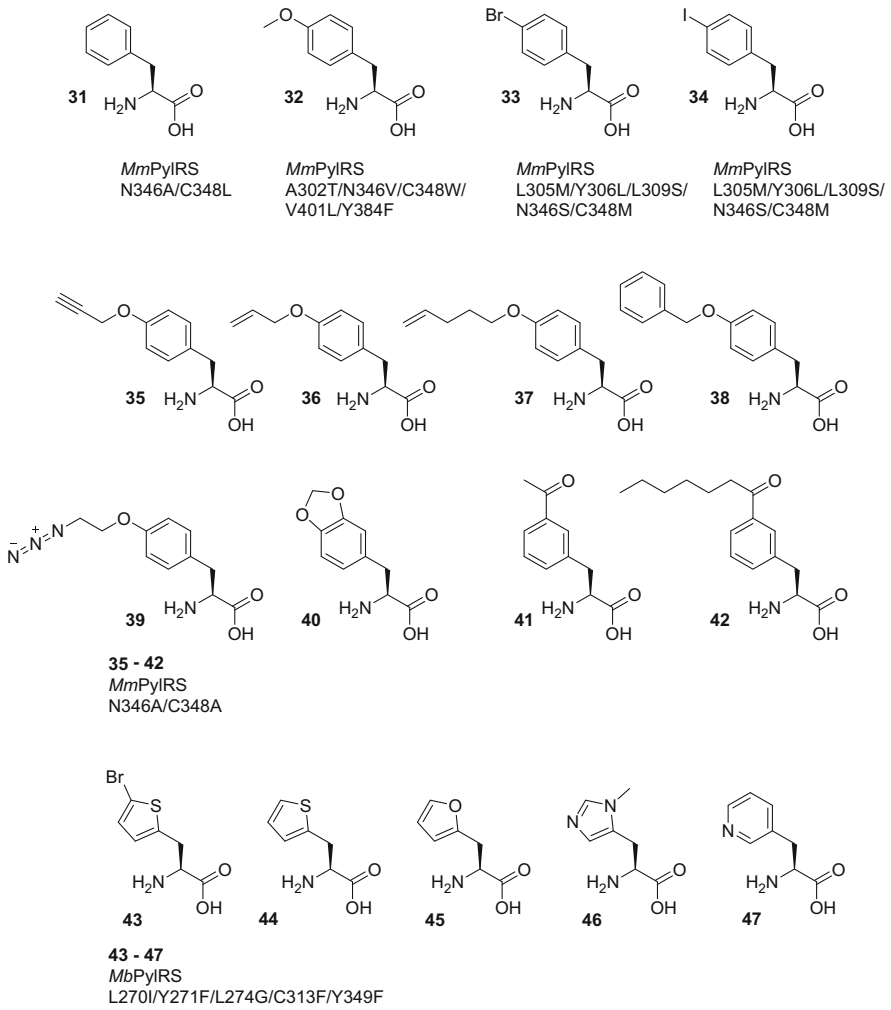

Fig. 6 Aromatic amino acids. *Top row*: Phenylalanine, methyltyrosine and halophenylalanines. *Middle rows*: Phenylalanine derivatives with bulky para and meta substituents. *Bottom row*: Histidine analogs. References: **31, 33, 34**: [49]; **32**: [18]; **35–38**: [45]; **39–42**: [50]; **43–47**: [51]

8 Aromatic Amino Acid Analogs (Phe, His, Tyr)

With the basic capability to transfer several pyrrolysine derivatives to the partner tRNA in vitro and in vivo, PylRS variants were successfully evolved toward structurally different aromatic ncAAs with chemical structures as shown in Fig. **6**.

A broad-specificity double-alanine mutant of PylRS allowed efficient production of recombinant proteins site-specifically modified with *p*-propargyloxy phenylalanine (**35**) supplied at 5 mM concentration [45]. Bearing short aromatic side chains,

L-phenylalanine, *p*-iodo- and *p*-bromo-L-phenylalanine were shown to be used for tRNA aminoacylation by an evolved *Methanosarcina mazei* PylRS enzyme. The iodinated amino acid analog is envisioned for X-ray crystallography (as a marker heavy atom otherwise incorporated via crystal soaking or chemical treatment) and for protein crosslinking via Suzuki–Miyaura reactions. Although orthogonal translation using tyrosine analogs is frequently accomplished using the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/$tRNA^{Tyr}$ pair, this system bears different structural ncAA substrate requirements compared to the pyrrolysine system [49]. Further expansion of the structural substrate diversity of the PylRS substrate spectrum focused on *O*-methyl-L-tyrosine (Ome, **32**). Using X-ray crystallography and non-hydrolyzable ATP derivatives, the evolved *Mm*PylRS was shown to exhibit a decreased active site volume and compensatory mutations for lost substrate interactions. The aforementioned "gatekeeper" residue, Asn346, was specifically targeted for mutagenesis and directed evolution yielded a high-fidelity synthetase enzyme. Including genetic adjustments for tRNA processing, the Ome-specific enzyme also proved functional in mammalian HEK292 and HeLa cells [18]. Targeting PylRS residue Asn346 for mutagenesis, a N346S:C348I double mutant of the *M. mazei* aminoacyl-tRNA synthetase was also found to be "polyspecific", enabling the incorporation of several meta-substituted phenylalanine-based aromatic ncAAs [52]. With chemical ncAA synthesis starting from tyrosine, red-shifted photoswitchable azobenzenes were incorporated into sfGFP [53] via an *Mm*PylRS variant evolved in a previous study for azobenzenes photoswitchable via 365-nm light [54].

9 Multiple Noncanonical Amino Acid Mutagenesis

In principle, incorporating any of the above-mentioned noncanonical amino acids at multiple defined sites in a target protein via stop codon suppression is feasible. Incorporation of two different ncAAs can also be achieved, for example by combining PylRS-based amber suppression with a quadruplet-decoding *Mj*TyrRS-based orthogonal pair [55]. Introduction of more and more stop codons in the target gene, however, decreases the final protein yields obtainable via recombinant expression. Toward more efficient single and multiple ncAA incorporation via stop codon suppression, $tRNA^{Pyl}_{CUA}$ has been rationally evolved, with efficiency improvements expected to stem from interactions with *E. coli* elongation factor-Tu (EF-Tu) [56]. As introduced above, *E. coli* strains with attenuated [57] or deleted RF1 facilitate amber suppression at multiple sites, both in vivo and in cell-free systems [58–60]. To minimize toxicity resulting from ncAA incorporation at off-target sites in the host cell proteome, genomic recoding has yielded RF1-free *E. coli* strains deprived of the amber stop codons of 95 essential or all protein-coding genes, respectively [6, 61]. Whereas the unique structural features of $tRNA^{Pyl}_{CUA}$ establish the orthogonality of PylRS-based amber suppression, recent

in vitro studies revealed multiple steps in orthogonal translation which limit the efficiency [62]. Consequently, further expression strain and plasmid setup engineering can be expected to yield improvements for ncAA incorporation.

10 Outlook and Perspectives

During the past decade, orthogonal protein translation has been well-established. Orthogonal tRNA/synthetase pairs have been made compatible with host organisms such as *E. coli*, *S. cerevisiae*, mammalian cell culture, *C. elegans*, and *D. melanogaster*. Specialized fields of biological research, for example pathogen microbiology and virology (see previous sections), now employ the diversity of ncAAs to modify proteins site specifically with high efficiency. Using orthogonal pairs, diverse methods have become available for precise fluorescent protein labeling or selective crosslinking in vitro and in vivo.

Limitations revealed during the development of synthetase enzymes which selectively charge their partner tRNA with ncAAs have been recognized and addressed. For instance, in vitro assays reveal catalytic efficiencies of evolved synthetases and allow fine-tuning toward higher orthogonal pair efficiency [63]. Protein structures have been determined not only for the wild-type enzymes but also for several variants generated by directed evolution. In the near future these data should allow the generation of precisely designed PylRS active site libraries as a new method to obtain an even more diverse ncAA substrate spectrum. As illustrated by the incorporation of pyrroline-carboxy-lysine (Pcl, **1**), the production of the ncAA from less complex and thus more affordable precursor molecules can be achieved by the same cellular host which is able to incorporate it genetically into proteins. The development of crosslinking agents and photo-induced reactions follows a route toward high biocompatibility, at the same time maintaining selectivity and reaction speed. Consequently, many new applications of orthogonal translation are expected in the near future.

Because of the large and constantly increasing number of orthogonal pairs reported so far, it should be noted that this work cannot completely cover all developments in the PylRS research field. For the same reason, developments of other orthogonal pairs such as those based on *Mj*TyrRS could not be covered herein. Consequently, readers are referred to reviews such as those of Neumann or Liu and Schultz for further references [64, 65].

References

1. Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, Visel A, Woyke T, Kyrpides NC, Rubin EM (2014) Stop codon reassignments in the wild. Science 344 (6186):909–913

2. Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of Escherichia coli. Science 292(5516):498–500
3. Craigen WJ, Caskey CT (1987) The function, structure and regulation of E. coli peptide chain release factors. Biochimie 69(10):1031–1041
4. Johnson DBF, Wang C, Xu J, Schultz MD, Schmitz RJ, Ecker JR, Wang L (2012) Release factor one is nonessential in Escherichia coli. ACS Chem Biol 7(8):1337–1344
5. Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K (2010) Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res 38(22):8188–8195
6. Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342(6156):357–360
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
8. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis tool web services from the EMBL-EBI. Nucleic Acids Res 41(Web Server issue): W597–W600
9. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27(2):221–224
10. Kavran JM, Gundllapalli S, O'Donoghue P, Englert M, Söll D, Steitz TA (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci U S A 104(27):11268–11273
11. Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, Umehara T, Söll D, Nureki O (2009) Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. Nature 457(7233):1163–1167
12. James CM, Ferguson TK, Leykam JF, Krzycki JA (2001) The amber codon in the gene encoding the monomethylamine methyltransferase isolated from Methanosarcina barkeri is translated as a sense codon. J Biol Chem 276(36):34252–34258
13. Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296 (5572):1462–1466
14. Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 296(5572):1459–1462
15. Borrel G, Gaci N, Peyret P, O'Toole PW, Gribaldo S, Brugère J-F (2014) Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. Archaea 2014:374146
16. Polycarpo C, Ambrogelly A, Bérubé A, Winbush SM, McCloskey JA, Crain PF, Wood JL, Söll D (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proc Natl Acad Sci U S A 101(34):12450–12454
17. Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. Chem Biol 15 (11):1187–1197
18. Takimoto JK, Dellas N, Noel JP, Wang L (2011) Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. ACS Chem Biol 6(7):733–743
19. Kwok Y, Wong JT (1980) Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Can J Biochem 58(3):213–218
20. Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. Protein Sci 7(2):419–426

21. Katayama H, Nozawa K, Nureki O, Nakahara Y, Hojo H (2012) Pyrrolysine analogs as substrates for bacterial pyrrolysyl-tRNA synthetase in vitro and in vivo. Biosci Biotechnol Biochem 76(1):205–208
22. Ou W, Uno T, Chiu H-P, Grünewald J, Cellitti SE, Crossgrove T, Hao X, Fan Q, Quinn LL, Patterson P, Okach L, Jones DH, Lesley SA, Brock A, Geierstanger BH (2011) Site-specific protein modifications through pyrroline-carboxy-lysine residues. Proc Natl Acad Sci U S A 108(26):10437–10442
23. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nat Chem 4(4):298–304
24. Lin S, Zhang Z, Xu H, Li L, Chen S, Li J, Hao Z, Chen PR (2011) Site-specific incorporation of photo-cross-linker and bioorthogonal amino acids into enteric bacterial pathogens. J Am Chem Soc 133(50):20581–20587
25. Lee Y-J, Wu B, Raymond JE, Zeng Y, Fang X, Wooley KL, Liu WR (2013) A genetically encoded acrylamide functionality. ACS Chem Biol 8(8):1664–1670
26. Gattner MJ, Vrabel M, Carell T (2013) Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. Chem Commun (Camb) 49(4):379–381
27. Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, Fox JM, Chin JW (2012) Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. J Am Chem Soc 134(25):10317–10320
28. Plass T, Milles S, Koehler C, Schultz C, Lemke EA (2011) Genetically encoded copper-free click chemistry. Angew Chem Int Ed Engl 50(17):3878–3881
29. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. J Am Chem Soc 133(30):11418–11421
30. Lee MM, Fekner T, Tang T-H, Wang L, Chan AH-Y, Hsu P-H, Au SW, Chan MK (2013) A click-and-release pyrrolysine analogue. Chembiochem 14(7):805–808
31. Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S (2009) Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. J Mol Biol 385(5):1352–1360
32. Li Y-M, Yang M-Y, Huang Y-C, Li Y-T, Chen PR, Liu L (2012) Ligation of expressed protein α-hydrazides via genetic incorporation of an α-hydroxy acid. ACS Chem Biol 7(6):1015–1022
33. Neumann H, Hancock SM, Buning R, Routh A, Chapman L, Somers J, Owen-Hughes T, van Noort J, Rhodes D, Chin JW (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell 36 (1):153–163
34. Ai H-W, Lee JW, Schultz PG (2010) A method to site-specifically introduce methyllysine into proteins in E. coli. Chem Commun (Camb) 46(30):5506–5508
35. Virdee S, Kapadnis PB, Elliott T, Lang K, Madrzak J, Nguyen DP, Riechmann L, Chin JW (2011) Traceless and site-specific ubiquitination of recombinant proteins. J Am Chem Soc 133 (28):10708–10711
36. Umehara T, Kim J, Lee S, Guo L-T, Söll D, Park H-S (2012) N-Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. FEBS Lett 586(6):729–733
37. Hancock SM, Uprety R, Deiters A, Chin JW (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J Am Chem Soc 132(42):14819–14824
38. Ai H, Shen W, Sagi A, Chen PR, Schultz PG (2011) Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. Chembiochem 12(12):1854–1857
39. Schmidt MJ, Summerer D (2013) Red-light-controlled protein-RNA crosslinking with a genetically encoded furan. Angew Chem Int Ed Engl 52(17):4690–4693

40. Wang Y-S, Wu B, Wang Z, Huang Y, Wan W, Russell WK, Pai P-J, Moe YN, Russell DH, Liu WR (2010) A genetically encoded photocaged Nepsilon-methyl-L-lysine. Mol Biosyst 6 (9):1557–1560
41. Gautier A, Nguyen DP, Lusic H, An W, Deiters A, Chin JW (2010) Genetically encoded photocontrol of protein localization in mammalian cells. J Am Chem Soc 132(12):4086–4088
42. Nguyen DP, Mahesh M, Elsässer SJ, Hancock SM, Uttamapinant C, Chin JW (2014) Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. J Am Chem Soc 136(6):2240–2243
43. Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. J Am Chem Soc 131(25):8720–8721
44. Hao Z, Song Y, Lin S, Yang M, Liang Y, Wang J, Chen PR (2011) A readily synthesized cyclic pyrrolysine analogue for site-specific protein 'click' labeling. Chem Commun (Camb) 47 (15):4502–4504
45. Wang Y-S, Fang X, Wallace AL, Wu B, Liu WR (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. J Am Chem Soc 134(6):2950–2953
46. Kaya E, Vrabel M, Deiml C, Prill S, Fluxa VS, Carell T (2012) A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. Angew Chem Int Ed Engl 51(18):4466–4469
47. Yu Z, Pan Y, Wang Z, Wang J, Lin Q (2012) Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells. Angew Chem Int Ed Engl 51(42):10600–10604
48. Li Y, Pan M, Li Y, Huang Y, Guo Q (2013) Thiol-yne radical reaction mediated site-specific protein labeling via genetic incorporation of an alkynyl-L-lysine analogue. Org Biomol Chem 11(16):2624–2629
49. Wang Y-S, Russell WK, Wang Z, Wan W, Dodd LE, Pai P-J, Russell DH, Liu WR (2011) The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. Mol Biosyst 7(3):714–717
50. Tuley A, Wang Y-S, Fang X, Kurra Y, Rezenom YH, Liu WR (2014) The genetic incorporation of thirteen novel non-canonical amino acids. Chem Commun (Camb) 50(20):2673–2675
51. Xiao H, Peters FB, Yang P-Y, Reed S, Chittuluru JR, Schultz PG (2014) Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. ACS Chem Biol 9 (5):1092–1096
52. Guo L, Wang Y, Nakamura A, Eiler D, Kavran JM, Wong M, Kiessling LL, Steitz TA, O'Donoghue P, Söll D (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. Proc Natl Acad Sci U S A 111(47):16724–16729
53. John AA, Ramil CP, Tian Y, Cheng G, Lin Q (2015) Synthesis and site-specific incorporation of red-shifted azobenzene amino acids into proteins. Org Lett 17(24):6258–6261
54. Hoppmann C, Lacey VK, Louie GV, Wei J, Noel JP, Wang L (2014) Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. Angew Chem Int Ed Engl 53(15):3932–3936
55. Lammers C, Hahn LE, Neumann H (2014) Optimized plasmid systems for the incorporation of multiple different unnatural amino acids by evolved orthogonal ribosomes. Chembiochem 15 (12):1800–1804
56. Fan C, Xiong H, Reynolds NM, Söll D (2015) Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. Nucleic Acids Res 43(22):e156
57. Wu I-L, Patterson MA, Carpenter Desai HE, Mehl RA, Giorgi G, Conticello VP (2013) Multiple site-selective insertions of noncanonical amino acids into sequence-repetitive polypeptides. Chembiochem 14(8):968–978
58. Hong SH, Ntai I, Haimovich AD, Kelleher NL, Isaacs FJ, Jewett MC (2014) Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth Biol 3(6):398–409

59. Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7(11):779–786
60. Zheng Y, Lajoie MJ, Italia JS, Chin MA, Church GM, Chatterjee A (2016) Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. Mol Biosyst 12(6):1746–1749
61. Mukai T, Hoshi H, Ohtake K, Takahashi M, Yamaguchi A, Hayashi A, Yokoyama S, Sakamoto K (2015) Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. Sci Rep 5:9699
62. Wang J, Kwiatkowski M, Forster AC (2016) Kinetics of tRNA(Pyl)-mediated amber suppression in Escherichia coli translation reveals unexpected limiting steps and competing reactions. Biotechnol Bioeng 113(7):1552–1559
63. Wang N, Ju T, Niu W, Guo J (2015) Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids. ACS Synth Biol 4(3):207–212
64. Neumann H (2012) Rewiring translation - genetic code expansion and its applications. FEBS Lett 586(15):2057–2064
65. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444

Adv Biochem Eng Biotechnol (2018) 162: 21–44
DOI: 10.1007/10_2016_8

Published online: 9 June 2016

Promoter and Terminator Discovery and Engineering

Matthew Deaner and Hal S. Alper

Abstract Control of gene expression is crucial to optimize metabolic pathways and synthetic gene networks. Promoters and terminators are stretches of DNA upstream and downstream (respectively) of genes that control both the rate at which the gene is transcribed and the rate at which mRNA is degraded. As a result, both of these elements control net protein expression from a synthetic construct. Thus, it is highly important to discover and engineer promoters and terminators with desired characteristics. This chapter highlights various approaches taken to catalogue these important synthetic elements. Specifically, early strategies have focused largely on semi-rational techniques such as saturation mutagenesis to diversify native promoters and terminators. Next, in an effort to reduce the length of the synthetic biology design cycle, efforts in the field have turned towards the rational design of synthetic promoters and terminators. In this vein, we cover recently developed methods such as hybrid engineering, high throughput characterization, and thermodynamic modeling which allow finer control in the rational design of novel promoters and terminators. Emphasis is placed on the methodologies used and this chapter showcases the utility of these methods across multiple host organisms.

Keywords Biotechnology, Gene expression, Metabolic engineering, Promoters, Synthetic biology, Terminators

M. Deaner
McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton St. Stop C0400, Austin, TX 78712, USA

H.S. Alper (✉)
McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton St. Stop C0400, Austin, TX 78712, USA

Institute for Cellular and Molecular Biology, The University of Texas at Austin, 2500 Speedway Avenue, Austin, TX 78712, USA
e-mail: halper@che.utexas.edu

Contents

1 Introduction

Promoters and terminators play an indispensable role in metabolic engineering and synthetic biology applications for controlling gene expression. These critical elements play a part in regulating both the strength of transcription and the longevity of the transcript. Together, these two forces dictate the overall abundance of mRNA within the cell and ultimately play a significant role in determining protein contents within cells. At the same time, optimizing microorganisms for chemical production via metabolic engineering often requires the use of these elements to create highly regulated intracellular flux [1], often through high-strength promoters [2]. Fine-level control, inducibility, and expression range are all quite important in these endeavors, as has been seen with large strain engineering efforts such as rewiring the yeast *Saccharomyces cerevisiae* for industrial-level heterologous artemisinin production [3]. Fortunately, our understanding and cataloging of synthetic control elements such as promoters and terminators is continuously improving. In this chapter we consider the selection and engineering of both promoters and terminators for a variety of possible host organisms. Initially, we describe early strategies which mainly relied on genome mining and semi-rational mutagenesis techniques to improve sequence diversity and function. Next, we describe recent advances in the design of these parts using techniques such as hybrid engineering, high-throughput characterization, thermodynamic modeling, synthetic part

development, and rational design. In each of these cases, both our understanding and the utility of these parts are enhanced, thus increasing the rate of design cycles within cells.

2 Early Efforts of Promoter Identification and Diversification

2.1 Native Promoter Mining

The initial set of catalogued promoters for synthetic use was derived from the genome of the host organism or a phage that targets the host organism [4–8]. These promoters were often uncovered as a result of genomic dissections. The advent of genome sequencing and annotation (especially of hosts such as *Escherichia coli* and *S. cerevisiae*) allowed for the rapid discovery of endogenous promoters, especially when coupled with mRNA quantification methods. In a similar fashion, promoters for more complex systems such as mammalian hosts have largely been discovered via high-throughput screening methods such as "promoter trapping [9–11]." This approach typically involves random integration of a promoter-less vector containing GFP followed by fluorescence-based selection to determine adjacent, upstream regions of the genome that enable transcription. In similar fashion to other hosts, the sequencing of genomes (such as the CHO genome [12]) allowed for the discovery of novel, dynamic promoters such as pTXnip, which expresses proportionally to cell density [13].

Libraries of native promoters serve an important role as major synthetic parts and are among the most highly characterized [14, 15]; however, they remain limited in their ability to sample complete gene expression ranges. Although multiple gene overexpression techniques have been used in *E. coli* [16–18] and *S. cerevisiae* [19–22], among other organisms, this approach can be limited and leads to the build-up of toxic intermediates that reduce productivity [23]. In some cases – including commonly-used native promoters in *S. cerevisiae* – dependencies such as carbon-source metabolism [24] can impact part performance. Such a conditional function is exacerbated in mammalian hosts, as commonly-used viral promoters vary widely in performance between cell lines and are often unstable after many cell generations [25–27]. As a result, further engineering of promoters is necessary to obtain desired fine-tuned expression, stability, and conditional performance.

2.2 Mutagenesis Techniques to Diversify Promoter Strength

Random mutagenesis is a powerful approach to augment promoter function without explicitly requiring extensive knowledge of sequence-to-function mapping.

Specifically, because mutagenesis techniques such as error-prone PCR (Ep-PCR) indiscriminately target both consensus and non-consensus promoter regions, libraries with a large dynamic range of promoter function can be easily obtained. For instance, error-prone PCR was used to generate a mutant library of the prokaryotic P_L-λ bacteriophage-derived promoter, enabling a 196-fold dynamic range of expression in *E. coli* [28]. The utility of this library was demonstrated by optimizing the expression of phosphoenolpyruvate carboxylase (*ppc*) for biomass yield and deoxy-xylulose-P-synthase (*dxs*) for maximal lycopene production. The importance of an expression continuum was highlighted by the fact that optimal *dxs* expression was dependent on strain genetic background. Similar mutagenesis of the strong constitutive *S. cerevisiae TEF1* promoter yielded a library exhibiting a 15-fold dynamic range [28, 29]. Likewise, this library was used to optimize glycerol 3-phosphate dehydrogenase (*GPD1*) expression for glycerol overproduction in yeast.

As an alternative to Ep-PCR, serial deletion of promoter regions has been used to modulate expression, especially for mammalian hosts. Initially, serial deletion was used as a genetic tool to systematically remove portions of a promoter sequence to better understand function [30–32]. As these deletions often tend to dampen promoter activity, this approach has recently been used to generate libraries of weaker promoters [33, 34]. In this regard, serial deletion has been used to create knockdown libraries of glutamine synthetase (*GS*) expression for the GS-CHO expression system [35]. Moreover, serial deletion can also identify promoter variants that are cell-line specific. For example, the human cytomegalovirus (hCMV) promoter was optimized for transgene expression in both CHO-K1 and HEK-293 cells [36]. This study found that the full-length promoter gave the highest stable expression in CHO-K1 cells whereas the addition of the first exon to the minimal enhancer and core promoters was optimal for expression in HEK293 cells.

Although Ep-PCR and serial deletion are effective at creating a large dynamic range of promoter strength, these approaches suffer from two major deficiencies: (1) higher level expression is hard to achieve and (2) large pools of inactive mutants are generated because of aberrant mutagenesis of elements critical for transcription [2]. Newer techniques (described in the sections below) are required to gain higher expression consistently. To address the second limitation of large inactive pools, more targeted approaches that make use of molecular understanding of promoter function can be employed. As an example, a saturation mutagenesis approach (Fig. 1a) was used to specifically modulate the sequence between consensus −35 "TTGACA" and −10 "TATAAT" motifs [37]. As these two motifs are both necessary and sufficient for the recruitment of the σ^{70} factor of RNA polymerase II (RNAP II) to initiate transcription [38], a randomized linker region was generated that resulted in a promoter library with a 400-fold dynamic range in *Lactococcus lactis* [39]. To improve the dynamic range further, a library including mutations of the −35 and −10 motifs exhibited another three orders of magnitude in range, thus demonstrating the importance of the entire promoter sequence [39].

Eukaryotic promoters, although more complex and less rigidly defined than prokaryotic counterparts, can be broken down into a core promoter [40, 41] and

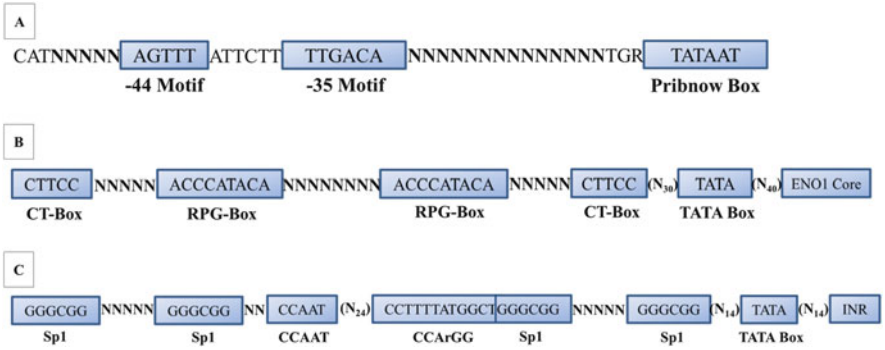

Fig. 1 Saturation mutagenesis strategies used to diversify promoters and improve understanding of promoter design rules. (**a**) Prokaryotic promoters have a highly constrained architecture with consensus −35 and −10 motifs spaced by exactly 17 base pairs for optimal function. (**b**, **c**) Eukaryotic promoters lack a rigidly-defined consensus architecture. (**b**) In yeast, promoters can be broken down into an upstream activating sequence (UAS) containing transcription factor binding sites (TFBSs), such as those for GCR1p (CT-Box) and Rap1p (RPG-box), and a core promoter which serves to recruit RNA Polymerase II. (**c**) In mammalian hosts, promoters follow a similar general architecture but contain additional consensus motifs such as the initiator element (INR, shown above), transcription factor IIB recognition sequence (BRE), motif ten element (MTE), and downstream promoter element (DPE)

upstream enhancer element(s) [42, 43] located 5′ of the core promoter. Efforts to engineer these distinct elements have been successful. For example, Jeppsson et al. created an *ENO1*-based promoter scaffold (Fig. 1b) containing two GCR1p TFBSs, two Rap1p TFBSs, and a TATA box coupled by spacers whose length was based on the architecture of native promoters [44, 45]. Randomization of these spacer regions afforded 37 synthetic promoters that spanned 3 orders of magnitude in strength. The utility of this library was demonstrated for the controlled knockdown of *ZWF1* expression, resulting in a 16% increase in yeast ethanol production from xylose fermentation. Finally, this same approach of creating synthetic promoter scaffolds followed by saturation mutagenesis has been applied to mammalian promoters (Fig. 1c) in which mutagenesis of regions between TFBSs in the JeT promoter afforded a weakened synthetic promoter library with a tenfold range [46].

Collectively, these early mutagenesis techniques demonstrate that utilizing native promoters (prokaryotes) or constructing synthetic promoters (eukaryotes) followed by randomization of spacer regions can provide a promoter library marked by downregulation. Although efforts continue to use these approaches, a greater understanding of promoter architecture and high-throughput characterization techniques have yielded new methods to design promoters rationally with highly specific expression characteristics as described in the following sections.

3 Rational Construction of Promoters with Desired Characteristics

3.1 Hybrid Promoter Engineering

Once essential components of promoter architecture are defined, it is possible to combine disparate elements in a "hybrid promoter engineering" scheme. Importantly, in contrast to Ep-PCR and saturation mutagenesis, the construction of hybrid promoters often yields synthetic promoters which are stronger than the core scaffold [2]. Thus, this technique serves as a potent way to amplify the expression of promoters – an important goal of many engineering endeavors. The first instance of hybrid promoter engineering involved the fusion of the *trp* and *lac* promoters to create the *tacI* and *tacII* promoters [47]. Notably, this resulted in promoters that were between 7 and 11 times stronger than the derepressed *lac* promoter although maintaining the same regulation. Similar approaches in *E. coli* have been utilized to generate regulated promoters. For instance, a strong binding site for the FadR transcription factor was placed upstream of the strong phage promoters P_L and P_{T7} to create a dynamic biosensor-regulator for acyl-CoA conversion to fatty acids in *E. coli* [48]. A similar concept was used to produce a malonyl-CoA responsive hybrid promoter that controlled flux from acyl-CoA to malonyl-CoA [49]. However, prokaryotic promoters may also be limited by promoter escape after transcript initiation, meaning that the addition of redundant hybrid elements is not guaranteed to improve transcription and can reduce transcription in some cases [50].

Unlike prokaryotic promoters, eukaryotic promoters are largely enhancer-limited, meaning that the addition of enhancer elements (by including additional binding sites) can both regulate and amplify promoter activity (Fig. 2a) [51]. Combining previously isolated Upstream Activating Sequences (UASs) from *CYC1* [52, 53], *CLB2* (UAS_{CLB}) [54], *CIT1* (UAS_{CIT}) [55], *GAL1-10* (UAS_{GAL}) [56], and *TEF1* (UAS_{TEF}) [51] with core promoters such as *GPD* (P_{GPD}) [24], *TEF1* (P_{TEF}) [4], *LEU2* (P_{LEUM}) [52], and *CYC1* (P_{CYC}) [57] can result in a predictable increase in transcriptional activity [51]. Ultimately, the strongest constitutive promoter in yeast was generated which had mRNA levels 2.5-fold higher than the *GPD* promoter [24]. Hybrid yeast promoters can also be designed for altered regulation. For example, linking various elements of UAS_{GAL} to a constitutive core results in a functional, galactose inducible promoter [51]. A similar approach has been conducted with regulated regions of the *ARO9* UAS [58]. Collectively, these approaches resulted in a library of galactose-inducible promoters with a 40-fold range in induced expression strength, and a tryptophan-inducible promoter with a 29-fold range in induced expression strength. This hybrid promoter approach has been extended to non-conventional yeasts such as the host *Yarrowia lipolytica*. For example, hybrid engineering on the LEU2 core promoter resulted in a constitutive promoter library with 400-fold range in expression [49]. Most importantly, this work demonstrated the generalizability of the hybrid promoter approach to multiple core promoters and alternative UAS elements [59]. Such strong promoters were

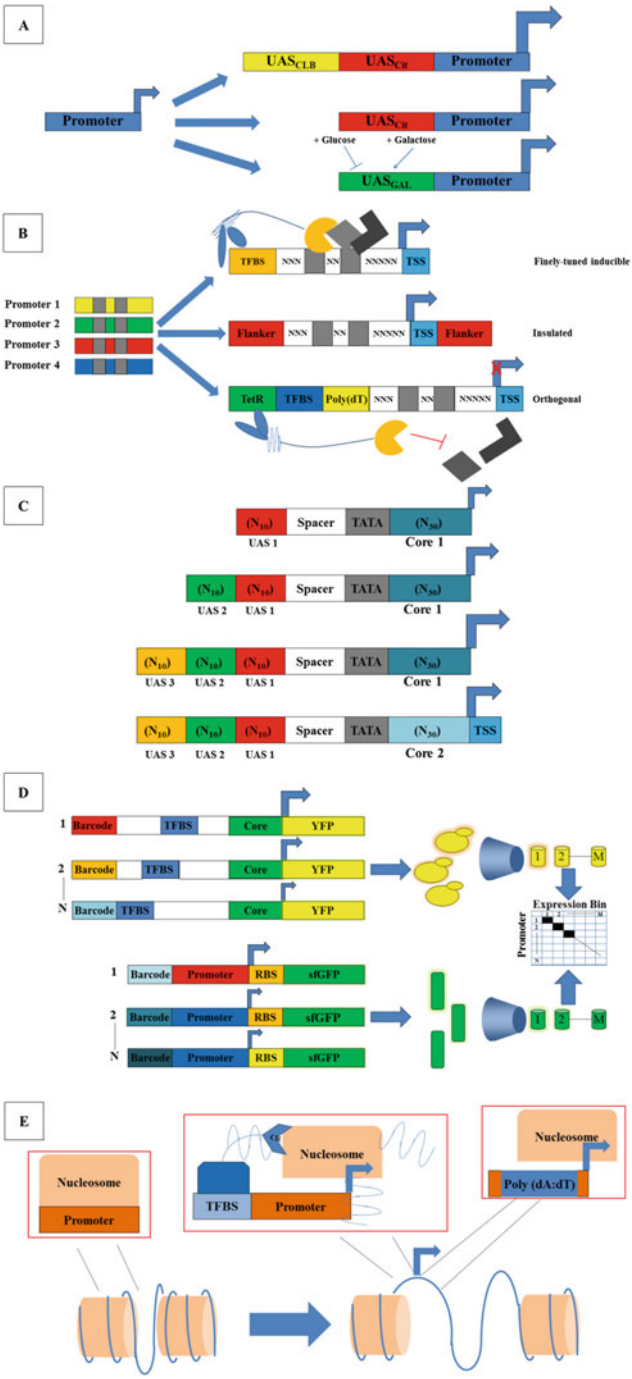

Fig. 2 Promoter engineering strategies. (**a**) Hybrid promoter engineering uses combinations of sequence motifs to modify expression and regulation. (**b**) Synthetic promoter scaffolds may be constructed based on native promoters with desired characteristics. These scaffolds can then be

used in the rewiring of *Y. lipolytica*, in which constitutive overexpression of *DGA1* using the UAS1B$_{16}$-*TEF1* hybrid promoter (among other genetic changes) resulted in a 60-fold improvement in lipogenesis [60].

Finally, the hybrid promoter approach has been further generalized to mammalian systems. For instance, the binding site of repressor PDX1 in the hCMV promoter was removed, enhancing expression fourfold in transient luciferase experiments [61]. The traditional additive hybrid approach has also been generalized to mammalian hosts to increase expression [62], improve transgene expression in specific hosts [63, 64], and impart novel regulation on promoters. As an example, a strong, cold-inducible promoter was created by combining a mild-cold responsive enhancer (MCRE) to the hCMV promoter [65]. Using this promoter and shifting temperature from 37°C to 32°C afforded sixfold higher erythropoietin production. Collectively, these results indicate that the hybrid promoter approaches are useful in both increasing net expression and imparting unique regulation.

3.2 Synthetic Promoter Scaffolds and Libraries

More recently, efforts have been made to establish synthetic and/or orthogonal [66, 67] promoters. Certainly bacterial systems can take advantage of the T7 RNA polymerase system [68] to generate short, synthetic, and orthogonal promoters for usage in logic gates [69–71]. However, the diversity of synthetic prokaryotic promoters is limited by the strict consensus promoter architecture not found in eukaryotes. To create a library of orthogonal core promoters in *S. cerevisiae*, native promoters were screened over a wide range of growth conditions to find a promoter scaffold that would exhibit the least amount of natural regulation [67]. The resulting candidate promoter, *PFY1* (P_{PFY1}), was then de-constructed to produce a minimal promoter scaffold (Fig. 2b) containing the ~100-bp core promoter, a Reb1p binding site, and a poly-dT element that maintained nucleosome depletion and constant DNA bending for constitutive RNA polymerase II access. By randomizing the spacer regions within this core promoter, a library of 36 minimally-regulated promoters with a 10-fold dynamic range in expression was created. This same methodology has been generalized to other organisms including *Pichia pastoris*, where four natively regulated promoters were sequence aligned to create a set of minimal

Fig. 2 (continued) diversified using saturation mutagenesis and modified via hybrid promoter engineering. (**c**) Minimal synthetic core and enhancer elements may be selected using randomization followed by FACS. (**d**) Promoter elements have been fully characterized via high-throughput oligo library synthesis followed by FACS sorting into different expression bins. (**e**) Expression can be tuned by altering nucleosome occupancy using a nucleosome prediction model or via addition of nucleosome-disfavoring poly (dA:dT) tracts. Chromatin regulators (CRs) can program a diverse range of transcriptional logic when targeted to synthetic promoters, thus creating more efficient synthetic circuits

core promoters from which sequence elements were transferred to modify the native *AOX1* promoter [72]. This same approach has been applied to human liver cells where a synthetic promoter scaffold with enhanced TF binding was created via the alignment of the hCMV and *HEF1α* promoters [64].

In an effort to generate more minimal, synthetic promoters using a library-based approach, Redden and Alper [73] developed an *S. cerevisiae* minimal core promoter scaffold (Fig. 2c) by dissecting both the core element and the UAS element and identifying functional, minimal units using a library-based approach involving FACS analysis and a series of robustness tests. Ultimately, a series of nine generic core elements were isolated which have limited homology to the genome. The same methodical workflow was used to isolate six synthetic 10-bp UAS sequences that activated these synthetic core promoters. Finally, these elements were combined to generate a minimal promoter with 70% the activity of GPD with an 80% reduction in size. Importantly, these promoters represent a minimal scaffold with highly defined consensus regions similar to those of prokaryotic promoters and thus these elements may be further rationally engineered for desired characteristics. Finally, in HeLa cells, synthetic 100-bp enhancers were created via construction of a library containing tandem repeats of random, micro-array printed 10-bp oligonucleotides [74]. This approach resulted in an enhancer with twice the strength of the hCMV enhancer. Thus, rationally constructing purely synthetic libraries can result in novel promoters with prescribed function across multiple hosts.

4 Sequence-Level Prediction and Specification of Promoters

Most of the methods described above rely heavily on repeated iterations of the synthetic biology design-build-test cycle [75, 76]. In contrast, the ability to specify promoter function at the DNA level would rapidly accelerate the field of synthetic biology by reducing the number of design cycles. This section describes many of the efforts that have been made toward this end.

4.1 Promoter Characterization and Standardization

Promoters, composed of a vast array of distinct regulatory elements, behave as a system that integrates an input from the host to produce an output: gene expression. As high-throughput oligo synthesis [77] and quantification of DNA, mRNA, and protein levels have improved, large combinatorial libraries may be generated to measure promoter performance across a wide range of contexts (Fig. 2d). For instance, in prokaryotes, the Ribosome Binding Site (RBS) controls the binding of the ribosome to the mRNA transcript, thus regulating gene expression at the

translational level whereas the promoter regulates expression at the transcriptional level. The independent function of these two regulatory elements has been thoroughly characterized and modeled via the construction of a library containing combinations of 114 promoters and 111 RBSs [78]. Although the model could explain 96% of RNA levels, its prediction of 82% of protein levels demonstrates the complex regulation of prokaryotic gene expression at the translational level. Thus, it is important to consider RBS performance when designing expression cassettes in pathways.

Eukaryotic transcription is regulated by a complex "program" of TF binding and RNAP II recruitment, and thus underlying "design rules" can be extracted that determine how the orientation, copy number, and context of TFBSs affect transcription. To parse these design rules, Sharon et al. [79] created a combinatorial library varying these parameters for 75 transcription factors. Fluorescence-activated cell sorting (FACS) coupled with high-throughput sequencing of 6,500 barcoded promoters generated a large dataset that uncovered regulatory design rules for TFs. For instance, in promoters that contained a Gcn4p binding site, expression and binding site location were related via a periodic function. Using a similar high-throughput characterization technique in mouse liver cells, it was possible to rapidly screen thousands of rationally designed enhancer haplotype variants [80]. This study found that enhancers are highly robust to single nucleotide variation (SNV), but that combinations of SNVs have an additive negative effect on function. This study also determined novel expression-enhancing motifs and characterized predicted TFBSs, thus laying the foundation for future enhancer design rules. In mammalian hosts, a similar predictive model has been used to identify K-mers that denote enhancers recognized by certain TFs [81, 82]. This model can be trained on CHIP-seq data [83] to predict enhancers throughout the genome.

Whereas TFBSs with a well-characterized function may be added to tune expression rationally, sequence-function mapping for core promoters is less understood. The core promoter sequence determines how RNAP II binds in the TATA region, forms the pre-initiation complex to unwind the DNA directly downstream, scans for a TSS, and initiates transcription [84–86]. Moving towards rational design, 859 native *S. cerevisiae* promoters were characterized using flow cytometry to generate a model relating maximal expression to short oligo motifs (K-mers) which impact these steps [86]. Although this model only accounted for 25% of the variance in an aggregate test promoter set, it nonetheless mapped expression-enhancing and repressing characteristics to short motifs in the core promoter to allow prediction of novel synthetic promoters. These results were improved upon via construction and high-throughput characterization of 13,000 specifically designed synthetic core promoters [87], leading to a model relating expression to the presence and orientation of consensus core promoter regions. However, despite analysis of thousands of systematically designed core promoters, the design rules for sequence level specification of core promoter activity are much less understood than those for UAS manipulation.

4.2 Thermodynamic Modeling and Prediction of Promoters

To fully expedite the synthetic biology design cycle, it is desirable to develop methods to design entire promoters de novo for predictable expression. In prokaryotes, thermodynamic models of ribosome interaction with mRNA secondary structure have been constructed to calculate the proportion of bound RBS-mRNA complexes, and thus translation rate [88, 89]. A thermodynamics-based RBS calculator was able to predict expression levels within a factor of 2.3 over an expression range of five orders of magnitude. Most importantly, this RBS calculator takes into account variations in translation rate depending on the genetic context of the RBS, thus allowing a "forward engineering" approach for novel applications.

Although eukaryotic transcriptional regulation involves countless protein factor binding events prior to transcription initiation, it is nonetheless possible to thermodynamically model individual steps as a surrogate for transcription initiation rate. A thermodynamic model incorporating both TF-DNA and TF-TF interactions was trained upon a promoter library containing different TFBS combinations using "effective TF concentration" as a floating parameter to fit the data [90]. Overall, the model predicted 56% of the variance in expression across a wide variety of TFBS arrangements, thus laying a foundation for de novo design of regulatory logic at the DNA sequence level.

To generalize this model further, other events in transcription initiation have been considered. Thermodynamic modeling of the TATA–TATA-binding protein (TBP) complex formed as a first step in the recruitment of RNAP II [91] and re-design of promoters with different consensus TATA boxes created a promoter library which predictably scaled with the thermodynamic affinity of TBP to each TATA Box [92]. Incorporating the thermodynamic model for the TBP–TATA complex with the previously developed model for TF-RNA Polymerase II and TF-TF binding [90] explained 75% of variance in promoter expression across a wide variety of genetic contexts. These examples demonstrate the utility of thermodynamically modeling transcription initiation steps as a means to predict expression. Since discovering promoters is highly important for uncharacterized mammalian hosts, thermodynamic sequence-level approaches have been used to predict novel promoters based on DNA structural properties such as duplex stability and bendability [93, 94]. In addition, mammalian promoter regions have been modeled at the sequence level using an "alpha score," which describes the likelihood that a genomic region contains a promoter based on its nucleotide composition. Remodeling the X-linked gene cancer/testis antigen 1A promoter to have twice the alpha score improved expression in a non-quantitative manner [95]. Although predictive of high expression, these techniques are limited as they cannot design promoters de novo with prescribed expression. Nevertheless, they demonstrate the potential to use heuristic models for the design and prediction of DNA function.

4.3 Prediction and Rational Modulation of Promoter Nucleosome Occupancy

In eukaryotes, the secondary structure of promoter DNA wound around nucleosomes controls access to the transcription machinery [96]. As a result, the rational design of novel promoters must consider how primary sequence contributes to DNA secondary structure. Nucleosome occupancy at promoters strongly regulates gene expression because nucleosome binding can occlude TFBSs and RNAP II recruitment to the core promoter [97]. Accordingly, rational addition of a tunable nucleosome-disfavoring poly(dA:dT) element [91, 98, 99] upstream of the natural Gcn4p binding site in a synthetic *His3*-based promoter library afforded predictable control over nucleosome occupancy and thus expression [100]. Similarly, mutation of CpG islands known to be prone to methylation and silencing by histones eliminated promoter silencing during long-term transgene expression in embryonic stem cells [101]. Thus, nucleosome-disfavoring sequences may be considered part of the rational eukaryotic promoter engineering toolbox along with the addition of hybrid enhancers (Fig. 2e).

To map nucleosome occupancy to primary sequence for predictive engineering of promoters, a Hidden Markov Model (HMM) was trained on a genome-wide nucleosome map [102]. This model was utilized to investigate nucleosome occupancy of the previously mentioned *TEF1* promoter library, demonstrating that expression correlated inversely with predicted cumulative nucleosome occupancy in a very robust manner. To create a predictive model, a greedy algorithm was developed which allowed re-design of native promoters for up to 16-fold greater strength [103]. Furthermore, this approach was used for the successful de novo design of synthetic yeast promoters. Importantly, sequence-level prediction of nucleosome occupancy affords a predictive method to optimize native promoters fully regardless of genetic context. As a result, future efforts in this area must consider the precise control of nucleosome occupancy to modulate expression.

4.4 Design of Synthetic Promoters with Controlled Chromatin Environment

Moving forward from nucleosome models, the context of eukaryotic DNA is important in considering promoter function. Specifically, eukaryotic DNA is wound around histone octamers in 147 base pair increments and packaged together tightly to create the “bead-on-a-string” backbone of the chromatin [104]. This structure is not composed randomly; in fact, the structure of chromatin surrounding genes has a direct impact on their regulation [105–111]. Thus, any endeavor to engineer promoters rationally as synthetic biology “parts” that exhibit defined functions in any genetic context must take into account the chromatin environment of the promoter.

The first step towards any rational bottom-up synthetic biology engineering approach is to parse design rules from the native system. To create design rules for chromatin-based control, a combinatorial library of zinc finger-based synthetic transcription factors was created with specific yeast chromatin regulators (CRs) tethered as the activation domain [112]. These CRs impact gene expression by regulating PIC formation, remodeling and assembly of nucleosomes, chromatin accessibility via histone modification, and transcriptional elongation. From this library screening approach, many different classes of CRs were delineated: activators and repressors, synergistic regulators, spatially encoded regulators that could repress transcription from a non-canonical position downstream of genes, and CRs that could activate or repress multiple genes simultaneously over a long range of genomic space. These minimal chromatin-based components can thus act as synthetic "parts" to create a diverse array of transcriptional logic and predictably tune expression by altering chromatin state. These initial efforts demonstrate the first work towards considering greater genetic context for promoters.

In closing, promoter discovery and characterization has progressed from genome mining to random mutagenesis to combinatorial and rational design. In some of these later cases, the use of computational models has been able to speed the design-build-test cycle. Although limitations still exist with respect to inducible promoters, pure synthetic design, and maximal expression levels, the field has progressed rapidly in recent years.

5 Terminator Discovery and Characterization

In addition to promoters, terminators serve as an important control point when tuning expression in circuits and pathways [113, 114]. Unlike promoters, terminator cataloguing has not been as extensive until recently. In fact, most commonly used terminators have been relics from past experiments and are not often the most efficient. As an example, commonly used terminators such as the native bacteriophage T7 terminator exhibit low termination efficiencies, meaning that transcriptional flux continues through the expression cassette and affects the regulation of downstream genes and limits polymerase recycling [113–115]. Furthermore, the collection of terminators available to researchers has traditionally been much smaller in breadth than promoters [116], thus limiting large-scale pathways and circuits because of the fear of genetic instability via homologous recombination [117, 118]. Terminators also serve as a control point to tune expression in eukaryotes via the stability of the 3′ end of the mRNA transcript [119–121]. Thus, the base of commonly used terminators must be diversified to meet pathway specifications via both discovery and engineering techniques. We highlight various approaches from terminator mining to synthetic design and models in the following sections.

5.1 Native Terminator Mining

To diversify initially from the commonly used terminator library in *E. coli*, an extensive library of 582 natural and synthetic terminators [122, 123] was constructed and analyzed for its termination efficiency [124]. To enable further terminator engineering, the study also delineated terminator design rules based on a mechanism where RNAP stalls at the U:A tract, allowing an RNA hairpin to form within the RNA exit channel and terminating transcription. It was shown that the composition of the terminator U-tract effectively controls polymerase dissociation and can thus be rationally designed to impact terminator strength. This work served as one of the more exhaustive studies for bacteria to determine alternative terminators for synthetic constructs.

In contrast to prokaryotic intrinsic termination, eukaryotic mRNA transcript stability is regulated by recruited protein factors such as the cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) [125]. Thus, terminators must be characterized not only by their termination efficiency but also by their impact on mRNA and protein levels. Yamanishi et al. undertook the first genome-scale flow cytometry characterization of yeast terminators, determining that the majority of terminators enabling higher expression from a synthetic construct came from ribosomal protein genes [120]. A separate, high-capacity terminator library was constructed by selecting a subset of terminators originating from genes shown to have higher mRNA half-lives [121]. Characterization of this library established a direct relationship between terminator strength and mRNA half-life, thus laying the groundwork for terminator design rules. In addition, the utility of these alternative terminators was proven by improved pathway flux with similar or lower promoter strength as those originally paired with a "traditional" terminator. Thus, terminators clearly serve as an important synthetic part that must be rationally specified to tune expression for metabolic engineering applications.

6 Rational Construction of Terminators with Desired Characteristics

6.1 Hybrid Terminator Engineering

Similar to promoters, the hybrid engineering approach has yielded synthetic terminators with enhanced efficiencies. Multiple combinations of both native and synthetic termination signals were used to enhance the termination efficiency of the T7 terminator while retaining its orthogonality [126]. However, this hybrid approach faces limitations in eukaryotes because termination is a highly concerted process regulated by multiple disparate elements (Fig. 3a).

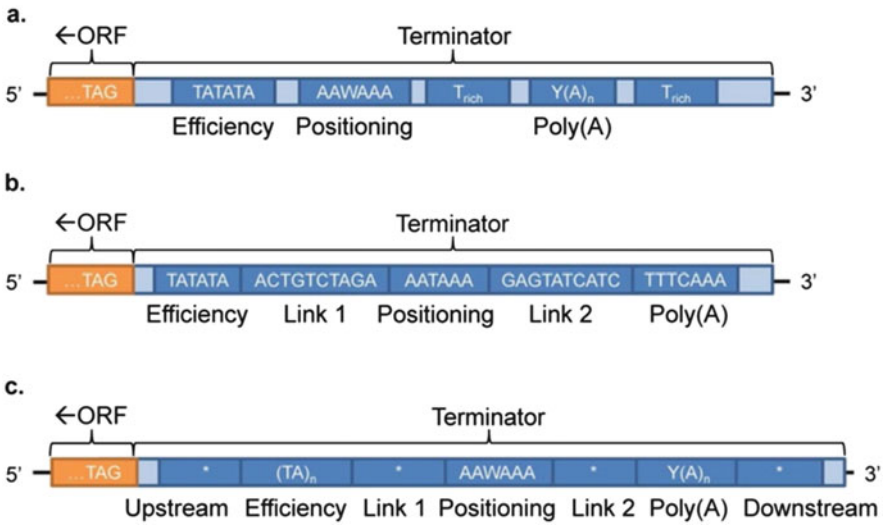

Fig. 3 Evolution of terminators in *S. cerevisiae*. (**a**) Unlike promoters, native terminators consist of many defined consensus motifs. (**b**) A minimal terminator scaffold was created by spacing these terminator motifs 10 bp apart. (**c**) This scaffold was engineered by rationally modifying the linkers between consensus motifs, adding upstream and downstream elements, and changing the length and sequence of consensus motifs. Licensing: Reprinted with permission from Curran KA, Morse NJ, Markham KA et al. (2015) Short, Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol.* 2015, 4, 824–832. doi:10.1021/sb5003357. Copyright © 2015 American Chemical Society

6.2 *Synthetic Terminator Scaffolds and Libraries*

To overcome the limitations of hybrid terminator engineering in yeast, a synthetic minimal terminator scaffold (T_{Guo}) was constructed by stringing together defined consensus efficiency, positioning, and poly-adenylation elements which cooperate in the cleavage and 3′ polyadenylation of the mRNA transcript (Fig. 3b) [127]. This minimal scaffold was both diversified and enhanced using modified consensus termination elements and mRNA stability elements [128] to produce a library of rationally designed synthetic terminators (Fig. 3c) which were functional in multiple hosts and improved *CAD1* expression for itaconic acid production [129]. Importantly, this technique allowed delineation of design rules based on consensus element identity and spacing, enabling potential rational design of synthetic terminators. These resulting terminators were much shorter in size than native terminators with the additional benefit of enhanced mRNA stability and increased protein production. Thus, in a similar fashion as described with promoters above, once a fundamental understanding of molecular function is obtained, synthetic part design can proceed.

7 Sequence-Level Prediction and Specification of Terminators

Although the previously described methods of synthetic terminator design allow rational diversification of the terminator library, they are nonetheless limited by the natural sequence space. Pure de novo design of terminators requires a fundamental understanding of the constraints underlying terminator function. Very early studies have begun to elucidate underlying design principles for terminators; however, this area is lagging behind the progress made with promoters as described above.

7.1 Terminator Characterization and Standardization

To this end, high-throughput studies have been carried out to measure quantitatively the performance of terminators and determine predictive sequence features for design in both prokaryotes and eukaryotes. For instance, systematic variation of terminator U-tract and hairpin stem-loop sequences in the aforementioned *E. coli* terminator library [124] afforded optimal expression-enhancing consensus sequences for rational construction of synthetic terminators.

Both native and synthetic terminator libraries have been constructed and characterized to tease apart the functions of different terminator motifs [130] in regulating mRNA abundance in yeast [131, 132]. Characterization of these libraries showed that the AU-rich efficiency element upstream of the poly(A) site plays a major role in 3′ end processing and transcription termination. In addition, terminators were broken down into mono- and di-nucleotide K-mers, leading to identification of dA:dT elements as a major determinant in terminator strength. From these studies, it appears that terminators can be broken down into a collection of tunable elements for rational design.

7.2 Thermodynamic Modeling and Prediction of Terminators

To generate a finer continuum of terminator function, it has become necessary to engineer entirely synthetic terminator sequences based on known design rules and thermodynamic prediction. In prokaryotes, multiple biophysical models have been developed to predict terminator strength based on elementary steps in termination, including U:A hybrid formation, hairpin formation, and mRNA transcript dissociation [122, 133, 134]. Training one of these models on a set of natural and synthetic

terminators over a large dynamic range in termination efficiencies afforded a linear sequence-function model with a high coefficient of determination ($R^2 = 0.81$) [134].

In *S. cerevisiae*, however, terminator function is much less predictable based simply on distinct sequence elements whose function is determined by biophysical models. In fact, characterization of the aforementioned rationally designed synthetic library [129] demonstrated that consensus termination motifs were not entirely additive. This suggests there is a fundamental code underlying termination in yeast which remains to be uncovered before thermodynamic prediction becomes feasible. However, with a more rigidly defined architecture than promoters, yeast terminators are highly amenable to rational engineering for desired characteristics. Thus, creating fundamental models to describe eukaryotic termination and half-life stabilization are required to advance the field of terminator engineering.

8 Future Directions in Promoter and Terminator Engineering

Improved promoters and terminators help minimize the length of the design cycle. Optimal design of these elements must meet three criteria: robustness, orthogonality, and predictable tunability. Promoters and terminators must be robust in that they function consistently regardless of genetic background, genetic context, and cellular environment [135]. In this regard, unexpected deviation from desired promoter or terminator function is a severe hindrance to the rapid development of circuits and pathways leading to multiple iterations of the design cycle. To improve robustness, efforts have been made to create synthetic promoter scaffolds based on highly constitutive promoters which function consistently across many different cellular environments. However, to date, few significant efforts have been made to engineer eukaryotic promoters that are robust to differing genetic contexts. These efforts are also complicated by the fact that eukaryotic promoters are highly regulated by the chromatin environment in which they are placed. It is thus imperative to develop design rules that govern promoter and terminator chromatin environment to predict and control these factors for optimal gene expression. The promise of purely orthogonal elements can bypass some of the robustness issues as these promoters and terminators seem to function more ubiquitously. Overall, many strides have been made in the past 5 years to provide novel expression capabilities to promoters and terminators. However, because of the regulatory complexity of microorganism hosts, new techniques must be developed to predict and design promoters and terminators for desired function. Nevertheless, these new synthetic parts have greatly improved the ability to engineer strains for metabolic engineering and synthetic biology applications.

References

1. Blazeck J, Alper H (2010) Systems metabolic engineering: genome-scale models and beyond. Biotechnol J 5:647–659. doi:10.1002/biot.200900247
2. Blazeck J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol J 8:46–58. doi:10.1002/biot.201200120
3. Paddon CJ, Westfall PJ, Pitera DJ et al (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496:528–532. doi:10.1038/nature12051
4. Gatignol A, Dassain M, Tiraby G (1990) Cloning of *Saccharomyces cerevisiae* promoters using a probe vector based on phleomycin resistance. Gene 91:35–41
5. Hauf J, Zimmermann F, Müller S (2000) Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. Enzyme Microb Technol 26: 688–698
6. Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11:2237–2255
7. Reifenberger E, Boles E, Ciriacy M (1997) Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression. Eur J Biochem 245:324–333
8. Diderich JA, Schepper M, van Hoek P et al (1999) Glucose uptake kinetics and transcription of HXT genes in chemostat cultures of *Saccharomyces cerevisiae*. J Biol Chem 274: 15350–15359. doi:10.1074/jbc.274.22.15350
9. Pontiller J, Gross S, Thaisuchat H et al (2008) Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol 39:135–139. doi:10.1007/s12033-008-9044-9
10. Pontiller J, Maccani A, Baumann M et al (2010) Identification of CHO endogenous gene regulatory elements. Mol Biotechnol 45:235–240. doi:10.1007/s12033-010-9278-1
11. Chen J, Haverty J, Deng L et al (2013) Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J Biotechnol 167:255–261. doi:10.1016/j.jbiotec.2013.07.001
12. Xu X, Nagarajan H, Lewis NE et al (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29:735–741. doi:10.1038/nbt.1932
13. Le H, Vishwanathan N, Kantardjieff A et al (2013) Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metab Eng 20:212–220. doi:10.1016/j.ymben.2013.09.004
14. Partow S, Siewers V, Bjørn S et al (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. Yeast 27:955–964. doi:10.1002/yea.1806
15. Sun J, Shao Z, Zhao H et al (2012) Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. Biotechnol Bioeng 109:2082–2092. doi:10.1002/bit.24481
16. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72:211–222. doi:10.1007/s00253-006-0465-8
17. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189:113–130
18. Elvin CM, Thompson PR, Argall ME et al (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. Gene 87:123–126
19. Walfridsson M, Hallborn J, Penttilä M et al (1995) Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase. Appl Environ Microbiol 61:4184–4190
20. Lu C, Jeffries T (2007) Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. Appl Environ Microbiol 73:6072–6077. doi:10.1128/AEM.00955-07

21. Wisselink HW, Toirkens MJ, del Rosario Franco Berriel M et al (2007) Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose. Appl Environ Microbiol 73:4881–4891. doi:10.1128/AEM.00177-07
22. Alper H, Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol 7:715–723. doi:10.1038/nrmicro2186
23. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330: 1355–1358. doi:10.1126/science.1193990
24. Da Silva NA, Srikrishnan S (2012) Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. FEMS Yeast Res 12:197–214. doi:10.1111/j.1567-1364.2011.00769.x
25. Addison CL, Hitt M, Kunsken D, Graham FL (1997) Comparison of the human versus murine cytomegalovirus immediate early gene promoters for transgene expression by adenoviral vectors. J Gen Virol 78(Pt 7):1653–1661
26. Xia W, Bringmann P, McClary J et al (2006) High levels of protein expression using different mammalian CMV promoters in several cell lines. Protein Expr Purif 45:115–124. doi:10.1016/j.pep.2005.07.008
27. Kim M, O'Callaghan PM, Droms KA, James DC (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108:2434–2446. doi:10.1002/bit.23189
28. Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683
29. Nevoigt E, Kohnke J, Fischer CR et al (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl Environ Microbiol 72: 5266–5273. doi:10.1128/AEM.00530-06
30. Boshart M, Weber F, Jahn G et al (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41:521–530
31. Dorsch-Häsler K, Keil GM, Weber F et al (1985) A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. Proc Natl Acad Sci U S A 82:8325–8329
32. Nelson JA, Reynolds-Kohler C, Smith BA (1987) Negative and positive regulation by a short segment in the 5′-flanking region of the human cytomegalovirus major immediate-early gene. Mol Cell Biol 7:4125–4129
33. Prentice HL, Tonkin CJD, Caamano L, Sisk WP (2007) High level expression of proteins using sequences from the ferritin heavy chain gene locus. J Biotechnol 128:50–60. doi:10.1016/j.jbiotec.2006.09.021
34. Thaisuchat H, Baumann M, Pontiller J et al (2011) Identification of a novel temperature sensitive promoter in CHO cells. BMC Biotechnol 11:51. doi:10.1186/1472-6750-11-51
35. Fan L, Kadura I, Krebs LE et al (2013) Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J Biotechnol 168:652–658. doi:10.1016/j.jbiotec.2013.08.021
36. Mariati NYK, Chao S-H et al (2010) Evaluating regulatory elements of human cytomegalovirus major immediate early gene for enhancing transgene expression levels in CHO K1 and HEK293 cells. J Biotechnol 147:160–163. doi:10.1016/j.jbiotec.2010.02.022
37. Nair TM, Kulkarni BD (1994) On the consensus structure within the *E. coli* promoters. Biophys Chem 48:383–393
38. Gruber BTM, Gross CA (2003) Assay of *Escherichia coli* RNA polymerase: sigma–core interactions. Methods Enzymol 370:206–212
39. Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences. Appl Environ Microbiol 64:82–87
40. Juven-Gershon T, Hsu J-Y, Kadonaga JT (2006) Perspectives on the RNA polymerase II core promoter. Biochem Soc Trans 34:1047–1050. doi:10.1042/BST0341047

41. Juven-Gershon T, Hsu J-Y, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter – the gateway to transcription. Curr Opin Cell Biol 20:253–259. doi:10.1016/j.ceb.2008.03.003
42. Struhl K (1984) Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. Proc Natl Acad Sci U S A 81:7865–7869
43. Struhl K (1995) Yeast transcriptional regulatory mechanisms. Annu Rev Genet 29:651–674. doi:10.1146/annurev.ge.29.120195.003251
44. Jeppsson M, Johansson B, Jensen PR et al (2003) The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. Yeast 20:1263–1272. doi:10.1002/yea.1043
45. Drazinic CM, Smerage JB, López MC, Baker HV (1996) Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). Mol Cell Biol 16:3187–3196
46. Tornøe J, Kusk P, Johansen TE, Jensen PR (2002) Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297:21–32. doi:10.1016/S0378-1119(02)00878-8
47. De Boer HA, Comstock LJ, Vasser M (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. Proc Natl Acad Sci U S A 80:21–25
48. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotechnol 30:354–359. doi:10.1038/nbt.2149
49. Blazeck J, Liu L, Redden H, Alper H (2011) Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. Appl Environ Microbiol 77:7905–7914. doi:10.1128/AEM.05763-11
50. Hsu LM (2002) Promoter clearance and escape in prokaryotes. Biochim Biophys Acta 1577:191–207. doi:10.1016/S0167-4781(02)00452-9
51. Blazeck J, Garg R, Reed B, Alper HS (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. Biotechnol Bioeng 109:2884–2895. doi:10.1002/bit.24552
52. Guarente L, Hoar E (1984) Upstream activation sites of the CYC1 gene of *Saccharomyces cerevisiae* are active when inverted but not when placed downstream of the "TATA box". Proc Natl Acad Sci U S A 81:7860–7864
53. Guarente L, Lalonde B, Gifford P, Alani E (1984) Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S. cerevisiae*. Cell 36:503–511
54. Van Slyke C, Grayhack EJ (2003) The essential transcription factor Reb1p interacts with the CLB2 UAS outside of the G2/M control region. Nucleic Acids Res 31:4597–4607
55. Rosenkrantz M, Kell CS, Pennell EA et al (1994) Distinct upstream activation regions for glucose-repressed and derepressed expression of the yeast citrate synthase gene CIT1. Curr Genet 25:185–195
56. West RW, Yocum RR, Ptashne M (1984) *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. Mol Cell Biol 4:2467–2478
57. Guarente L, Ptashne M (1981) Fusion of *Escherichia coli* lacZ to the cytochrome c gene of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 78:2199–2203
58. Kim D, Kim JD, Baek K et al. (2003) Improved mammalian expression systems by manipulating transcriptional termination regions. Biotechnol Prog 19:1620–1622. doi:10.1021/bp0341186
59. Blazeck J, Reed B, Garg R et al (2013) Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. Appl Microbiol Biotechnol 97:3037–3052. doi:10.1007/s00253-012-4421-5
60. Blazeck J, Hill A, Liu L et al (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. Nat Commun 5:3131. doi:10.1038/ncomms4131

61. Chao S-H, Harada JN, Hyndman F et al (2004) PDX1, a cellular homeoprotein, binds to and regulates the activity of human cytomegalovirus immediate early promoter. J Biol Chem 279: 16111–16120. doi:10.1074/jbc.M312304200
62. Berg DT, Mooney PQ, Baez M, Grinnell BW (1988) Tandem promoter/enhancer units create a versatile regulatory element for the expression of genes in mammalian cells. Nucleic Acids Res 16:1635
63. Gehrke S, Jérôme V, Müller R (2003) Chimeric transcriptional control units for improved liver-specific transgene expression. Gene 322:137–143
64. Magnusson T, Haase R, Schleef M et al (2011) Sustained, high transgene expression in liver with plasmid vectors using optimized promoter-enhancer combinations. J Gene Med 13: 382–391. doi:10.1002/jgm.1585
65. Sumitomo Y, Higashitsuji H, Higashitsuji H et al (2012) Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. BMC Biotechnol 12:72. doi:10.1186/1472-6750-12-72
66. Garg A, Lohmueller JJ, Silver PA, Armel TZ (2012) Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res 40:7584–7595. doi:10.1093/nar/gks404
67. Blount BA, Weenink T, Vasylechko S, Ellis T (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One 7:1–11. doi:10.1371/journal.pone.0033279
68. Rong M, He B, McAllister WT, Durbin RK (1998) Promoter specificity determinants of T7 RNA polymerase. Proc Natl Acad Sci U S A 95:515–519
69. Temme K, Hill R, Segall-Shapiro TH et al (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res 40:8773–8781. doi:10.1093/nar/gks597
70. Shis DL, Bennett MR (2013) Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. Proc Natl Acad Sci U S A 110:5028–5033. doi:10.1073/pnas.1220157110
71. Segall-Shapiro TH, Meyer AJ, Ellington AD et al (2014) A "resource allocator" for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol 10:742
72. Vogl T, Ruth C, Pitzer J et al (2014) Synthetic core promoters for *Pichia pastoris*. ACS Synth Biol 3:188–191. doi:10.1021/sb400091p
73. Redden H, Alper HS (2015) The development and characterization of synthetic minimal yeast promoters. Nat Commun 6:7810. doi:10.1038/ncomms8810
74. Schlabach MR, Hu JK, Li M, Elledge SJ (2010) Synthetic design of strong promoters. Proc Natl Acad Sci U S A 107:2538–2543. doi:10.1073/pnas.0914803107
75. Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11: 367–379. doi:10.1038/nrg2775
76. Way JC, Collins JJ, Keasling JD, Silver PA (2014) Integrating biological redesign: where synthetic biology came from and where it needs to go. Cell 157:151–161. doi:10.1016/j.cell.2014.02.039
77. Kosuri S, Eroshenko N, LeProust EM et al (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat Biotechnol 28:1295–1299. doi:10.1038/nbt.1716
78. Kosuri S, Goodman DB, Cambray G et al (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc Natl Acad Sci U S A 110: 14024–14029. doi:10.1073/pnas.1301301110
79. Sharon E, Kalma Y, Sharp A et al (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 30:521–530. doi:10.1038/nbt.2205
80. Patwardhan RP, Hiatt JB, Witten DM et al (2012) Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol 30:265–270. doi:10.1038/nbt.2136
81. Lee D, Karchin R, Beer MA (2011) Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res 21:2167–2180. doi:10.1101/gr.121905.111

82. Fletez-Brant C, Lee D, McCallion AS, Beer MA (2013) kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets. Nucleic Acids Res 41: W544–W556. doi:10.1093/nar/gkt519
83. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497–1502. doi:10.1126/science.1141319
84. Giardina C, Lis JT (1993) DNA melting on yeast RNA polymerase II promoters. Science 261: 759–762
85. Sugihara F, Kasahara K, Kokubo T (2011) Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of *Saccharomyces cerevisiae*. Nucleic Acids Res 39:59–75. doi:10.1093/nar/gkq741
86. Lubliner S, Keren L, Segal E (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res 41:5569–5581. doi:10.1093/nar/gkt256
87. Lubliner S, Regev I, Lotan-pompan M et al (2015) Core promoter sequence in yeast is a major determinant of expression level. 1008–1017. doi:10.1101/gr.188193.114.1008
88. Salis HM, Mirsky EA, Voigt CA (2010) Automated design of synthetic ribosome binding sites to precisely control protein expression. Nat Biotechnol 27:946–950. doi:10.1038/nbt.1568.Automated
89. Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498:19–42. doi:10.1016/B978-0-12-385120-8.00002-4
90. Gertz J, Cohen BA (2009) Environment-specific combinatorial cis-regulation in synthetic promoters. 1–9. doi:10.1038/msb2009.1
91. Iyer V, Struhl K (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J 14:2570–2579
92. Mogno I, Vallania F, Mitra RD, Cohen BA (2010) TATA is a modular component of synthetic promoters. Genome Res 20:1391–1397. doi:10.1101/gr.106732.110
93. Abeel T, Saeys Y, Bonnet E et al (2008) Generic eukaryotic core promoter prediction using structural features of DNA. Genome Res 18:310–323. doi:10.1101/gr.6991408
94. Gan Y, Guan J, Zhou S (2012) A comparison study on feature selection of DNA structural properties for promoter prediction. BMC Bioinformatics 13:4. doi:10.1186/1471-2105-13-4
95. Grabherr MG, Pontiller J, Mauceli E et al (2011) Exploiting nucleotide composition to engineer promoters. PLoS One 6, e20136. doi:10.1371/journal.pone.0020136
96. Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98:285–294. doi:10.1016/S0092-8674(00)81958-3
97. Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. Cell 116:699–709. doi:10.1016/S0092-8674(04)00205-3
98. Anderson JD, Widom J (2001) Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol Cell Biol 21:3830–3839. doi:10.1128/MCB.21.11.3830-3839.2001
99. De Boer CG, Hughes TR (2014) Poly-dA:dT tracts form an in vivo nucleosomal turnstile. PLoS One 9, e110479. doi:10.1371/journal.pone.0110479
100. Raveh-Sadka T, Levo M, Shabi U et al (2012) Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet 44:743–750. doi:10.1038/ng.2305
101. Swindle CS, Kim HG, Klug CA (2004) Mutation of CpGs in the murine stem cell virus retroviral vector long terminal repeat represses silencing in embryonic stem cells. J Biol Chem 279:34–41. doi:10.1074/jbc.M309128200
102. Xi L, Fondufe-Mittendorf Y, Xia L et al (2010) Predicting nucleosome positioning using a duration hidden Markov model. BMC Bioinformatics 11:346. doi:10.1186/1471-2105-11-346
103. Curran KA, Crook NC, Karim AS et al (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. Nat Commun 5:4002. doi:10.1038/ncomms5002

104. Keung AJ, Joung JK, Khalil AS, Collins JJ (2015) Chromatin regulation at the frontier of synthetic biology. Nat Rev Genet. doi:10.1038/nrg3900
105. Ellis L, Atadja PW, Johnstone RW (2009) Epigenetics in cancer: targeting chromatin modifications. Mol Cancer Ther 8:1409–1420. doi:10.1158/1535-7163.MCT-08-0860
106. Ernst J, Kheradpour P, Mikkelsen TS et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49. doi:10.1038/nature09906
107. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M (2011) Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 12:36–47. doi:10.1038/nrm3036
108. Onder TT, Kara N, Cherry A et al (2012) Chromatin-modifying enzymes as modulators of reprogramming. Nature 483:598–602. doi:10.1038/nature10953
109. Rheinbay E, Louis DN, Bernstein BE, Suvà ML (2012) A tell-tail sign of chromatin: histone mutations drive pediatric glioblastoma. Cancer Cell 21:329–331. doi:10.1016/j.ccr.2012.03.001
110. Schuster-Böckler B, Lehner B (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature 488:504–507. doi:10.1038/nature11273
111. Wang X, Chen J, Quinn P (2012) Reprogramming microbial metabolic pathways. Zhurnal Eksp i Teor Fiz 181–201. doi:10.1007/978-94-007-5055-5
112. Keung AJ, Bashor CJ, Kiriakov S et al (2014) Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell 158:110–120. doi:10.1016/j.cell.2014.04.047
113. Du L, Gao R, Forster AC (2009) Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase. Biotechnol Bioeng 104:1189–1196. doi:10.1002/bit.22491
114. Du L, Villarreal S, Forster AC (2012) Multigene expression in vivo: supremacy of large versus small terminators for T7 RNA polymerase. Biotechnol Bioeng 109:1043–1050. doi:10.1002/bit.24379
115. Carter AD, Morris CE, McAllister WT (1981) Revised transcription map of the late region of bacteriophage T7 DNA. J Virol 37:636–642
116. Redden H, Morse N, Alper HS (2014) The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Res 1–12. doi:10.1111/1567-1364.12188
117. Sleight SC, Bartley BA, Lieviant JA, Sauro HM (2010) Designing and engineering evolutionary robust genetic circuits. J Biol Eng 4:12. doi:10.1186/1754-1611-4-12
118. Renda BA, Hammerling MJ, Barrick JE (2014) Engineering reduced evolutionary potential for synthetic biology. Mol Biosyst 10:1668–1678. doi:10.1039/c3mb70606k
119. Abe H, Aiba H (1996) Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. Biochimie 78:1035–1042
120. Yamanishi M, Ito Y, Kintaka R et al (2013) A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a "Terminatome" toolbox BT. ACS Synth Biol 2:337–347
121. Curran KA, Karim AS, Gupta A, Alper HS (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab Eng 19:88–97. doi:10.1016/j.ymben.2013.07.001
122. Von Hippel PH, Yager TD (1991) Transcript elongation and termination are competitive kinetic processes. Proc Natl Acad Sci U S A 88:2307–2311
123. Gama-Castro S, Jiménez-Jacinto V, Peralta-Gil M et al (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res 36:D120–D124. doi:10.1093/nar/gkm994
124. Chen Y-J, Liu P, Nielsen AAK et al (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 10:659–664. doi:10.1038/nmeth.2515

125. Nag A, Narsinh K, Martinson HG (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat Struct Mol Biol 14:662–669. doi:10.1038/nsmb1253
126. Mairhofer J, Wittwer A, Cserjan-puschmann M, Striedner G (2014) Synthetic termination signal capable of improving bioprocess. ACS Synth Biol. doi:10.1021/sb5000115
127. Guo Z, Sherman F (1996) Signals sufficient for 3′-end formation of yeast mRNA. Mol Cell Biol 16:2772–2776
128. Geisberg JV, Moqtaderi Z, Fan X et al (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156:812–824. doi:10.1016/j.cell.2013.12.026
129. Curran KA, Morse NJ, Markham KA et al (2015) Short, synthetic terminators for improved heterologous gene expression in yeast. ACS Synth Biol 4(7):824–832. doi:10.1021/sb5003357
130. Mischo HE, Proudfoot NJ (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim Biophys Acta 1829:174–185. doi:10.1016/j.bbagrm.2012.10.003
131. Shalem O, Carey L, Zeevi D et al (2013) Measurements of the impact of 3′ end sequences on gene expression reveal wide range and sequence dependent effects. PLoS Comput Biol. doi:10.1371/journal.pcbi.1002934
132. Shalem O, Sharon E, Lubliner S et al (2015) Systematic dissection of the sequence determinants of gene 3′ end mediated expression control. PLoS Genet 11:e1005147. doi:10.1371/journal.pgen.1005147
133. Yager TD, von Hippel PH (1991) A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. Biochemistry 30:1097–1118
134. Cambray G, Guimaraes JC, Mutalik VK et al (2013) Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res 41:5139–5148. doi:10.1093/nar/gkt163
135. Leavitt JM, Alper HS (2015) Advances and current limitations in transcript-level control of gene expression. Curr Opin Biotechnol 34:98–104. doi:10.1016/j.copbio.2014.12.015

Adv Biochem Eng Biotechnol (2018) 162: 45–76
DOI: 10.1007/10_2016_9

Published online: 21 March 2016

Engineering Biomolecular Switches for Dynamic Metabolic Control

Cheng-Wei Ma, Li-Bang Zhou, and An-Ping Zeng

Abstract Living organisms have been exploited as production hosts for a large variety of compounds. To improve the efficiency of bioproduction, metabolic pathways in an organism are usually manipulated by various genetic modifications. However, bottlenecks during the conversion of substrate to a desired product may result from cellular regulations at different levels. Dynamic regulation of metabolic pathways according to the need of cultivation process is therefore essential for developing effective bioprocesses, but represents a major challenge in metabolic engineering and synthetic biology. To this end, switchable biomolecules which can sense the intracellular concentrations of metabolites with different response types and dynamic ranges are of great interest. This chapter summarizes recent progress in the development of biomolecular switches and their applications for improvement of bioproduction via dynamic control of metabolic fluxes. Further studies of bioswitches and their applications in industrial strain development are also discussed.

Keywords Biomolecular engineering, Bioswitches, Cellular regulation, Dynamic metabolic control, Strain development

Contents

C.-W. Ma, L.-B. Zhou, and A.-P. Zeng (✉)
Institute of Bioprocess and Biosystems Engineering, Hamburg University of Technology, Denickestrasse 15, 21073 Hamburg, Germany
e-mail: aze@tu-harburg.de

1 Introduction

Industrial biotechnology focuses on the development of living organisms, especially microorganisms, as production hosts for a large variety of compounds with potential usages in the chemical, food, pharmaceutical, agriculture, and health care industries. With the occurrence of synthetic biology, compounds that previously could not be synthesized by natural microorganisms can now be produced by combining metabolic pathways from different organisms into a single host [1–3]. However, although microorganisms have been developed and used to produce various chemicals and materials, their production efficiency is often not high enough to reduce the production cost to a competitive level with the traditional fossil-based routes. Approaches that can improve the production efficiency of organisms are therefore of great importance. As a widely used strategy, metabolic engineering has served as a powerful approach to overcome bottlenecks in the bioproduction processes with microbial hosts such as *Escherichia coli* and *Saccharomyces cerevisiae*.

Metabolic engineering has so far been successfully applied to construct efficient bioproduction strains by manipulating metabolic pathways in an organism. Generally speaking, strategies traditionally used in metabolic engineering to generate high producers [4] are (1) enhancement of precursor supply and the transport of both substrates and products by gene overexpression, (2) removal of enzymes involved in competing pathways and enzymes that may cause degradation of products by gene knockout, and (3) down-regulated gene expression of enzymes taking part in competing but essential pathways via promoter engineering. These strategies are “static” because of their permanent and unchangeable modifications at the genetic level [5]. As a consequence, these modifications cause unwanted burdens on the organisms when the cellular and environmental conditions are changed. In particular, the “static” strategies are not efficient to deal with competing pathways which are essential for the growth of cells, because there is no definite ratio between the production pathways and the competing pathways.

On the other hand, bottlenecks during the conversion of substrate to a desired product in microorganisms may result from cellular regulations at different levels. Improvement of production strains by manipulating the dynamics of metabolic

pathways and fluxes to account for changing cellular and environmental conditions is more desirable and challenging, whereas this cannot be realized using static control methods as found in most of the current metabolic engineering praxis. Moreover, synthesis of a desired product always involves multiple metabolic pathways and these pathways often exhibit specific properties suitable for production under distinct conditions and host organisms [6]. Concerted dynamic control thus arises when synergy is required among different metabolic pathways, i.e., regarding reducing equivalent demand, cofactor preferences, and intermediate utilization, etc. [7] Thus, dynamic metabolic control of related pathways, as a complementary strategy used in current metabolic engineering, may lead to an increased productivity and yield. This chapter summarizes recent progress in the development of biomolecular switches and their applications for improvement of bioproduction via dynamic metabolic control.

2 Natural Bioswitches

For dynamic control of metabolic fluxes, cellular entities or devices able to regulate metabolic activities by response to input signals are required and biomolecules that can fulfill such a requirement are called bioswitches. Because organisms are exposed to a variety of conditions in their environment, such as varying temperatures, availability of different nutrients, exposure to toxins, and products of their own metabolism, they need to be able to adjust rapidly to the changing conditions. Various bioswitches have thus been evolved in nature and discovered by researchers. They can be proteins that function in signaling pathways or participate in the transcription process, or allosteric enzymes that catalyze the metabolic reactions. They can also be non-coding RNAs such as riboswitches. These natural bioswitches can respond to various input signals and regulations can occur at different cellular levels (Table 1). Specifically, environmental signals are transduced by two-component regulatory systems and corresponding gene expressions are modulated at DNA level, whereas intracellular metabolites can be sensed by one-component regulatory systems at transcription level, by RNA-based bioswitches at either transcription or translation level, or even at protein level via allosteric enzymes.

For bioproduction of a particular compound encountered in industrial biotechnology, bioswitches able to bind at least two molecules are of particular interest. For these bioswitches, the function of the downstream target can be modulated by the binding of small molecules which are usually the intermediates involved in the biosynthetic pathways or the target product. In this chapter, only recent studies on bioswitches sensitive to small molecules and used for dynamic metabolic control are covered (Fig. 1). For discussion of biosensors usually coupled with reporter genes such as green fluorescent protein (GFP) for strain screening, readers are referred to other recent reviews (e.g., [8, 9]).

Table 1 Natural bioswitches that regulate cellular metabolism

Input signals	Bioswitches	Output regulations
• Light, temperature, pressure, et al. • Extracellular molecules (e.g., nutrients) • Intracellular molecules (e.g., metabolites)	• Signaling proteins (e.g., two-component regulatory system) • Transcription factors (e.g., one-component regulatory system) • RNA (e.g., riboswitches) • Allosteric enzymes	• Transcription • Translation • Metabolic reaction

2.1 DNA-Level Bioswitches

Regulation at DNA level is mainly mediated via transcription factors. Besides RNA polymerase, transcription factors include a wide number of proteins that play roles in initiating and regulating the transcription of genes. One distinct feature of transcription factors from other proteins is that they have DNA-binding domains and so are able to bind to a specific sequence of DNA. Regulation of transcription is the most common form for the control of gene expression, which allows for unique expression of each gene according to changing environments. By taking advantage of their DNA binding specificity, transcription factors can be employed as bioswitches to regulate metabolic fluxes once the expression of target enzymes are under control of a unique transcription factor and this strategy is facilitated by the diversity of transcription factors existing in nature (Table 2).

As a key mechanism to link environmental signals to cellular responses, two-component regulatory systems (Fig. 1a) enable organisms to sense, respond, and adapt to a wide range of environmental factors including nutrients, cellular redox state, changes in osmolarity, quorum signals, antibiotics, pH, and even physical factors such as light and temperature [10]. Although only a few two-component systems have been identified in eukaryotic organisms, they are widely distributed in prokaryotes. Some bacteria can contain as many as 200 - two-component systems to transfer different input signals to adequate outputs [11]. Different from two-component regulatory systems, a type of bacterial transcription regulators known as one-component regulatory systems consists of proteins that serve both as metabolite sensors and transcription regulators because they include both an "input domain" and an "output domain" in their structure (Fig. 1b). Because control of gene expression via one-component systems is more common and more diverse in bacteria and archaea than two-component systems of transcription regulation, it is speculated that one-component systems are evolved before two-component systems and may even have served as their evolutionary precursors [12]. The mechanisms of many families of one-component transcription regulation systems have been characterized on a structural level [13, 14]. One-component

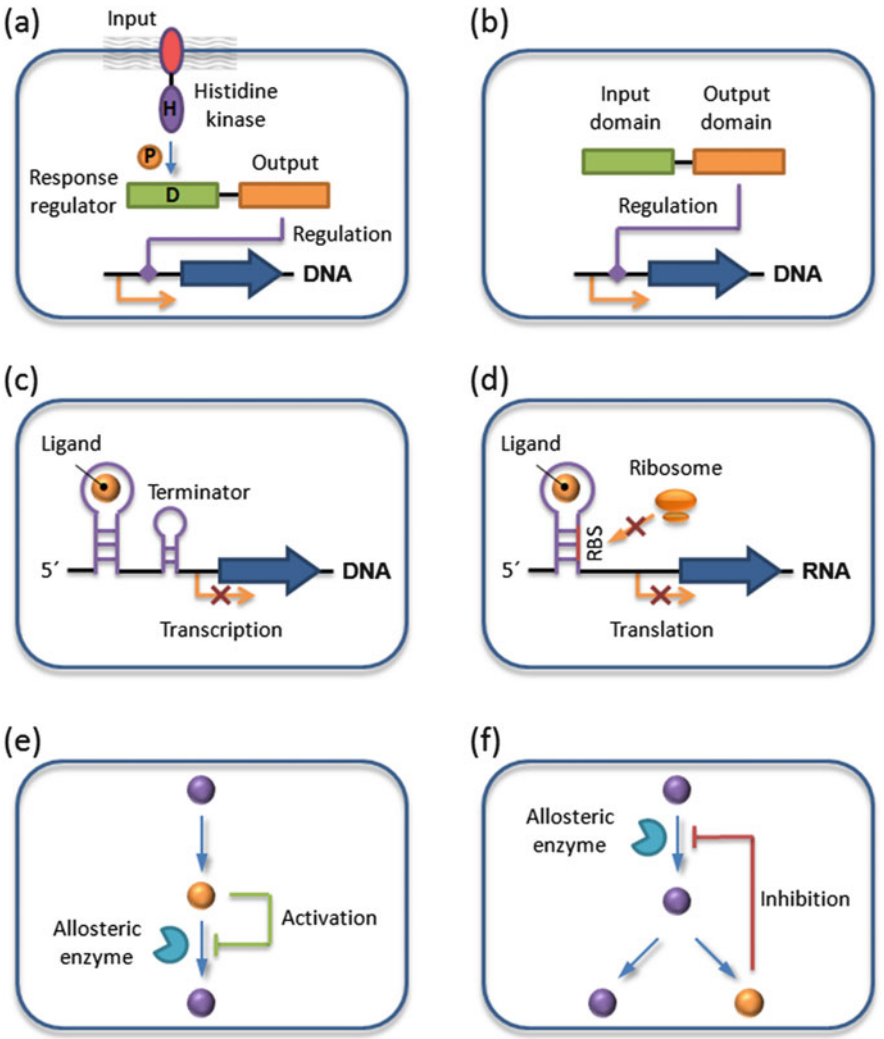

Fig. 1 Regulation of metabolic fluxes by bioswitches at different molecular levels. (**a**) Two-component regulatory system. *H* histidine; *D* aspartate acid; *P* phosphorization. (**b**) One-component regulatory system. (**c**) Riboswitches that function by stimulating transcription termination. (**d**) Riboswitches that function by interrupting translation initiation. (**e**) Allosteric enzymes with feedforward activation. (**f**) Allosteric enzymes with feedback inhibition

transcription regulation systems that have been identified in sequenced genomes can be assembled into families based on their sequence similarity, predominantly in their DNA-binding helix-turn-helix (HTH) domain. These families often include regulators with significant similarity in DNA-binding but divergent metabolite-sensing domains, and they tend to control expression of genes involved in related functions.

Table 2 Ligand-sensitive DNA regulators

Transcription factor	Source	Ligand	Refs.
PcaU	*Acinetobacter*	3,4-Dihydroxybenzoate	[80]
FapR	*B. subtilis*	Malonyl-CoA	[67, 81]
QdoR	*B. subtilis*	Kaempferol, quercetin	[82]
CysR	*C. glutamicum*	*O*-Acetyl (homo-) serine	[83]
Lrp	*C. glutamicum*	L-Valine, L-leucine, L-isoleucine, L-methionine	[84–86]
LysG	*C. glutamicum*	L-Lysine, L-arginine, L-histidine	[87, 88]
AraC	*E. coli*	Lycopene	[89]
DcuS	*E. coli*	Succinate	[90]
DcuR	*E. coli*	Succinate	[90]
FadR	*E. coli*	Fatty acyl-CoA	[65]
LacI	*E. coli*	IPTG, lactose	[91]
SoxR	*E. coli*	NADPH	[92]
TyrR	*E. coli*	L-Tyrosine	[89]
FdeR	*Herbaspirillum seropedicae*	Naringenin	[82]
BenR	*P. putida*	Benzoate	[93]
NahR	*P. putida*	Benzoic acids	[94]
PcaR	*P. putida*	β-Ketoadipate	[90]
BmoR	*Thauera butanivorans*	1-Butanol	[90]

The most abundant type of transcriptional regulator in the prokaryotic kingdom is the LysR family of transcriptional regulators [15]. The conserved overall structure includes an N-terminal DNA-binding domain, linked to two C-terminal effector-binding domains which are made of two α/β subdomains connected by two short polypeptide fragments. These connecting fragments form a hinge or cleft which accommodates the small molecule effector. Despite considerable conservation both structurally and functionally, LysR-type transcriptional regulators regulate a diverse set of genes, including primary and secondary metabolisms. The effector molecules have been identified for some of the LysR-type regulators. They include substrates, products, and intermediates of pathways under their control and related metabolites [16]. In particular, members of the LysR family play diverse roles in controlling amino acid biosynthesis. In addition to LysR, which acts as an activator in lysine biosynthesis, biosynthesis of glutamate can be activated by GltC. Biosynthesis of isoleucine/valine are under the control of IlvR and IlvY. The transport of arginine is regulated by ArgP. Besides these local regulators, some transcriptional regulators can function globally. For example, MetR is able to regulate biosynthesis/transport of both cysteine and methionine.

For discovery of transcription factor binding sites, chromatin immunoprecipitation (ChIP) is an important experimental technique. This can be used for studying interactions between specific proteins and DNA in the cell and determining their localization on a specific genomic locus. In recent years, the combination of ChIP

with the second generation DNA-sequencing technology (ChIP-seq) allows precise genomic functional assay, especially in genome-wide mapping of transcription factor binding sites, the revelation of underlying molecular mechanisms of differential gene regulation governed by specific transcription factors, and the identification of epigenetic marks [17]. For the analysis of ChIP-seq data, a novel approach called ChIPModule has been developed to discover systematically transcription factors and their cofactors [18]. Given a ChIP-seq dataset and the binding patterns of a large number of transcription factors, ChIPModule can efficiently identify groups of transcription factors whose binding sites significantly co-occur in the ChIP-seq peak regions. By testing ChIPModule on simulated data and experimental data, it was shown that ChIPModule can not only identify known cofactors of transcription factors but also predict new cofactors. Although the ChIP-seq experiments provide an unprecedented opportunity to discover binding motifs, which is important for the study of gene transcriptional regulation, de novo motif discovery methods often neglect underrepresented motifs in ChIP-seq peak regions. To address this issue, a novel approach called SIOMICS has been developed to discover motifs from ChIP-seq data [19, 20]. Compared with other methods for motif discovery, SIOMICS showed advantages in terms of speed, the number of known cofactor motifs predicted in experimental data sets, and the number of false motifs predicted in random data sets.

2.2 RNA-Level Bioswitches

In recent years it has become evident that posttranscriptional regulation mediated by RNA regulators is critical to many cellular processes in both prokaryotic and eukaryotic kingdoms [21]. Cells are able to use these small molecules to respond rapidly to various environmental signals and stresses. Among the RNA regulators, riboswitches are attracting great interest from researchers [22]. A riboswitch is a regulatory segment of a messenger RNA molecule which binds a small molecule, resulting in a change in production of the proteins encoded by the mRNA in response to the concentration of its effector molecule [23, 24].

Riboswitches are composed of an aptamer domain and an expression platform. Aptamers are short nucleic acid sequences capable of binding specific ligands with high affinity and specificity. Upon binding of a small molecule, the structure of the expression platform changes in response to the alteration of the aptamer. Because of the diversity of expression platforms, the same type of riboswitches may be capable of regulating gene expression at different levels. Riboswitch control of transcriptional termination exists most commonly in Gram-positive bacteria (Fig. 1c) whereas riboswitch control of translational initiation is often found in most Gram-negative bacteria (Fig. 1d). A variety of metabolite-binding riboswitches has been discovered and characterized [23, 25]. As listed in Table 3, the ligands range in complexity from metal ions to enzymatic cofactors such as flavin mononucleotide (FMN), *S*-adenosyl methionine (SAM), and coenzyme B_{12}.

Table 3 Classification and characterization of riboswitches

Riboswitch	Ligand	Length[a]	Affinity[b]	Refs.
Enzymatic cofactor				
TPP	TPP	100–120	100 nM	[95, 96]
FMN	FMN	120–140	5 nM	[97]
B_{12}	AdoCbl	200–220	300 nM	[98, 99]
SAH	SAH	65–80	20 nM	[100, 101]
SAM-I	SAM	100	4 nM	[102]
SAM-II	SAM	60	1 μM	[63]
SAM-III	SAM	80	/[c]	[103]
SAM-IV	SAM	60	15 μM	[104]
Moco	Moco	140	/	[105]
THF	THF	100–120	70 nM	[106]
Amino acid				
Glycine	Glycine	100–120	30 μM	[107]
Lysine	Lysine	165–190	1 μM	[108]
Glutamine	Glutamine	60–80	150 μM	[109]
GlcN6P	GlcN6P	170	200 μM	[110, 111]
Nucleotide				
Adenine	Adenine	70	300 nM	[112, 113]
Guanine	Guanine	70	5 nM	[113, 114]
dG	dG	70	80 nM	[114]
$preQ_1$-I	$preQ_1$	40	50 nM	[115]
$preQ_1$-II	$preQ_1$	25–45	100 nM	[115, 116]
c-di-GMP-I	c-di-GMP	110	1 nM	[117]
c-di-GMP-II	c-di-GMP	90	200 pM	[118]
Ion				
Mg^{2+}	Magnesium	70	200 μM	[119]
F	Fluoride	110	60 μM	[120, 121]

Abbreviations: *TPP* thiamine pyrophosphate, *FMN* flavin mononucleotide, *B_{12}* coenzyme B_{12}, *AdoCbl* adenosyl-cobalamine, *SAH* *S*-adenosylhomocysteine, *SAM* *S*-adenosylmethionine, *Moco* molybdenum cofactor, *THF* tetrahydrofolate, *GlcN6P* glucosamine-6-phosphate, *dG* deoxyguanosine, *preQ1* prequeuosine, *c-di-GMP* cyclic dimeric guanosine monophosphate

[a]The size of aptamer sequence

[b]The binding affinity is given according to the results of corresponding publications

[c]Not detected

Riboswitch as a *cis*-acting regulatory molecule has potential applications in dynamic control of metabolic pathways. This is dependent on several factors. First, the extreme structural flexibility of RNA aptamers enables highly specific recognition of a wide range of regulatory signals. A second factor is that riboswitches are RNA-derived regulatory molecules. This means the engineering work on riboswitches is relatively easy compared to that on the protein level. A third key factor is the feasibility of integration of riboswitches with different target genes, which allows one type of riboswitch to impact expression of genes located at different

pathways simultaneously. All these factors make riboswitches ideal genetic devices for realization of dynamic metabolic control, especially in cases where concerted regulation of multiple pathways is required.

2.3 Protein-Level Bioswitches

Compared with regulation at DNA and RNA levels, the control of metabolic fluxes through allosteric enzymes is realized without the processes of transcription and translation [5]. The main feature of allosteric enzymes is that they possess at least two stereospecifically distinct ligand binding sites: the active site where the substrate binds and the allosteric site where an allosteric effector binds. The binding of regulatory molecules at the allosteric site results in the modification of properties of the distinct active site. For example, the apparent change in binding affinity at the active site may result in either an increase of enzyme activity in the case of the binding of an activator or decrease of enzyme activity in the case of the binding of an inhibitor.

For metabolic control, allosteric regulations conducted by allosteric enzymes are natural examples of control loops, including both feedforward from upstream substrates (Fig. 1e) and feedback from downstream products (Fig. 1f). This is found in the negative feedback loops of many biosynthetic pathways where one of the products of the pathway inhibits further production of the product by closing down an enzyme involved in one of early steps of the pathway. Alternatively, a pathway can be activated by the presence of a specific molecule which switches on one of its crucial enzymes. In general, the first enzyme or a key branch point of a pathway is down-regulated by the pathway's product. The examples listed in Table 4 are collected from *E. coli* and categorized according to their functions in different metabolic modules. It can be seen that there are both feedforward activation and feedback inhibition during the generation of precursor metabolites and energy. Feedback inhibition is the main mechanism used in the biosynthetic pathways of amino acids whereas feedforward activation plays key roles in the biosynthesis of nucleosides and nucleotides.

Among the regulatory domains existing in allosteric proteins, the ACT domain is a motif that was first identified as a regulatory module in a number of diverse proteins. The name originates from three of the proteins in the domain family: aspartokinase, chorismate mutase, and TyrA (prephenate dehydrogenase). It is a structural motif in proteins of 70–80 amino acids. The archetypical ACT domain is composed of four β strands and two α helices arranged in a βαββαβ fold [26]. It is one of a growing number of different intracellular small molecule binding domains that function in the control of metabolism, solute transport, and signal transduction. Particularly, most of the proteins containing the ACT domain are found to be involved in amino acid and purine synthesis and in many cases they are allosteric enzymes regulated by the binding of ligands [27]. For instance, the archetypical ACT domain protein *E. coli* D-3-phosphoglycerate dehydrogenase (3-PGDH)

Table 4 Allosteric enzymes discovered in *E. coli*

Enzyme	Gene	Activator(s)	Inhibitor(s)
Generation of precursor metabolites and energy			
6-Phosphofructokinase I	*pfkA*	GDP, ADP	Phosphoenolpyruvate
6-Phosphofructokinase II	*pfkB*		ATP
Pyruvate kinase I	*pykF*	Fructose-1,6-bisphosphate	
Pyruvate kinase II	*pykA*	AMP	
Citrate synthase	*gltA*	Acetyl-CoA	NAD+, oxaloacetate, NADH
Phosphoenolpyruvate carboxylase	*ppc*	A long-chain fatty acid, GTP, fructose-1,6-bisphosphate, acetyl-CoA	(S)-Malate, L-aspartate
Amino acids biosynthesis			
Aspartate kinase III	*lysC*		L-Lysine
Dihydrodipicolinate synthase	*dapA*		L-Lysine
Homoserine O-succinyltransferase	*metA*		*S*-Adenosyl-L-methionine, L-methionine
Chorismate mutase	*pheA*		L-Phenylalanine
Prephenate dehydratase	*pheA*		L-Phenylalanine
D-3-phosphoglycerate dehydrogenase	*serA*		Glycine, L-serine
Carbamoyl phosphate synthetase	*carB*, *carA*	Ammonium, inosine-5′-phosphate, L-ornithine	Uridine-5′-phosphate
ATP phosphoribosyltransferase	*hisG*		L-Histidine
γ-Glutamyl kinase	*proB*		L-Proline
Nucleosides and nucleotides biosynthesis			
Amidophosphoribosyl transferase	*purF*		Guanosine-5′-phosphate, AMP
Aspartate carbamoyltransferase	*pyrI*, *pyrB*	ATP	CTP
Uridylate kinase	*pyrH*	GTP	
CTP synthetase	*pyrG*	GTP	
Uridylate kinase	*pyrH*	GTP	

Data collected from Ecocyc (http://ecocyc.org/)

catalyzes the first step in the biosynthesis of serine and its activity is regulated by the binding of glycine and serine. Aspartokinase III from *E. coli* is the first and key switch of pathways for the synthesis of aspartate-derived amino acids and it is inhibited by its end product lysine. The bifunctional chorismate mutase/prephenate dehydratase (P-protein) from *E. coli* catalyzes the first two steps in the biosynthesis of phenylalanine and its function is inhibited by the binding of phenylalanine.

Although it is well known that the majority of proteins bind specific metabolites and that such interactions are relevant to metabolic and gene regulation, there are so far no efficient methods to identify functional allosteric protein-metabolite

interactions systematically. Based on dynamic metabolite data, an integrated approach combining both experiments and computations has recently been presented for discovery of allosteric regulations relevant in vivo [28]. In this approach, the culture conditions of *E. coli* were switched every 30 s between media containing either pyruvate or ^{13}C-labeled fructose or glucose. The reversal of flux through glycolysis pathways was observed and the rapid changes in metabolite concentration were measured. Then these data were fitted to a kinetic model of glycolysis and the consequences of 126 putative allosteric interactions on metabolite dynamics were systematically tested. As a result, allosteric interactions governing the reversible switch between gluconeogenesis and glycolysis were identified, including one through which pyruvate activates fructose-1,6-bisphosphatase. It has been shown that this approach can identify the most likely interactions and provide hypotheses about their function from large sets of putative allosteric interactions.

3 Engineering of Bioswitches

Bioswitches that can sense different signals are interesting biological components with potential usage for realization of dynamic control of metabolic fluxes. Although a variety of natural bioswitches has been discovered so far, bioswitches, especially those that can respond to metabolites involved in the target biosynthetic pathways, are still needed in the practice. For example, the construction of artificial bioswitches that can respond to non-natural signals is both challenging and highly desirable for a precise and dynamic control of fluxes of growth-essential but competing pathways in metabolic engineering of industrial microorganisms. From the perspective of engineering, both the DNA-level and protein-level bioswitches are protein based although they function at different molecular levels. Thus, they are classified into the group of protein-based bioswitches in this section whereas bioswitches that function at the RNA level as described above belong to the group of RNA-based bioswitches.

3.1 Engineering of Protein-Based Bioswitches

As seen from the natural bioswitches, allosteric regulation is used as a very efficient mechanism to control metabolism in most biological processes. It is an important mechanism to maintain metabolic fluxes and limit accumulation of metabolic intermediates by binding effector molecules which are considered to function in a purely structural manner by selectively stabilizing a specific conformational state [29]. Understanding the mechanisms of allosteric regulation, especially the pathways that mediate signal transduction from the allosteric site to the active site upon effector binding, can provide useful information for engineering bioswitches with

novel properties. However, proteins are inherently dynamic molecules which undergo structural fluctuations over a wide range of timescales. A thorough knowledge of the principles governing protein dynamics is therefore of fundamental importance for functional study and design of new protein functions.

Computational Modeling of Allosteric Regulation

Rapid advances have been made during the past few years in the investigation of protein dynamics. Besides experimental approaches, such as NMR relaxation, ultra-high resolution low-temperature X-ray crystallography, and ultra-fast laser technologies, computational tools such as molecular dynamics simulations of protein dynamics and allostery offer the opportunity to explore mechanistic details that are difficult to observe experimentally. There are two key challenges in the computational modeling of allostery. One is to predict the structure of one allosteric state starting from the structure of the other and the transition states between or to sample conformational states existing in the allosteric ensemble. The other is to elucidate the mechanisms underlying the conformational coupling of the effector and active sites and to identify residues that mediate the allosteric process. In practice, these challenges can be overcome by developments of novel modeling approaches and computational procedures.

Prediction of Transition States

Characterization of the conformational states of allosteric proteins requires access to long-time-scale motions, currently inaccessible by standard molecular dynamics simulations. In addition, large-scale conformational changes in proteins involve barrier-crossing transitions on the complex free energy surfaces of high-dimensional space. Such rare events cannot be efficiently captured by conventional molecular dynamics simulations. Special computational approaches are therefore needed to explore protein dynamics underlying allosteric regulation. To this end, advanced accelerated molecular dynamics approaches that extend the effective simulation time and capture large-scale motions of functional relevance have been explored and were employed to investigate the conformational changes associated with substrate binding to *Trypanosoma cruzi* proline racemase enzyme (TcPR), which are believed to expose critical residues that elicit a host mitogenic B-cell response [30]. Potential conformational epitopes located in the vicinity of newly identified transient binding pockets were also illustrated by subsequent conservation and fragment mapping analyses. To characterize the free energy profile of a conformational transition pathway in a high-dimensional space, the on-the-fly string method and the multi-state Bennett acceptance ratio (MBAR) method were combined by Matsunaga and coworkers [31]. In the study of *E. coli* adenylate kinase, the minimum free energy paths of the conformational transitions were explored by the on-the-fly string method in 20-dimensional space spanned by

the 20 largest-amplitude principal modes. Moreover, evaluation of the free energy and various kinds of average physical quantities along the pathways are also possible with this combined approach.

Challenges encountered in the study of long-time-scale motions can also be solved by simplified modeling approaches such as the normal mode model and a combination of network construction with coarse-grained model. To study the rigor to post-rigor transition in myosin, a consequence of ATP binding, a normal mode superposition model has been developed to predict the transition path between the two states obtained from the X-ray structures [32]. It was shown that rigid-body motions of the various subdomains and specific residues at the subdomain interfaces are key elements in the transition. The allosteric communication between the nucleotide binding sites resulted from local changes upon ligand binding, and this induced large amplitude motions in the structure of the protein. It is hypothesized that allosteric communication in proteins relies upon networks of quaternary (collective, rigid-body) and tertiary (residue–residue contact) motions, and cyclic topology of these networks is necessary for allosteric communication. To prove this, a novel procedure was proposed by Daily and Gray [33]. In this procedure, rigid bodies were first identified from the displacement between the inactive and the active structures and "quaternary networks" were constructed from these rigid bodies. Finally, "global communication networks" were formed by integrating quaternary networks with a coarse-grained representation of contact rearrangements.

Elucidation of Signal Transduction Pathways

To discover signal transduction pathways that mediate the allosteric communication and key residues involved in the allosteric process, varied approaches and algorithms have been reported, most of which are based on the results of molecular dynamics simulations. For example, an interaction-correlation analysis of the trajectories obtained from molecular dynamics simulations has been proposed and applied to the PDZ2 domain to identify the possible signal transduction pathways [34]. In this approach, a residue correlation matrix is constructed from the interaction energy correlations between all residue pairs along the trajectories of the simulations. With the residue correlation matrix, it is possible to discover continuous interaction pathways by a hierarchical clustering analysis as well as the energetic origin of the long-range coupling associated with allosteric regulation. In another study by Ma and coworkers [35] to reveal the anticooperative mechanism of PII protein from *Synechococcus elongatus* upon binding of 2-oxoglutarate, the binding pocket size was first defined by identifying residues that contributed greatly to the ligand binding. It was then found that the anticooperativity was realized through population shift of the binding pocket size in an asymmetric manner. Based on dynamic correlation analysis, a new algorithm was developed and utilized to discover residues that mediated the anticooperative process with high probability. Chen et al. [36, 37] took aspartokinase, an important allosteric

enzyme for industrial amino acids production, as a model system, and a predictive approach combining protein dynamics and evolution was demonstrated for rational reengineering of enzyme allostery. In this method, molecular dynamic simulations of aspartokinase and statistical coupling analysis of protein sequences of the aspartokinase family were combined to identify a cluster of residues which are correlated during protein motion and coupled during the evolution. This cluster of residues was believed to form an interconnected network that mediated the allosteric regulation. Experimental verifications with mutations of the key residues demonstrated the high efficiency and reliability of the combined approach for deregulation of aspartokinase from both *E. coli* and *Corynebacterium glutamicum*.

To get more insight into how intramolecular communication occurs within an allosteric protein, a perturbation response scanning method has been developed. The key of this method is that it couples elastic network models with linear response theory to predict critical residues in allosteric transitions [38]. In the study of PDZ domain, it was found that the residues with the highest mean square fluctuation response upon perturbing the binding sites agreed well with experimentally determined residues involved in allosteric transitions. Allosteric pathways can then be constructed by linking the residues giving the same directional response upon perturbation of the binding sites. The idea of perturbation has also been used in the energy dissipation model of allosteric regulation proposed by Ma et al. [39]. With *E. coli* aspartokinase III as a model system, a novel approach to reveal the intramolecular signal transduction network was developed based on the energy dissipation model [40]. A key feature of this approach is that direction information is specified after inferring the protein residue–residue interaction network involved in the process of signal transduction. This enables fundamental analysis of the regulation hierarchy and identification of regulation hubs of the signaling network. The energy dissipation model and network construction method have also been successfully applied to a heteromultimeric allosteric protein, *C. glutamicum* aspartokinase, to explore the signal transduction involved in intersubunit interactions and allosteric communication with emphasis on the intersubunit signaling process [41].

Strategies for Engineering Protein-Based Bioswitches

For developments of new protein-based bioswitches, different strategies have been reported (Fig. 2). They can be engineering the binding pocket of an existing allosteric protein, fusing a naturally existing allosteric domain to the protein of interest, or directly modifying the structure of a non-allosteric enzyme.

Engineering Bioswitches for New Ligand Binding

Because there are at least two distinct binding sites in a typical bioswitch (one is the effector binding site responsible for the recognition of signal molecules and the

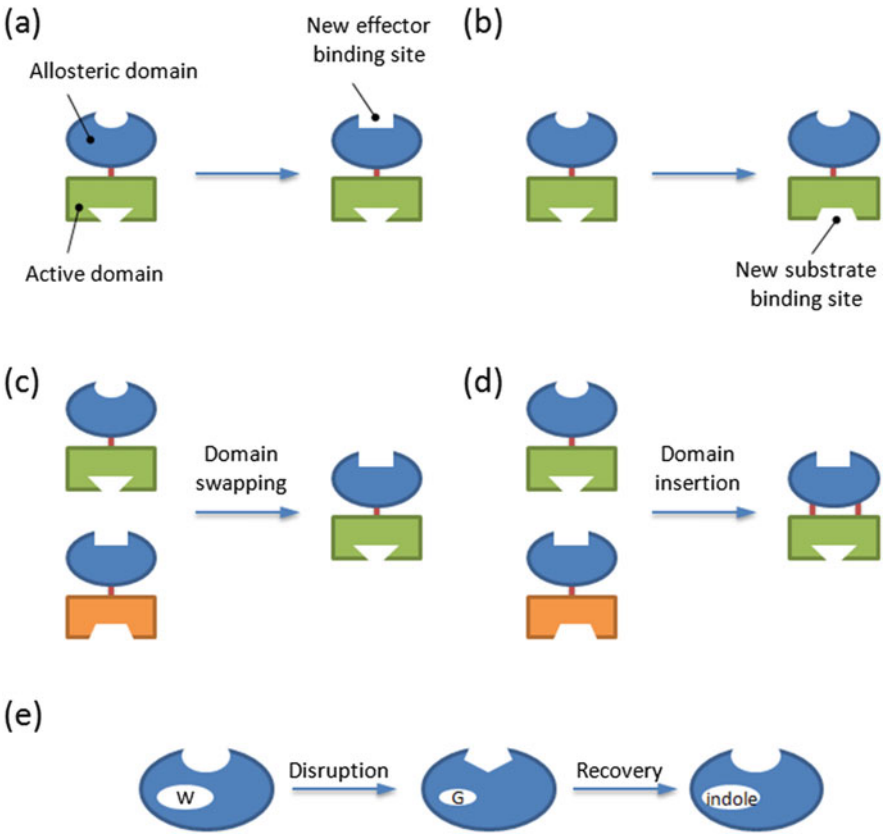

Fig. 2 Different strategies used for engineering of protein-based bioswitches. (**a**) Engineering bioswitches for new effector binding. (**b**) Engineering bioswitches for new substrate binding. (**c**) Engineering of bioswitches via domain swapping. (**d**) Engineering of bioswitches via domain insertion. (**e**) Creation of de novo bioswitches. *W* tryptophan; *G* glycine. Adapted from Deckert et al. [50]

other is the active site responsible for the binding of modulated molecules), the most straightforward approach to construct novel bioswitches is to engineer the binding site for new ligands based on existing bioswitches. Both the method of rational design and the approach of directed evolution or their combination can be employed to obtain new binding sites. For engineering protein-based bioswitches, there are many individual experiences from protein engineering which have shown great success in practice. Nevertheless, the challenge of this strategy is to keep the allosteric function of the bioswitches after the modification.

In an example by Tang and Cirino [42], the AraC regulatory protein from the *E. coli* ara operon was engineered to activate transcription in response to D-arabinose and not in response to its native effector, L-arabinose. To achieve this, two different AraC mutant libraries, each with four randomized binding pocket

residues, were subjected to fluorescence-activated cell sorting (FACS)-mediated dual screening using a GFP reporter. Both libraries yielded mutants with the desired switch in effector specificity, and one mutant was found to maintain tight repression in the absence of effector. This example demonstrated the power of dual screening for altering the ligand binding specificity of a protein inducer and represents steps toward the design of customized in vivo molecular reporters and genetic switches for metabolic engineering.

In another study by Chen et al. [43], homoserine dehydrogenase (HSDH) of *C. glutamicum*, which is naturally allosterically regulated by L-threonine and L-isoleucine, was used as an example to demonstrate the feasibility of reengineering an allosteric enzyme to respond to a non-natural inhibitor L-lysine. To this end, the natural L-threonine binding sites of HSDH were first predicted and verified by mutagenesis experiments. Then the L-threonine binding sites were engineered to an L-lysine binding pocket by replacing a key loop responsible for the ligand binding specificity, which resulted in a reengineered HSDH which only responded to L-lysine inhibition but not to L-threonine. Because the L-threonine biosynthetic pathway is essential for cell growth and its formation is competing with the biosynthesis of L-lysine, this study represents a significant step toward the construction of artificial molecular circuits for dynamic control of the growth-essential byproduct formation pathway for L-lysine biosynthesis.

Engineering of Bioswitches via Domain Fusion

This strategy is based on the fact that the structures of proteins are often organized in functional domains and this is also true for allosteric proteins. In allosteric proteins, the signal recognition function is conducted by the regulatory domain and the active domain is a DNA binding domain in transcription factors or a catalytic domain in the case of allosteric enzymes. Thus, an efficient approach to construct protein-based bioswitches is to create hybrid proteins with switch-like behavior via domain fusion of two existing domains. For this purpose, the domain containing the function to be modulated is fused with the signal recognition domain. The challenge of this strategy is that the fusion should be conducted in a proper manner so that the input signal can be transmitted from the regulatory domain to the active domain, thereby modulating its activity. In practice, two approaches are usually employed to link the two functional domains. One is domain swapping which just connects the terminals of the two domains to create the recombinant allosteric protein. The other is domain insertion which involves circular permutation of one domain [44].

The strategy of domain swapping has recently been demonstrated with 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAHPS), the first enzyme of the aromatic amino acid biosynthesis. DAHPS shows remarkable variation in allosteric response and machinery and both contemporary regulated and unregulated orthologs have been reported. A chimeric protein was generated by joining the catalytic domain of an unregulated DAHPS with the regulatory domain of a

regulated enzyme [45]. It was shown that this simple gene fusion event on its own was sufficient to confer functional allostery to the unregulated enzyme. The fusion protein shared structural similarities with its regulated parent protein and underwent an analogous major conformational change in response to the binding of allosteric effector L-tyrosine to the regulatory domain. Domain swapping can also be used to create novel bioswitches by connecting domains from heterologous proteins. This has been illustrated by recombining the genes coding for TEM1β-lactamase (BLA) and the *E. coli* maltose binding protein (MBP) to create a family of MBP–BLA hybrids in which maltose was a positive or negative effector of β-lactam hydrolysis [46]. Some of the constructed MBP–BLA switches were effectively "on–off" in nature, with maltose altering catalytic activity by as much as 600-fold.

The strategy of domain insertion is to engineer regulatory activities into proteins through interface design at conserved allosteric sites by creating a chimeric protein. A hybrid protein named PAS-DHFR has been constructed by connecting a light-sensing signaling domain from a plant member of the Per/Arnt/Sim (PAS) family of proteins with the dihydrofolate reductase (DHFR) from *E. coli* [47]. With no optimization, the hybrid protein exhibited light-dependent catalytic activity which depended on the site of connection and on known signaling mechanisms in both proteins. This example demonstrated that the intramolecular networks of two proteins can be joined across their surface sites such that the activity of one protein can be controlled by the activity of the other. In a recent study, a protein with a unique topology, called uniRapR, was constructed with the aid of computational protein design [48]. The key feature of this chimeric protein is that its conformation is controlled by the binding of a small molecule and the conformational change can be used as an artificial regulatory domain to control activity of kinases. To prove this, activation of Src kinase using uniRapR was demonstrated in both single cells and whole organisms. The rational creation of uniRapR not only offers a powerful means for targeted activation of many pathways to study signaling in living organisms but also exemplifies the strength of computational protein design. The more recent work by Feng et al. [49] attempts to provide a general methodology to develop biosensors for a broad range of molecules in eukaryotes. In this method, the ligand-binding domain is fused to either a fluorescent protein or a transcriptional activator and the key feature is that the protein is destabilized by mutation so that the fusion accumulates only in cells containing the target ligand. When this method was employed to develop biosensors for digoxin and progesterone, it was found that transcription was activated with a dynamic range of up to ~100-fold upon addition of ligand to the cells.

Creation of De Novo Bioswitches

Disruption and recovery of protein structure may represent a general technique for introducing allosteric control into proteins, and thus serves as a starting point to build a variety of protein-based bioswitches. This strategy has recently been demonstrated by designing a de novo allosteric effector site directly into the

catalytic domain of an enzyme and it is distinct from traditional chemical rescue of enzymes in that it relies on disruption and restoration of structure, rather than active site chemistry, as a means to achieve modulate function. In the two examples given by Deckert and coworkers [50], W33G in a β-glycosidase enzyme and W492G in a β-glucuronidase enzyme, indole-dependent activities were engineered into enzymes by removing a buried tryptophan side chain which served as a buttress for the active site architecture. In both cases the loss of function can be restored by the subsequent addition of indole. In particular, the rescued β-glycosidase was fully functionally equivalent to the corresponding wild-type enzyme and its activity can be modulated in living cells using indole as an input signal.

3.2 Engineering of RNA-Based Bioswitches

RNAs are ideal for the design of gene switches that can monitor and program cellular behavior. Because of the modular composition of riboswitches, engineered riboswitches can be made by first exploiting RNA aptamers as core component and then combining different aptamers and expression platforms. Novel riboswitches can also be constructed and identified through model-driven approaches or even rationally designed via structure-based methods (Fig. 3).

In Vitro Selection Technology

Engineered riboswitches can be made by exploiting RNA aptamers as the core component. They can be generated by an in vitro directed evolution technology called SELEX (systematic evolution of ligands by exponential enrichment). In this approach, a pool of randomized sequences is mixed with an immobilized target. Non-binding molecules are removed by washing whereas bound molecules are specifically eluted, amplified, and subjected to further rounds of selection. Gradually increasing the stringency during the following cycles can lead to aptamers that bind with affinities in the picomolar range and discriminate between closely related compounds. Aptamers against a plethora of different ligands have been generated including ions, organic compounds such as amino acids or antibiotics, proteins, viruses, and even whole cells [51].

However, most naturally found riboswitches down-regulate gene expression on metabolite binding, probably because of their roles in negative feedback regulation within the metabolic pathways. Selections of gene switches have been performed by alternately employing separate ON (positive) and OFF (negative) selection markers. The use of independent selection markers for the ON and OFF selections significantly complicates the selection process by requiring plasmid isolation steps and increases the chances of isolating false positives in each step. To overcome this, an efficient platform to select engineered riboswitches and logic gates from complex libraries using a single selection marker has already been established by taking

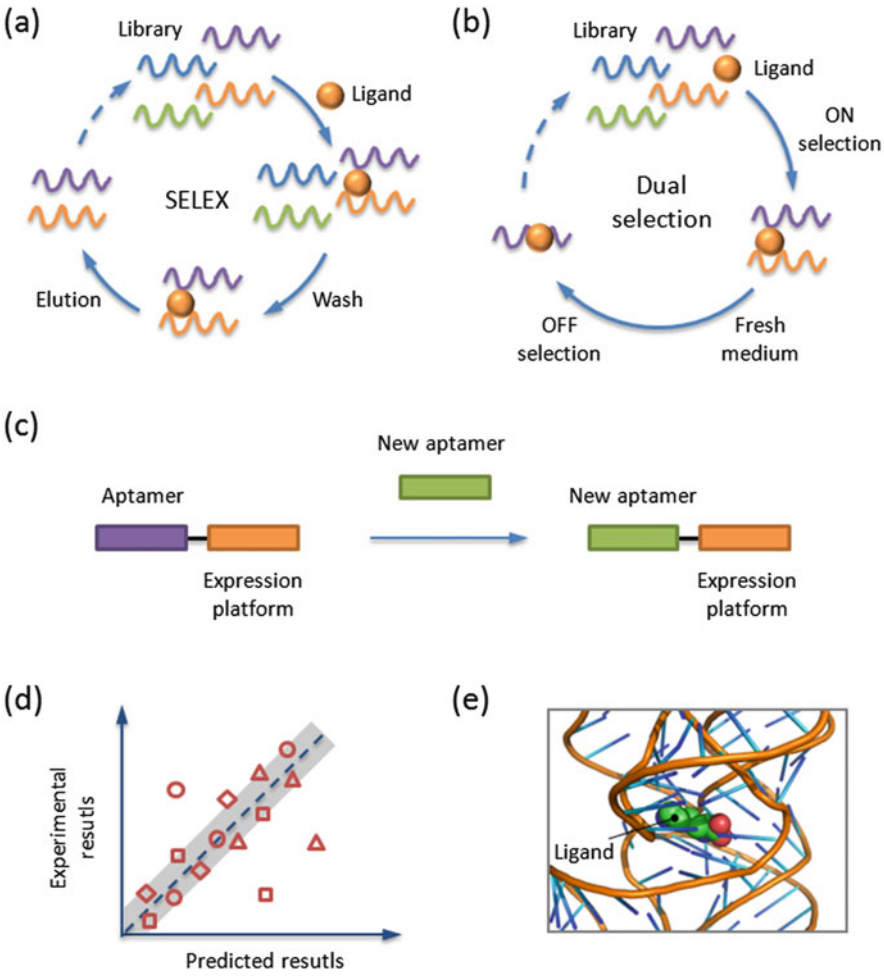

Fig. 3 Different approaches for engineering of riboswitches. (**a**) In vitro selection of new aptamers. (**b**) In vitro dual selection of riboswitches. (**c**) Module-based construction of riboswitches with new aptamer. (**d**) Investigation of riboswitches with mathematical models. (**e**) Structure-based design of riboswitches

advantage of tetA, which encodes a tetracycline/H+ antiporter as both a positive and a negative selection marker [52]. Expression of TetA confers tetracycline resistance on the cells (ON selection) whereas the overexpression of the membrane-bound protein renders them more sensitive to toxic metal salts such as $NiCl_2$ and other compounds (OFF selection). Use of a single selection marker for ON and OFF selections not only simplifies the selection procedure but also makes the process more robust against false positives. With the dual selection approach, a lysine ON riboswitch was recently successfully obtained from the lysine OFF riboswitch and used for improving L-lysine bioproduction [53].

Although in vitro selection technology is a versatile experimental tool for the discovery of novel RNA molecules, finding complex RNA molecules is difficult because most RNAs identified from random sequence pools are simple motifs. Thus, enriching in vitro selection pools with complex structures could increase the probability of discovering novel RNAs. Recently, a computational approach was presented for designing a starting library of RNA sequences with increased formation of complex structural motifs and enhanced affinity to a desired target molecule [54]. This approach consists of two steps: (1) generation of RNA sequences based on customized patterning of nucleotides with increased probability of forming a base pair and development of a set of criteria used for selection of a sequence with potential binding affinity; (2) with a protocol for RNA 3D structure prediction, a high-throughput virtual screening of the generated library is carried out to select aptamers with binding affinity to a small-molecule target. With integration of in vitro selection technology, this approach is expected to accelerate the experimental screening and selection of high-affinity aptamers by significantly reducing the search space of RNA sequences.

Module-Based Construction

Although in vitro selection technology is efficient to explore aptamers, it has been pointed out that only a few aptamers have the potential to be exploited as sensing domains for the engineering of riboswitches because a conformational change upon ligand binding has to occur and the association of the ligand has to be fast [55]. It has been demonstrated that riboswitches are modular in that they can host a variety of natural and synthetic aptamers to create novel chimeric RNAs that regulate transcription both in vitro and in vivo [56]. Modularity of riboswitches therefore enables facile engineering of novel genetic regulatory devices from aptamers. Moreover, this technique does not require selection of device-specific "communication modules" required to transmit ligand binding to the regulatory domain, enabling rapid engineering of novel functional RNAs. With the module-based approach, it has been proved that transcriptional "ON" riboswitches are also capable of hosting foreign aptamers [57].

The design criteria for synthetic riboswitches acting on transcription have recently been examined by Wachsmuth et al. [58] using theophylline-dependent riboswitches as model systems. It was shown that terminator hairpin stability and folding traps had a major impact on the functionality of the designed constructs. Furthermore, a combination of several copies of individual riboswitches led to a much improved activation ratio between induced and uninduced gene activity and to a linear dose-dependent increase in reporter gene expression. By taking advantage of the modularity of riboswitches, novel riboswitches that work in a eukaryotic cell-free translation system has also be constructed [59]. In these riboswitches, translation mediated by an internal ribosome entry site was promoted only in the presence of a specific ligand, whereas it was inhibited in the absence of the ligand. The riboswitch, which was regulated by theophylline, showed a high switching

efficiency and dependency on theophylline. In addition, another three kinds of riboswitches controlled by FMN, tetracycline, and sulforhodamine B were constructed only by calculating the ΔG value of one stem-loop structure.

Computational Approaches

It is known that the function of riboswitches can be modulated through sequence alteration, but there are no quantitative frameworks to investigate or guide riboswitch tuning. It remains unclear how their sequence controls the physics of riboswitch switching and activation, particularly when changing the ligand-binding aptamer domain. To this end, mathematical modeling was combined with experimental approaches to investigate the relationship between riboswitch function and their performance [60]. Modeling results showed that the competition between reversible and irreversible rate constants dictated the performance for different regulatory mechanisms. It was also found that practical system restrictions, such as an upper limit on ligand concentration, can significantly alter the requirements for riboswitch performance, necessitating alternative tuning strategies. In another study, a statistical thermodynamic model was reported to predict the sequence-structure-function relationship for translation-regulating riboswitches that activate gene expression, characterized inside cells and within cell-free transcription-translation assays [61]. With this model, automated computational design was carried out for 62 synthetic riboswitches that used 6 different RNA aptamers to sense diverse chemicals (theophylline, tetramethylrosamine, fluoride, dopamine, thyroxine, 2,4-dinitrotoluene) and activated gene expression by up to 383-fold.

Because the three-dimensional structures are available for a growing subset of RNAs, structure-based techniques have been employed to study the mechanism of riboswitch and to guide the prediction and design for specific functions and new characteristics. For instance, guided by 3D structures, Wilson-Mitchell et al. [62] examined the recognition and specificity mechanisms of lysine riboswitches. In another report, structure-based design approach was combined with a fluorescence binding assay for development of SAM-II riboswitch aptamer and identification of a SAM analogue that selectively binds to SAM-II riboswitch aptamer with comparable binding affinity to its native metabolite [63]. In an attempt to de novo design a synthetic riboswitch that regulates gene expression at the transcriptional level, an in silico pipeline was developed to design the actuator part as RNA sequences that can fold into functional intrinsic terminator structures [64]. Using the well-characterized theophylline aptamer as sensor, several of the designed constructs showed ligand-dependent control of gene expression in *E. coli*, demonstrating that it is possible to engineer riboswitches not only for translational but also for transcriptional regulation.

4 Applications of Bioswitches for Dynamic Metabolic Control

Metabolic burden and imbalance caused by uncontrolled or deregulated metabolic pathways result in suboptimal productivity. Applications of bioswitches for dynamic control of metabolism can prevent the accumulation of temporarily unnecessary intermediates produced by heterologous pathways by fine-tuning the metabolic fluxes. Redirection of the endogenous resources into heterologous pathways can be further tuned by dynamic control of competing pathways. In addition, modulation of bacterial behavior by manipulating molecular communication finds use in a variety of applications, particularly those employing natural or synthetic bacterial consortia (Fig. 4).

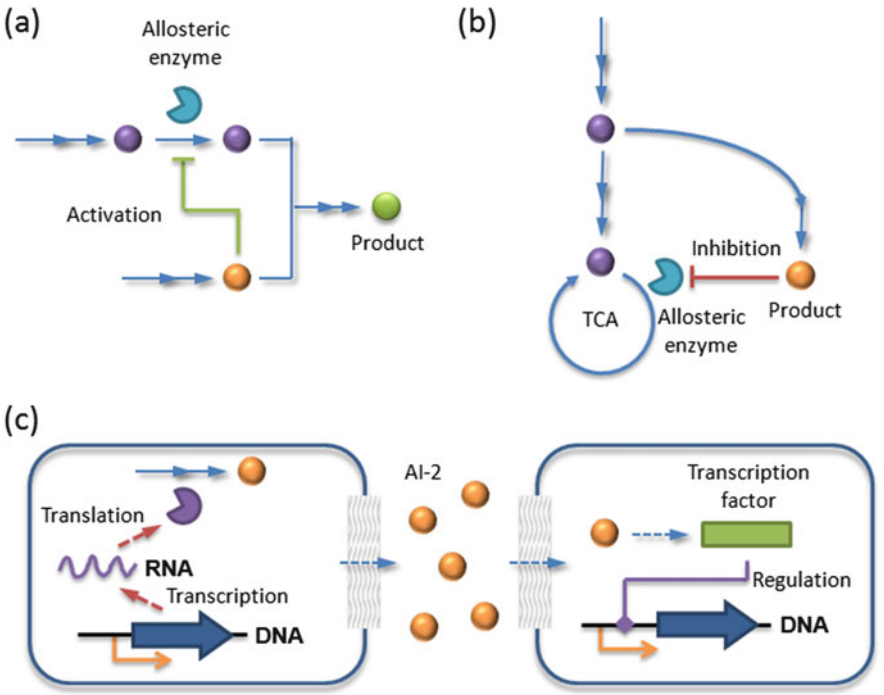

Fig. 4 Application examples of bioswitches for dynamic metabolic control. (**a**) Dynamic control of bioproduction pathways. (**b**) Dynamic control of competing pathways. (**c**) Dynamic control of cell–cell communications

4.1 Dynamic Control of Bioproduction Pathways

To demonstrate that product titers and conversion yields of heterologous pathways can be improved by dynamic control of bioproduction pathways, a dynamic sensor-regulator system was developed by Zhang et al. [65] to produce fatty acid-based products in *E. coli* for biodiesel production. In this dynamic system, the transcription factor FadR which senses fatty acyl-CoA was employed and the expression of genes involved in biodiesel production were dynamically regulated. With this implementation the stability of biodiesel-producing strains can be substantially improved. Additionally, the titer was increased to 1.5 g/L and the yield threefold to 28% of the theoretical maximum.

For fatty acid biosynthesis, the formation of malonyl-CoA, which is biosynthesized from acetyl-CoA by the acetyl-CoA carboxylase, is the rate-limiting step. However, overexpression of acetyl-CoA carboxylase improves fatty acid production, but the cell growth is negatively influenced. This is expected to be solved by dynamical compensation of the critical enzymes involved in the supply and consumption of malonyl-CoA for efficient redirection of carbon flux toward fatty acids biosynthesis. As shown in the study by Xu et al. [66], implementation of this metabolic control resulted in an oscillatory malonyl-CoA pattern and a balanced metabolism between cell growth and product formation, yielding 15.7- and 2.1-fold improvement in fatty acids titer compared with the wild-type strain and the strain carrying the uncontrolled metabolic pathway. Recently, another malonyl-CoA sensor-actuator that controls gene expression levels based on intracellular malonyl-CoA concentrations was devised [67]. With this sensor-actuator, the expression of acetyl-CoA carboxylase can be negatively controlled, which means that the expression of acetyl-CoA carboxylase is able to be up-regulated when the malonyl-CoA concentration is low, and the expression is down-regulated when excess amounts of malonyl-CoA are accumulated. It was shown that the toxicity associated with acetyl-CoA carboxylase overexpression can be effectively alleviated by the regulatory circuit. When the feedback circuit was used to regulate the fatty acid pathway, the fatty acid titer and productivity were increased by 34% and 33%, respectively. The malonyl-CoA sensor can also be used in the production of other malonyl-CoA-derived products. In the work by David et al. [68], a hierarchical dynamic control system is developed around the key pathway intermediate malonyl-CoA. The upper level of the control system ensures down-regulation of the endogenous use of malonyl-CoA for fatty acid biosynthesis whereas the lower level of the control system is based on the use of a novel biosensor for malonyl-CoA to activate expression of a heterologous pathway that uses this metabolite for production of 3-hydroxypropionic acid (3-HP). It was shown that the production of 3-HP was increased by tenfold after introduction of the dual pathway control.

Dynamic control of pathway enzymes requires sensors that can detect and respond to pathway products or intermediates, but these are largely unknown. In a recent attempt to improve the production of non-native isoprenoids, dynamic control of ERG9 expression was explored by using different ergosterol-responsive

promoters [69]. For this purpose, several ergosterol-responsive promoters were identified using quantitative real-time PCR analysis in an engineered strain with relatively high mevalonate pathway activity. It was found that the expression levels for ERG11, ERG2, and ERG3 were significantly lower in the engineered strain over the reference strain, indicating that these genes were transcriptionally down-regulated when ergosterol was in excess. Replacement of the native ERG9 promoter with these ergosterol-responsive promoters revealed that all engineered strains improved amorpha-4,11-diene by two- to fivefold over the reference strain with ERG9 under its native promoter. Promoters that respond to the accumulation of toxic intermediates can also be identified via whole-genome transcript arrays and used to improve the final titers of a desired product by controlling accumulation of the intermediate. This approach was recently demonstrated by regulating farnesyl pyrophosphate production in the isoprenoid biosynthetic pathway in *E. coli* [70]. It was shown that this strategy was able to improve production of amorphadiene, the final product, by twofold over that from inducible or constitutive promoters with reduced acetate accumulation and improved growth.

4.2 Dynamic Control of Competing Pathways

Dynamical tuning of endogenous processes is another efficient approach for redirection of endogenous resources into heterologous pathways. In the study given by Solomon et al. [71], a metabolite valve was proposed to balance the demands of cell health and pathway. To realize it, a control node of glucose utilization, glucokinase (Glk), was exogenously manipulated through either engineered antisense RNA or an inverting gene circuit. Results showed that these techniques were able to control glycolytic flux directly by redirection of glucose into a model pathway, leading to an increase in the pathway yield and reduced carbon waste to acetate. Moreover, the specific growth rate of engineered *E. coli* can be reduced by up to 50% without altering final biomass accumulation. The same strategy was then employed to develop a metabolite valve in *S. cerevisiae* for control of glycolytic flux through the central carbon metabolism [72]. This was demonstrated by diverting glucose flux away from glycolysis into a model pathway, gluconate, in a hexokinase 2 and glucokinase 1 deleted strain. A maximum tenfold decrease in hexokinase activity was achieved by controlling the transcription of hexokinase 1 with the tetracycline transactivator protein, resulting in a 50-fold increase in gluconate yields, from 0.7% to 36% mol/mol of glucose. The reduction in glucose flux also led to a significant decrease in ethanol by production that extended to semianaerobic conditions shown in the production of isobutanol. It is worth noting that these applications involve control of a production pathway by external supplementation of inducers/repressors, which are different from the endogenous dynamic regulation as illustrated in the other applications.

The deletion of a pathway responsible for growth and cell maintenance has seldom been employed, as conditional knockout is required to optimize

intracellular metabolism at each fermentation phase for bacterial growth and production. In this regard, a metabolic toggle switch was constructed in *E. coli* as a novel conditional knockout approach and applied in isopropanol production [73]. The resulting redirection of excess carbon flux caused by interruption of the TCA cycle via switching gltA OFF improved isopropanol production titer and yield up to 3.7 and 3.1 times, respectively. To control the competing but essential metabolic by-pathways of lysine biosynthesis, similar strategy was employed to control the TCA cycle activity by using lysine riboswitch with intracellular L-lysine as a signal [74]. Lysine riboswitches from both *E. coli* (ECRS) and *Bacillus subtilis* (BSRS) were used to control the *gltA* gene and thus the TCA cycle activity in a lysine-producing strain *C. glutamicum* LP917. Compared with the strain LP917, the lysine production was 63% higher in the mutant ECRS-*gltA* and 38% higher in the mutant BSRS-*gltA*, indicating a higher metabolic flux into the lysine synthesis pathway. A lysine-ON riboswitch library was constructed using *tetA*-based dual genetic selection based on the natural *E. coli* lysine-OFF riboswitch. Selected lysine-ON riboswitches were linked with the *lysE* gene to achieve a dynamic control of lysine transport in a recombinant lysine-producing strain, *C. glutamicum* LPECRS, which bears a deregulated aspartokinase and a lysine-OFF riboswitch controlled citrate synthase. Batch fermentation results showed that, with the additional control of *lysE* by a lysine-ON riboswitch, the strain achieved an increase in yield by 21% compared to that of the strain *C. glutamicum* LPECRS, and the concerted control by both OFF and ON type lysine riboswitches led to an increase in yield by 89% compared to that of the strain embedded with only deregulated aspartokinase [53].

4.3 Dynamic Control of Cell–Cell Communications

Coordination between cell populations via prevailing metabolic cues has been noted as a promising approach to connect synthetic devices and drive phenotypic or product outcomes. To demonstrate this, "controller cells" have been developed by manipulating the molecular connection between cells via modulating the bacterial signal molecule, autoinducer-2 (AI-2), which is secreted as a quorum-sensing signal in many bacterial species [75]. Specifically, *E. coli* was engineered to overexpress components responsible for autoinducer uptake (*lsrACDB*), phosphorylation (*lsrK*), and degradation (*lsrFG*). To characterize the dynamic balance among the various uptake mechanisms, a simple mathematical model was established by recapitulating experimental data. Two controller "knobs" were found to affect the increase of AI-2 uptake. One is the overexpression of the AI-2 transporter, *LsrACDB*, which controls removal of extracellular AI-2. The other is the overexpression of the AI-2 kinase, *LsrK*, which increases the net uptake rate by limiting secretion of AI-2 back into the extracellular environment.

With the quorum sensing system for cell density-dependent regulation of gene expression, a self-induced metabolic state switching was developed for microbial

isopropanol production [76]. To this end, a synthetic quorum sensing system was constructed using a synthetic *lux* promoter and a positive feedback loop and used as a tunable cell density sensor-regulator in *E. coli*. In this system, self-induction of a target gene expression is driven by quorum-sensing signals, and its threshold cell density can be changed depending on the concentration of a chemical inducer. This study demonstrated that auto-redirection of metabolic flux from central metabolic pathways toward a synthetic isopropanol pathway at a desired cell population led to a significant increase in isopropanol production.

5 Perspectives

Bioswitches are of great interest for the development of industrial strains with productivity high enough to compete with traditional chemical routes, especially when heterologous pathways are integrated into the host cells. Allosteric regulation and riboswitches, as the fundamental mechanisms in biology to control cellular metabolism and gene regulation, provide a variety of candidates with potential applications in dynamic control of metabolic fluxes. However, the dynamic response range of natural bioswitches needs to be engineered so that they can be used in industrial strain development where the effector concentration is usually much higher than that in a non-producer. Meanwhile, concerted dynamic regulation of metabolic pathways according to the need of cultivation process is often necessary for developing effective bioproduction strains and processes. Switchable regulating biomolecules that can sense the intracellular concentration of metabolites with different response types and dynamic ranges are required to enable concerted dynamic control of multiple pathways. In addition, novel bioswitches that can respond to non-natural effectors are demanded to fulfill the diversity of regulatory targets encountered in metabolic engineering.

As illustrated above, engineering work can be conducted on different types of bioswitches to obtain satisfactory characteristics. In the case where allosteric enzymes are used as bioswitches, the engineering work has to be carried out at the protein level, which could be challenging in some cases because of the complicated relationship between protein structure and its function. In addition, modifications on the protein level are often not feasible to be transferred among different allosteric proteins and thus their applications for concerted control of multiple pathways are limited. As for protein-based DNA regulators, besides modifications of the effector binding site at the protein level, their DNA binding specificity also has to be engineered to avoid the cross-regulation of unexpected genes, which may involve engineering at both the DNA and the protein level. From the perspective of engineering, the response profile of RNA-level bioswitches can be easily engineered and its applications in dynamic control of multiple metabolic pathways are expected to attract more attention.

In practice, rational approaches are desired for more efficient engineering of bioswitches although some of the engineering work can be accomplished by

random mutagenesis and in vivo/in vitro screening. In this regard, computational modeling with novel algorithms continue to contribute to the understanding of the underlying mechanisms and to developing bioswitches with novel properties. If necessary, they can be further improved by directed evolution methods [77]. On the other hand, to guide the design and optimization of bioswitches, techniques such as fast sampling and integrated on-line fast cell separation and quenching [78] can be used to access the intracellular metabolites concentrations more accurately. As for the regulatory points that should be controlled by the embedded bioswitches, they can be identified by using approaches such as time-resolved ^{13}C-labeled metabolic flux analysis [79], which can also be utilized to investigate the regulatory effects of these dynamic controls.

References

1. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD (2015) Complete biosynthesis of opioids in yeast. Science 349(6252):1095–1100
2. Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol 12(5):355–367
3. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7(7):445–452
4. Lee SY, Kim HU (2015) Systems strategies for developing industrial microbial strains. Nat Biotechnol 33(10):1061–1072
5. Holtz WJ, Keasling JD (2010) Engineering static and dynamic control of synthetic pathways. Cell 140:19–23
6. Shen CR, Liao JC (2013) Synergy as design principle for metabolic engineering of 1-propanol production in *Escherichia coli*. Metab Eng 17:12–22
7. Nielsen J, Fussenegger M, Keasling J, Lee SY, Liao JC, Prather K, Palsson B (2014) Engineering synergy in biotechnology. Nat Chem Biol 10:319–322
8. Eggeling L, Bott M, Marienhagen J (2015) Novel screening methods--biosensors. Curr Opin Biotechnol 35:30–36
9. Zhang J, Jensen MK, Keasling JD (2015) Development of biosensors and their application in metabolic engineering. Curr Opin Chem Biol 28:1–8
10. Casino P, Rubio V, Marina A (2010) The mechanism of signal transduction by two-component systems. Curr Opin Struct Biol 20:763–771
11. Laub MT, Goulian M (2007) Specificity in two-component signal transduction pathways. Annu Rev Genet 41:121–145
12. Ulrich LE, Koonin EV, Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 13:52–56
13. Hou C, Tsodikov OV (2015) Structural basis for dimerization and DNA binding of transcription factor FLI1. Biochemistry 54(50):7365–7374
14. Makthal N, Rastegari S, Sanson M, Ma Z, Olsen RJ, Helmann JD, Musser JM, Kumaraswami M (2013) Crystal structure of peroxide stress regulator from Streptococcus pyogenes provides functional insights into the mechanism of oxidative stress sensing. J Biol Chem 288 (25):18311–18324
15. Knapp GS, Hu JC (2010) Specificity of the *E. coli* LysR-type transcriptional regulators. PLoS One 5(12):e15189

16. Maddocks SE, Oyston PC (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology 154(Pt 12):3609–3623
17. Mundade R, Ozer HG, Wei H, Prabhu L, Lu T (2014) Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle 13(18):2847–2852
18. Ding J, Cai X, Wang Y, Hu H, Li X (2013) ChIPModule: systematic discovery of transcription factors and their cofactors from ChIP-seq data. Pac Symp Biocomput 2013:320–331
19. Ding J, Hu H, Li X (2014) SIOMICS: a novel approach for systematic identification of motifs in ChIP-seq data. Nucleic Acids Res 42(5):e35
20. Ding J, Dhillon V, Li X, Hu H (2015) Systematic discovery of cofactor motifs from ChIP-seq data by SIOMICS. Methods 79–80:47–51
21. Oliva G, Sahr T, Buchrieser C (2015) Small RNAs, 5′ UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. FEMS Microbiol Rev 39(3):331–349
22. Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152(1-2):17–24
23. Breaker RR (2012) Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 4(2): pii: a003566
24. Fürtig B, Nozinovic S, Reining A, Schwalbe H (2015) Multiple conformational states of riboswitches fine-tune gene regulation. Curr Opin Struct Biol 30:112–124
25. Mellin JR, Cossart P (2015) Unexpected versatility in bacterial riboswitches. Trends Genet 31(3):150–156
26. Grant GA (2006) The ACT domain: a small molecule binding domain and its role as a common regulatory element. J Biol Chem 281(45):33825–33829
27. Liberles JS, Thórólfsson M, Martínez A (2005) Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. Amino Acids 28(1):1–12
28. Link H, Kochanowski K, Sauer U (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nat Biotechnol 31(4):357–361
29. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339
30. de Oliveira CA, Grant BJ, Zhou M, McCammon JA (2011) Large-scale conformational changes of *Trypanosoma cruzi* proline racemase predicted by accelerated molecular dynamics simulation. PLoS Comput Biol 7:e1002178
31. Matsunaga Y, Fujisaki H, Terada T, Furuta T, Moritsugu K, Kidera A (2012) Minimum free energy path of ligand-induced transition in adenylate kinase. PLoS Comput Biol 8:e1002555
32. Cecchini M, Houdusse A, Karplus M (2008) Allosteric communication in myosin V: from small conformational changes to large directed movements. PLoS Comput Biol 4:e1000129
33. Daily MD, Gray JJ (2009) Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. PLoS Comput Biol 5:e1000293
34. Kong Y, Karplus M (2009) Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 74:145–154
35. Ma CW, Lüddecke J, Forchhammer K, Zeng AP (2014) Population shift of binding pocket size and dynamic correlation analysis shed new light on the anticooperative mechanism of PII protein. Proteins 82:1048–1059
36. Chen Z, Rappert S, Sun J, Zeng AP (2011) Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production. J Biotechnol 154(4):248–254
37. Chen Z, Bommareddy RR, Frank D, Rappert S, Zeng AP (2014) Deregulation of feedback inhibition of phosphoenolpyruvate carboxylase for improved lysine production in *Corynebacterium glutamicum*. Appl Environ Microbiol 80(4):1388–1393
38. Gerek ZN, Ozkan SB (2011) Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. PLoS Comput Biol 7:e1002154
39. Ma CW, Xiu ZL, Zeng AP (2011) A new concept to reveal protein dynamics based on energy dissipation. PLoS One 6:e26453

40. Ma CW, Xiu ZL, Zeng AP (2012) Discovery of intramolecular signal transduction network based on a new protein dynamics model of energy dissipation. PLoS One 7:e31529
41. Ma CW, Xiu ZL, Zeng AP (2015) Exploring signal transduction in heteromultimeric protein based on energy dissipation model. J Biomol Struct Dyn 33:134–146
42. Tang SY, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. Angew Chem Int Ed Engl 50:1084–1086
43. Chen Z, Rappert S, Zeng AP (2015) Rational design of allosteric regulation of homoserine dehydrogenase by a nonnatural inhibitor L-lysine. ACS Synth Biol 4:126–131
44. Yu Y, Lutz S (2011) Circular permutation: a different way to engineer enzyme structure and function. Trends Biotechnol 29:18–25
45. Cross PJ, Allison TM, Dobson RC, Jameson GB, Parker EJ (2013) Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. Proc Natl Acad Sci U S A 110:2111–2116
46. Guntas G, Mansell TJ, Kim JR, Ostermeier M (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci U S A 102:11224–11229
47. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R (2008) Surface sites for engineering allosteric control in proteins. Science 322:438–442
48. Dagliyan O, Shirvanyants D, Karginov AV, Ding F, Fee L, Chandrasekaran SN, Freisinger CM, Smolen GA, Huttenlocher A, Hahn KM, Dokholyan NV (2013) Rational design of a ligand-controlled protein conformational switch. Proc Natl Acad Sci U S A 110:6800–6804
49. Feng J, Jester BW, Tinberg CE, Mandell DJ, Antunes MS, Chari R, Morey KJ, Rios X, Medford JI, Church GM, Fields S, Baker D (2015) A general strategy to construct small molecule biosensors in eukaryotes. Elife 4:pii:e10606
50. Deckert K, Budiardjo SJ, Brunner LC, Lovell S, Karanicolas J (2012) Designing allosteric control into enzymes by chemical rescue of structure. J Am Chem Soc 134:10055–10060
51. Sinha J, Topp S, Gallivan JP (2011) From SELEX to cell dual selections for synthetic riboswitches. Methods Enzymol 497:207–220
52. Nomura Y, Yokobayashi Y (2014) Dual genetic selection of synthetic riboswitches in *Escherichia coli*. Methods Mol Biol 1111:131–140
53. Zhou LB, Zeng AP (2015) Engineering a lysine-ON riboswitch for metabolic control of lysine production in *Corynebacterium glutamicum*. ACS Synth Biol 4(12):1335–1340
54. Chushak YG, Martin JA, Chávez JL, Kelley-Loughnane N, Stone MO (2014) Computational design of RNA libraries for in vitro selection of aptamers. Methods Mol Biol 1111:1–15
55. Wittmann A, Suess B (2012) Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. FEBS Lett 586(15):2076–2083
56. Ceres P, Garst AD, Marcano-Velázquez JG, Batey RT (2013) Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth Biol 2(8):463–472
57. Ceres P, Trausch JJ, Batey RT (2013) Engineering modular 'ON' RNA switches using biological components. Nucleic Acids Res 41(22):10449–10461
58. Wachsmuth M, Domin G, Lorenz R, Serfling R, Findeiß S, Stadler PF, Mörl M (2015) Design criteria for synthetic riboswitches acting on transcription. RNA Biol 12(2):221–231
59. Ogawa A (2011) Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. RNA 17(3):478–488
60. Beisel CL, Smolke CD (2009) Design principles for riboswitch function. PLoS Comput Biol 5(4):e1000363
61. Espah Borujeni A, Mishler DM, Wang J, Huso W, Salis HM (2016) Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. Nucleic Acids Res 44(1):1–13
62. Wilson-Mitchell SN, Grundy FJ, Henkin TM (2012) Analysis of lysine recognition and specificity of the Bacillus subtilis L box riboswitch. Nucleic Acids Res 40:5706–5717

63. Ham YW, Humphreys DJ, Choi S, Dayton DL (2011) Rational design of SAM analogues targeting SAM-II riboswitch aptamer. Bioorg Med Chem Lett 21(17):5071–5074
64. Wachsmuth M, Findeiß S, Weissheimer N, Stadler PF, Mörl M (2013) De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res 41 (4):2541–2551
65. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotechnol 30(4):354–359
66. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proc Natl Acad Sci U S A 111(31):11299–11304
67. Liu D, Xiao Y, Evans BS, Zhang F (2015) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. ACS Synth Biol 4(2):132–140
68. David F, Nielsen J, Siewers V (2016) Flux control at the malonyl-CoA node through hierarchical dynamic pathway regulation in *Saccharomyces cerevisiae*. ACS Synth Biol. doi:10.1021/acssynbio.5b00161
69. Yuan J, Ching CB (2015) Dynamic control of ERG9 expression for improved amorpha-4,11-diene production in *Saccharomyces cerevisiae*. Microb Cell Fact 14:38
70. Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD, Keasling JD (2013) Engineering dynamic pathway regulation using stress-response promoters. Nat Biotechnol 31(11):1039–1046
71. Solomon KV, Sanders TM, Prather KL (2012) A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng 14(6):661–671
72. Tan SZ, Manchester S, Prather KL (2016) Controlling central carbon metabolism for improved pathway yields in *Saccharomyces cerevisiae*. ACS Synth Biol. doi:10.1021/acssynbio.5b00164
73. Soma Y, Tsuruno K, Wada M, Yokota A, Hanai T (2014) Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. Metab Eng 23:175–184
74. Zhou LB, Zeng AP (2015) Exploring lysine riboswitch for metabolic flux control and improvement of L-lysine synthesis in *Corynebacterium glutamicum*. ACS Synth Biol 4 (6):729–734
75. Zargar A, Quan DN, Emamian M, Tsao CY, Wu HC, Virgile CR, Bentley WE (2015) Rational design of 'controller cells' to manipulate protein and phenotype expression. Metab Eng 30:61–68
76. Soma Y, Hanai T (2015) Self-induced metabolic state switching by a tunable cell density sensor for microbial isopropanol production. Metab Eng 30:7–15
77. Packer MS, Liu DR (2015) Methods for the directed evolution of proteins. Nat Rev Genet 16 (7):379–394
78. da Luz JA, Hans E, Zeng AP (2014) Automated fast filtration and on-filter quenching improve the intracellular metabolite analysis of microorganisms. Eng Life Sci 14:135–142
79. Guo W, Sheng J, Feng X (2016) ^{13}C-Metabolic flux analysis: an accurate approach to demystify microbial metabolism for biochemical production. Bioengineering 3(1):3
80. Jha RK, Kern TL, Fox DT, M Strauss CE (2014) Engineering an Acinetobacter regulon for biosensing and high-throughput enzyme screening in *E. coli* via flow cytometry. Nucleic Acids Res 42(12):8150–8160
81. Xu P, Wang W, Li L, Bhan N, Zhang F, Koffas MA (2014) Design and kinetic analysis of a hybrid promoter-regulator system for malonyl-CoA sensing in *Escherichia coli*. ACS Chem Biol 9(2):451–458
82. Siedler S, Stahlhut SG, Malla S, Maury J, Neves AR (2014) Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. Metab Eng 21:2–8

83. Hoffmann K, Grünberger A, Lausberg F, Bott M, Eggeling L (2013) Visualization of imbalances in sulfur assimilation and synthesis of sulfur-containing amino acids at the single-cell level. Appl Environ Microbiol 79(21):6730–6736
84. Mustafi N, Grünberger A, Kohlheyer D, Bott M, Frunzke J (2012) The development and application of a single-cell biosensor for the detection of L-methionine and branched-chain amino acids. Metab Eng 14(4):449–457
85. Mustafi N, Grünberger A, Mahr R, Helfrich S, Nöh K, Blombach B, Kohlheyer D, Frunzke J (2014) Application of a genetically encoded biosensor for live cell imaging of L-valine production in pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* strains. PLoS One 9(1):e85731
86. Mahr R, Gätgens C, Gätgens J, Polen T, Kalinowski J, Frunzke J (2015) Biosensor-driven adaptive evolution of L-valine production of *Corynebacterium glutamicum*. Metab Eng 32:184–194
87. Binder S, Schendzielorz G, Stäbler N, Krumbach K, Hoffmann K, Bott M, Eggeling L (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. Genome Biol 13(5):R40
88. Schendzielorz G, Dippong M, Grünberger A, Kohlheyer D, Yoshida A, Binder S, Nishiyama C, Nishiyama M, Bott M, Eggeling L (2014) Taking control over control: use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways. ACS Synth Biol 3(1):21–29
89. Chou HH, Keasling JD (2013) Programming adaptive control to evolve increased metabolite production. Nat Commun 4:2595
90. Dietrich JA, Shis DL, Alikhani A, Keasling JD (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. ACS Synth Biol 2(1):47–58
91. Kiviet DJ, Nghe P, Walker N, Boulineau S, Sunderlikova V, Tans SJ (2014) Stochasticity of metabolism and growth at the single-cell level. Nature 514(7522):376–379
92. Siedler S, Schendzielorz G, Binder S, Eggeling L, Bringer S, Bott M (2014) SoxR as a single-cell biosensor for NADPH-consuming enzymes in *Escherichia coli*. ACS Synth Biol 3 (1):41–47
93. Uchiyama T, Miyazaki K (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomics libraries for enzyme-encoding genes. Appl Environ Microbiol 76(21):7029–7035
94. Van Sint FS, van Beilen JB, Witholt B (2006) Selection of biocatalysts for chemical synthesis. Proc Natl Acad Sci U S A 103(6):1693–1698
95. Miranda-Rios J, Navarro M, Soberon M (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc Natl Acad Sci U S A 98:9736–9741
96. Winkler WC, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419:952–956
97. Winkler WC, Cohen-Chalamish S, Breaker RR (2002) An mRNA structure that controls gene expression by binding FMN. Proc Natl Acad Sci U S A 99:15908–15913
98. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002) Genetic control by a metabolite binding mRNA. Chem Biol 9:1043
99. Nou X, Kadner RJ (2000) Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. Proc Natl Acad Sci U S A 97:7190–7195
100. Wang JX, Breaker RR (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. Biochem Cell Biol 86:157–168
101. Wang JX, Lee ER, Morales DR, Lim J, Breaker RR (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol Cell 29:691–702
102. Fuchs RT, Grundy FJ, Henkin TM (2006) The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. Nat Struct Mol Biol 13:226–233

103. Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A (2008) Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. Nat Struct Mol Biol 15:1076–1083
104. Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, Ruzzo WL, Breaker RR (2008) The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. RNA 14:822–828
105. Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL, Breaker RR (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. Mol Microbiol 68:918–932
106. Ames TD, Rodionov DA, Weinberg Z, Breaker RR (2010) A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. Chem Biol 17:681–685
107. Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306:275–279
108. Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev 17:2688–2697
109. Ames TD, Breaker RR (2011) Bacterial aptamers that selectively bind glutamine. RNA Biol 8:82–89
110. Roth A, Nahvi A, Lee M, Jona I, Breaker RR (2006) Characteristics of the *glmS* ribozyme suggest only structural roles for divalent metal ions. RNA 12:607–619
111. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428:281–286
112. Mandal M, Breaker RR (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol 11:29–35
113. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chem Biol 11:1729–1741
114. Kim JN, Roth A, Breaker RR (2007) Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2′-deoxyguanosine. Proc Natl Acad Sci U S A 104:16092–16097
115. Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R, Iwata-Reuyl D, Breaker RR (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat Struct Mol Biol 14:308–317
116. McCown PJ, Liang JJ, Weinberg Z, Breaker RR (2014) Structural, functional, and taxonomic diversity of three preQ1 riboswitch classes. Chem Biol 21:880–889
117. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science 321:411–413
118. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science 329:845–848
119. Ramesh A, Winkler WC (2010) Magnesium-sensing riboswitches in bacteria. RNA Biol 7:77–83
120. Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. Science 335:233–235
121. Li S, Breaker RR (2012) Fluoride enhances the activity of fungicides that destabilize cell membranes. Bioorg Med Chem Lett 22:3317–3322

Adv Biochem Eng Biotechnol (2018) 162: 77–116
DOI: 10.1007/10_2016_12

Published online: 16 September 2016

Pathway Design, Engineering, and Optimization

Eva Garcia-Ruiz, Mohammad HamediRad, and Huimin Zhao

Abstract The microbial metabolic versatility found in nature has inspired scientists to create microorganisms capable of producing value-added compounds. Many endeavors have been made to transfer and/or combine pathways, existing or even engineered enzymes with new function to tractable microorganisms to generate new metabolic routes for drug, biofuel, and specialty chemical production. However, the success of these pathways can be impeded by different complications from an inherent failure of the pathway to cell perturbations. Pursuing ways to overcome these shortcomings, a wide variety of strategies have been developed. This chapter will review the computational algorithms and experimental tools used to design efficient metabolic routes, and construct and optimize biochemical pathways to produce chemicals of high interest.

Keywords Cell factories, DNA assembly, Metabolic engineering, Pathway construction, Pathway design, Pathway optimization, Synthetic biology

Contents

E. Garcia-Ruiz and M. HamediRad
Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

H. Zhao (✉)
Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: zhao5@illinois.edu

1 Introduction

During the last few decades, intensive exploitation of natural sources and increasing concerns on environmental pollution have motivated a growing interest in developing sustainable processes to produce fuels, commodity chemicals, and natural products [1, 2]. Microorganisms have emerged as suitable platforms for sustainable, environmentally friendly, and cost-effective processes to produce a whole range of compounds [1, 3–5]. In nature, microorganisms have exhibited a wide metabolic versatility, allowing them to produce a variety of chemicals. This ability has been exploited by the scientific community to develop microbial cell factories to synthesize desired chemicals.

In some cases, the chemical of interest is an endogenous metabolite and can be produced in the original organism. However, native pathways are usually tightly regulated and do not fulfill industrial productivity expectations. Therefore, overproduction of the desired compound can be achieved by metabolic engineering of the native host by, for example, channeling cellular fluxes toward the desired pathway or modulating cellular regulatory networks. In other cases, natural pathways or synthetic pathways combining enzymes from different organisms or even new enzymes can be inserted in a more suitable heterologous host to produce the chemical of interest. Nevertheless, multi-enzymatic pathways from different species may not function optimally in the desired host. Causes for low or no production of the desired molecule are often multifactorial. In some circumstances, it is because of an inherent failure of the pathway. In others, such pathways into the cell usually generate different cell perturbations such as growth impairment, accumulation of metabolites, generation of toxic intermediates, and oxidative stress to name a few [1, 2, 6, 7]. Production of the target chemical can be achieved not only by optimizing the biochemical pathway but also by engineering the host microorganism. In this case, the overall metabolic performance of a cell may be improved by modulating gene expression on a genome scale using traditional gene deletion

methods or more recent techniques involving small regulatory ribonucleic acids (RNAs) [1, 2, 8–10]. However, methods for engineering the host microorganism are beyond the scope of this chapter and are not discussed here.

In this chapter we discuss recent strategies used to design, engineer, and optimize biochemical pathways to produce chemicals of high interest. We describe computational algorithms used to design efficient metabolic routes and experimental tools to construct and improve the efficiency of the designed pathway (Table 1).

Table 1 List of tools for pathway engineering

Tool	Description	Advantages	References
Pathway design tools			
Retropath	Pathway design containing circuits and self-regulation based on the specifications given to the program	Especially useful when the regulatory elements are being included into the pathway design	[26, 152]
OptForce	Find the metabolic engineering modifications on the flux of each reaction to improve the production of the target	Overproduction of the target molecule by optimizing the flux of each reaction	[153, 154]
CORBAPy	Network based algorithm which designs the elements of the network based on biological hypothesis	The network based algorithm allows the discovery of unknown pathways	[155]
XTMS	Design and score the possible pathways for production of the target chemical	Scoring system for ranking the pathways reduces the number of constructs needed for characterization	[156]
Metabolic tinker	Search for all of the thermodynamically possible paths between two compounds (source and target)	Enabling tool for discovering thermodynamically possible metabolic pathways	[157]
GEM-Path	Specifically in *E. coli* and eliminate unfavorable pathways in each step of the search	Improved and fast searching algorithm for pathways in *E. coli*	[158]
Pathway construction tools			
BioBrick-based	Sequential assembly based on restriction digestions using standardized suffixes and prefixes	The availability of a comprehensive library of standardized BioBrick parts coupled with its modularity makes this method very powerful and flexible	[33–35]
Gibson assembly-based	Overlapping sequences at/near the end of the DNA parts are simultaneously chewed back and repaired	Scarless, fast and reliable assembly of multiple parts	[30, 38–40]

(continued)

Table 1 (continued)

Tool	Description	Advantages	References
Ligase chain reaction	DNA bridges put the DNA parts next to each other and a thermostable ligase assembles them together	Modular method especially useful for combinatorial assemblies	[42]
Golden gate assembly-based	Iterative cycles of restriction and ligation using Type IIS endonucleases that cleavage outside of the recognition site releasing tunable 4 bp overhangs	Scarless, fast and reliable assembly of multiple parts	[43, 45–48]
DNA assembler	Leveraging yeast homologous recombination machinery for assembling parts with designed homology regions	Flexible, reliable and recommended for large constructs	[49, 50]
Pathway optimization tools			
Gene expression			
Plasmid copy number	Modulate copy number of plasmid to reduce metabolic burden	Balancing of different genes expression can be easily modulated	[60, 61]
Chromosomal integration (RAGE, CasEMBLR, Di-CRISPR)	Integrate the pathway into a specific region of the genome	Increase protein expression, genetic stability and reproducibility. Also reduce metabolic burden	[64, 69–71, 75]
Promoter strength	Engineer promoters with spanned strength to modulate gene expression	Fine tuning of one or multiple gene expression	[76–82]
Transcriptional terminators	Engineer terminators with different strengths to modulate gene expression	Increase mRNA stability allowing increase protein expression. Also allow fine tuning of gene expression	[86, 87]
CRISPR-based methods	Use modified CRISPR-Cas system to modulate gene expression	Allow precise temporal repression or activation of a gene	[88–93]
Codon optimization	Replacement of codons to meet the host codon bias	Increase gene expression	[94–96]
	Change of codons to modify mRNA structure	Favor translation efficiency	[100–102]
	Randomization of codons	Disable hidden control elements	[27, 102]
RBS optimization	Optimization of RBS by computational tools	Increase translation efficiency	[28, 36, 107–109]
Protein activity			
Protein engineering	In vitro protein engineering to increase activity or modify substrate specificity	Increased activity bypasses low gene expression. Substrate specificity increases catalytic efficiency and avoid side reactions	[13, 112]

(continued)

Table 1 (continued)

Tool	Description	Advantages	References
Library of homologous proteins	Screen different protein homologues with different traits	Allow finding proteins with best features	[117–119]
Cofactors	Increase cofactor expression levels or swap cofactor specificity	Reduce competition for cofactors	[118, 121–123]
Spatial localization			
Scaffolds	Anchor the proteins of the pathway to a scaffold	Favor metabolite tunneling avoiding diffusion	[127, 128]
Compartmentalization	Encapsulation of pathway enzymes in cellular organelles or bacterial microcompartments	Reduce metabolite diffusion, avoid metabolite transport and regulation, reduce toxicity, prevent competition for intermediates	[129, 130, 135]

2 Pathway Design

Designing pathways for chemical biosynthesis in microorganisms requires an in-depth knowledge of the enzymes catalyzing the reactions and of the physiology of microorganisms themselves. In many cases, this information is incomplete because of the complex nature of biological systems. Traditionally, the design process consists of surveying the literature to find the candidate genes and assembling those demonstrated to have the desired activities into a biochemical pathway. This is then followed by the characterization and optimization of the designed pathways (Fig. 1). However, because of the small number of genes that could be analyzed by a single person and the suboptimal decisions, this process is usually inefficient. Designing the pathways for the production of those chemicals is therefore difficult and time consuming in many cases. Thus a considerable number of software packages have been developed to overcome this shortcoming. These packages in most cases generate a large list of series of enzymes (pathways) that can potentially convert one or more of the abundant precursors available in the cell to the desired products [11]. These pathways are then sorted based on a wide range of criteria and the best candidate pathways for this conversion are reported to the user. The chosen pathways are then constructed and characterized to find the most efficient. Each pathway comprises regulatory elements such as promoters, Ribosome Binding Sites (RBS), terminators, and the genes coding the protein of interest. Because all these parts greatly depend on the host in which the pathway is expressed, we first discuss the criteria for choosing the proper host before looking into the intricacies of pathway design.

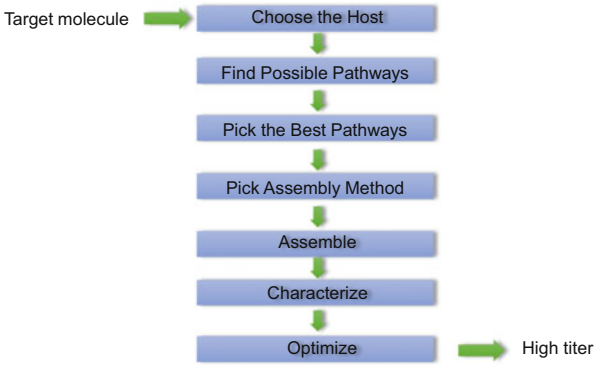

Fig. 1 The common workflow from choosing the host to chemical production. After the host is chosen, the possible pathways are found and ranked and a few of the highest ranked pathways are constructed and characterized. After optimization of the best pathway, the production in high titers is achieved

2.1 Choosing the Target Molecule and Host Organism

As mentioned before, many different chemicals have been produced in microorganisms and these chemicals range from antibiotics and natural products to commodity chemicals and fuels. For example, ethanol is produced on a very large scale for different applications from beverages to fuel. The target molecule is determined by the market, but the decision on the production host for that molecule is the key.

Depending on its ecological niche, each organism has evolved and achieved some fitness advantages over others to ensure its survival and proliferation. This survival strategy is different from one organism to the next. For example, *Escherichia coli* has an astonishingly fast growth rate and consumes available nutrients very quickly, rapidly outperforming competing strains in the culture. On the other hand, *Saccharomyces cerevisiae* grows more slowly but produces ethanol which kills most of the bacteria present in the same culture, after which the alcohol is consumed as a carbon source. *Yarrowia lipolitica* has a rather interesting strategy and stores energy as intracellular lipids constituting up to 36% of its dry weight [12]. Bearing these differences in mind, there is no super host that is best for production of all target molecules. Therefore, the identity of the target molecule plays a very important role in the choice of the production host. As an example, *S. cerevisiae* is the ideal host for the production of ethanol and ethanol-derived chemicals [13–15] whereas *Y. lipolitica* is a great host for production of fatty acid-derived products [16].

Another consideration when choosing the host is the danger it might pose to the end user of the product. If the final product is intended to be used as a food additive, the microbial host is preferred to have been granted GRAS status (Generally Regarded as Safe) by FDA (American Food and Drug Administration), QPS status (Qualified Presumption of Safety) by EFSA (European Food Safety Authority), or similar. Therefore, it can fulfill safety requirements, such as ensuring the absence of adverse health effects arising from the presence of endotoxins and emetic toxins.

Host choice greatly affects the design of the pathway and the performance of the strain in the production setting. The availability of metabolic engineering tools is also another factor to take into account for selecting the host. Two of the most

Table 2 Guidelines for choosing a proper host

Metabolic resources	Abundant precursors and co-factors for the pathway of interest
Minimum metabolic adjustments	Ideally choose hosts with characteristics desired for the production of the final product, for example, produce ethanol in yeast and taking advantage of robust endogenous pathways
Product secretion	Good secretion ability in the host for product of interest
Toxicity of products	Ideally, none of the products or intermediates are toxic to the cell
Genomic toolset and cultivation conditions	Facile tools are available for genetic modification and engineering. Cultivation conditions are not too difficult on an industrial scale, which includes the oxygen demand and optimum growth media and temperature
Proper enzyme folding	The protein of interest can be properly expressed and folded in the host of choice

commonly used industrial microorganisms are *E. coli* and *S. cerevisiae*. The extensive metabolic engineering toolkit for these organisms is one of the major factors for their preference as the production host. These hosts have been extensively studied and engineered to produce a wide variety of products from different feedstocks [17–20]. A detailed comparison between them and other alternative hosts, including the advantages and disadvantages of each system, is reviewed elsewhere [21].

The choice of host can have significant impact on the choice of pathways and enzymes for the production of the desired chemical. Even though it was shown that over half of the gene products involved in small molecule metabolism of *E. coli* and yeast carry out common reactions [22], the regulatory elements are widely different between the two organisms [21]. Even different strains of the same microorganism can have different behaviors [23]. The host also determines the regulatory elements (promoters, terminators, and RBS), codon preferences, maturation modifications, and the secretion machinery. Fisher and co-workers suggested six factors for choosing the host [21], which we summarized in Table 2.

2.2 In Silico Pathway Design

Once the host is chosen, the pathway design process begins. Engineering the host to produce the target molecule in industrial quantities is challenging and requires careful considerations. Traditionally, a few pathways are selected based on similar pathways in the literature. However, because of the large and growing number of possible pathways, manually picking the best is inefficient and impractical. To solve this problem, a myriad of bioinformatics tools have been developed which can search public databases to design and rank possible pathways producing the target molecule. These models search through enzyme databases such as BRENDA [24] and, by finding the enzymes that can possibly catalyze the reaction, they generate a large number of potential pathways, many of which do not exist in nature.

Assembling all of these pathways is neither practical nor necessary. So the next step is to find the potential pathways with higher chances of success and constructing them. The best pathway is not necessarily a well-known pathway in nature and it may be a combination of genes from different organisms. Not being constrained to using native genes or obtaining all of the genes from one source has its advantages. The best pathway is usually chosen based on specificity of the enzymes for the desired reaction, the number of enzymes involved in the pathway, the thermodynamic favorability of the reactions in the pathway, and the toxicity of the intermediates to the cell [25].

There are many programs designed for pathway design and each uses a different search algorithm and ranking strategy. A list of some of these programs can be found in Table 1. These computational tools have been successfully used in an extensive range of applications. In one example [26], Retropath was used to find pathways for flavonoid pinocembrin production. By searching the enzymes in the database, nine million pathways were predicted that could potentially produce this compound. This list was then narrowed down to 12 highly ranked pathways which were then constructed and characterized. The metabolic network was then optimized using Retropath and a 17-fold improvement in the final titer was achieved.

Other elements of the pathway have been modeled and characterized too. For example, with the modeling of different RBSs, translation initiation rates can now be predicted with high accuracy [27]. This model was used to create a library of RBSs with different strengths and achieve a wide dynamic range of translation levels for proteins of interest. This RBS library calculator was then combined with a system level kinetic model and 73 different variations of a pathway were designed, built, and characterized [28].

3 Pathway Construction

After the pathways of interest are selected and designed, they have to be assembled and constructed. Deoxyribonucleic acid (DNA) assembly strategies have been developed for a long time and have progressed from restriction digestion/ligation to more sophisticated seamless multi-part assembly methods [29]. Using the newly developed techniques, DNA constructs as large as the size of entire genomes and with as many as 25 parts have been assembled [30]. Because of the need for high throughput assembly and because of the sophistication of some of these DNA assembly techniques, many online tools have been developed to facilitate and optimize the design and assembly strategy for a specific construct. DNA assembly methods have been extensively reviewed [29, 31] and in this section we mostly focus on the more recent assembly methods and web tools that help select and plan the best assembly strategy.

3.1 Methods Based on Restriction Digestion/Ligation

3.1.1 BioBrick-Based Methods

Because of the complexity of biological parts and assembly strategies, extensive efforts have been put into modularizing the biological parts. The idea of these DNA parts or bricks and the tools to assemble them together easily was first introduced in 1996 [32], but the term BioBrick was first used by Tom Knight and the assembly strategy was published later [33]. In the BioBrick method, all parts are stored in circular plasmids that are easily amplified. Restriction endonucleases *EcoR*I and *Xba*I are used as prefixes and *Sph*I and *Pst*I as suffixes to create two compatible sticky ends between the parts being assembled. The parts are subsequently assembled by digestion and ligation. Iterative digestion and ligation allows the assembly of multiple parts (Fig. 2a). These standard parts are commonly used and stored in databases that are continuously updated. Over 2,000 parts are now available with more parts being designed and added to the database by researchers around the world.

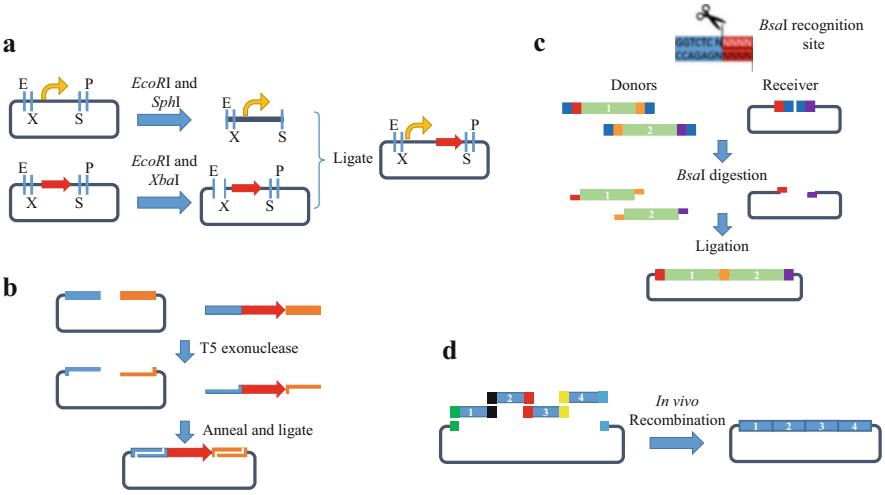

Fig. 2 DNA assembly methods. (**a**) In BioBrick assembly, prefix and suffixes are used to assemble parts in order. Four enzymes – *EcoR*I (E), *Xba*I (X), *Sph*I (S), and *Pst*I (P) – are used as the double digestions sites in the prefix and suffix regions. The correct set of prefix/suffix has to be chosen for each step and the final DNA molecule contains the suffix and prefix for further assembly rounds and addition of new parts. (**b**) In Gibson assembly, the T5 exonuclease digests the 5′ ends of the DNA parts and the digested pieces are ligated to form the final construct. (**c**) In the Golden Gate assembly method, a Type IIS endonuclease such as the most commonly used *Bsa*I is used to digest the region next to the recognition site (shown here in *blue*) generating a 4-bp overhang. In the ligation step the matching overhangs ligate together, resulting in the assembly of the DNA fragments in the designed order. (**d**) In DNA assembler method, the DNA parts with homology are recombined by the cellular homologous recombination machinery in *S. cerevisiae*

There have been many modifications and improvements over the BioBrick method to make the strategy more flexible and useful. ePathBrick is one of these methods in which the same principle of irreversible digestion/ligation was expanded to four restriction enzymes. This change makes it easier to assemble combinatorial pathways and in one example, they assembled seven parts on a single ePathBrick vector. Different variations of the same pathway were assembled to generate 54 different vectors [34]. This method was later widely used to construct combinatorial pathways to investigate the effect of each of the DNA parts on the whole pathway. Changing different parts of a pathway and seeing the effect on the production provides valuable insight on the function contribution of the individual parts. The ePathBrick method was used to assemble 18 plasmids with different combinations of a 3-gene catechin biosynthetic pathway. Three variants of the first two genes and two variants of the last genes were tested and characterized, and the best combination was found [35]. In another study, this ePathBrick was used to optimize the transcription rate of genes involved in the fatty acid biosynthetic pathway [36]. The entire pathway was divided into three modules and each module was transcribed from a different promoter. Changing the promoter regulating each of the modules enabled the researchers to identify the bottlenecks of the pathway and to reduce them by fine tuning the transcription levels. It is evident from these applications that modularity of an assembly method is very important and can lead to useful applications.

One of the problems with BioBrick assembly is the use of site-specific restriction enzymes. Because the recognition sequence of these enzymes is usually rather short, it is likely that they are present in genes that are going to be cloned. A six-base pair recognition site, for example, can randomly appear roughly every 4 kb which makes this method troublesome for longer constructs. Traditionally synonymous point mutations are introduced to replace the pre-existing cut sites in the genes such that they are no longer recognized by the restriction enzymes [33].

An alternative method called iBrick is described in a recent paper which solves this problem to a great extent [37]. In this method, two restriction enzymes of I-SceI and PI-PspI with very long (>18 bp) recognition sites were used. Using these enzymes greatly reduces the probability of restriction sites found within genes and enables the users to construct longer pathways without the need for modulating the sequence of the genes involved. Using iBrick, a carotenoid (~4 kb) and actinorhodin (~20 kb) biosynthetic cluster was constructed without introducing point mutations.

3.1.2 Gibson Assembly-Based Methods

The Gibson assembly method was developed by Daniel Gibson in 2009 [30, 38, 39]. This method allows for scarless single-pot assembly of multiple parts at the same time. The parts being assembled usually have around 25-bp homology which guides the assembly (Fig. 2b). After mixing the parts with T5 exonuclease, this enzyme starts digesting one strand of the parts (chew back) and the Phusion

polymerase starts repairing the DNA parts following the exonuclease. In this process, the flanking regions anneal to each other and, with the exonuclease being heat inactivated and Phusion polymerase catching up, the reaction is completed in the same buffer at a constant temperature of 50 °C. This isothermal assembly method was used to assemble 25 DNA fragments, constituting the entire *Mycoplasma genitalium* genome [40]. The Gibson Assembly Kit is commercially available at New England BioLabs (Ipswich, MA) and many web tools are available for designing the overhangs between the DNA parts.

One of the shortcomings of the Gibson assembly-based methods is that the two adjacent parts must have homology regions with each other. The promise of synthetic biology is modular design and a lot of protocols depend on this modularity. This modular design allows for better and easier construction of combinatorial assemblies. For example, in many studies different homologs of a gene have to be cloned in a pathway in multiple assemblies. However, using Gibson assembly-based methods, because there is a small homology between each two adjacent parts, changing one part in the assembly requires changing its adjacent parts as well, which becomes problematic in large libraries of constructs. This inherent shortcoming can be overcome by designing linkers between the parts. By adding a short DNA sequence (linker) before and after each part, the assembly becomes independent of the sequence of parts and anything with the appropriate linkers can be inserted in the proper location. Designing linkers can be tricky because orthogonality of the linkers can make a huge difference in the assembly strategy. Decreasing the homology between the linkers can reduce the percentage of misassembly. R2oDNA Designer [33] is an online tool to design these with improved efficiency. The optimized linkers were used with three homology-based assembly methods and efficiency of more than 75% was reported [33].

There have been many modifications and improvements on the Gibson assembly protocol. One of them sought a combination of BioBricks and Gibson assembly which results in both multi-part assembly of the Gibson method and modularity of BioBricks [41]. In this method, a long linker was designed between each of two parts to be assembled and, using Gibson assembly, all of the parts were assembled together. It is noteworthy that because of the sensitivity of the sequence before the start codon, the whole RBS region was used as the overhang but an overhang sequence was added between the terminator and the coding sequence. Using this method, PCR-amplified parts with BioBrick style linkers were generated and a randomized library with different promoters, genes, and terminators for the lycopene biosynthetic pathway was constructed with a 200-fold expression level difference between the constructs [41].

3.1.3 Ligase Chain Reaction-Based Methods

Ligase Chain Reaction (LCR) is an innovative scarless ligation based method optimized by scientists from Amyris (Emeryville, CA). In this method, a “bridge” is designed with homology between the parts to be assembled. The temperature of

the reaction is increased and the DNA is denatured. By decreasing the temperature, the bridge anneals to the fragments, and the two fragments are ligated together by a thermostable DNA ligase. This cycle is then repeated and the assembled fragments serve as a template for the next ligation reaction. This assembly technique is very versatile, and any combination of the parts can be assembled without pre-processing and designing specific overhangs. Because this process does not involve amplification in the assembly, the mutation rate is very low (less than 1 per >50 kb). It was reported that Gibson assembly cannot assemble constructs of 4 or more parts with efficiency of more than 50%, but LCR could assemble up to 12 DNA pieces with more than 60% efficiency. Thirteen factors affecting the efficiency of LCR method have been experimentally optimized as a condition for LCR assembly [42].

3.1.4 Golden Gate-Based Assembly Methods

The Golden Gate method relies on digestion with Type IIS endonucleases whose recognition site is adjacent to the cut site. The advantage of this mode of cleavage is that the sequence of the recognition site is independent of that of the cut site, and hence the resulting four base overhang can be customizable. This flexibility in choosing specific overhang sequences enables the user to design different overhangs for each junction. Then, in the subsequent ligation step, the sticky ends are exposed, and complementary overhangs are ligated. Consequently, the assembly of the desired parts in the desired order can be achieved (Fig. 2c) [43].

Because of its flexibility and modularity, this method was quickly adopted as a gold standard for DNA assembly. Researchers have used it for a myriad of applications from large-scale TALEN synthesis [44] to natural product discovery [31]. In the first work, one-pot single step assembly of 13 DNA fragments was performed, achieving an efficiency of ~98%, which demonstrates the capability and robustness of the Golden Gate assembly method.

Because of the widespread use of this method, many variations and improvements have been developed. One prominent example is MoClo [45, 46]. In this approach, a modular cloning system based on hierarchical assembly has been proposed. Here, in a first step of assembly, CDSs, promoters, RBSs, and terminators are assembled in a plasmid (level 0 assembly). Subsequently, these cassettes are assembled together in a second level of assembly (level 2 assembly). Iterative cycles of higher levels of assembly would produce larger cassettes making MoClo a powerful method for hierarchical assembly of large plasmids. Similarly, a comprehensive toolkit was also developed for *S. cerevisiae*. In this case, a set of characterized parts such as promoters, CDSs, and terminators are available in a Golden Gate-ready plasmid [47]. Similar to BioBrick assembly, standardized part libraries based on these methods have been deposited in Addgene and are available to the public [47, 48].

Despite the advantages of the Golden Gate method, there is one major limitation which may hamper its extensive use. The presence of the recognition site of the Type IIS endonuclease in the parts to be assembled greatly reduces the efficiency of

the assembly reaction. Thereby, these sequences should be avoided when possible. However, similar to BioBrick-based methods, this has to be performed only once and the repaired part can be reused for multiple assemblies.

3.2 *In Vivo Recombination-Based Methods*

Homologous recombination allows the assembly of multiple DNA parts with homology to each other inside the cells. This process is known to be used by the cells for repairing the unwanted DNA breaks which sometimes occur in the genome. By transforming the DNA pieces with homology, yeast cells recognize this homology and assemble the plasmids of interest by yeast homologous recombination [49]. This method, also known as DNA assembler, works by extending each part for ~40 bp to have a homology region with the adjacent part. This homology region is then recognized by the homologous recombination machinery and the parts get assembled. When all of the parts are assembled and the selection is applied, only the cells with circular plasmids survive (Fig. 2d). It should be noted that because this method is based on sequence homology, the parts are similar to what they would be if they were to be assembled by the Gibson assembly. The flexibility and ability to construct large plasmids using this method is a great advantage, but the slow growth of yeast cells and possible misassemblies because of the similarity of the homologous parts are the limitations of this method. Using this method, Shao and coworkers could assemble large constructs including a ~9-kb xylose utilization pathway, ~11 kb of zeaxanthin pathway, and a plasmid containing both of these pathways with more than 70% efficiency [49].

This method was later modified to improve the efficiency. In one report the origin of replication and marker were disconnected and each of them served as another part in the DNA assembly. The idea is that, because both of these parts have to be present in the assembled construct, some of the transformants harboring misassembled plasmids are unable to survive and fewer false positives are observed. This strategy resulted in 100-fold decrease in false positive transformants compared to the original DNA assembler method [50].

Because DNA assembler is a powerful method for assembling large constructs, many studies have used it for constructing large plasmids, many of which are larger than 20 kb [17, 51–55]. Nonetheless, by increasing the number of genes in the pathway, the percentage of correct constructs decreases and more colonies have to be picked to find the correct construct. It seems that having larger but fewer DNA pieces is a good strategy for getting less false positives with the DNA assembler method. One way to solve this problem is to combine in vitro and in vivo assembly methods. Yuan and coworkers [56] were able to assemble large constructs of ~13, 22, and 44 kb plasmids by assembling small pieces of their construct using the LCR method and then assembling the larger pieces using DNA assembler. By combining these two methods they achieved the impressive fidelity of 71% for the 44-kb construct.

3.3 In Silico Design of DNA Assembly

Given the variety and differences between the above-mentioned DNA assembly methods, it can be tricky to choose the right one. Some of the assembly methods perform the best for assemblies with larger pieces and some others with larger numbers of pieces. Sometimes it is easier and even more cost-effective to synthesize some of the parts, but sometimes not. The j5 DNA design software is available that suggests when DNA synthesis is cost-effective. The success rate of the assembly in some cases also depends on the sequence of the parts being assembled. For instance, if the parts have high sequence homology with each other, DNA assembler may not be the ideal strategy because misassemblies are likely to happen. These sequence dependencies are difficult to detect manually and computational tools are required to suggest the best assembly strategy. If not considered carefully, many problems may arise because of these intricacies.

One of the most widely used DNA assembly automation packages is j5 which designs combinatorial libraries and hierarchical assemblies with its extensive design rules [57]. It also takes advantage of the ever decreasing cost of DNA synthesis and suggests synthesis when it is cost effective to do so. j5 has an extensive cost optimization option which not only helps with the assembly protocol but also optimizes the cost, making it a useful tool for construction of a large number of pathways [58].

Another software package called Raven has an interactive learning function and can interact with the user [59]. This package designs the assembly strategy but gets feedback from the user and if one of the steps of the assembly doesn't work for any reason, it changes the strategy to avoid that specific step. This package was reported to outperform the non-optimized assemblies with the p value of <0.0001.

4 Pathway Optimization

A designed pathway often does not function optimally in a desired host. Therefore, it is crucial to optimize a number of factors to obtain a functional and efficient pathway. Pathway optimization tools can be classified in three different groups: gene expression, protein function, and spatial localization.

4.1 Gene Expression

The introduction of a set of heterologous genes usually entails a metabolic burden for the host. As a result, the chemical of interest may not be produced in yields that fulfill the expectations of industrial implementation. To alleviate the metabolic stress and hence increase the yield of the chemical, the expression of a heterologous

pathway can be improved by tuning a number of parameters that usually affect either the transcription and/or the translation processes.

4.1.1 Transcriptional Level

Plasmid, Chromosomal Integration and Copy Number

The introduction of metabolic pathways into a host is usually based on three different platforms: high-copy number plasmids (HCP), low-copy number plasmids (LCP), and chromosomal integration (CI). They perform differently and show pros and cons. Thus, when engineering a metabolic pathway, selecting the right cloning platform can be an arduous task.

Many pathway designs rely on the use of plasmids because of their availability and variability. Traditionally it was believed that using HCP would benefit the expression of a pathway, as more copies would lead to higher protein expression, and thus overproduce the chemical of interest overall. This is the case of the salicylate biosynthesis pathway in *E. coli*. When the EntC (isochorismate synthase) from *E. coli* and the PchB (isochorismate pyruvate lyase) from *Pseudomonas fluorescens* were expressed in an HCP, the salicylic acid (SA) titer reached ~770 mg/L, whereas in LCP the production dropped to ~200 mg/L [60]. Nevertheless, it was shown that LCP could also result in better expression levels than using HCP [61]. Possible explanations can be cellular toxicity of expressed heterologous proteins or limited availability of cellular expression machinery such as transcription factors, and therefore increasing the DNA copy number does not increase expression [47, 62]. Recently, Wu and coworkers engineered *E. coli* to produce resveratrol from L-tyrosine. The authors divided the pathway in three modules that were expressed in individual plasmids with different copy numbers (from 10 to 100) to modulate and alleviate bottlenecks in the pathway. A combination of low and middle copy number plasmids resulted in higher resveratrol production (~35 mg/L). Interestingly, when the higher copy number plasmid was used in any module, the resveratrol titers dropped dramatically [63].

Furthermore, the use of plasmids is usually associated with a metabolic burden on the cells being particularly obvious with HCP. It has been shown that cells carrying plasmids have generally lower growth rates than cells without plasmids [47, 64, 65]. It is believed that this burden is partially linked to the cost of the maintenance of the plasmid in the cell. Recently, Karim and coworkers who were intrigued by the factors that influence the "plasmid burden" in yeast cells investigated the effect of a number of plasmid elements, for instance origin of replication, selection markers, promoters, and copy number in haploid and diploid cells [66]. Interestingly, this study unraveled interactions between different elements that somehow could mask individual effects of plasmid elements. For example, increased plasmid loads are correlated with decreased growth rates. However, this impact is more evident in diploid cells than in haploid cells. Selection markers, especially auxotrophic ones, can also impair the growth significantly. In addition,

plasmid copy number can be modified up to a certain range by all the factors evaluated. These data reflect an intricate contribution to the plasmid burden of different factors which should be taken into account to make a knowledgeable decision when choosing a plasmid in metabolic pathway engineering [66].

Although plasmids are easy to handle and allow flexibility, they suffer from genetic instability forcing to maintain a selective pressure by using strict formulated media or addition of antibiotics that can boost the cost of the production of a chemical [47, 64, 67]. On the other hand, plasmid-based protein expression is not consistent from cell to cell, indicating that limited copy number regulation compromises the reproducibility [47]. These disadvantages are turning chromosomal integration into the method of choice.

Chromosomal integration overcomes all of these drawbacks. It was proved that gene integration in host genome produced reliable protein expression patterns, unaffected growth rates, and also bypassed the use of selective compounds. Additionally, integration in the genome is usually in single or low copy number, so protein toxicity and competition for metabolites can be buffered [47, 64, 67]. Nevertheless, the genetic context where heterologous genes are integrated also seems to influence protein expression [64, 68]. Recently, Yin and coworkers studied the impact of the chromosomal location in the polyhydroxybutyrate (PHB) production in *E. coli*. The *pha*CAB operon (PHB synthesis pathway) from *Ralstonia eutropha* and a red fluorescent protein (RFP) were integrated downstream of 13 different chromosomal locations, some of them with high transcriptional activities. They found similar results for both *rfp* and *pha*CAB; the *asnB* (asparagine synthetase B) location showed the highest transcriptional levels out of the 13 locations evaluated, measured by real-time PCR. However, in the case of *pha*CAB, a single copy of the operon did not produce detectable levels of PHB. Then the *pha*CAB operon was integrated in multiple copies in the chromosome via Cre-loxP system. PHB was only detected when four copies of the operon were introduced, and PHB levels increased with the number of copies of the pathway, showing maximum levels (~34.1 wt%) with 50 copies [64]. At the same time, a plasmid-based *pha*CAB expression was evaluated. Despite the high levels of 43.68 wt% obtained, the production of PHB dropped dramatically to 8.08 wt% when the antibiotic pressure was removed, proving the instability of the plasmid system [64]. This work illustrated the importance of chromosomal integration, genomic context, and copy number to reach high production levels of a chemical.

Bearing in mind these results, chromosomal integration strategies should allow efficient multiple copies integration in specific regions of the genome. However, traditional methods showing some limitations are tackled by recent studies. Gu and coworkers used a flippase recombinase (FLP) from yeast to optimize gene overexpression for amino acid production in *E. coli*. The FLP can recombine two DNA sequences containing a 34-bp recombination site (FRT). Accordingly, one main requirement is that the host genome should contain an FRT site. The authors determined that increasing the concentration of donor plasmid in the transformation and the number of FRT sites in the chromosome led to an increased number of copies inserted into the genome [67]. Using this strategy, the production of L-tryptophan was optimized. The introduction of two copies of *aroK* gene (shikimate

kinase) in the chromosome increased L-tryptophan levels ~87%, although more number of copies decreased the production. In a similar way, the three genes *serA*^{FR}, *serB*, and *serC* for overproduction of L-serine were inserted in the chromosome in different copy numbers. The highest L-serine strain producer contained 10:4:4 copy number combination of *serA*^{FR}, *serB*, and *serC* genes, respectively. This FLP/FRT recombination strategy allowed optimizing and balancing integrated gene copy number of a metabolic pathway in a single step [67].

Chromosomal integration of an alginate-degrading pathway through recombinase-assisted genome engineering (RAGE) showed 40-fold improved ethanol production titers from brown macroalgae over its corresponding plasmid-based counterparts [69, 70]. This study again highlighted the instability of plasmid-based pathways. In addition, the study revealed that the distance between the chromosomal origin of replication and the integration point impacted the growth, indicating the important role that the chromosomal location has in the expression of the pathway. In this case, modified enzymatic pathways from 34 to 59 kb were efficiently integrated in *E. coli* chromosome using the Cre-lox recombination system. Moreover, the authors could apply the FLP/FRT recombination strategy to remove the antibiotic marker to generate a markerless strain for further chromosomal modifications. Through this approach it was also possible to balance the pathway copy number that allowed higher cell densities [69]. These recombinase-based methods are efficient and permit genomic integrations in specific regions marked by recombination sites (loxP and/or FRT for instances) but this advantage turns into a limitation as these sites have to be previously introduced in the genome by other approaches.

CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/associated protein-9 nuclease) is a powerful tool to generate double strand breaks (DSBs) in yeast chromosome in a single locus or multiple loci with high efficiency [71–74]. By combining the CRISPR/Cas9 editing properties and the yeast in vivo DNA assembly efficiency, it is possible to insert a multi-gene enzymatic pathway in a high efficiency, reliable, and marker-free fashion. By using the CasEMBLR method, Jakočiūnas and coworkers integrated a carotenoid pathway composed of 15 parts, and also developed a tyrosine-producing yeast strain by insertion of 10 parts. The advantage of this method is that a set of linear DNA parts (promoters, ORFs, terminators for instance) with sequence homology in their ends can be assembled and integrated in a single step in any desired location in the genome with efficiencies ranging from 30 to 90% without the need for using selectable markers [71]. In spite of the fact that introduction of multiple copies of the same element can be challenging, CasEMBLR may allow swapping between biological parts easily, and reduce the effort in constructing donor plasmids containing different combinations of elements.

Similarly, Shi and coworkers have also recently exploited the CRISPR/Cas9-based DSBs combined with yeast in vivo recombination. The authors have developed Di-CRISPR, delta integration CRISPR, which targets delta sites in *S. cerevisiae* chromosome to integrate multiple pathway copies. Di-CRISPR enabled the integration of 18 copies of a large cassette (24 kb) consisting of a

xylose utilization and (*R*,*R*)-2,3-butanediol (BDO) production pathways in a single step with high efficiency [75].

Promoter Strength

Promoter is a control element generating great modifications in gene expression as strong promoters usually result in increased mRNA levels. Therefore, increasing the promoter strength is a successful approach to enhance protein expression. Nevertheless, in multi-gene pathways this approach can lead to transcriptional/translational stress, accumulation of metabolites, and toxicity. To prevent these problems, balancing the promoter strength between genes is an option. There have recently been much effort to characterize and develop libraries of natural/hybrid/synthetic promoters with a wide dynamic range in terms of promoter strength that allow the precise regulation of each gene in the pathway [47, 76–79].

For example, Liang and coworkers developed a set of inducible hybrid promoters based on the GAL promoter in yeast. The new group of promoters was tightly regulated in the presence of minimal concentrations of estradiol (10 nM). By refactoring a zeaxanthin biosynthetic pathway in yeast using this set of promoters, the authors reported a production improvement of 50-fold over the pathway with constitutive promoters [80].

Lee and coworkers characterized a set of yeast constitutive promoters that allowed them to develop a linear regression model to engineer pathways in a predictable fashion. By this approach the authors achieved the production of violacein for the first time in yeast [81].

Similarly, Zhang and coworkers optimized the production of amorphadiene (AD), a precursor of artemisinin, by the experimental design-aided systematic pathway optimization (EDASPO) method. Basically, the pathway was divided into four modules, and the genes were under the control of T7 and T7-variant promoters. By characterizing a few combinations, the authors developed a linear regression model that enabled further optimization and achieved a threefold enhanced AD titer [82]. Balancing the promoter strength has been successfully used to engineer multi-gene pathways [19, 63, 83, 84].

Transcriptional Terminators

Although terminators have an important role in the transcription termination and in the mRNA half-life [85], there have been fewer studies in terminator development and characterization for metabolic engineering applications. Recently, the impact of a number of terminators on gene expression in yeast was studied, revealing their capacity to modulate expression as much as promoters [86]. The authors found a strong correlation between the expression and the increased mRNA half-life, suggesting that terminators influence the stability of mRNA [86]. In another study, the same group developed a set of short synthetic terminators which performed similarly to those common in yeast. To evaluate the utility of these

terminators in pathway engineering they expressed the codon-optimized *cis*-aconitic acid decarboxylase (CAD1) gene from *Aspergillus terreus* under the control of the weak promoter *TEFmut3*, followed by the collection of synthetic terminators. Constructs containing the synthetic terminators allowed similar or even higher itaconic acid titers than those containing the yeast CYC1 terminator [87]. In addition to this, the synthetic terminators were functional in a different yeast species, suggesting a generalized use of these terminators [87]. These studies highlighted the potential utilization of terminators to modulate gene expression and metabolic pathway balancing for optimal pathway function in the designated host.

CRISPR-Based Modulation

A recent strategy to modulate gene expression is to exploit the properties of the CRISPR system. Qi and coworkers developed a CRISPR-based system for gene repression on a genome scale, CRISPR interference (CRISPRi) [88]. A dCas9 (a mutant RNA-guided DNA endonuclease defective in nuclease activity) was used to bind a small guide RNA (sgRNA) that targets specific DNA sequences. The dCas9 binds sgRNA, and the complex sgRNA-dCas9 binds these DNA regions. A precise design of the sgRNA can target different DNA elements such as promoters or ORFs. Therefore, sgRNA-dCas9 complex can efficiently block the transcription process at different levels, for instance it can impede the transcription factor binding, the RNA polymerase binding, or the transcriptional elongation [88]. CRISPRi has been successfully used in metabolic engineering to modulate multiple genes of a polyhydroxyalkanoate (PHA) biosynthesis pathway in *E. coli*. By engineering sgRNA to different targets that produce a range of expression levels, it was possible to modulate the 4-hydroxybutyrate (4HB) content in poly (3-hydroxybutyrate-*co*-4-hydroxybutyrate) [P(3HB-*co*-4HB)] [89].

In a similar approach, dCas9 can be fused to a repressor or activator module, thus allowing the dCas9 guided by gRNA to silence or activate a gene [90–92]. Zalatan and coworkers further modified the CRISPR system to convert the gRNA into an RNA scaffold (scRNA). The scRNA was designed to contain sequences recognized by RNA-binding protein modules. Then, transcriptional activators or repressors were fused to RNA-binding proteins, so these activators or repressors could be recruited and bound to the RNA scaffold at a desired DNA location to activate or silence a gene [93]. It was anticipated that these CRISPR-based gene modulation approaches could be effectively used in multi-enzyme pathway optimization. It makes it feasible to turn ON/OFF enzyme expression to maximize pathway productivity by generating predictable and flexible metabolite flux. This approach was validated with a highly branched violacein biosynthetic pathway in yeast. This pathway consisted of five genes (VioABEDC) producing violacein as a final product. Nevertheless, different modulation of the last two steps (VioD and VioC) can generate four colored products. Thus, by switching on and off the expression of VioA, VioD, and VioC, all possible pathway routes were achieved in a predictable manner [93].

4.1.2 Translational Level

Codon Optimization

It is widely believed that different organisms have different codon usage depending on the abundance and availability of tRNAs. Codon optimization strategies usually replace rare codons with those that meet the host codon bias. Therefore this can be an efficient tool to favor the translation of heterologous proteins. There are a number of studies where codon optimization successfully culminated in improved protein expression [94–96]. It is suggested that optimal codons improved the mRNA stability [97]. However, there is also empirical evidence that using frequent codons is sometimes detrimental [98]. There is great controversy about why this strategy is not consistent from protein to protein. Lanza and coworkers noticed that codon optimization is usually based on data from the whole genome, but growth conditions and other factors can modify tRNA abundance, so traditional approaches omit relevant information that can impact the translation process in specific environments [99]. To overcome this drawback, the authors developed a "condition-specific codon optimization" method consisting of using codon bias based on genes expressed under a desired condition. This approach increased 2.9-fold the catechol 1,2-dioxygenase gene expression in yeast over a commercial optimized version [99].

It was recently suggested that the codon bias has low effect in translation efficiency [100, 101], pointing out that mRNA structure, especially in the first 15–20 amino acids, is the main factor that affects the translation efficiency [100–102], and thereby secondary structures of mRNA may impede binding of ribosome and pause elongation [102]. On the other hand, it is known that control elements may appear embedded in the coding region that are difficult to identify. By randomizing the codon sequence, it is possible to disable these hidden elements [102]. Computational tools can assist in designing optimized genes that prevent the drawbacks arising from mRNA structures and *cis*-regulation [27, 102, 103].

Optimization of RBS

The initiation of translation in prokaryotes occurs when the 16S rRNA of the small ribosome subunit binds to the Shine–Dalgarno (SD) sequence in the RBS in the mRNA. It is usually located 5–15 bases upstream from the start codon, and changes in its sequence-dictated affinity can change the expression levels several orders of magnitude, enabling fine-tuning of the pathway expression [27, 77, 102, 104–106].

Recent examples demonstrated the effectiveness of RBS optimization to increase the productivity of metabolic pathways; for instances, astaxanthin, fatty acids, and riboflavin titers were enhanced in *E. coli* [36, 107, 108]. However, screening a combinatorial library of RBS in a multi-gene pathway can be tedious and impracticable even with high-throughput screening methods. Also many of the

combinations may have detrimental effects bearing in mind that the initiation of translation can be affected by many factors: the structure of the mRNA can interfere with the binding of the ribosome to the RBS, so weak RBS can lead to low expression levels; strong interactions with the RBS can also cause stalling of translation, and distance between SD and start codon has been shown to be critical, for instance [27, 102]. It is obvious that a rational design is highly desirable. Current online tools use computational methods that circumvent these drawbacks, considering all of the potential molecular interactions, and design RBSs with a wide range of initiation translational rates [27, 28, 109].

4.2 Protein Activity

Despite the many efforts in enhancing and balancing gene expression in multienzyme pathways, in some cases the production of a desired chemical is still difficult to accomplish. In these cases, either the intrinsic activity/specificity or suboptimal environmental conditions can be a limiting factor. To achieve improved pathway outcomes, it is critical to modulate protein properties.

4.2.1 Protein Engineering

There are two general approaches to alter the intrinsic properties of a protein: directed evolution and rational design. Both strategies have been applied successfully in protein engineering for pathway optimization. Whereas rational design requires a thorough knowledge of structure-function protein characteristics, directed evolution explores the whole protein sequence, and circumvents the limitations of incomplete structure-protein information [110, 111].

Lian and coworkers [13] constructed a cellobiose utilization pathway to produce ethanol from cellulosic biomass in yeast. It consisted of a cellodextrin transporter and a β-glucosidase. In this study the cellodextrin transporter 2 (CDT2) from *Neurospora crassa* was engineered by directed evolution to increase its cellobiose uptake activity. CDT2 is a facilitator, and thus does not consume ATP for cellobiose uptake, and it may provide energetic benefits in anaerobic cultures, although it is less efficient than others transporters such as CDT1. After three rounds of directed evolution, the best CDT2 evolved variant enabled over fourfold increased cellobiose consumption rate and ethanol productivity in anaerobic conditions. More rational design experiments showed that both specific activity and transporter expression levels were ameliorated. By this approach, the total ethanol yield was increased by more than 25% [13].

In another recent example, a biosynthetic pathway for *cis,cis*-muconic acid (*cc*MC) production consisting of three enzymes, AroZ, AroY, and CatA, was engineered in *E. coli*. The authors observed the accumulation of a metabolic intermediate, catechol, the substrate of CatA (catechol 1,2-dioxygenase). Replacing

inducible promoters by constitutive ones did not solve the bottleneck. Accordingly, rational design to alter enzyme activity mitigated the limiting step. Higher enzymatic activities were obtained by widening the channel of the catalytic pocket. Improved CatA variant also led to ~26% improved *cc*MC productivity [112].

Introduction of unnatural amino acids (UAAs) in protein sequences can also diversify the biochemical properties of an enzyme or even lead to new functionalities. Although incorporation of UAAs has been used in protein engineering, resulting in improved biocatalysts, it has not been applied in pathway optimization probably because the introduction of orthogonal pairs of aminoacyl-tRNA synthetase/tRNA in the desired host is still challenging and needs further optimization [113–116]. The use of engineered enzymes containing UAA in pathway optimization may increase the spectrum of catalytic reactions that can be performed by engineered hosts to address biosynthetic bottlenecks.

4.2.2 Homologous Proteins

Modification of protein properties by protein engineering to meet pathway requirements can be challenging and often fails. Thus it sounds more feasible to find the appropriate protein among the current proteins available. In nature there exists proteins capable of executing the same function in a variety of organisms. Despite playing similar catalytic roles, they may exhibit diversified features such as different optimal pH, temperatures, higher activities, specificities, promiscuities, and regulation among others. Thus, the selection of the proper subset of enzymes with better performance in the desired host is essential in the construction of an efficient pathway. Nevertheless, limited information about the biochemical properties of proteins can hinder the design. In these circumstances, a less rational design such as a combinatorial library can bypass the lack of information. For example, *Gluconobacter oxydans* WSH-003 was engineered to produce 2-keto-L-gulonic acid (2KLG), a precursor of vitamin C. The heterologous pathway consisted of L-sorbose dehydrogenases (SDH) and L-sorbosone dehydrogenases (SNDH) from *Ketogulonicigenium vulgare* WSH-001. In this study, five SDH and two SNDH enzymes from *K. vulgare* WSH-001 with different features [117] were combined. Ten combinations were analyzed and the best one achieved 4.9 g/L of 2KLG [118].

Recently, more rational design has been used to engineer *S. cerevisiae* to produce taxadiene. The catalysis of farnesyl diphosphate (FPP) to geranylgeranyl diphosphate (GGPP) by geranylgeranyl diphosphate synthase (GGPPS) is a limiting step in taxadiene production. Thus the optimization of this enzyme may increase the productivity. A computational approach was used to predict the binding affinity of six GGPPSs from different organisms with its substrate FPP. The protein modeling and docking predicted that the GGPPSbc (from *Taxus baccata* × *Taxus cuspidate*) may benefit the limiting reaction. The authors proved the model empirically, and observed that the taxadiene titer was improved over tenfold using GGPPSbc [119].

4.2.3 Cofactors

A large number of enzymes involved in metabolic reactions depend on cofactors for proper functionality. When an exogenous pathway is introduced in a host, competition for the cofactors and/or redox imbalance can emerge and cause metabolic stress, impair cellular growth, and an overall reduction in the productivity of the pathway [120]. Thus, tuning the concentration of cofactors [121, 122] or swapping the cofactor specificity [123] can be used to enhance pathway efficiency.

Lim and coworkers elegantly compensated the redox imbalance created by the introduction of a synthetic *n*-butanol pathway in *E. coli* [122]. In this study, the *E. coli* host was previously engineered for production of butyrate where the cofactor regeneration pathway was redirected to use butyrate as the final electron acceptor [83]. Introduction of the heterologous *n*-butanol pathway in this host generated NADH deficiency, highlighting the need for further engineering. To supply more NADH, the authors modulated the pyruvate dehydrogenase complex (PDH enzymatic complex) which catalyzes the decarboxylation of pyruvate into acetyl-CoA, producing CO_2 and NADH. To overcome the limitation conferred by strong inhibition of the complex under anaerobic conditions, a mutant PDH complex active in anaerobic conditions and driven by strong control elements was integrated into the chromosome, yielding a 12% improvement in *n*-butanol titers. On the other hand, some pyruvate could still be catalyzed by NAD^+-independent pyruvate formate lyase (PFL), producing acetyl-CoA and formate. The NAD^+-dependent formate dehydrogenase (FDH) from yeast converts formate into CO_2 and produces NADH. The *fdh1* gene expression was fine-tuned using synthetic 5′-UTRs. The optimal engineered strain showed 35% increased *n*-butanol titers achieving 6.8 g/L [122].

In another example, Gao and coworkers engineered *G. oxydans* WSH-003 strain to produce 2KLG. The heterologous pathway consisted of SDH and SNDH from *K. vulgare* WSH-001. These dehydrogenases require pyrroloquinoline quinine (PQQ) for functionality, and may compete with native PQQ-dependent proteins. Thereby, the biosynthetic PQQ cluster was overexpressed to avoid a PQQ bottleneck. Increasing the supply of the cofactor resulted in an increase of 20% of 2KLG [118].

In a similar way, Cui and coworkers observed that increased NADPH concentrations favored the production of shikimic acid (SA) in *E. coli* [121]. The shikimate dehydrogenase reduces 3-dehydroshikimate to shikimate using NADPH as a cofactor; therefore the availability of NADPH may limit the productivity of the pathway. The authors proved that overexpression of transhydrogenase (*pntAB*) or/and NAD kinase (*nadK*), two native enzymes involved in NADPH regeneration, increased the SA titer by more than twofold [121].

4.3 Spatial Localization

The efficiency of a pathway sometimes depends on factors unrelated to protein expression or catalytic activity. Toxicity of metabolic intermediates, reduced

availability of intermediates because of diffusion or consumption in other metabolic pathways, and reduced local enzyme concentrations are some factors that hinder pathway efficiency. Colocalization of pathway enzymes can efficiently decrease intermediate loss, increase protein concentration, and reduce toxicity by metabolite tunneling. Spatial colocalization can be achieved by anchoring the enzymes in a scaffold or by enzyme sequestration into cellular compartments. This approach has been extensively reviewed elsewhere [4, 7, 124]. Here we describe a few recent successful examples.

4.3.1 Scaffold Strategies

This strategy for spatial colocalization of enzymes is based on the interaction between the proteins of interest and a synthetic protein [125], RNA [126], or DNA [127] scaffold. Proteins are fused to a binding domain that recognizes and anchors enzymes to the scaffold.

A recent example is the improvement of butyrate production in *E. coli*. Three enzymes of the pathway, 3-hydroxybutyryl-CoA dehydrogenase (Hbd), 3-hydroxybutyryl-CoA dehydratase (Crt), and *trans*-enoyl-coenzyme A reductase (Ter), were fused to ligands for GBD, SH3, and PDZ domains. When these constructions were expressed in *E. coli*, the butyrate production increased from 1.22 to 3.51 g/L [128]. The authors also observed a decline of by-product acetate production. It was suggested that the scaffold approach directed the carbon flux efficiently through the immobilized enzymes [128].

Another approach is to use DNA molecules as a scaffold. In this case the proteins of interest are fused to zinc-finger (ZF) domains that bind specific DNA motifs [127]. Thus, a plasmid DNA can be designed to contain a number of different recognitions sites for different ZF domains. Conrado and coworkers proved the feasibility of DNA scaffolds in metabolic engineering by increasing the productivity of trans-resveratrol, 1,2-propanediol, and mevalonate in *E. coli*. In this study the authors corroborated the hypothesis that improved yields were the result of optimal proximity between the enzymes of the pathway optimizing metabolites channeling. For this purpose, the ZF binding motifs in the DNA scaffold for each enzyme were located far from each other (no proximity between enzymes) or with 2–12 bp spacers (proximity between enzymes). The improvements were annulled when the enzymes were far apart [127].

4.3.2 Compartmentalization

The use of scaffolds to organize enzymes spatially helps to improve metabolites channeling, but it can impede the proper folding of multimer enzymes or cause a metabolic burden by consuming additional cellular sources to synthesize the scaffold [4]. Pathway encapsulation can overcome these issues and benefit pathway engineering. Expression of all the enzymes in a pathway scaffold-free in a specific cellular organelle avoids metabolites transport, diffusion and leakage, prevents

competition for intermediates with other pathways, escapes from regulation, and increases the concentration of enzymes and proximities between them (small compartments compared with cytoplasm) [129].

One striking example that represents the benefits of compartmentalization targeted the Ehrlich pathway into yeast mitochondria to produce isobutanol. The isobutanol pathway consisted of five enzymes divided into two sets: (1) acetolactate synthase (ALS), ketolacid reductoisomerase (KARI), and dehydroxyacid dehydratase (DADH) which are present in mitochondria and (2) α-ketoacid decarboxylase (α-KDC) and alcohol dehydrogenase (ADH) which are usually in cytoplasm. In this study, α-KDC and ADH were directed to mitochondria by fusion with the N-terminal mitochondrial localization signal from subunit IV of the yeast cytochrome c oxidase (CoxIV). α-Ketoisovalerate (α-KIV) is produced by DADH in the mitochondria and has to be transported to the cytoplasm to be further modified by α-KDC. The authors found that one limiting factor in the pathway was the availability of α-KIV in the cytoplasm. Thus avoiding the transport of this intermediate from mitochondria to cytoplasm may increase the availability of the intermediate, and hence increase the titer of the pathway. The overexpression of the five genes together with the targeting of α-KDC and ADH to mitochondria enabled titers of 635 mg/L of isobutanol, which represented ~twofold improvement over the same pathway with α-KDC and ADH directed to the cytoplasm (380 mg/L) and over ninefold compared with the control with an empty plasmid (67 mg/L) [129]. Additionally, the authors reported an increment in the production of other branched-chain alcohols as isopentanol and 2-methyl-1-butanol. One suggestion that the authors proposed to support these phenomena is that the first three enzymes of the pathway are also involved in isoleucine, leucine, and valine biosynthetic pathways generating metabolic intermediates that eventually can be a substrate for α-KDC and ADH producing isopentanol and 2-methyl-1-butanol [129].

In another example, the production of penicillin was enhanced by targeting part of the biosynthetic penicillin pathway to the peroxisome in *Aspergillus nidulans*. Three enzymes are involved in the pathway; the last one (isopenicillin N acyltransferase, AatA) is located in the peroxisome whereas the other two (δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase, AcvA and isopenicillin N synthase, IpnA) are in the cytoplasm. As the intermediates need to be transported into the peroxisome, it was suggested that colocalizing all the enzymes in the same compartment may benefit the production of penicillin. The authors found that targeting AcvA into the organelle increased the penicillin production by 3.2-fold. Interestingly, targeting IpnA to the peroxisome dropped penicillin production drastically. One reason could be that the redox state of the peroxisome did not provide the appropriate environmental conditions for activity and stability of IpnA [130].

The compartmentalization strategy also enabled the enhancement of itaconic acid production in *Aspergillus niger*. Overexpression of two enzymes, *cis*-aconitate decarboxylase and aconitase, in mitochondria led to a twofold improvement of itaconic acid production compared with the overexpression of the two enzymes in the cytoplasm [131]. However, a similar approach by targeting *cis*-aconitic acid

decarboxylase (CAD) from *Aspergillus terreus* to mitochondria failed to improved itaconic acid production in *S. cerevisiae* [132].

Compartmentalization strategies can definitely be used in metabolic engineering to improve pathway efficiency. Nevertheless, organelle environmental considerations should be taken into account as the conditions may not be favorable for specific enzymatic reactions.

Harnessing the cellular organelles has been used for pathway engineering in eukaryotes. In the case of prokaryotes, the use of bacterial microcompartments (MCPs) for metabolic engineering is a promising strategy [133, 134]. Bacterial MCPs are metabolic enzymes involved in a specific process encapsulated in protein shells that encase metabolic intermediates which can be volatile or toxic for the cell [133, 134]. Although the use of MCPs in pathway optimization is in its earliest stages, and needs further characterization, recently Lawrence and coworkers proved its potential in metabolic engineering. The authors reproduced an MCP from *Citrobacter freundii* to generate a bioreactor to produce ethanol in *E. coli*. The pyruvate decarboxylase (encoded by the *pdc* gene) and alcohol dehydrogenase (encoded by the *adh* gene) from *Zymomonas mobilis* were targeted to the heterologous MCP. The ethanol production was almost doubled in those strains expressing PDC and ADH targeted to the MCP [135].

5 Applications

Metabolic engineering and synthetic biology tools have enabled the engineering of microorganisms to produce a wide spectrum of chemicals with applications in several fields. The examples described below highlight recent advances in the design, construction, and optimization of pathways for biosynthesis of valuable chemicals.

5.1 Production of Pharmaceutical Products

Natural products are the main source of drugs and pharmaceuticals. However, recovery of these products from their natural source is usually tedious, time consuming, and inefficient, and their chemical synthesis is not always available. Thus, there is an increasing interest in developing new manufacturing platforms based on model microorganisms that are easy to manipulate and are usually able to reproduce numerous enzymatic steps in mild conditions which are more environmentally friendly than the chemical synthesis.

The biosynthetic production of many pharmaceuticals has been accomplished. Some examples are the production of: (2*S*)-pinocembrim, suggested for treatment of cerebral ischemic injury [136, 137]; shikimic acid, precursor of an anti-influenza drug [121]; catechins, precursor of anthocyanins and tannins [35]; resveratrol, as a

therapeutic compound [63, 127]; penicillin [130]; and *N*-acetylglucosamine, as treatment for cartilage disease [84]. Here we discuss the production of artemisinic acid [138] and opioids [139].

The synthesis of artemisinic acid, a precursor of the antimalarial artemisinin, in large quantities is a remarkable example. To date, the unique source of artemisinin has been its natural plant producer *Artemisia annua*. Nevertheless, the supply of this plant to the pharmaceutical industry was environment-dependent, generating fluctuation in the price from year to year. Since 2004 many attempts have been made to produce artemisinin commercially affordable, especially in the developing world [138]. For that purpose it was proposed to develop a microorganism-based platform capable of synthesizing at least 25 g/L of artemisinin [138]. There are two key steps in the biosynthesis of artemisinin: (1) synthesis of amorphadiene and (2) synthesis of artemisinic acid which can be chemically converted to artemisinin [138]. Thus, the main objective was to overproduce amorphadiene. Although the production of the intermediate metabolite amorphadiene in *E. coli* was improved up to levels of 27.4 g/L [140], the following steps in the pathway to obtain artemisinic acid dissuaded researchers from continuing to engineer *E. coli*. The main reason for this decision was the general limited ability of *E. coli* to express heterologous eukaryotic P450, an enzyme involved in the conversion of amorphadiene into artemisinic acid. As the success of the project at this point was compromised, it was concluded that a change of production host would benefit overall the final productivity of the pathway. Then *S. cerevisiae* was engineered to produce artemisinic acid [141, 142] (Fig. 3). The first step to increase the production of amorphadiene was to overexpress the mevalonate pathway genes responsible for conversion of acetyl-CoA to FPP by using galactose-inducible strong promoters. The copy number of tHMG1gene (truncated HMG-CoA reductase) was also triplicated as its expression was found to be a rate-limiting factor. The heterologous gene from *A. annua* amorphadiene synthase (ADS) expressed in a high-copy plasmid was codon-optimized for *S. cerevisiae*, although the production of amorphadiene was not improved compared with the non-codon-optimized version of the gene. Finally, the optimization of the fermentation conditions led to 37–41 g/L amorphadiene titers [142]. The next step was to introduce the amorphadiene oxidase cytochrome P450 (CYP71AV1) and its cognate reductase (CPR1) together with the ADS gene from *A. annua* driven by galactose-inducible strong promoters in a high-copy plasmid to convert amorphadiene into artemisinic acid. To reduce the toxicity generated by high levels of CPR1, it was integrated in the genome in a single copy under a weak promoter (GAL3 promoter). Finally, the cytochrome b_5 from *A. annua* (CYB5), the artemisinic aldehyde dehydrogenase (ALDH1), and alcohol dehydrogenase (ADH1) from *A. annua* were also integrated in the genome under a strong promoter (GAL7p) (Fig. 3). The artemisinic acid titers obtained were 25 g/L [143]. These levels allow the economical production of artemisinin through a photochemical transformation process developed by Sanofi [138, 143, 144]. Bioengineering of yeast has enabled the current cost-effective industrial production of artemisinin, making the antimalarial treatment affordable in developing countries, and hence saving lives [138].

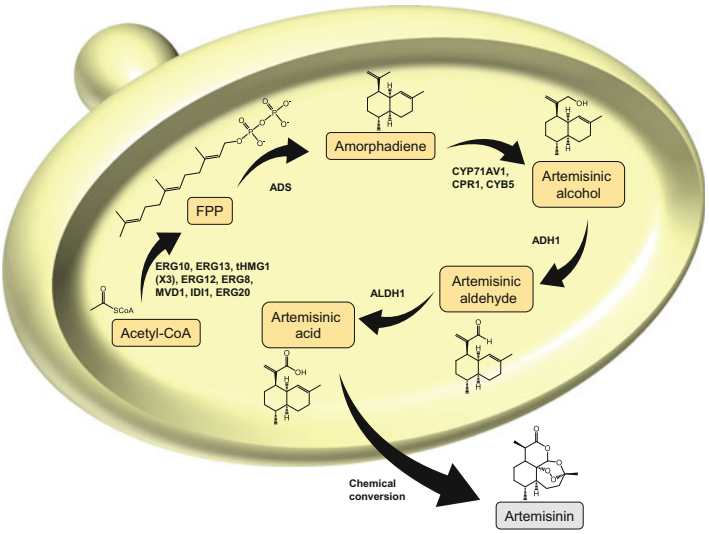

Fig. 3 Scheme of pathway engineering for semi-synthetic artemisinin production in yeast. The first step was to increase the levels of *amorphadiene* by overexpression of the mevalonate pathway (from *acetyl-CoA* to *FPP*), and it was also necessary to introduce ADS. The second step was to convert *amorphadiene* into *artemisinic acid* by introduction of additional genes (see text for more details). Finally, purified *artemisinic acid* was converted into *artemisinin* through a chemical process. *ERG10* acetoacetyl-CoA thiolase, *ERG13* HMG-CoA synthase, *tHMG1* truncated HMG-CoA reductase, *ERG12* mevalonate kinase, *ERG8* phosphomevalonate kinase, *MVD1* mevalonate diphosphate decarboxylase, *IDI1* isopentenyl diphosphate isomerase, *ERG20* farnesyl diphosphate synthase, *FPP* farnesyl diphosphate, *ADS* amorphadiene synthase, *CYP71AV1* cytochrome P450 enzyme, *CYB5* cytochrome b_5, *ADH1* artemisinic alcohol dehydrogenase, *ALDH1* artemisinic aldehyde dehydrogenase

Another interesting example is the production of opioids in *S. cerevisiae* [139, 145]. Opioid drugs are used in the medical treatment of severe pain. Currently these drugs are derived from the opium poppy (*Papaver somniferum*). As in the case of artemisinin, poppy agribusiness is susceptible to environmental factors, and is also subjected to strict governmental control. This study described the engineering efforts to produce thebaine and hydrocodone in baker's yeast. The biosynthetic pathway genes were divided into modules to facilitate their optimization (Fig. 4). The first step was to increase the carbon flux to (*S*)-reticuline biosynthesis. Four modules containing 17 genes from a variety of organisms (plants, bacteria, yeast, and mammals) were integrated in the genome. Module I was designed to increase L-tyrosine and 4-hydroxyphenylacetaldehyde (4-HPAA), precursors of the (*S*)-reticuline. Module II contained the genes to synthesize and recycle the tetrahydrobiopterin (BH4) redox cofactor. Module III included the genes to synthesize (*S*)-norcoclaurine. Module IV contained the genes to synthesize (*S*)-reticuline. The integration of these modules in yeast genome gave rise to 20 μg/L of (*S*)-reticuline. Module V contained additional copies of three genes (mutated tyrosine hydroxylase, $TyrH^{WR}$; 4′-*O*-methyltransferase, 4′OMT; and norcoclaurine

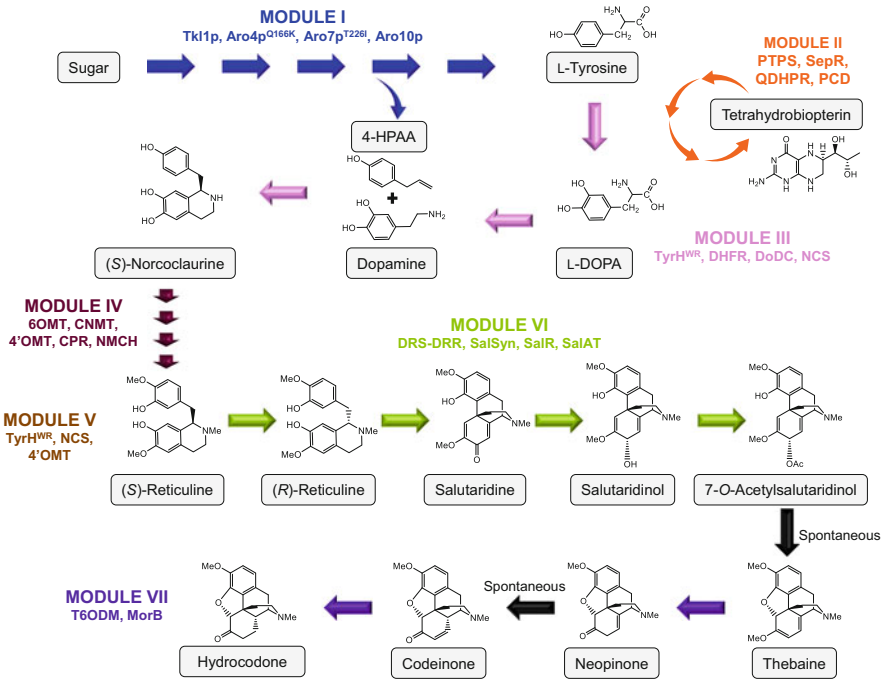

Fig. 4 Scheme of pathway engineering for opioids (thebaine and hydrocodone) production in yeast. Genes included in the same module are represented by the same color. Module I contains the genes to increase precursors of (*S*)-reticuline: *Tkl1p* transketolase, $Aro4p^{Q166K}$ 3-deoxy-D-arabino-2-heptulosonic acid-7-phosphate (DAHP) synthase (mutation Q166K), $Aro7p^{T226I}$ chorismate mutase (mutation T226I), *Aro10p* phenylpyruvate decarbosylase. Module II contains the genes to synthesize and recycle the tetrahydrobiopterin (BH4) redox cofactor: *PTPS* 6-pyruvoyl tetrahydrobiopterin synthase, *SepR* sepiapterin reductase, *QDHPR* quinonoid dihydropteridine reductase, *PCD* pterin carbinolamine. Module III contains the genes to synthesize (*S*)-norcoclaurine: $TyrH^{WR}$ tyrosine hydroxylase (mutations R37E, R38E, W166Y), *DHFR* dihydrofolate reductase, *DoDC* L-DOPA decarboxylase, *NCS* norcoclaurine synthase. Module IV contains the genes to synthesize (*S*)-reticuline: *6OMT* norcoclaurine 6-*O*-methyltransferase, *CNMT* coclaurine *N*-methyltransferase, *4′OMT* 3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase, *NMCH N*-methylcoclaurine hydroxylase. Module V contains additional copies of $TyrH^{WR}$, *NCS*, and *4′OMT*. Module VI contains the genes to synthesize thebaine: *DRS-DRR* 1,2-dehydroreticuline synthase-1,2-dehydroreticuline reductase, *SalSyn* salutaridine synthase, *SalR* salutaridine reductase, *SalAT* salutaridinol 7-*O*-acetyltransferase. Module VII contains the genes to synthesize hydrocodone: *T6ODM* thebaine 6-O-demethylase, *MorB* morphine reductase

synthase, NCS) which were suggested to produce a bottleneck in the pathway flux. The introduction of module V led to a fourfold improvement of the (*S*)-reticuline titers. The second step was the production of thebaine. The four enzymes involved in this process were engineered. First, the discovery of the 1,2-dehydroreticuline synthase/reductase (DRS/DRR), an epimerase that converts (*S*)-reticuline to (*R*)-reticuline, was a key step for the production of thebaine. This epimerase was identified by bioinformatic analysis of genomic and transcriptomic databases. The next enzyme in the pathway, salutaridine synthase (SalSyn) exhibited

N-glycosylation, which resulted in reduced activity of the enzyme. Protein engineering was used to create a chimeric SalSyns with different N-terminal ends from cheilanthifoline synthase (CFS), a plant P450 enzyme which was heterologously expressed in yeast with high activity, to prevent *N*-glycosylation. Additionally, the codon-optimized salutaridine reductase (SalR) and salutaridinol acetyltransferase (SalAT) homologues from different *Papaver sp*. were compared. The best combination of the four engineered enzymes in module VI included: *P*. *bracteatum* DRS-DRR, PbDRS-DRR; yeast codon-optimized *P*. *bracteatum* N-terminal variant SalSyn, yEcCFS1-83-yPbSalSyn92-504; yeast codon-optimized *P*. *bracteatum* SalR, PbSalR; and yeast codon-optimized *P*. *somniferum* SalAT, PsSalAT. The strain containing all 6 modules (with 24 genes cassettes) integrated in the chromosome was able to produce 6.4 μg/L of thebaine, the first morphinan alkaloid in the biosynthetic pathway. Then the pathway was extended to produce hydrocodone by introduction of a seventh module containing thebaine 6-*O*-demethylase (T6ODM) and morphine reductase (MorB) (Fig. 4). The final strain harbored 26 genes in 7 modules and produced 0.3 μg/L of hydrocodone from glucose for the first time, as the poppy cannot produce this compound [139]. The levels of opioids obtained in this work do not support industrial implementation as one dose of this drug requires the fermentation of thousands of liters [139]. More engineering studies are therefore needed to increase the production levels. However, this study demonstrated the potential of synthetic biology and metabolic engineering to design organisms beyond the limits of nature.

5.2 Production of Fuels and Chemicals

Increasing energy demand is pushing to the limit the use of non-renewable fossil fuel sources. Thus, there is a special interest in developing microbial cell factories able to produce fuels and alternative petroleum-derived chemicals. Some recent examples are the microbial synthesis of isobutanol, isopentanol, 2-methyl-1-butanol [129], *n*-butanol [122], ethanol [69, 146], (*2R*,*3R*)-butanediol [147], fatty acids [36, 148], and fatty acid ethyl esters (FAEE) [149].

Baker's yeast *S*. *cerevisiae* was engineered for the production of FAEE. First, a wax ester synthase (WS), responsible for the synthesizing of FAEE from acyl-CoA and ethanol, was introduced. Five different WSs from different organisms were evaluated. The WSs from *Marinobacter hydrocarbonoclasticus* DSM 8798 (WS2) allowed the highest FAEE titers of all five, 6.3 mg/L [150]. The synthesis of acyl-CoA requires acetyl-CoA, an essential intermediate metabolite involved in several pathways. Thus, acetyl-CoA availability could hinder the FAEE production. Two different strategies were used to increase acetyl-CoA levels in the cytoplasm. The first was the introduction of an ethanol degradation pathway to redirect the carbon flux to the synthesis of acetyl-CoA. This pathway consisted of the endogenous alcohol dehydrogenase 2 (ADH2) and acetaldehyde dehydrogenase (ALD6), and a mutated variant of the acetyl-CoA synthetase (ACS_{SE}^{L641P}) from *Salmonella*

enterica which cannot be acetylated. These three enzymes were overexpressed in a high-copy plasmid together with WS2. The yield obtained was 408 ± 270 μg $gCDW^{-1}$, which is three times the productivity of the strain carrying only the WS2. However, the reproducibility was compromised, probably because of variations related to the large high-copy plasmid. To circumvent plasmid number fluctuations between replicates, the WS2 was detached from the ethanol degradation pathway in a different plasmid. Then the FAEE productivity was significantly improved by 2.7-fold [149]. Integration of five or six copies of *ws2* in yeast chromosome increased the FAEE titer more than fivefold compared with its plasmid-based counterpart [151]. The second strategy to increase cytosolic acetyl-CoA levels and NADPH cofactor levels was to introduce a heterologous phosphoketolase (PHK) pathway by expressing *xpkA* (encoding xylulose-5-phosphate phosphoketolase) and *ack* (encoding acetate kinase) from *A. nidulans*. Replacement of *ack* by *pta* (phosphotransaketylase) from *Bacillus subtilis* that catalyzed the direct conversion of acetyl phosphate into acetyl-CoA was also evaluated (Fig. 5). Both PHK pathways, *xpkA*/*pta* and *xpkA*/*ack*, together with the integration of *ws2*, increased the production of FAEE by 1.6- to 1.7-fold (4,670 and 5,100 μg FAEE $gCDW^{-1}$), compared with the strain with only *ws2*

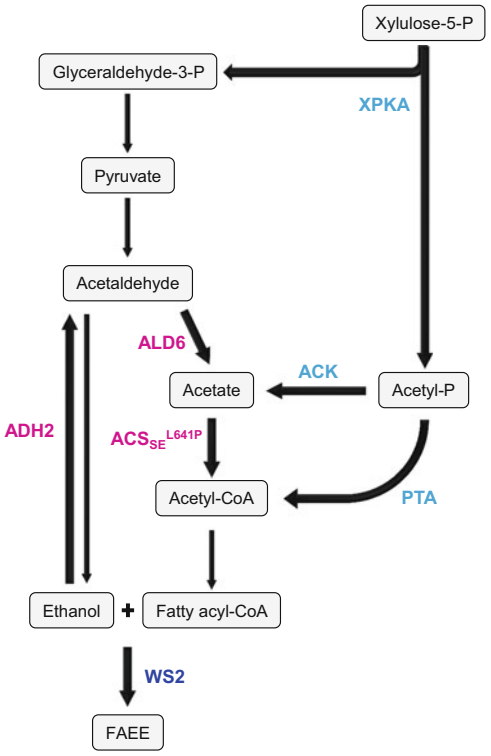

Fig. 5 Scheme of pathway engineering for FAEE production in yeast. Two pathways to increase acetyl-CoA levels were used: the ethanol degradation pathway and the phosphoketolase (PHK) pathway. Enzymes involved in the ethanol degradation pathway are presented in *pink*, and enzymes involved in the PHK are presented in *light blue*. The reaction steps catalyzed by the introduced enzymes are represented by *thick arrows*. *ADH2* alcohol dehydrogenase 2, *ALD6* acetaldehyde dehydrogenase, ACS_{SE}^{L641P} acetyl-CoA synthase, *XPKA* xylulose-5-phosphate phosphoketolase, *ACK* acetate kinase, *PTA* phosphotransacetylase, *WS2* wax ester synthase, *FAEE* fatty acid ethyl esters

[149]. Increasing precursor and cofactor levels increased the FAEE production, although further engineering is needed to achieve higher FAEE levels for commercial applications.

6 Conclusions

Many attempts have been made to establish microbial platforms for the production of valuable compounds. Although there are some commercially successful examples of microbial production of bio-based chemicals on an industrial scale, there are still a number of challenges remaining. Recent advances in synthetic biology and metabolic engineering have enabled the production of a wide range of chemicals in heterologous hosts. The examples described in this chapter have identified and overcome a variety of bottlenecks that can arise when a heterologous pathway is introduced into a host microorganism. However, when a bottleneck was bypassed a new one emerged. Iterative cycles of optimization are needed to achieve an efficient pathway, which can be tedious and time-consuming. This highlights the need for new approaches to expedite the process. It is anticipated that the increasing genomic, metagenomic, and metabolic information available can permit the development of accurate computational algorithms that can eventually help predict efficient biosynthetic pathways. Improved pathway designs combined with new experimental tools are expected to reduce efforts and to facilitate pathway construction and optimization. It is envisaged that in future years the number of chemicals efficiently produced in microbial platforms will increase dramatically.

Acknowledgements We thank the National Institutes of Health (GM077596), the National Academies Keck *Futures Initiative* on Synthetic Biology, the Energy Biosciences Institute, the Department of Energy under Advanced Research Projects Agency-Energy (ARPA-E) (DE-AR0000206), and the National Science Foundation as part of the Center for Enabling New Technologies through Catalysis (CENTC), CHE-0650456 for financial support in our protein and pathway engineering projects.

References

1. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol 8(6):536–546
2. Dai Z, Nielsen J (2015) Advancing metabolic engineering through systems biology of industrial microorganisms. Curr Opin Biotechnol 36:8–15
3. Chen Y, Nielsen J (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. Curr Opin Biotechnol 24 (6):965–972
4. Liu Y, Shin H-d, Li J, Liu L (2015) Toward metabolic engineering in the context of system biology and synthetic biology: advances and prospects. Appl Microbiol Biotechnol 99 (3):1109–1118

5. Du J, Shao Z, Zhao H (2011) Engineering microbial factories for synthesis of value-added products. J Ind Microbiol Biotechnol 38(8):873–890
6. Eriksen DT, Li S, Zhao H (2013) Chapter 3 - Pathway engineering as an enabling synthetic biology tool. In: Zhao H (ed) Synthetic biology. Academic, Boston, pp 43–61
7. Jones JA, Toparlak ÖD, Koffas MAG (2015) Metabolic pathway balancing and its role in the production of biofuels and chemicals. Curr Opin Biotechnol 33:52–59
8. Soma Y, Tsuruno K, Wada M, Yokota A, Hanai T (2014) Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. Metab Eng 23:175–184
9. Lee SY, Kim HU (2015) Systems strategies for developing industrial microbial strains. Nat Biotechnol 33(10):1061–1072
10. Si T, HamediRad M, Zhao H (2015) Regulatory RNA-assisted genome engineering in microorganisms. Curr Opin Biotechnol 36:85–90
11. Fernández-Castané A, Fehér T, Carbonell P, Pauthenier C, Faulon J-L (2014) Computer-aided design for metabolic engineering. J Biotechnol 192 Part B:302–313
12. Tai M, Stephanopoulos G (2013) Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. Metab Eng 15:1–9
13. Lian J, Li Y, HamediRad M, Zhao H (2014) Directed evolution of a cellodextrin transporter for improved biofuel production under anaerobic conditions in *Saccharomyces cerevisiae*. Biotechnol Bioeng 111(8):1521–1531
14. Sun J, Wen F, Si T, Xu J-H, Zhao H (2012) Direct conversion of xylan to ethanol by recombinant *Saccharomyces cerevisiae* strains displaying an engineered minihemicellulosome. Appl Environ Microbiol 78(11):3837–3845
15. Kim SR, Skerker JM, Kang W, Lesmana A, Wei N, Arkin AP, Jin Y-S (2013) Rational and evolutionary engineering approaches uncover a small set of genetic changes efficient for rapid xylose fermentation in *Saccharomyces cerevisiae*. PLoS One 8(2), e57048
16. Rutter C, Zhang S, Rao C (2015) Engineering *Yarrowia lipolytica* for production of medium-chain fatty acids. Appl Microbiol Biotechnol 99(17):7359–7368
17. Eriksen DT, HamediRad M, Yuan Y, Zhao H (2015) Orthogonal fatty acid biosynthetic pathway improves fatty acid ethyl ester production in *Saccharomyces cerevisiae*. ACS Synth Biol 4(7):808–814
18. Feng X, Lian J, Zhao H (2015) Metabolic engineering of *Saccharomyces cerevisiae* to improve 1-hexadecanol production. Metab Eng 27:10–19
19. Freestone TS, Zhao H (2016) Combinatorial pathway engineering for optimized production of the anti-malarial FR900098. Biotechnol Bioeng 113(2):384–392
20. Lian J, Si T, Nair NU, Zhao H (2014) Design and construction of acetyl-CoA overproducing *Saccharomyces cerevisiae* strains. Metab Eng 24:139–149
21. Fisher AK, Freedman BG, Bevan DR, Senger RS (2014) A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. Comput Struct Biotechnol J 11(18):91–99
22. Jardine O, Gough J, Chothia C, Teichmann SA (2002) Comparison of the small molecule metabolic enzymes of *Escherichia coli* and *Saccharomyces cerevisiae*. Genome Res 12 (6):916–929
23. Feng X, Zhao H (2013) Investigating host dependence of xylose utilization in recombinant *Saccharomyces cerevisiae* strains using RNA-seq analysis. Biotechnol Biofuels 6(1):1–17
24. Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, Scheer M, Schomburg D (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res 41(D1):D764–D772
25. Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng 106(3):462–473

26. Fehér T, Planson A-G, Carbonell P, Fernández-Castané A, Grigoras I, Dariy E, Perret A, Faulon J-L (2014) Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. Biotechnol J 9(11):1446–1457
27. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27(10):946–950
28. Farasat I, Kushwaha M, Collens J, Easterbrook M, Guido M, Salis HM (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol Syst Biol 10(6):731
29. Chao R, Yuan Y, Zhao H (2015) Recent advances in DNA assembly technologies. FEMS Yeast Res 15(1):1–9
30. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345
31. Cobb R, Ning J, Zhao H (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. J Ind Microbiol Biotechnol 41(2):469–477
32. Rebatchouk D, Daraselia N, Narita JO (1996) NOMAD: a versatile strategy for *in vitro* DNA manipulation applied to promoter analysis and vector design. Proc Natl Acad Sci U S A 93 (20):10891–10896
33. Shetty RP, Endy D, Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2(1):1–12
34. Xu P, Vansiri A, Bhan N, Koffas MAG (2012) ePathBrick: a synthetic biology platform for engineering metabolic pathways in *E. coli*. ACS Synth Biol 1(7):256–266
35. Zhao S, Jones JA, Lachance DM, Bhan N, Khalidi O, Venkataraman S, Wang Z, Koffas MAG (2015) Improvement of catechin production in *Escherichia coli* through combinatorial metabolic engineering. Metab Eng 28:43–53
36. Xu P, Gu Q, Wang W, Wong L, Bower AGW, Collins CH, Koffas MAG (2013) Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. Nat Commun 4:1409
37. Liu J-K, Chen W-H, Ren S-X, Zhao G-P, Wang J (2014) iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. PLoS One 9(10), e110852
38. Gibson DG, Smith HO (2010) Method for in vitro recombination. US 7,776,532 B2
39. Gibson DG, Smith HO, Hutchison CA, Young L, Venter JC (2015) Methods for in vitro joining and combinatorial assembly of nucleic acid molecules. US 8,968,999 B2
40. 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, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319(5867):1215–1220
41. Sleight SC, Sauro HM (2013) Randomized BioBrick assembly: a novel DNA assembly method for randomizing and optimizing genetic circuits and metabolic pathways. ACS Synth Biol 2(9):506–518
42. Sd K, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, Platt D, Shapland EB, Serber Z, Dean J, Newman JD, Chandran SS (2014) Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 3(2):97–106
43. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3(11), e3647
44. Liang J, Chao R, Abil Z, Bao Z, Zhao H (2014) FairyTALE: a high-throughput TAL effector synthesis platform. ACS Synth Biol 3(2):67–73
45. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6(2), e16765
46. Weber E, Werner S, Engler C, Gruetzner R, Marillonnet S (2011) System and method of modular cloning. EP 2,395,087 A1
47. Lee ME, DeLoache WC, Cervantes B, Dueber JE (2015) A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol 4(9):975–986

48. Iverson SV, Haddock TL, Beal J, Densmore DM (2016) CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. ACS Synth Biol 5(1):99–103
49. Shao Z, Zhao H, Zhao H (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37(2):e16
50. Kuijpers NGA, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran J-M, Daran-Lapujade P (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. Microb Cell Fact 12:47
51. Shao Z, Zhao H (2014) Manipulating natural product biosynthetic pathways via DNA assembler. Curr Protoc Chem Biol 6(2):65–100
52. Shao Z, Zhao H (2013) Construction and engineering of large biochemical pathways via DNA assembler. In: Polizzi KM, Kontoravdi C (eds) Synthetic biology. Methods in molecular biology, vol 1073. Humana Press, New York, pp 85–106
53. Shao Z, Zhao H (2012) Chapter Ten – DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. In: David AH (ed) Methods in enzymology, vol 517. Academic Press, San Diego, pp 203–224
54. Shao Z, Luo Y, Zhao H (2012) DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*. In: Barredo J-L (ed) Microbial carotenoids from fungi. Methods in molecular biology, vol 898. Humana Press, New York, pp 251–262
55. Mitchell LA, Chuang J, Agmon N, Khunsriraksakul C, Phillips NA, Cai Y, Truong DM, Veerakumar A, Wang Y, Mayorga M, Blomquist P, Sadda P, Trueheart J, Boeke JD (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. Nucleic Acids Res 43(13):6620–6630
56. Yuan Y, Andersen E, Zhao H (2016) Flexible and versatile strategy for the construction of large biochemical pathways. ACS Synth Biol 5(1):46–52
57. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1(1):14–21
58. Hillson N (2014) j5 DNA assembly design automation. In: Valla S, Lale R (eds) DNA cloning and assembly methods. Methods in molecular biology, vol 1116. Humana Press, New York, pp 245–269
59. Appleton E, Tao J, Haddock T, Densmore D (2014) Interactive assembly algorithms for molecular cloning. Nat Methods 11(6):657–662
60. Lin Y, Sun X, Yuan Q, Yan Y (2014) Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*. Metab Eng 23:62–69
61. Jones KL, Kim S-W, Keasling JD (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab Eng 2(4):328–338
62. Lee TH, Maheshri N (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. Mol Syst Biol 8(1)
63. Wu J, Liu P, Fan Y, Bao H, Du G, Zhou J, Chen J (2013) Multivariate modular metabolic engineering of *Escherichia coli* to produce resveratrol from L-tyrosine. J Biotechnol 167 (4):404–411
64. Yin J, Wang H, Fu X-Z, Gao X, Wu Q, Chen G-Q (2015) Effects of chromosomal gene copy number and locations on polyhydroxyalkanoate synthesis by *Escherichia coli* and *Halomonas* sp. Appl Microbiol Biotechnol 99(13):5523–5534
65. Da Silva NA, Srikrishnan S (2012) Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. FEMS Yeast Res 12(2):197–214
66. Karim AS, Curran KA, Alper HS (2013) Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. FEMS Yeast Res 13(1):107–116
67. Gu P, Yang F, Su T, Wang Q, Liang Q, Qi Q (2015) A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies. Sci Rep 5:9684

68. Lee TJ, Parikh RY, Weitz JS, Kim HD (2014) Suppression of expression between adjacent genes within heterologous modules in yeast. G3 (Bethesda) 4(1):109–116
69. Santos CNS, Regitsky DD, Yoshikuni Y (2013) Implementation of stable and complex biological systems through recombinase-assisted genome engineering. Nat Commun 4
70. Santos CNS, Yoshikuni Y (2014) Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. Nat Protoc 9(6):1320–1336
71. Jakočiūnas T, Rajkumar AS, Zhang J, Arsovska D, Rodriguez A, Jendresen CB, Skjødt ML, Nielsen AT, Borodina I, Jensen MK, Keasling JD (2015) CasEMBLR: Cas9-Facilitated multiloci genomic integration of *in vivo* assembled DNA parts in *Saccharomyces cerevisiae*. ACS Synth Biol 4(11):1226–1234
72. Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, Lee ME, DeLoache W, Dueber JE, Arkin AP, Cate JH (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. eLife 3
73. Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, Si T, Zhao H (2015) Homology-Integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. ACS Synth Biol 4(5):585–594
74. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41(7):4336–4343
75. Shi S, Liang Y, Zhang MM, Ang EL, Zhao H (2016) A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. Metab Eng 33:19–27
76. Curran KA, Crook NC, Karim AS, Gupta A, Wagman AM, Alper HS (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. Nat Commun 5
77. Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai Q-A, Tran AB, Paull M, Keasling JD, Arkin AP, Endy D (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10(4):354–360
78. Mutalik VK, Guimaraes JC, Cambray G, Mai Q-A, Christoffersen MJ, Martin L, Yu A, Lam C, Rodriguez C, Bennett G, Keasling JD, Endy D, Arkin AP (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods 10(4):347–353
79. Redden H, Alper HS (2015) The development and characterization of synthetic minimal yeast promoters. Nat Commun 6
80. Liang J, Ning JC, Zhao H (2013) Coordinated induction of multi-gene pathways in Saccharomyces cerevisiae. Nucleic Acids Res 41(4), e54
81. Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. Nucleic Acids Res 41 (22):10668–10678
82. Zhang C, Zou R, Chen X, Stephanopoulos G, Too H-P (2015) Experimental design-aided systematic pathway optimization of glucose uptake and deoxyxylulose phosphate pathway for improved amorphadiene production. Appl Microbiol Biotechnol 99(9):3825–3837
83. Lim JH, Seo SW, Kim SY, Jung GY (2013) Refactoring redox cofactor regeneration for high-yield biocatalysis of glucose to butyric acid in *Escherichia coli*. Bioresour Technol 135:568–573
84. Liu Y, Zhu Y, Li J, H-d S, Chen RR, Du G, Liu L, Chen J (2014) Modular pathway engineering of *Bacillus subtilis* for improved N-acetylglucosamine production. Metab Eng 23:42–52
85. Leavitt JM, Alper HS (2015) Advances and current limitations in transcript-level control of gene expression. Curr Opin Biotechnol 34:98–104
86. Curran KA, Karim AS, Gupta A, Alper HS (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab Eng 19:88–97

87. Curran KA, Morse NJ, Markham KA, Wagman AM, Gupta A, Alper HS (2015) Short synthetic terminators for improved heterologous gene expression in yeast. ACS Synth Biol 4(7):824–832
88. Qi Lei S, Larson Matthew H, Gilbert Luke A, Doudna Jennifer A, Weissman Jonathan S, Arkin Adam P, Lim Wendell A (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183
89. Lv L, Ren Y-L, Chen J-C, Wu Q, Chen G-Q (2015) Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab Eng 29:160–168
90. Gilbert Luke A, Larson Matthew H, Morsut L, Liu Z, Brar Gloria A, Torres Sandra E, Stern-Ginossar N, Brandman O, Whitehead Evan H, Doudna Jennifer A, Lim Wendell A, Weissman Jonathan S, Qi Lei S (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154(2):442–451
91. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31(9):833–838
92. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32(4):347–355
93. Zalatan Jesse G, Lee Michael E, Almeida R, Gilbert Luke A, Whitehead Evan H, La Russa M, Tsai Jordan C, Weissman Jonathan S, Dueber John E, Qi Lei S, Lim Wendell A (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160(1–2):339–350
94. Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS (2013) Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. Metab Eng 19:33–41
95. Tong Y, Yang H, Xin Y, Zhang L, Wang W (2015) Novel integration strategy coupling codon and fermentation optimization for efficiently enhancing sarcosine oxidase (SOX) production in recombinant *Escherichia coli*. World J Microbiol Biotechnol 31(5):707–716
96. Wei L, Liu J, Qi H, Wen J (2015) Engineering *Scheffersomyces stipitis* for fumaric acid production from xylose. Bioresour Technol 187:246–254
97. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker Kristian E, Graveley Brenton R, Coller J (2015) Codon optimality is a major determinant of mRNA stability. Cell 160(6):1111–1124
98. Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ (2013) Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol Biol Evol 30(3):549–560
99. Lanza AM, Curran KA, Rey LG, Alper HS (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst Biol 8:33
100. Allert M, Cox JC, Hellinga HW (2010) Multifactorial determinants of protein expression in prokaryotic open reading frames. J Mol Biol 402(5):905–918
101. Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol Syst Biol 10(12)
102. Welch M, Villalobos A, Gustafsson C, Minshull J (2011) Chapter three – designing genes for successful protein expression. In: Christopher V (ed) Methods in enzymology, vol 498. Academic Press, San Diego, pp 43–66
103. Villalobos A, Ness J, Gustafsson C, Minshull J, Govindarajan S (2006) Gene designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinf 7(1):285
104. Coussement P, Maertens J, Beauprez J, Van Bellegem W, De Mey M (2014) One step DNA assembly for combinatorial metabolic engineering. Metab Eng 23:70–77

105. Nowroozi F, Baidoo EK, Ermakov S, Redding-Johanson A, Batth T, Petzold C, Keasling J (2014) Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. Appl Microbiol Biotechnol 98(4):1567–1581
106. Oliver JWK, Machado IMP, Yoneda H, Atsumi S (2014) Combinatorial optimization of cyanobacterial 2,3-butanediol production. Metab Eng 22:76–82
107. Li Y, Gu Q, Lin Z, Wang Z, Chen T, Zhao X (2013) Multiplex iterative plasmid engineering for combinatorial optimization of metabolic pathways and diversification of protein coding sequences. ACS Synth Biol 2(11):651–661
108. Zelcbuch L, Antonovsky N, Bar-Even A, Levin-Karp A, Barenholz U, Dayagi M, Liebermeister W, Flamholz A, Noor E, Amram S, Brandis A, Bareia T, Yofe I, Jubran H, Milo R (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res 41(9), e98
109. Na D, Lee D (2010) RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. Bioinformatics 26(20):2633–2634
110. Cobb RE, Chao R, Zhao H (2013) Directed evolution: past, present, and future. AIChE J 59 (5):1432–1440
111. Eriksen DT, Lian J, Zhao H (2014) Protein design for pathway engineering. J Struct Biol 185 (2):234–242
112. Han L, Liu P, Sun J, Wu Y, Zhang Y, Chen W, Lin J, Wang Q, Ma Y (2015) Engineering catechol 1,2-dioxygenase by design for improving the performance of the cis, cis-muconic acid synthetic pathway in *Escherichia coli*. Sci Rep 5:13435
113. Lewis JC (2015) Metallopeptide catalysts and artificial metalloenzymes containing unnatural amino acids. Curr Opin Chem Biol 25:27–35
114. Neumann H, Neumann-Staubitz P (2010) Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. Appl Microbiol Biotechnol 87(1):75–86
115. Neumann H, Slusarczyk AL, Chin JW (2010) *De novo* generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. J Am Chem Soc 132(7):2142–2144
116. Ugwumba IN, Ozawa K, Xu Z-Q, Ely F, Foo J-L, Herlt AJ, Coppin C, Brown S, Taylor MC, Ollis DL, Mander LN, Schenk G, Dixon NE, Otting G, Oakeshott JG, Jackson CJ (2011) Improving a natural enzyme activity through incorporation of unnatural amino acids. J Am Chem Soc 133(2):326–333
117. Gao L, Du G, Zhou J, Chen J, Liu J (2013) Characterization of a group of pyrroloquinoline quinone-dependent dehydrogenases that are involved in the conversion of L-sorbose to 2-Keto-L-gulonic acid in *Ketogulonicigenium vulgare* WSH-001. Biotechnol Prog 29 (6):1398–1404
118. Gao L, Hu Y, Liu J, Du G, Zhou J, Chen J (2014) Stepwise metabolic engineering of *Gluconobacter oxydans* WSH-003 for the direct production of 2-keto-L-gulonic acid from D-sorbitol. Metab Eng 24:30–37
119. Ding M-z, Yan H-f, Li L-f, Zhai F, Shang L-q, Yin Z, Yuan Y-j (2014) Biosynthesis of taxadiene in *Saccharomyces cerevisiae*: selection of geranylgeranyl diphosphate synthase directed by a computer-aided docking strategy. PLoS One 9(10):e109348
120. Akhtar MK, Jones PR (2014) Cofactor engineering for enhancing the flux of metabolic pathways. Front Bioeng Biotechnol 2:30
121. Cui Y-Y, Ling C, Zhang Y-Y, Huang J, Liu J-Z (2014) Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. Microb Cell Fact 13:21
122. Lim JH, Seo SW, Kim SY, Jung GY (2013) Model-driven rebalancing of the intracellular redox state for optimization of a heterologous n-butanol pathway in *Escherichia coli*. Metab Eng 20:56–62
123. King ZA, Feist AM (2014) Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. Metab Eng 24:117–128

124. Siu K-H, Chen RP, Sun Q, Chen L, Tsai S-L, Chen W (2015) Synthetic scaffolds for pathway enhancement. Curr Opin Biotechnol 36:98–106
125. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KLJ, Keasling JD (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27(8):753–759
126. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333(6041):470–474
127. Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, Turnšek J, Tomšič N, Avbelj M, Gaber R, Koprivnjak T, Mori J, Glavnik V, Vovk I, Benčina M, Hodnik V, Anderluh G, Dueber JE, Jerala R, DeLisa MP (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic Acids Res 40(4):1879–1889
128. Baek J-M, Mazumdar S, Lee S-W, Jung M-Y, Lim J-H, Seo S-W, Jung G-Y, Oh M-K (2013) Butyrate production in engineered *Escherichia coli* with synthetic scaffolds. Biotechnol Bioeng 110(10):2790–2794
129. Avalos JL, Fink GR, Stephanopoulos G (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. Nat Biotechnol 31(4):335–341
130. Herr A, Fischer R (2014) Improvement of *Aspergillus nidulans* penicillin production by targeting AcvA to peroxisomes. Metab Eng 25:131–139
131. Blumhoff ML, Steiger MG, Mattanovich D, Sauer M (2013) Targeting enzymes to the right compartment: metabolic engineering for itaconic acid production by *Aspergillus niger*. Metab Eng 19:26–32
132. Blazeck J, Miller J, Pan A, Gengler J, Holden C, Jamoussi M, Alper H (2014) Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production. Appl Microbiol Biotechnol 98(19):8155–8164
133. Bobik TA, Lehman BP, Yeates TO (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Mol Microbiol 98 (2):193–207
134. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA (2014) Diverse bacterial microcompartment organelles. Microbiol Mol Biol Rev 78(3):438–468
135. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, Warren MJ (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS Synth Biol 3(7):454–465
136. Shi L-l, Chen B-n, Gao M, Zhang H-a, Li Y-j, Wang L, Du G-h (2011) The characteristics of therapeutic effect of pinocembrin in transient global brain ischemia/reperfusion rats. Life Sci 88(11–12):521–528
137. Wu J, Du G, Zhou J, Chen J (2013) Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. Metab Eng 16:48–55
138. Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol 12(5):355–367
139. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD (2015) Complete biosynthesis of opioids in yeast. Science 349(6252):1095–1100
140. Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD (2009) High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. PLoS One 4(2), e4489
141. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440(7086):940–943
142. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production of amorphadiene in yeast, and its

conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109(3):E111–E118
143. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Jiang H, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496 (7446):528–532
144. Dhainaut J, Dlubala A, Guevel R, Medard A, Oddon G, Raymond N, Turconi J (2013) Photochemical process for producing artemisinin. Patent US 8,507,697 B2
145. Smolke CD, Thodey C, Trenchard I, Galanie S (2014) Benzylisoquinoline alkaloids (bia) producing microbes, and methods of making and using the same. US Patent 20,140,273,109 A1
146. Enquist-Newman M, Faust AME, Bravo DD, Santos CNS, Raisner RM, Hanel A, Sarvabhowman P, Le C, Regitsky DD, Cooper SR, Peereboom L, Clark A, Martinez Y, Goldsmith J, Cho MY, Donohoue PD, Luo L, Lamberson B, Tamrakar P, Kim EJ, Villari JL, Gill A, Tripathi SA, Karamchedu P, Paredes CJ, Rajgarhia V, Kotlar HK, Bailey RB, Miller DJ, Ohler NL, Swimmer C, Yoshikuni Y (2014) Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform. Nature 505(7482):239–243
147. Lian J, Chao R, Zhao H (2014) Metabolic engineering of a *Saccharomyces cerevisiae* strain capable of simultaneously utilizing glucose and galactose to produce enantiopure (2R,3R)-butanediol. Metab Eng 23:92–99
148. Torella JP, Ford TJ, Kim SN, Chen AM, Way JC, Silver PA (2013) Tailored fatty acid synthesis via dynamic control of fatty acid elongation. Proc Natl Acad Sci U S A 110 (28):11290–11295
149. de Jong BW, Shi S, Siewers V, Nielsen J (2014) Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. Microb Cell Fact 13:39
150. Shi S, Valle-Rodriguez J, Khoomrung S, Siewers V, Nielsen J (2012) Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. Biotechnol Biofuels 5(1):7
151. Shi S, Valle-Rodríguez JO, Siewers V, Nielsen J (2014) Engineering of chromosomal wax ester synthase integrated *Saccharomyces cerevisiae* mutants for improved biosynthesis of fatty acid ethyl esters. Biotechnol Bioeng 111(9):1740–1747
152. Carbonell P, Parutto P, Baudier C, Junot C, Faulon J-L (2014) Retropath: automated pipeline for embedded metabolic circuits. ACS Synth Biol 3(8):565–577
153. Chowdhury A, Zomorrodi AR, Maranas CD (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. PLoS Comput Biol 10(2):e1003487
154. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6(4):e1000744
155. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for python. BMC Syst Biol 7:74
156. Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon J-L (2014) XTMS: pathway design in an eXTended Metabolic Space. Nucleic Acids Res 42(W1):W389–W394
157. McClymont K, Soyer OS (2013) Metabolic tinker: an online tool for guiding the design of synthetic metabolic pathways. Nucleic Acids Res 41(11), e113
158. Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. Metab Eng 25:140–158

Adv Biochem Eng Biotechnol (2018) 162: 117–146
DOI: 10.1007/10_2016_13

Published online: 19 October 2016

Synthetic Biology for Cell-Free Biosynthesis: Fundamentals of Designing Novel In Vitro Multi-Enzyme Reaction Networks

Gaspar Morgado, Daniel Gerngross, Tania M. Roberts, and Sven Panke

Abstract Cell-free biosynthesis in the form of in vitro multi-enzyme reaction networks or enzyme cascade reactions emerges as a promising tool to carry out complex catalysis in one-step, one-vessel settings. It combines the advantages of well-established in vitro biocatalysis with the power of multi-step in vivo pathways. Such cascades have been successfully applied to the synthesis of fine and bulk chemicals, monomers and complex polymers of chemical importance, and energy molecules from renewable resources as well as electricity. The scale of these initial attempts remains small, suggesting that more robust control of such systems and more efficient optimization are currently major bottlenecks. To this end, the very nature of enzyme cascade reactions as multi-membered systems requires novel approaches for implementation and optimization, some of which can be obtained from in vivo disciplines (such as pathway refactoring and DNA assembly), and some of which can be built on the unique, cell-free properties of cascade reactions (such as easy analytical access to all system intermediates to facilitate modeling).

Keywords Cascade reaction, Combinatorial optimization, DNA assembly, Rational optimization, Scaling, System assembly

Contents

G. Morgado, D. Gerngross, T.M. Roberts, and S. Panke (✉)
Bioprocess Laboratory, Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland
e-mail: sven.panke@bsse.ethz.ch

1 Introduction

Biocatalysis – the catalytic conversion of a chemical compound by an enzyme – has made major contributions to the development of the bulk chemical, fine chemical, and pharmaceutical industries [1]. Even though it has become possible to use enzymes and homogeneous catalysts concomitantly [2], biocatalysis is usually used in isolation as one or a few steps in a chemical sequence. However, enzymes have one unique advantage over chemical catalysts, which is the similarity of reaction conditions – a large fraction of enzymes evolved to operate under the same set of environmental conditions: aqueous media, neutral pH, and ambient temperature. This enables – in principle – the easy installation of systems of enzyme reactions, in which multiple reactions are going on at the same time in the same vessel and thus enables large molecular modifications or the exploitation of a larger set of reactions, as thermodynamically less favored reactions can be combined with thermodynamically favored reactions to obtain high yield.

Of course this is also the operating principle of the metabolism of the living cell, and in fact cells excel at the generation of an amazing molecular diversity from only a few starting materials [3–6]. However, in terms of applications, cell-free systems feature a number of advantages: they do not suffer from additional mass transfer barriers such as cytoplasmic membranes, they can often handle non-conventional solvents better, they do not suffer to the same extent from toxic effects or unproductive reactions with starting material, intermediates or products, and finally they are easier to control, as the composition of the system is under the control of the operator (Fig. 1). In fact, such biocatalytic systems or "cascade reactions" have been in use for synthetic purposes for quite a long time, in particular in the fields of cofactor regeneration [7], the production of monosaccharides [8, 9] or activated monosaccharides [10], oligosaccharides [11], or enantiomerically pure compounds in high yield [12], both in the academic and the industrial domains. Here, we discuss the emerging synthetic biology of such cascade reactions, specifically some recent examples of cascade reactions and methods to design, implement, and optimize them.

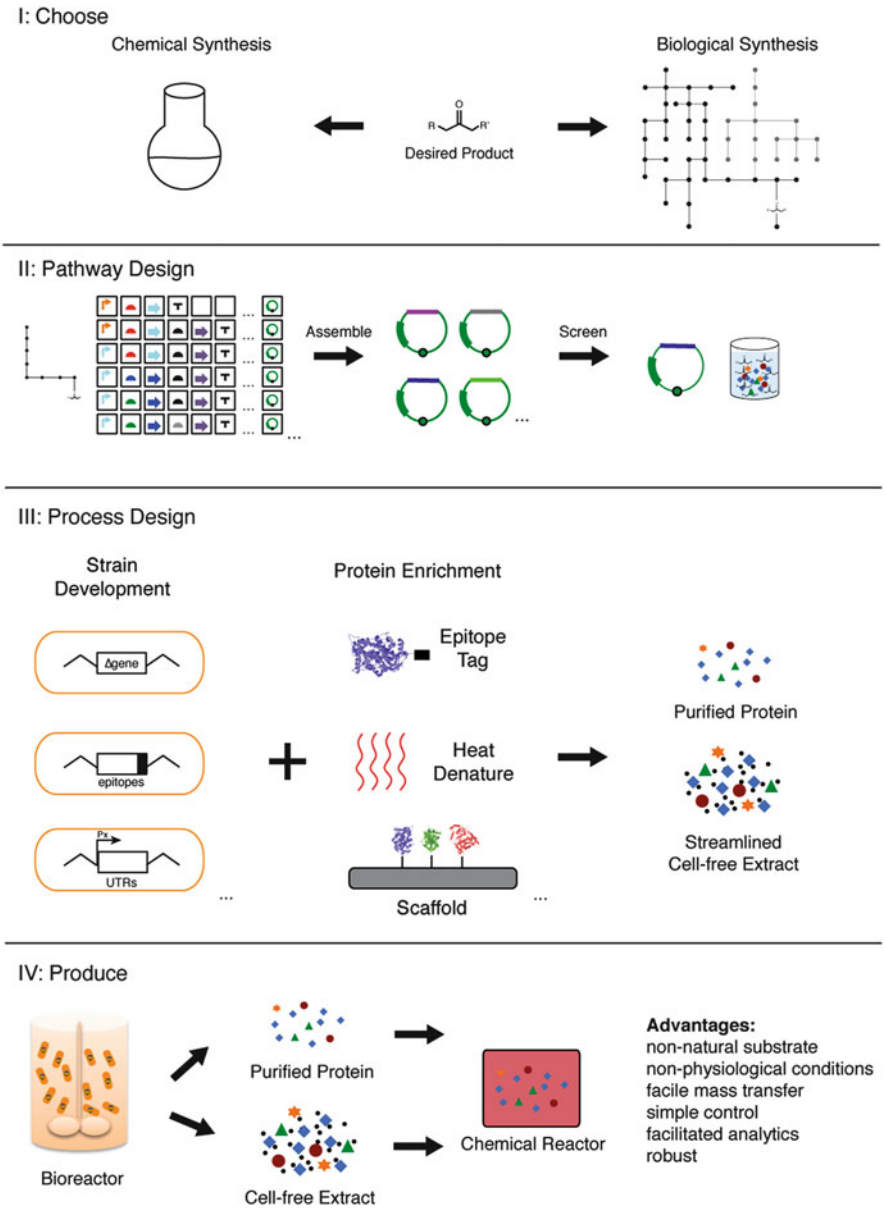

Fig. 1 Design and Implementation of multi-enzyme cascades in fine chemical production. (**a**) To make a given product from a desired starting material one must choose whether to make the product via chemical or biological means. (**b**) If the biological route is chosen, the next step is to identify enzymes that could synthesize the chemical of interest and to choose parts and design constructs accordingly. Here, to obtain specifically the desired enzymes, the host strain can be engineered by inserting the sequence for purification tags into genes encoding desired proteins (e.g., His_6-tag, an epitope for antibody recognition, or a binding domain to arrange enzyme to a given scaffold), or by introducing mutations into target genes to confer thermostability to allow enrichment of thermostable enzymes during heat treatment. Other means of engineering include up or down regulation or deletion of genes, for example the knockout of enzymes catalyzing sink

2 Recent Examples of Cascade Reactions

Obviously, one of the laborious elements of implementing cascade reactions is that several enzymes have to be used and obtaining enzymes can be a very laborious activity, including gene cloning, overproduction, and purification. Consequently, different approaches have been developed that either work with purified enzymes, thus enjoying a maximum of control over the system, or reduce the effort that goes into assembly in a variety of ways. We discuss examples for different approaches.

2.1 Cascade Reactions with Purified Enzymes

A natural field of application for cascade reactions is the fine chemical or pharmaceutical industry, where additional effort (e.g., protein purification) might be a minor inconvenience when compared to system control (including reproducibility, control over yield, or optical purity). An important class of compounds in the field of bioactive substances is the isoprenoids, an inexhaustible source of natural products [13]. An essential element of all isoprenoid pathways are the steps from a starting material – such as phosphoenolpyruvate (PEP) – to a central intermediate in isoprenoid synthesis – such as isopentenyl-pyrophosphate (IPP) or dimethylallyl pyrophosphate (DMAPP) – along one of two possible pathways, the mevalonate or the methylerythritol-4-phosphate pathway. The route along the mevalonate intermediate was implemented as an 11 enzyme pathway to DMAPP including the steps to recycle ATP, NADH, and acetyl-CoA [14]. Remarkably, a 100% yield on the carbon derived from PEP was achieved, and the pathway could be expanded by two more enzymes to produce isoprene. Clearly, making the bulk chemical isoprene from PEP is not a meaningful concept for large-scale chemicals, but it would be for a number of bioactive compounds, and the achieved efficiency was indeed remarkable. A somewhat shorter version of this pathway was implemented starting from mevalonate in four steps to DMAPP and then in three further steps to amorphadiene, a precursor to the antimalarial drug artemisinin [15].

Another important class of compounds as intermediates for drugs is saccharides. As pointed out above, there is a long history of using cascade reactions to synthesize monosaccharides or activated monosaccharides as cascade reactions, for example to obtain unnatural saccharides [8] or as precursors to antiviral drugs [16]. Along these lines, the nucleoside analog 2′,3′-dideoxynosine (didanosine),

Fig. 1 (continued) reactions or the upregulation of genes encoding enzymes involved in bottleneck reactions. (**c**) This is coupled with the design of the process for generating streamlined cell-free extracts or purified proteins. (**d**) Once conditions and constructs have been chosen, the resulting strains can be used for production of the enzymes required for production of a desired chemical

an antiviral precursor, was produced from dideoxyribose using a five-enzyme cascade reaction. Three of the enzymes were required to produce the didanosine and two to recycle ATP from PEP. The pathway was optimized by directed evolution of each of the three main pathway enzymes, which led to a substantial increase in nucleoside production selectivity and productivity. Interestingly, it also allowed shortening the pathway to a four-step cascade, as a mutant version of one of the pathways had a changed selectivity and allowed the elimination of an isomerase from the main pathway [17]. D-Fagomine, an iminosugar with antihyperglycemic effect, can also be obtained in a multi-step one-pot reaction consisting of a four-step cascade and a subsequent separate chemocatalytic step: in the cascade, PP_i-based phosphorylation of glycerol by an acid phosphatase is followed by oxidation of the resulting glycerol phosphate L-glycerol-3-phosphate oxidase (GPO) to obtain dihydroxyacetone phosphate (DHAP) with concomitant inactivation of the side-product hydrogen peroxide by catalase [18]. This is an elegant solution of the DHAP synthesis problem, which has been investigated many times in the past, as DHAP can act as an aldol donor for a variety of enzymes which allow the production of vicinal diols of complementary diastereoselectivity [19]. In the present cascade, DHAP was also used to produce the immediate precursor to D-fagomine by addition of an aldehyde acceptor and using a fructosebisphosphate aldolase.

Finally, on a more preparative scale, D-psicose, a potential replacement for traditional sugar, was produced from sucrose in a three-step cascade employing a hydrolase and two subsequent enzymatically catalyzed isomerization steps [20]. The resulting equilibrium problem was solved by operating the cascade integrated with a continuous chromatography step.

Cascade reactions were even used to produce antibiotics, such as the polyketide-based bacteriostatic agents enterocin and wailupemycin. They were generated fully in vitro by an 11-membered enzyme cascade using malonate and benzoate as starting substrates and reconstituting a polyketide synthase [21].

Although a number of these cascade schemes were carried out merely on an analytical scale, such schemes also play an important role in the design of high-yield reaction schemes for the synthesis of optically pure intermediates for pharmaceuticals. These cascades are typically shorter and therefore also easier to optimize and scale. Recent examples include the production of berbines from racemic benzylisoquinolones by employing an enantiospecific berberine bridge enzyme and the similarly enantiospecific monoamine oxidase-catalyzed oxidation of the unwanted enantiomer to a prochiral precursor, which can again be converted in situ to the racemate [22]. Likewise, a three-enzyme cascade was used to improve the optical purity of 2,5-disubstituted pyrrolidines. Here, an asymmetrically substituted diketone was first converted by an enantioselective ω-transaminase to an amine by reductive amination, coupled to removal of the side-product lactate by two further enzymatic reactions. The amine formed a substituted pyrroline after cyclization and was then converted without diastereoselectivity to a diastereomeric mix of 2,5-disubstituted pyrrolidines. However, by integrating an enantiospecific monamine oxidase, one of the pyrrolidine diastereomers from the mix could be re-oxidized to the pyrroline, which led to a steady enrichment of one pyrrolidine

diastereomer, up to a final diastereomeric excess of 99% [23]. Such examples of resolving enantiomeric or diastereomeric mixes are complemented by cascades in which prochiral starting materials can be converted into diastereomers with some flexibility in terms of stereoconfiguration, as exemplified by a two-enzyme cascade for the formation of either norpseudoephedrine or norephidrine [24].

Cascade reactions are also explored outside the fine chemical/pharmaceutical realm, in particular in the energy domain, underlining the great potential of such systems and also the need for ensuring that the cascades are operating optimally. As discussed further below, problems of providing cascades of multiple enzymes can be addressed by employing thermophilic enzymes. However, moving to in vitro schemes allows moving to chemical strategies which are more suitable than those realized in cells. For example, a cascade comprising 11 enzymes was implemented for the production of ethanol and isobutanol from glucose [25]. The cascade is an optimized version of the Entner–Doudoroff pathway requiring only one cofactor, NAD^+, and not containing any phosphorylated intermediates. Depending on the desired product, the initial four-step cascade (from glucose to pyruvate) is supplemented by a two-step cascade to ethanol or a four-step cascade to isobutanol. The pathway is balanced in view of NAD^+ reduction and NADH oxidation, and resulted in yields on glucose of more than 50% for both products.

An alternative biofuel or polymeric intermediate, 1,3-propanediol, was available from glycerol in a three-step cascade which used hydrogen to close the redox balance [26]. However, more frequently, hydrogen is a target of cascade reactions which aim to provide it from renewable resources, either directly [27–29] or as a reducing equivalent [30]. At the core of these complex cascades with 10–12 enzymes lies a smart combination of the enzymes from glycolysis and the pentose-phosphate cycle to convert glucose and water to hydrogen (through NADPH) and CO_2, allowing quite impressive yields beyond 95% for the case of conversion of xylose to xylitol as a biofuel precursor.

In terms of energy transformation, enzymatic cascade reactions are applicable not only to the synthesis of biofuels or biohydrogen but also to the generation of electricity. The principles outlined above can also be applied to the transfer of electrons to an electrode, generating sugar-based biobatteries of considerable energy-storage densities. As an illustration, a cascade of 13 enzymes was used for the complete oxidation of maltodextrin to CO_2 and water, in the process donating electrons to NAD^+, which in turn transferred them via diaphorase to a vitamin-based electron mediator inside an aerated fuel cell [31].

Finally, enzymatic cascade reactions are also used for the formation of monomers for bulk chemical applications, such as lactic acid from glucose as a renewable starting material. Lactic acid production is straightforward with standard glycolysis expanded by a lactate dehydrogenase to convert pyruvate to lactate. Standard glycolysis would, however, not be balanced in terms of cofactors (netting 2 ATP per consumed glucose), and therefore this specific cascade was equipped with a non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase, resulting in a balanced ten-enzyme cascade with a yield of lactate on glucose of up to 100% [32]. This core cascade was later further extended by another seven enzymes for the

production of *n*-butanol with a molar yield of 82% [33] or by only one additional enzyme to produce malate [34].

2.2 Cascade Reactions Using Cell-Free Extracts (CFXs)

Despite the advantages of cascades of purified enzymes, their laborious implementation might prevent their use. This prompts the question of whether enzymes would actually need to be separated from the cell-free extract (CFX) from which they are obtained after cultivation of the host organism, and in fact many enzyme cascade reactions are implemented with CFXs or perforated cells [25, 35]. This is obviously most relevant for enzymes that are intracellularly produced, for example, in the Gram-negative bacterium *Escherichia coli*, which is in practical terms still a very widely distributed method. Using recombinant enzymes as part of a CFX also has the potential benefit that additional enzymes in the CFX because of the cultivation can also be used for the cascade reaction. In fact, this principle has been widely exploited in cell-free protein synthesis and the synthesis of activated mono- and of oligosaccharides. In the former method, central catabolism, oxidative phosphorylation, transcription, and translation machineries were exploited to synthesize proteins [36], with a broad variety of applications [37]. In fact, CFX-based protein synthesis was used to produce a variety of biopharmaceuticals such as vaccines against influenza [38] and lymphoma [39], antibodies [40, 41], cytokines [42], and natural [43] and synthetic viruses [44, 45].

In saccharide synthesis, the surviving metabolism of perforated cells was used to regenerate cofactors and provide starting materials by conversion of orotic acid and glucose [10]. Of course, the additional functionality can also have negative effects. For example, the intrinsically complex nature of CFXs makes it difficult to achieve reproducibility. However, careful optimization can reduce the corresponding problems [46].

Interestingly, a conceptually novel enzyme cascade to DHAP was also implemented, starting from CO_2 and employing a novel, computationally designed formolase to convert formaldehyde into DHAP [47]. Even though this cascade was not balanced with respect to cofactor regeneration, it provides an interesting perspective on the integration of computationally designed enzymes (which bring novel reactions to biochemistry) into in vitro pathways.

Bioactive compounds were also synthesized with CFXs, specifically to confirm the biosynthesis of specific natural compounds such as the antitumor agent azinomycin B [48]. In this study, cell lysates from *Streptomyces sahachiroi* were sufficiently powerful to complete full one-pot biosynthesis of both naphthoate (a known pathway intermediate) and azinomycin B itself. This was a remarkable achievement given the chemical complexity of both compounds. The authors also tested a battery of inhibitors and amino acids to clarify the substrate and cofactor requirements within the pathway and gained some insight into the mechanism of azabicyclic ring formation, which is believed to arise from ornithine.

3 Assembling Cascade Reactions

As already mentioned, using enzyme cascades instead of, for example, living cells entails the additional effort of assembling the cascade. This means at least perforating cells or producing cell lysates, and can go as far as purifying enzymes and then combining them to the desired system. Although perforating cells has turned out to be useful even on an industrial scale [11], we do not discuss it further here but rather refer to seminal reviews that summarize efforts nicely [49, 50]. A number of alternative strategies are discussed below (Fig. 1).

3.1 Heat Purification

A straightforward and cheap method to purify enzymes to a useful extent is to rely on recombinant thermophile enzymes in a mesophilic host, such as *E. coli*. Here, cell lysates are prepared after the induction of expression from recombinant genes and a heating step is applied, during which most native proteins are deactivated and precipitated, but not the heterologous enzymes. An interesting side effect of this approach is that an increase in enzyme thermostability is often associated with an increase in process stability, which is of course beneficial and often essential to process economy [51, 52]. On the other hand, the specific activity of thermophilic enzymes tends to be optimal under those environmental conditions which are optimal for the host and which might or might not coincide with the optimal temperature for operating the cascade reaction. For example, the previously discussed DHAP is a rather labile product of a cascade reaction and producing it at higher temperatures would not be favorable. Consequently, processes relying on thermostable enzymes are particularly useful if the actual cascade reaction is also expected to operate at higher temperatures, for example, to prevent microbial contamination in a large-scale process, as would be the case for production of biofuels or rare sugars. In agreement with this, the already mentioned cascade reaction for the formation of ethanol or isobutanol was assembled from such thermostable enzymes and the process operated at 50 °C [25]. Similarly, the lactate, malate, and butanol production cascades discussed above were assembled from thermophilic enzymes which were separately cloned in *E. coli* strains and then prepared by heat treatment [32–34].

3.2 Affinity Tagging

An alternative that allows purifying multiple proteins in only one step is systematically equipping cascade members with tags that allow affinity purification of another form of separation, such as precipitation [53]. However, purification is

often either expensive (e.g., when considering the popular six-histidine tag (His-tag), which requires adsorption on Ni-nitrilotriacetic acid-coated surfaces) or associated with constructing protein fusions with large domains, which enforce purification of a substantial part of unwanted material. Inactivation of the target enzyme by the tag may also occur. Nevertheless, a number of studies made extensive use of the His-tag, including the biosynthesis of UDP-galactose through a seven-enzyme cascade reaction, in which all the enzymes carried a His-tag [54]. After purification, the enzymes were also immobilized on Ni-covered agarose beads and a higher production yield (~50%) was observed compared to the free protein counterpart. Another example of exhaustive usage of His-tags was provided by the reconstitution of complete biosynthetic pathways for purine [55] and pyrimidine [56] from glucose, ammonia, carbonate, creatine phosphate, α-ketoglutarate (and, for purines, serine) as one-pot cascades comprising 28 and 18 enzymes, respectively, including up to five cofactor regeneration cycles. Along the same lines, the approach was used to provide an alternative to the classical CFX-based cell-free protein synthesis systems already mentioned by providing all protein elements in a purified form. For that, 38 essential genes were extended (distributed over multiple strains) to include the sequence for the His-tag [57]. This initial step was facilitated by applying oligonucleotide-mediated mutagenesis [58], which allows the expansion of genes directly on the chromosome.

3.3 Streamlining CFXs

When considering one of the main disadvantages of CFXs – increased complexity – then one alternative to purifying enzymes from complex CFXs is simplifying the CFX, for example by removing known interfering functions to prevent the consumption of starting material, intermediates, or products. Over the years, many interfering functions, for example in cell-free protein synthesis [59] or saccharide synthesis [9], were identified and removed. This was also done recently for another attempt at addressing the already discussed DHAP synthesis problem: DHAP can be produced from glucose in four steps by means of standard glycolysis plus an enzyme such as glucokinase that allows the phosphorylation of glucose with ATP. ATP can be regenerated by employing the lower part of glycolysis (five enzymes) and lactate dehydrogenase (to regenerate NAD^+). However, in CFXs ATP is degraded to ADP and AMP and then hydrolyzed to adenine and ribose phosphate by AMP nucleosidase. Deleting the corresponding gene substantially improved ATP regeneration [60].

Gene deletion is straightforward and a variety of methods are available to implement it [61], but it cannot be used if the interfering function is either essential or of major importance for normal growth behavior. Then conditional removal becomes interesting, as exemplified for in vivo conditions for improving *myo*-inositol formation after inducing degradation of a key enzyme of glycolysis [62]. One suitable strategy to achieve this is to tag proteins genetically with a

"degradation tag", in fact a copy of the sequence that *E. coli* cells use to mark proteins, whose translation was prematurely terminated, for degradation by the ClpXP machinery [63] with the help of a adaptor protein called SspB. The intracellular level of proteins whose functional half-life was shortened in this manner can then be controlled either by stopping induction of the corresponding gene and relying on accelerated clearing of the gene product from the cytoplasm, or by inducing the adaptor and in this way accelerating degradation [62]. By timing the preparation of CFX suitably, such strategies can also be used for streamlining CFXs. Similarly, proteins could be equipped with specific proteolysis tags and the tagged proteins removed by selective proteolysis. This strategy is well-established for TEV-directed intracellular [64] or extracellular [65] hydrolysis.

3.4 Scaffolding

Multistep cascades generate intermediates which need to reach high concentration levels before the subsequent enzyme can operate under conditions of maximum rate, requiring the cascade to operate at a high overall concentration of chemical compounds, which might interfere with enzyme stability. This problem can be reduced by providing the separate enzymes of a cascade as part of a spatially organized complex on a scaffold rather than as independent units. In such complexes, apparent concentrations of starting materials and intermediates are higher as the active sites of two enzymes are in close proximity, so faster catalytic rates are reached at (in terms of averaging across the reactor volume) lower concentrations. Organized complex formation can be achieved by fusing enzymes to proteins that bind to a suitable scaffold, made of either protein [66], DNA [67], or RNA [68]. Exploitation for cascade reactions has focused on the highly versatile cellulosome scaffold [69] which allows interactions with different fusion partners and was exploited to scaffold a three-step cascade to produce fructose-6-phosphate from glyceraldehyde 3-phosphate [70].

4 Encoding Reaction Cascades

A crucial aspect of cascade reactions is that they operate as a system in which different feedback mechanisms lead to behavior that is not necessarily intuitively accessible. This might be on the level of productivity, where it turns out that the intermediate from a downstream reaction acts as an inhibitor of an upstream reaction, as is often the case in metabolic pathways. This might, however, also be across different levels of the implementation process, when it turns out that by assembling genes into an operon the specific combination of DNA sequences inadvertently introduced an additional promoter structure that changes expression behavior [71] and thus influences the composition of the cascade after heat

purification. One aspect of this system concept is that scaling is far from trivial for cascade reactions, as problems with single elements that are relatively straightforward to overcome in a one- or two-member reaction can aggregate in a system to an extent that it becomes difficult to deconvolute and address them efficiently. In other words, setting up cascade reactions can benefit from a rational design approach, in which the various steps of the implementation proceed along a rational, possibly standardized, and computationally supported sequence of steps (Fig. 1). Of course, this design process does not yet exist [72, 73]. However, across the different steps a variety of tools have been introduced which provide valuable assistance, and we discuss the design tools and strategies for the optimal assembly of cascade reactions in the following sections.

In these sections we follow the process from selecting the enzymes for the pathway to the final construction of a DNA molecule (or a few DNA molecules) that actually program the synthesis of the cascade into a bacterial cell (Fig. 2). For this, we rely heavily on the methods that have been implemented to set up in vivo pathways, for the simple reason that in a first approximation the required steps are very similar. We make this argument on two levels. The first level is enzyme selection. When considering in detail the examples for successful cascade reactions, it becomes clear that the long cascades in particular exist in the majority of enzymes that were not particularly engineered for a specific cascade. In other words, although the engineering of single enzymes remains crucial for the success of the overall cascade, the "backbone" usually consists of enzymes used in a function well known in standard biochemistry. In this contribution we therefore acknowledge the crucial roles that computational enzyme design [74] and directed evolution [75] play in the conversion of crucial reactions into cascade reactions, but we focus on the systems aspect and discuss tools that allow the assembly of systems of enzymes rather than the design of single enzymes.

The second level is system composition. Ultimately, the performance of a cascade reaction depends on its composition. As an in vitro pathway, the productivity of the pathway is subject to metabolic control by the various members [76], which in turn is exerted through kinetic parameters such as substrate and enzyme concentrations, affinities, allosteric interactions, cooperative behavior, degradation constants, etc. [77]. When considering the different methods to assemble cascade reactions (except for scaffolding), it becomes clear that the composition of the cascade is determined by the genetic construct assembled to express the encoded genes. Furthermore, it seems reasonable to assume that it is preferable to produce a cascade reaction with only one cultivation, in which the cells synthesize all the different enzymes to optimal levels, over many separate cultivations, in which each cell overproduces only one enzyme. In other words, the task at hand is to assemble the genes of a cascade reaction in one or a few bacterial operons, tune the gene expression to the optimal level, and, if necessary, optimize the performance.

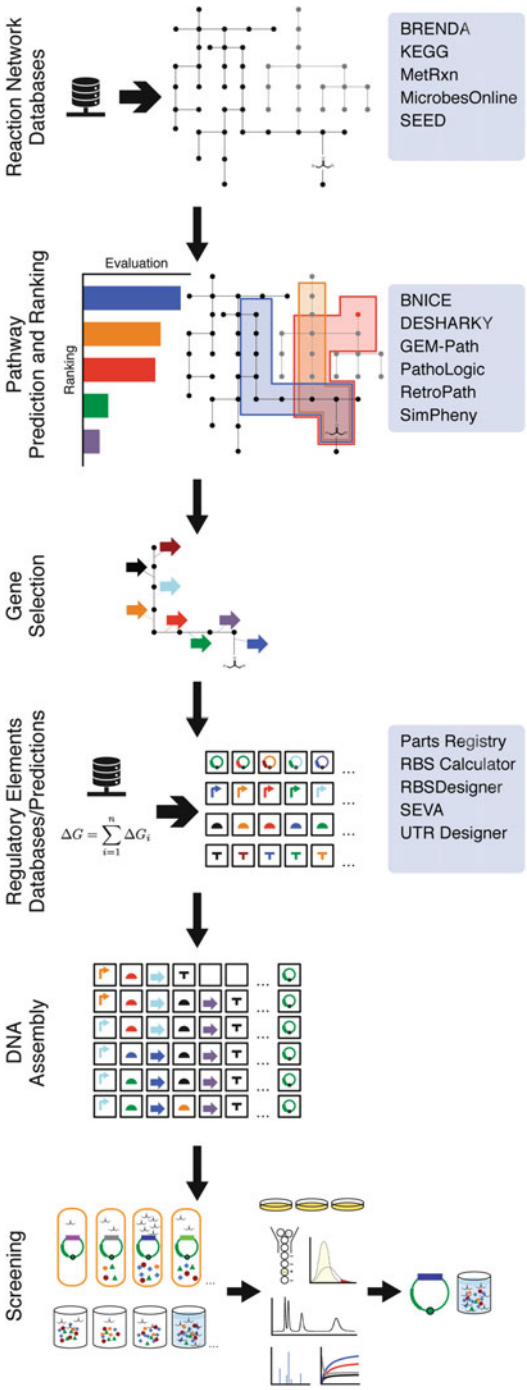

Fig. 2 Workflow of computer aided design of cascade reactions. Reaction network databases allow the exploration of known and characterized pathways and provide a basis for pathway prediction algorithms which propose an optimal route to a compound of interest. Algorithms that operate with reaction types allow in different ways the introduction of novel reactions and suggest potential enzymes that could be evolved toward this step and help to rank different possible pathways. When the pathway is to be assembled in one cell, genes are selected and tools are collected to allow the variation of the relative and absolute amounts of pathway members. Tools to support the generation of variation are available. Finally, the different elements are assembled into a DNA molecule and the resulting library is screened for optimal performance. For details and references, see main text

4.1 Computational Design of Novel Reaction Pathways

An increasing number of computational tools are available to obtain the enzymes that formally allow to connect a starting material with a desired product through a set of already demonstrated or hypothetical reactions [5, 78–80]. Databases such as BRENDA, KEGG, MetRxn, MicrobesOnline, and SEED [81–85] allow the exploration of known organism-specific metabolic routes and navigation through non-native combinations of these individual reactions that lead from a starting metabolite to an anticipated destination compound. This allows the quick identification of potential reaction networks based on known and well-described metabolites and enzymatic reactions.

However, there are a number of challenges involved: a variety of pathways are often possible from starting material to product and the optimal path is not necessarily obvious. Sometimes the opposite is true and chemical conversions are required for which no biological counterpart is known. When the enzymes for the cascade reactions are produced in one host, then they can effectively be seen as an in vivo pathway, and possible intermediates and products of this artificial pathway might interfere with the reactions in a host organism. The selected enzymes (or the host's enzymes) might display poor selectivity, leading to unanticipated consequences in the host. Finally, if CFXs are applied, cellular pathways might direct intermediates into unproductive side reactions [86]. These influences require tools that allow pathway enumeration together with evaluation of potential and possible consequences, and ideally inclusion of novel reactions, which are similar to but distinct from those already available in databases.

A number of methods have been proposed to identify these types of de novo pathways, including enzymes which are supposed to catalyze the new reaction steps, based on similar known reactions. They all revolve around a concept of a formalized representation of enzyme-catalyzed reaction which allows abstraction from a specific substrate/product couple (to include novel reactions based on known reactions but with novel substrate specificity) and concomitantly mathematical representation to allow computer-supported network generation. BNICE [87], for instance, predicts pathways by combining pathways according to the first three of the four possible layers of the enzyme classification system [88], which identifies enzymes by reaction type but not by detailed substrate specificity. Effectively, this allows representation of starting materials, products, and enzymes/reactions by bond-electron matrices, which can be systematically transformed and existing reaction types can be applied to novel substrates. The system was expanded to include thermodynamic considerations for ranking [89] and additional layers in which predictions about particularly suited enzyme scaffolds and specific suggestions for enzyme engineering were included [90].

Alternatively, reactions that are supposed to be considered for generating the network of possible reactions connecting a starting material and a product can be formulated by a limited number of reaction rules, which represent a large fraction of reactions collected in a central database such as KEGG at a level below substrate specificity [91]. Enzymes are also associated with reaction rules, and in this way a

network of potential enzyme reactions is implemented and in a next step evaluated based on local similarity between molecules, similarity between entire structures, thermodynamic feasibility, pathway distance, and the network of the host organism. For scoring, the importance of the different elements can be calibrated with training sets.

Finally, conversions of compounds can also be implemented by transformation of molecular signatures. Depending on the resolution with which molecules are to be represented, the molecular signatures can be computed at different heights, resulting in reaction networks of increasing size for decreasing height. Once the reaction space is defined in this way, pathways are ranked again according to thermodynamics, enzyme availability or selectivity, and product toxicity [92]. Combinations and variations of these approaches which use different criteria for pathway scoring have also become available (Fig. 2) [93–97].

In summary, a number of tools to predict feasible pathways are available which integrate a large number of criteria into selecting an appropriate pathway and apply at least formal methods to suggest novel enzyme reactions that are required to obtain a pathway. These suggestions can then be followed up by more sophisticated methods, or alternatively novel reactions, such as computationally designed reactions, can be included in the definition of the network which is subsequently scored to identify the most promising pathway. Although criteria such as impact on the growth behavior might not be a prime concern for implementing cascade reactions, the integrated scoring criteria in general are very helpful for pathway selection.

4.2 Controlling and Optimizing System Composition

Practical implementations of novel reaction systems do not easily reach satisfying productivities and require optimization when progressing from an initial concept of a biocatalytic network eventually to an applied system creating economic value. As pointed out before, the crucial factor for the optimization is control over relative and absolute protein levels of each introduced enzyme [98]. Conceptually, there are two different approaches: optimization by rational or combinatorial approaches, and both have been implemented.

Rational approaches include mostly straightforward analyses to identify rate-limiting steps by evaluating the effect of systematically increasing the concentration of one of the cascade members. Even though the idea of the rate-limiting step can be misleading in pathways [99], the approach as such is frequently applied, particularly if the cascade members are available in purified form. For example, in the aforementioned formolase cascade for the formation of DHAP [47], titration experiments were used to identify rate-limiting steps. When the enzymes are understood well enough, kinetic modeling might be used to support the identification of the limiting step. For example, the production of hydrogen from cellulose was optimized using a model built on rate equations for the involved enzymes and adapting the kinetic parameters to the observed experimental data

[29]. Alternatively, statistical approaches can be followed: the seven-step cascade to produce amorphadiene from mevalonate discussed above was optimized under two-phase reaction conditions for several variables such as absolute and relative enzyme levels, different types and concentrations of monovalent ions and magnesium ions, and the influence of pH [15]. The authors systematically evaluated 16 different combinations of enzyme levels designed with a factorial orthogonal array and response surface methodology, and also identified the reactions consuming farnesyl pyrophosphate and producing amorphadiene as bottlenecks.

A different rational approach was taken for the discussed cascade to produce lactate. Here, the authors conducted a series of initial experiments in which they expressed all cascade genes separately in a recombinant strain and determined mRNA level and specific productivity of the recombinant enzyme. In this way, they could obtain a rough indication of which cascade member would require most overproduction, and then they used this information for arranging the corresponding genes into an operon under the hypothesis that the first gene in an operon is the most heavily expressed [100]. The resulting CFX was indeed fourfold higher than when the same biomass was assembled from strains which overexpressed only one gene.

Although all these approaches are straightforward and also effective, in their focus on the formation of the end product they do not fully reflect the systems character of the cascade and it remains unclear whether the improvements they suggest can be transferred to larger scales. That would be helped if the analyses included a more comprehensive record of concentrations, including those of the intermediates. This has become possible by integrating online mass spectrometry into a continuous reactor setup, which has allowed monitoring of the compositions of substrates, most of the intermediates, and products in response to additions of enzymes to optimize the formation of DHAP [101] (Fig. 3). After three rounds of optimization of the operon design, DHAP productivity was improved 2.5-fold.

The alternative to rational approaches to system optimization are combinatorial approaches, which are meant to refer to approaches in which the performance space of a system is explored by a non-biased recombination of cascade elements and the subsequent identification of the best performing composition. Although this can also be done, in principle, with purified enzymes, our ever increasing proficiency in manipulating DNA molecules provides another straightforward approach which transfers the laborious implementation of combinatorial schemes from, for example, a robot (variation of enzyme concentrations according to a pre-programmed pipetting scheme) to chemistry (variation of enzyme concentration by varying transcriptional and translational signals). This coincides with the fact that controlling intracellular protein level has been a main target in synthetic biology in recent years.

The corresponding efforts were undertaken at various levels (Fig. 2), including efforts to standardize the construction process as well as the construction of suites of parts and tools to implement diversified protein production signals. The Registry of Standard Biological Parts (http://partsregistry.org) with its BioBrick™ standard for biological parts, for instance, provides a collection of genetically encoded parts

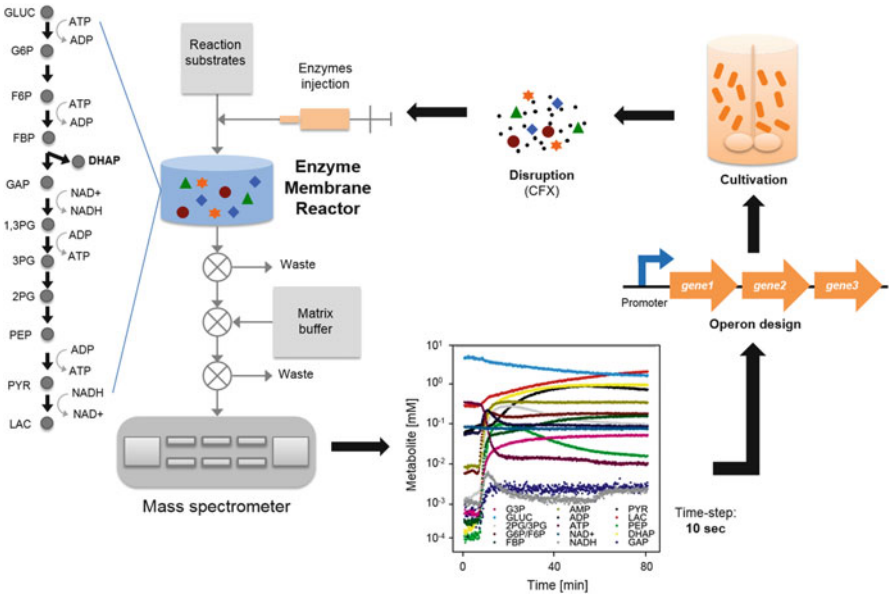

Fig. 3 Workflow for a semi-rational cascade optimization. A cascade (here for the production of dihydroxyacetone phosphate (DHAP) based on standard glycolysis with recycling of ATP and NAD^+) is assembled in an enzyme membrane reactor, where it can be freely perturbed with substrates and compounds. In this example, the cascade is part of a cell-free extract. The effluent is immediately analyzed by mass spectrometry at a high time resolution (multiple measurements per minute). This way, not just the changes in concentration of the target compound (DHAP) but of the entire spectrum of compounds can be followed. This system can be used to add enzymes selectively to identify cascade compositions that operate with an improved productivity, which can be used to inform the change in expression signals of key genes, so that the next cell-free extract approximates more closely the previously identified best composition

together with rules for physical composition and guidelines for functional composition and characterization [102, 103]. Another community driven standard is the Synthetic Biology Open Language (SBOL) [104, 105], which is a data standard aiming toward facilitating design and exchange of novel biological systems. Vectors were also subjected to standardization efforts, leading to the Standard European Vector Architecture (SEVA) [106, 107], which provides rules for the construction and nomenclature of prokaryotic plasmids together with an online and a physical database of various characterized vector designs.

The SEVA format allows a quick exchange of plasmid origins of replication, and the resulting changes in plasmid copy number (between 1 and several 100s [108]) and consequently gene dosage are a first important factor in governing protein levels [105, 109–111]. However, gene dosage only sets a baseline for DNA levels. Promoter activity controls the rate at which mRNAs are produced, and several attempts were undertaken to systematize the measurement of the somewhat poorly defined parameter “promoter strength” [112–117]. Ultimately, this resulted in the creation of several sets of promoters, which allow the tuning of protein expression

on the transcriptional level [116, 118–122]. Having access to characterized promoters with different strengths facilitated combinatorial screening approaches to optimize the expression levels of a multi-enzyme pathway [118, 119, 123].

Ultimately, large cascade reactions need to be encoded in oligo-gene operons, which calls for another layer of regulation at a finer resolution than the promoter, for example translational signals specific to each gene. Advances in biophysical models describing interactions of the ribosome with the 5′-UTR, that is, the ribosome binding site (RBS) involved in translation initiation, provide the means to control protein production over several orders of magnitude in a predictive manner [124–127]. On the basis of the Gibbs free energy difference between the folded mRNA and the assembled translation initiation complex between the 30S ribosomal subunit and the mRNA transcript, those thermodynamic models predict relative expression levels only based on the mRNA sequence. Consequently, providing a broad variety of protein synthesis signals can be achieved with reduced effort by using these models for RBS design, even if they are not accurate enough to allow direct design (Fig. 2).

One of the advantages of targeting the RBS for optimization of expression levels is that large changes in protein translation result from only small changes in the RBS sequence [128], thus simplifying the generation of libraries covering large portions of the accessible protein expression space. Moreover, based on available model predictions, the RBSs can be forward engineered and evaluated in silico, allowing a focused search for optimal expression level combinations [129]. For instance, Zelcbuch et al. [130] used the forward engineering capabilities of the RBS Calculator [125] to define a small set of RBSs that potentially covers a large range of expression levels. Applying this set to a combinatorial library for a branched carotenoid biosynthesis pathway revealed a large diversity of carotenoid productivities, in one instance outperforming previous pathway engineering efforts by a factor of four for the production of the industrially valuable compound astaxanthin. Nowroozi et al. [131] also included the use of the RBS Calculator predictions in their efforts to construct a combinatorial operon library of isoprenoid production pathways to improve the production of amorphadiene in *E. coli*. Again because only a few base pairs need to be changed, this method is also very suited to changing the expression level of genes located on the chromosome. For example, a synthetic Entner–Doudoroff pathway was introduced into the genome of *E. coli* [132] and the RBS signals were optimized in a combinatorial fashion using oligo-mediated recombineering-based combinatorial RBS screening to optimize the pathway's operon, which led to an increased NADPH regeneration rate. Similar approaches of combinatorial modulation of RBSs within an operon-encoded pathway had also been previously applied [133–135].

Another method to introduce individuality into operons is to produce transcripts of different lengths from the same promoter by inserting transcriptional terminators of intermediate strength into intergenic regions, so that downstream recombinant genes are sometimes part of the transcript and sometimes not, again contributing to the control of the cellular protein level. On the other end of the scale, preventing read through from a heavily transcribed operon into downstream sections of the

genome or a plasmid is desirable and requires efficient terminators. This led to the development of a variety of suites of terminators to address such needs. Cambray et al. [136] developed a genetic architecture that enabled a reliable determination of terminator efficiencies. Chen et al. [137] used a similar approach to characterize a large library of terminators for use in synthetic systems.

The tools mentioned above, combined with efforts to use DNA synthesis to remove known or opaque gene-internal regulatory sequences, are often considered the toolbox of pathway "refactoring" [138], in which a pathway of potentially diverse origins is taken out of its native regulatory context and recast into a format in which its performance is optimal from the point of view of the operator. Clearly, the criteria that apply to the in vivo setting in which this strategy is typically applied and to the in vitro setting of cascade reactions are very similar, so that refactoring is also a promising tool in cascade reaction optimization.

4.3 DNA Assembly

Once the different parts required for refactoring the pathway are available, they need to be assembled into one or a few operons [61]. Laboratory assembly of DNA has been around for several decades [139]. However, it has recently become faster and simpler, and has allowed for the construction of complex constructs [140]. Here, we contrast and compare some of the available methods. Current large-scale DNA assembly methods fall into three main categories: (1) those based on rounds of restriction digestion and ligation, and both (2) in vitro and (3) in vivo homology-based methods.

Methods that fall into category 1 include BioBrick, BglBrick, and Goldengate cloning [141–143]. In general, these methods work by the PCR of modules (or parts) with primers containing the required restriction enzyme recognition sites at the 5′ end. PCR products are then purified, cut, and ligated into a vector conforming to the BioBrick, BglBrick, or Goldengate standard. In the case of BioBricks and BglBricks standards it is critical that the 3′ end of part I is cut with a different enzyme than the 5′ end of part II, but that these two restriction enzymes yield compatible ends that, once ligated, no longer contain a restriction enzyme site. This allows for continual digestion and ligation of additional modules, which is amenable to the recursive methods of automation. However, one of the major disadvantages of the BioBrick and BglBrick methods is that a scar is formed at the junction of each module. In the case of BioBrick this is an 8-bp scar, limiting the use of BioBrick assembly to larger, non-coding regions. This issue has been improved upon with the BglBrick standard by utilizing *Bgl*II/*Bam*HI restriction sequences that result in a 6-bp scar and the introduction of a glycine-serine dipeptide when placed into a coding region [141]. Furthermore, the assembly of the parts is performed sequentially such that assembly of ten modules would require ten individual rounds of cloning, which can be costly and time-consuming. However, parallel assembly of parts (in multiple pots), the use of repeating parts

(promoters, RBSs), the simplified design of sequential additions of parts by algorithms [144], and liquid-handling robotics [145] can help to minimize these burdens.

Other in vitro methods, such as Goldengate cloning, can be performed in parallel and entirely avoid the introduction of a scar by using TypeIIS restriction enzymes which cut outside their recognition sequence [142]. This allows for assembly of multiple parts at once because the cut sequence can be unique to each module, allowing for ordered assembly of parts. As many as 68 parts have been successfully assembled by Goldengate cloning in three one-pot assembly reactions [146]. Similar to the other category 1 methods, Goldengate cloning requires that all parts being assembled are free internally of the enzyme recognition sites. In most cases this can be accommodated by introducing small changes in the nucleotide sequence of the modules to remove conflicts. However, if these sites occur in coding or regulatory regions, it can lead to changes in expression of proteins of interest.

So-called sequence independent methods overcome these restrictions by using homology-based assembly, which involves the in vitro resection and annealing of homologous regions (category 2) or in vivo homologous recombination (category 3). An advantage of these methods over category 1 methods is that it is unnecessary to alter the sequence of any of the parts being assembled. In vitro homology methods include circular polymerase extension cloning (CPEC) [147], sequence and ligase independent cloning (SLIC) [148], and Gibson or isothermal assembly (ITA) [149].

All category 2 in vitro methods require that the modules of interest end in sequences homologous to those on the ends of the neighboring modules. These "overlapping" modules are in general constructed by PCR with primers adding the homologous region to the 5′ end. All parts to be assembled are mixed together and incubated in an annealing mixture in which the homologous regions are exposed as single-stranded units and the opposing ssDNA strands are annealed.

In the case of CPEC, a form of overlap extension PCR, this occurs by rounds of denaturing, annealing, and extension in which the annealed vector and insert use each other as a template until the construct is circularized. The final plasmid contains two nicks that are repaired upon transformation. Successful assembly by CPEC relies on the identical melting temperature (T_m) of overlapping regions of adjacent modules [147]. Up to four modules generating an 8.4-kb construct were combined by CPEC. Assembly of a greater number or larger versions of modules would likely require other cloning methods [150]. The advantage of CPEC over other in vitro assembly methods is that it uses common PCR reaction components and does not require expensive kits or enzymes not already found in most laboratories.

In contrast to CPEC, SLIC and ITA use exonucleases to resect the dsDNA by T4 DNA polymerase and T5 exonuclease, respectively, thus generating extended single-stranded regions to support the annealing process of the homologous regions. In SLIC, once the homologous regions are annealed, only one dNTP is added to arrest the exonuclease activity of T4 DNA polymerase [148]. This results in a construct with nicks or gaps on either side of each fragment which are repaired upon

transformation. In ITA, a similar principle applies but ligation is integrated into the in vitro step. In one step, T5 exonuclease, Phusion polymerase, and Taq ligase are mixed. The T5 exonuclease resects the dsDNA to allow annealing, whereas Phusion polymerase fills in the gaps and the ligase seals the nicks. In essence, the method relies on the balance of activity between the polymerase and exonuclease (with the exonuclease being ultimately heat-inactivated at the standard temperature for this step) [149]. SLIC and ITA can be used to generate constructs containing ten modules, can vary greatly in size, and as such are preferable over CPEC for the generation of complicated constructs with many parts.

It is important to note that the USER method [151] and a hierarchical method similar to SLIC [152] are also very useful methods of ligation independent cloning; however they are not discussed in this chapter as they are not entirely sequence independent.

The main disadvantages of category 2 methods lie in the requirement for homologous regions. It is important that these overhanging regions are free from secondary structures which could otherwise hamper their annealing to the neighboring fragments. If secondary structures are unavoidable, it may be beneficial to choose a method with a higher reaction temperature, as the proposed annealing temperatures are different for the three presented methods (SLIC: 20 °C, ITA: 50 °C, CPEC:55–65 °C).

Category 3 or in vivo homology methods take advantage of the inherent DNA repair and homologous recombination machineries of the yeast *Saccharomyces cerevisiae*. The parts to be assembled are transformed into yeast together with a shuttle vector containing a yeast origin of replication and selection marker and cells are plated onto selective media. As with other in vitro homologous methods, the modules are generated by PCR with oligonucleotides containing the overlapping regions. Modules consisting of double-stranded DNA are then transformed into the yeast and ssDNA regions are exposed by the yeast exonuclease. The exposed ssDNA is then bound by RPA (yeast single-stranded DNA binding protein) resolving any secondary structures [153]; double-strand break repair mechanisms subsequently join the homologous regions together, generating a plasmid that can express the selection marker [154]. To reduce false positives it is important to choose a selection marker that does not contain homologous sequences in the host strain. This method has been successful with up to 38 pieces at once [155].

Although easy to design and relatively simple to implement, one of the major disadvantages of the use of yeast assembly for library generation and then expression in other hosts is the limited recovery of the plasmid DNA from the yeast cells (~1 ug of DNA per 10^{10}–10^{12} cells [156]). Additionally, it is critical that the assembled construct does not result in toxicity in yeast. Furthermore, it requires that the vectors contain replication sequences and resistance markers for both hosts, or require sub-cloning into an appropriate vector after assembly, which would require that the final assembled fragment is free from the enzymes required for sub-cloning.

For both category 2 and 3 methods it is critical that these homologous regions are unique to each part, otherwise unwanted assembly can occur leading to constructs

with parts lacking or assembled in an incorrect order. As such, the major advantage of category 1 methods over homology-based methods is that repeating elements can be easily used without any undesired fragment generation.

In general, each method has inherent advantages and disadvantages, and to generate complicated assemblies it may be beneficial to combine various methods to achieve the desired construct. ITA and yeast assembly have been used in combination to clone the *Mycoplasma genitalium* genome [157], and ITA, a "scarless-stitching method," and Goldengate were used in combination to refactor the nitrogen fixation pathway of *Klebsiella oxytoca* [138].

In summary, a great deal of effort has been invested in developing techniques for rapid, efficient, and easy assembly of desired constructs. Yeast assembly and ITA have been used for library generation of yeast plasmids to optimize production of xylose or components of the violacein pathway, respectively [118, 119, 158]. Overlap extension PCR has been used to generate a library of the mammalian calmodulin central linker for expression and purification in *E. coli* [159]. Fragment exchange, which combines restriction digest with TypeIIS restriction sites and annealing of homologous regions, has been used to screen for novel bioactive agents in *E. coli* [160, 161]. It remains to be seen whether these methods can also be used for efficient generation of libraries for production of desired products in *E. coli*.

In addition to methods that assemble plasmids, genome-based efforts have also contributed to the engineering of strains that can be exploited for use in in vitro systems. Traditionally these methods have included random radiation or chemical mutagenesis followed by rounds of screening. However, recent advances in large-scale genome engineering allow for a directed approach. The method relies on the co-expression of the lambda-red recombination system from bacteriophage and the transformation of short DNA sequences bearing the desired mutation into *E. coli* [162, 163]. These oligonucleotides can then act as Okazaki fragments on the lagging strand which introduce mutations during replication [58]. Efforts to automate oligo-mediated recombineering were undertaken, which allow targeting of multiple sites at once in a relatively high-throughput manner. This has been used to generate strains for enhanced product production by modification of ribosome binding sites [58] or promoters [164]. Furthermore, it has been used to insert hexa-histidine tags to allow for facile purification of the components of an entire pathway [57]. Furthermore, the efficiency of such an approach can be increased by co-usage of CRISPR-Cas9 to remove unwanted parent genotypes, thus specifically enriching engineered sequences [165]. Together, the described in vitro and in vivo DNA assembly and engineering efforts can result in optimization of pathways to be used for in vivo, cell-free extraction or in vitro applications.

5 Summary

In this chapter we have summarized a number of examples of cascade reactions for products with very diverse applications, ranging in scale from small (pharmaceutically active ingredients) to large (biofuels) and in nature from simple (molecular hydrogen) via stereochemically challenging (optically pure fine chemicals) to complex polymers. For all these examples, the attractiveness of the approach lies in the fact that multiple enzymes are brought together without undesired spatial separation in one vessel and under the same set of environmental circumstances, allowing the efficient performance of complex chemistry including the use of thermodynamically unfavorable reactions. This focus on multiple members of the reaction changes the nature of the operation to that of a system with emerging properties. Consequently, the methods applied to constructing and optimizing such enzyme cascade reactions need to be adapted, from identifying suitable members of such systems via assembling to optimizing them. Many of the required methods can be obtained from in vivo synthetic biology and its efforts of pathway refactoring. We have over the years become very good at manipulating DNA, and exploiting this for exploring the best performance of a cascade reaction by integrating its in vivo production with broadly different compositions seems a natural way to optimize cascades. However, the cell-free character also allows the introduction of novel elements such as advanced analytical strategies to track the performance of cascades.

Clearly, cascades, in particular larger cascades, need to be scaled beyond the available few examples [11, 12] to demonstrate their ultimate value for (bio) chemistry beyond proof of principle. However, very promising approaches, for example using thermophilic enzymes, are available, which suggests that, ultimately, scaling issues can be overcome. Here, it might serve to recall that enzyme processes are among those bioprocesses with the highest product volume (production of high fructose corn syrup employing glucose isomerase, annual production volume of 10^7 tons per year [166, 167]). This augurs well for the future of this promising approach.

Acknowledgment The authors wish to acknowledge financial support from the EU project ST-FLOW (#289326), and the Swiss National Science Foundation (NCCR Molecular Systems Engineering). G. M. was a fellow of Becas Chile-CONICYT.

References

1. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC et al (2012) Engineering the third wave of biocatalysis. Nature 485:185–194
2. Pamies O, Bäckvall JE (2003) Combined metal catalysis and biocatalysis for an efficient deracemization process. Curr Opin Biotechnol 14:407–413

3. Lee JW, Na D, Park JM, Lee J, Choi S et al (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol 8:536–546
4. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330:1355–1358
5. Woolston BM, Edgar S, Stephanopoulos G (2013) Metabolic engineering: past and future. Annu Rev Chem Biomol Eng 4:259–288
6. Chen Y, Nielsen J (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. Curr Opin Biotechnol 24:965–972
7. Chenault HK, Simon ES, Whitesides GM (1988) Cofactor regeneration for enzyme-catalysed synthesis. Biotechnol Bioeng 6:221–270
8. Fessner W-D, Walter C (1992) "Artificial metabolisms" for the asymmetric one-pot synthesis of branched-chain saccharides. Angew Chem Int Ed 31:614–616
9. Härle J, Panke S (2014) Synthetic biology for oligosaccharide production. Curr Org Chem 18:987–1004
10. Endo T, Koizumi S (2001) Microbial conversion with cofactor regeneration using genetically engineered bacteria. Adv Synth Catal 343:521–526
11. Koizumi S, Endo T, Tabata K, Ozaki A (1998) Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria. Nat Biotechnol 16:847–850
12. Altenbuchner J, Siemann-Herzberg M, Syldatk C (2001) Hydantoinases and related enzymes as biocatalysts for the synthesis of unnatural chiral amino acids. Curr Opin Biotechnol 12:559–563
13. Schrader J, Bohlmann J (2015) Biotechnology of isoprenoids. In: Scheper T (ed) Advances in biochemical engineering/biotechnology, vol 148. Springer, Heidelberg, p 475
14. Korman TP, Sahachartsiri B, Li D, Vinokur JM, Eisenberg D et al (2014) A synthetic biochemistry system for the *in vitro* production of isoprene from glycolysis intermediates. Proc Natl Acad Sci U S A 23:576–585
15. Chen X, Zhang C, Zou R, Zhou K, Stephanopoulos G et al (2013) Statistical experimental design guided optimization of one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene. PLoS ONE 8, e79650
16. Mahmoudian M, Noble D, Drake CS, Middleton RF, Montgomery DS et al (1997) An efficient process for the production of N-acetylneuraminic acid using N-acetylneuraminic aldolase. Enzyme Microb Technol 20:393–400
17. Birmingham WR, Starbird CA, Panosian TD, Nannemann DP, Iverson TM et al (2014) Bioretrosynthetic construction of a didanosine biosynthetic pathway. Nat Chem Biol 10:392–399
18. Babich L, van Hermet LJ, Bury A, Hartrog AF, Falcicchio P et al (2011) Synthesis of non-natural carbohydrates from glycerol and aldehydes in a one-pot four-enzyme cascade reaction. Green Chem 13:2895–2900
19. Schümperli M, Pellaux R, Panke S (2007) Chemcial and enzymatic routes to dihydroxyacetone phosphate. Appl Microbiol Biotechnol 75:33–45
20. Wagner N, Bosshart A, Failmezger J, Bechtold M, Panke S (2015) A separation-integrated cascade reaction to overcome thermodynamic limitations in rare sugar formation. Angew Chem Int Ed 54:4182–4186
21. Cheng Q, Xiang L, Izumikawa M, Meluzzi D, Moore BS (2007) Enzymatic total synthesis of enterocin polyketides. Nat Chem Biol 3:557–558
22. Schrittwieser JH, Groenendaal B, Resch V, Ghislieri D, Wallner S et al (2014) Deracemization by simultaneous bio-oxidative kinetic resolution and stereoinversion. Angew Chem Int Ed 53:3731–3734
23. O'Reilly E, Iglesias C, Ghislieri D, Hopwood J, Galman JL et al (2014) A regio- and stereoselective ω-transaminase/monoamine oxidase cascade for the synthesis of chiral 2,5-disubstituted pyrrolidines. Angew Chem Int Ed 53:2447–2450

24. Sehl T, Hailes HC, Ward JM, Wardenga R, Lieres E et al (2013) Two steps in one pot: enzyme cascade for the synthesis of nor(pseudo)ephedrine from inexpensive starting materials. Angew Chem Int Ed 52:6772–6775
25. Guterl J-K, Garbe D, Carsten J, Steffler F, Sommer B et al (2012) Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. ChemSusChem 5:2165–2172
26. Rieckenberg F, Ardao I, Rujananon R, Zeng A-P (2014) Cell-free synthesis of 1,3-propanediol from glycerol with a high yield. Eng Life Sci 14:380–386
27. Zhang YHP, Evans BR, Mielenz JR, Hopkins RC, Adams MW (2007) High-yield hydrogen production from starch and water by synthetic enzymatic pathway. PLoS ONE 2, e456
28. Martin del Campo JS, Rollin J, Myung S, Chun Y, Chandravan S et al (2013) High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. Angew Chem Int Ed 52:4587–4590
29. Rollin JA, del Campo JM, Myung S, Sun F, You C et al (2015) High-yield hydrogen production from biomass by *in vitro* metabolic engineering: mixed sugars coutilization and kinetic modeling. Proc Natl Acad Sci U S A 112:4964–4969
30. Wang Y, Huang W, Sathitsuksanoh N, YZhu Z, Zhang YHP (2011) Biohydrogenation from biomass sugar mediated by *in vitro* synthetic enzymatic pathways. Chem Biol 18:372–380
31. Zhu Z, Kin TT, Sun F, You C, Zhang YHP (2014) A high-energy-density sugar biobattery based on a synthetic enzymatic pathway. Nat Commun 5:3026
32. Ye X, Honda K, Sakai T, Okano K, Omasa T et al (2012) Synthetic metabolic engineering-a novel, simple technology for designing a chimeric metabolic pathway. Microb Cell Fact 11:120
33. Krutsakorn B, Honda K, Ye X, Imagawa T, Bei X et al (2013) *In vitro* production of n-butanol from glucose. Metabol Eng 20:84–91
34. Ye X, Honda K, Morimoto Y, Okano K, Ohtake H (2013) Direct conversion of glucose to malate by synthetic metabolic engineering. J Biotechnol 164:34–40
35. Dudley QM, Karim AS, Jewett MC (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J 10:69–82
36. Jewett M, Calhoun K, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4:220
37. Spirin AS, Swartz R (2008) Cell-free protein synthesis: methods and protocols. Wiley-VCH, Weinheim
38. Lu Y, Welsch JP, Swartz JR (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. Proc Natl Acad Sci U S A 111:125–130
39. Ng PP, Jia M, Patel KG, Brody JD, Swartz JR et al (2012) A vaccine directed to B cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. Proc Natl Acad Sci U S A 109:14526–14531
40. Takeda H, Ogasawara T, Ozawa T, Muraguchi A, Jih P-J et al (2015) Production of monoclonal antibodies against GPCR using cell-free synthesized GPCR antigen and biotinylated liposome-based interaction assay. Sci Rep 5:11333
41. Yin G, Garces ED, Yang J, Zhang J, Tran C et al (2012) Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. mAbs 4:217–225
42. Zawada JF, Yin G, Steiner AR, Yang J, Naresh A et al (2011) Microscale to manufacturing scale-up of cell-free cytokine production - a new approach for shortening protein production development timelines. Biotechnol Bioeng 108:1570–1578
43. Molla A, Paul AV, Wimmer E (1991) Cell-free, *de novo* synthesis of poliovirus. Science 254:1647–1651
44. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E et al (2008) Virus attenuation by genome-scale changes in codon pair bias. Science 320:1784–1787

45. Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297:1016–1018
46. Kwon YC, Jewett M (2015) High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Sci Rep 5:8663
47. Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A et al (2015) Computational protein design enables a novel one-carbon assimilation pathway. Proc Natl Acad Sci U S A 112:3704–3709
48. Liu C, Kelly GT, Watanabe CMH (2006) *In vitro* biosynthesis of the antitumor agent azinomycin B. Org Lett 8:1065–1068
49. Felix H (1982) Permeabilized cells. Anal Biochem 120:211–234
50. Krauser S, Weyler C, Blaß LK, Heinzle E (2013) Directed multistep biocatalysis using tailored permeabilized cells. Adv Biochem Eng/Biotechnol 137:185–234
51. Zhang Y-HP (2010) Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: challenges and opportunities. Biotechnol Bioeng 105:663–677
52. Zhang Y-HP, Sun J, Zhong J-J (2010) Biofuel production by in vitro synthetic enzymatic pathway biotransformation. Curr Opin Biotechnol 21:663–669
53. Wood DW (2014) New trends and affinity tag designs for recombinant protein purification. Curr Opin Plant Biol 26:54–61
54. Liu Z, Zhang J, Chen X, Wang PG (2002) Combined biosynthetic pathway for *de novo* production of UDP-galactose: catalysis with multiple enzymes immobilized on agarose beads. ChemBioChem 3:348–355
55. Schultheisz HL, Szymczyna BR, Scott LG, Williamson JR (2008) Pathway engineered enzymatic *de novo* purine nucleotide synthesis. ACS Chem Biol 3:499–511
56. Schultheisz HL, Szymczyna BR, Scott LG, Williamson JR (2011) Enzymatic *de novo* pyrimidine nucleotide synthesis. J Am Chem Soc 133:297–304
57. Wang HH, Huang P-Y, Xu G, Haas W, Marblestone A et al (2012) Multiplexed *in vivo* His-tagging of enzyme pathways for *in vitro* single-pot multienzyme catalysis. ACS Synth Biol 1:43–52
58. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G et al (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nat Biotechnol 460:894–898
59. Swartz J (2006) Developing cell-free biology for industrial applications. J Ind Microbiol Biotechnol 33:476–485
60. Bujara M, Schümperli M, Billerbeck S, Heinemann M, Panke S (2010) Exploiting cell free systems: Implementation and debugging of a system of biotransformations. Biotechnol Bioeng 106:376–389
61. Billerbeck S, Dietz S, Morgado G, Panke S (2012) Technologies for biosystems engineering. In: Wittmann C, Lee SY (eds) Systems metabolic engineering. Springer, Dordrecht, pp 83–115
62. Brockman IM, Prather KLJ (2015) Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. Metab Eng 28:104–113
63. McGinness KE, Baker TA, Sauer RT (2006) Engineering controllable protein degradation. Mol Cell 22:701–708
64. Taxis C, Stier G, Spadaccini R, Knop M (2009) Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. Mol Syst Biol 5:267
65. Billerbeck S, Calles B, Müller CL, de Lorenzo V, Panke S (2013) Towards functional orthogonalisation of protein complexes: individualisation of GroEL monomers leads to distinct quasihomogeneous single rings. ChemBioChem 14:2310–2321
66. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ et al (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27:753–759
67. Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY et al (2011) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic Acids Res 40:1879–1889

68. Sachdeva G, Garg A, Godding D, Way JC, Silver PA et al (2014) *In vivo* co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Res 42:9493–9503
69. Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554
70. You C, Myung S, Zhang Y-HP (2012) Facilitated substrate channeling in a self-assembled trifunctional enzyme complex. Angew Chem Int Ed 51:8787–8790
71. Yao AI, Fenton TA, Owsley K, Seitzer P, Larsen DJ et al (2013) Promoter element arising from the fusion of standard BioBricks parts. ACS Synth Biol 2:111–120
72. Kittleson JT, Wu GC, Anderson JC (2012) Successes and failures in modular genetic engineering. Curr Opin Chem Biol 16:329–336
73. Cardinale S, Arkin AP (2012) Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. Biotechnol J 7:856–866
74. Kiss G, Celebi-Olcum N, Moretti R, Baker D, Houk KN (2013) Computational enzyme design. Angew Chem Int Ed 52:5700–5725
75. Renata H, Wang ZJ, Arnold FH (2015) Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. Angew Chem Int Ed 54:3351–3367
76. Schilling CH, Schuster S, Palsson BO, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Progr 15:296–303
77. Hold C, Panke S (2009) Towards the engineering of *in vitro* systems. J R Chem Soc Interface 6:S507–S521
78. Medema MH, van Raaphorst R, Takano E, Breitling R (2012) Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol 10:191–202
79. Long MR, Ong WK, Reed JL (2015) Computational methods in metabolic engineering for strain design. Curr Opin Biotechnol 34C:135–141
80. Pasotti L, Zucca S (2014) Advances and computational tools towards predictable design in biological engineering. Comput Math Methods Med 2014:369681
81. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M et al (2015) BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Res 43:D439–D446
82. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK et al (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res 38:D396–D400
83. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B et al (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:977–982
84. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M et al (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–D205
85. Kumar A, Suthers PF, Maranas CD (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinf 13:6
86. Bujara M, Panke S (2012) In silico assessment of cell-free systems. Biotechnol Bioeng 109:2620–2629
87. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD et al (2005) Exploring the diversity of complex metabolic networks. Bioinformatics 21:1603–1609
88. Webb EC (1992) Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, 6th edn. Academic Press, San Diego
89. Finley SD, Broadbelt LJ, Hatzimanikatis V (2009) Computational framework for predictive biodegradation. Biotechnol Bioeng 104:1086–1097
90. Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U (2012) Integrating computational methods to retrofit enzymes to synthetic pathways. Biotechnol Bioeng 109:572–582

91. Cho A, Yun H, Park JH, Lee SY, Park S (2010) Prediction of novel synthetic pathways for the production of desired chemicals. BMC Syst Biol 4:35
92. Carbonell P, Planson A-G, Fichera D, Faulon J-L (2011) A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Syst Biol 5:122
93. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A et al (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7:445–452
94. Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. Metab Eng 25:140–158
95. Green ML, Karp PD (2004) A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinf 5:76
96. Rodrigo G, Carrera J, Prather KJ, Jaramillo A (2008) DESHARKY: automatic design of metabolic pathways for optimal cell growth. Bioinformatics 24:2554–2556
97. Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng 106:462–473
98. Jones JA, Toparlak ÖD, Koffas MA (2015) Metabolic pathway balancing and its role in the production of biofuels and chemicals. Curr Opin Biotechnol 33:52–59
99. Teusink B, Walsh MC, van Dam K, Westerhoff HV (1998) The danger of metabolic pathways with turbo design. Trends Biochem Sci 23:162–169
100. Ninh PH, Honda K, Sakai T, Okano K, Ohtake H (2015) Assembly and multiple gene expression of thermophilic enzymes in *Escherichia coli* for in vitro metabolic engineering. Biotechnol Bioeng 112:189–196
101. Bujara M, Schümperli M, Pellaux R, Heinemann M, Panke S (2011) Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. Nat Chem Biol 7:271–277
102. Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol 26:787–793
103. Knight T (2003) Idempotent vector design for standard assembly of Biobricks. Massachussetts Institute of Technology, Cambridge
104. Galdzicki M, Clancy KP, Oberortner E, Pocock M, Quinn JY et al (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. Nat Biotechnol 32:545–550
105. Myers C, Clancy K, Misirli G, Oberortner E, Pocock M et al (2015) The synthetic biology open language. Methods Mol Biol 1244:323–336
106. Martínez-García E, Aparicio T, Goñi-Moreno A, Fraile S, de Lorenzo V (2014) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. Nucleic Acids Res 43:D1193–D1198
107. Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A et al (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res 41:D666–D675
108. Mayer MP (1995) A new set of useful cloning and expression vectors derived from pBlueScript. Gene 163:41–46
109. French C, Ward JM (1996) Production and modification of E. coli transketolase for large-scale biocatalysis. Ann N Y Acad Sci 799:11–18
110. Friehs K (2004) Plasmid copy number and plasmid stability. Adv Biochem Eng/Biotechnol 86:47–82
111. Uhlin BE, Nordström K (1977) R plasmid gene dosage effects in *Escherichia coli* K-12: copy mutants of the R plasmid R1drd-19. Plasmid 1:1–7
112. Chappell J, Jensen K, Freemont PS (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. Nucleic Acids Res 41:3471–3481
113. Chappell J, Freemont P (2013) *In vivo* and *in vitro* characterization of σ70 constitutive promoters by real-time PCR and fluorescent measurements. Methods Mol Biol 1073:61–74

114. Davidson EA, van Blarcom T, Levy M, Ellington AD (2010) Emulsion based selection of T7 promoters of varying activity. In: Pacific symposium on biocomputing. 2010, World Scientific, pp 433–443
115. Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 39:1131–1141
116. Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J et al (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng 3:4
117. 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
118. Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. Nucleic Acids Res 41:10668–10678
119. Du J, Yuan Y, Si T, Lian J, Zhao H (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 40, e142
120. 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
121. Qin X, Qian J, Yao G, Zhuang Y, Zhang S et al (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol 77:3600–3608
122. Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683
123. Latimer LN, Lee ME, Medina-Cleghorn D, Kohnz RA, Nomura DK et al (2014) Employing a combinatorial expression approach to characterize xylose utilization in *Saccharomyces cerevisiae*. Metab Eng 25:20–29
124. Seo SW, Yang J-S, Kim I, Yang J, Min BE et al (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab Eng 15:67–74
125. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27:946–950
126. Na D, Lee D (2010) RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. Bioinformatics 26:2633–2634
127. Borujeni AE, Channarasappa AS, Salis HM (2013) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res 42:2646–2659
128. de Smit MH, Van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A 87:7668–7672
129. Farasat I, Kushwaha M, Collens J, Easterbrook M, Guido M et al (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol Syst Biol 10:731
130. Zelcbuch L, Antonovsky N, Bar-Even A, Levin-Karp A, Barenholz U et al (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res 41, e98
131. Nowroozi FF, Baidoo EEK, Ermakov S, Redding-Johanson AM, Batth TS et al (2014) Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. Appl Environ Microbiol 98:1567–1581
132. Ng CY, Farasat I, Maranas CD, Salis HM (2015) Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration. Metab Eng 29:86–96
133. Xu P, Gu Q, Wang W, Wong L, Bower AGW et al (2013) Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. Nat Commun 4:1409
134. Oliver JWK, Machado IMP, Yoneda H, Atsumi S (2014) Combinatorial optimization of cyanobacterial 2,3-butanediol production. Metab Eng 22:76–82
135. Ceroni F, Algar R, Stan G-B, Ellis T (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods 12:415–418

136. Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai Q-A et al (2013) Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res 41:5139–5148
137. Chen Y-J, Liu P, Nielsen AAK, Brophy JAN, Clancy K et al (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 10:659–664
138. Temme K, Zhao D, Voigt CA (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. Proc Natl Acad Sci U S A 109:7085–7090
139. Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids *in vitro*. Proc Natl Acad Sci U S A 70:3240–3244
140. Chao R, Yuan Y, Zhao H (2015) Recent advances in DNA assembly technologies. FEMS Yeast Res 15:1–9
141. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA et al (2010) BglBricks: a flexible standard for biological part assembly. J Biol Eng 4:1
142. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3, e3647
143. Shetty RP, Endy D, Jr TFK (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5
144. Densmore D, Hsiau TH-C, Kittleson JT, DeLoache W, Batten C et al (2010) Algorithms for automated DNA assembly. Nucleic Acids Res 38:2607–2616
145. Leguia M, Brophy J, Densmore D, Anderson JC (2011) Automated assembly of standard biological parts. Methods Enzymol 498:363–397
146. Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioeng Bugs 3:38–43
147. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS ONE 4, e6441
148. Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59
149. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
150. Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242–251
151. Bitinaite J, Rubino M, Varma KH, Schildkraut I, Vaisvila R et al (2007) USER friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res 35:1992–2002
152. Schmid-Burgk JL, Xie Z, Benenson Y (2014) Hierarchical ligation-independent assembly of PCR fragments. Methods Mol Biol 1116:49–58
153. Chen H, Lisby M, Symington LS (2013) RPA coordinates DNA end resection and prevents formation of DNA hairpins. Mol Cell 50:589–600
154. Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58:201–216
155. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res 37:6984–6990
156. Robzyk K, Kassir Y (1992) A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. Nucleic Acids Res 20:3790
157. Gibson DG, Benders G, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H et al (2008) Complete chemical synthesis, assembly, and cloning of a *Mcyoplasma genitalium* genome. Science 319:1215–1220
158. Kim B, Du J, Eriksen DT, Zhao H (2013) Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. Appl Environ Microbiol 79:931–941
159. Bradley LH, Bricken ML, Randle C (2011) Expression, purification, and characterization of proteins from high-quality combinatorial libraries of the mammalian calmodulin central linker. Protein Expr Purif 75:186–191

160. Tikhonova EB, Ethayathulla AS, Su Y, Hariharan P, Xie S et al (2015) A transcription blocker isolated from a designed repeat protein combinatorial library by *in vivo* functional screen. Sci Rep 5:8070
161. Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ et al (2005) Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. Nat Biotechnol 23:1171–1176
162. Wang J, Sarov M, Rientjes J, Fu J, Hollak H et al (2006) An improved recombineering approach by adding RecA to lambda Red recombination. Mol Biotechnol 32:43–53
163. He AS, Rohatgi PR, Hersh MN, Rosenberg SM (2006) Roles of *E. coli* double-strand-break-repair proteins in stress-induced mutation. DNA Repair 5:258–273
164. Wang HH, Kim H, Cong L, Jeong J, Bang D et al (2012) Genome-scale promoter engineering by coselection MAGE. Nat Methods 9:591–593
165. Ronda C, Pedersen LE, Sommer MOA, Nielsen AT (2016) CRMAGE: CRISPR optimized MAGE recombineering. Sci Rep 6:19452
166. Crabb WD, Shetty JK (1999) Commodity scale production of sugars from starches. Curr Opin Microbiol 2:252–256
167. DiCosimo R, McAuliffe J, Poulose AJ, Bohlmann G (2013) Industrial use of immobilized enzymes. Chem Soc Rev 42:6437–6474

Adv Biochem Eng Biotechnol (2018) 162: 147–174
DOI: 10.1007/10_2017_3

Published online: 1 June 2017

Synthetic Biology of Polyhydroxyalkanoates (PHA)

De-Chuan Meng and Guo-Qiang Chen

Abstract Microbial polyhydroxyalkanoates (PHA) are a family of biodegradable and biocompatible polyesters which have been extensively studied using synthetic biology and metabolic engineering methods for improving production and for widening its diversity. Synthetic biology has allowed PHA to become composition controllable random copolymers, homopolymers, and block copolymers. Recent developments showed that it is possible to establish a microbial platform for producing not only random copolymers with controllable monomers and their ratios but also structurally defined homopolymers and block copolymers. This was achieved by engineering the genome of *Pseudomonas putida* or *Pseudomonas entomophiles* to weaken the β-oxidation and in situ fatty acid synthesis pathways, so that a fatty acid fed to the bacteria maintains its original chain length and structures when incorporated into the PHA chains. The engineered bacterium allows functional groups in a fatty acid to be introduced into PHA, forming functional PHA, which, upon grafting, generates endless PHA variety. Recombinant *Escherichia coli* also succeeded in producing efficiently poly (3-hydroxypropionate) or P3HP, the strongest member of PHA. Synthesis pathways of P3HP and its copolymer P3HB3HP of 3-hydroxybutyrate and 3-hydroxypropionate were assembled respectively to allow their synthesis from glucose. CRISPRi was also successfully used to manipulate simultaneously multiple genes and control metabolic flux in *E. coli* to obtain a series of copolymer P3HB4HB of 3-hydroxybutyrate (3HB) and 4-hydroxybutyrate (4HB). The bacterial shapes were successfully engineered for enhanced PHA accumulation.

Keywords *Escherichia coli*, Inclusion bodies, PHB, Polyhydroxyalkanoates, Synthetic biology

D.-C. Meng and G.-Q. Chen (✉)
MOE Key Lab of Bioinformatics, School of Life Science, Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China
e-mail: chengq@mail.tsinghua.edu.cn

Contents

1 Introduction

Polyhydroxyalkanoates (PHA) are a family of structurally diverse intracellular biopolyesters accumulated by many microorganisms [1–3]. Because of their similar properties with traditional petroleum-based plastics, PHA have been developed for applications in the packaging, medicine, pharmacy, agriculture, and food industries [4–6]. Compared with other well-known biodegradable or biobased polymers with less CO_2 emission, such as polylactide (PLA), PHA have much wider diversity in monomers with over 150 structural variations reported [7, 8].

Based on monomer lengths, PHA monomers are divided into short-chain-length (scl) consisting of 3–5 carbon atoms, and medium-chain-length (mcl) of 6–14 carbon atoms (Fig. 1) [8, 9]. Based on the composition of the monomers and their arrangements, PHA have been classified into homopolymers consisting of one monomer, random copolymers of two or more different monomers, and block copolymers of at least two homopolymers connected by covalent bond(s) (Fig. 2) [9, 10]. The microstructures of PHA and monomer compositions affect the thermal and physical properties of PHA, which affects their applications (Table 1) [11, 12]. For example, the most studied PHA family member, poly (3-hydroxybutyrate) or P3HB, first reported in 1926 [13], is very brittle with high crystallinity which limits its applications [14]. In many cases, it is not easy to achieve precise control of PHA structure. For example, random copolymers consisting of 3-hydroxyhexanoate (3HHx or C6), 3-hydroxyoctanoate (3HO or

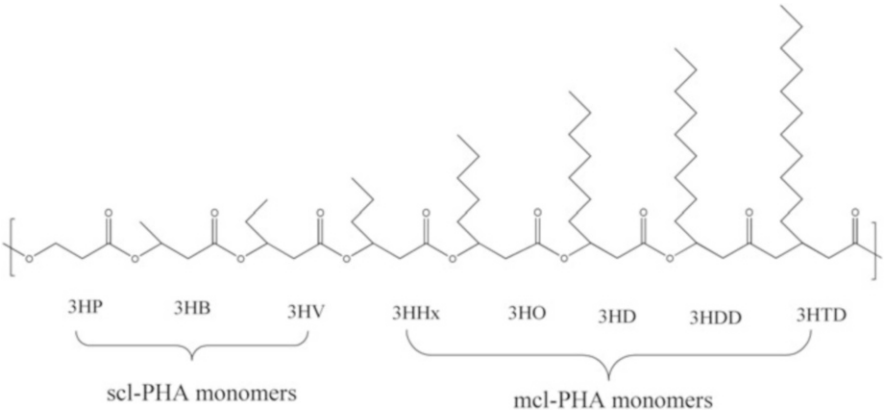

Fig. 1 Traditional PHA monomers. *3HP* 3-hydroxypropionate, *3HB* 3-hydroxybutyrate, *3HV* 3-hydroxyvalcrate, *3HHx* 3-hydroxyhexanoate, *3HO* 3-hydroxyoctanoate, *3HD* 3-hydroxydecanoate, *3HDD* 3-hydroxydodecanoate, *3HTD* 3-hydroxytetradecanoate

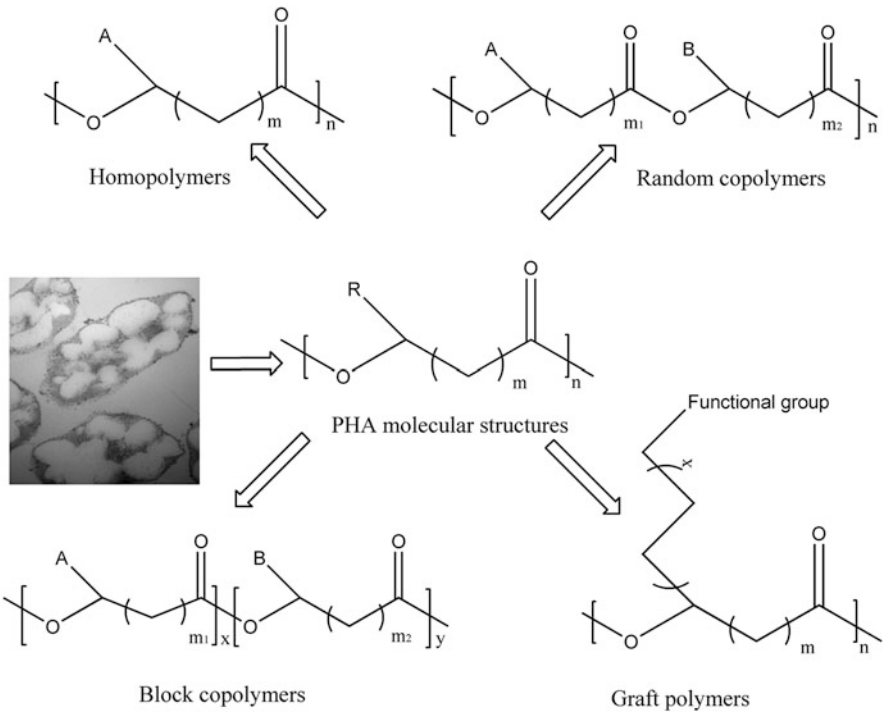

Fig. 2 PHA molecular structures [9]

Table 1 Physical characterization of PHA and traditional petroleum-based plastic [4, 11]

	Thermal properties		Mechanical properties	
PHA	T_m (°C)	T_g (°C)	σ_{mt} (MPa)	ε_b (%)
P3HP[a]	78.13	−17.85	21.54 ± 1.10	497.6 ± 6.2
P4HB[a]	61	−47	34.66 ± 0.98	1,000
P3HB[a]	171.8	3.1	18.0 ± 0.7	3.0 ± 0.4
PHBV[b]	114	−5	26	27
$PHBHH_X$[a]	125	0	7.0 ± 0.5	400 ± 36
Polypropylene[b]	170	–	34	400
Polystyrene[b]	110	–	50	–

P3HP poly(3-hydroxypropinoate), *P4HB* poly(4-hydroxybutyrate), *P3HB* poly (3-hydroxybutyrate), *PHBV* poly(3-hydroxybutyrate-*co*-20 mol% 3-hydroxyvalcrate), *PHBHHx* poly(3-hydroxybutyrate-*co*-12 mol% 3-hydroxyhexanoate), T_m melting temperature, T_g glass transition temperature, σ_{mt} maximum tension strength, ε_b elongation at break

[a]Physical properties of P3HB, P4HB, PHB, PHBHHx [11]

[b]Physical properties of PHBV, polypropylene, polystyrene [4]

C8), 3-hydroxydecanoate (3HD or C10), and 3-hydroxydodecanoate (3HDD or C12) are always formed when a fatty acid is added to cultures of *Pseudomonads* belonging to the rRNA homology group I, as β-oxidation in *Pseudomonas* spp. always shorten the C12 to C10, C8, and C6 [15]. On the other hand, the in situ fatty acid synthesis pathway, although lower in fatty acid synthesis rate for supplying PHA monomers than β-oxidation, also supplies various monomers for PHA synthesis [16], leading to PHA consisting of various monomers in random copolymers. The traditional PHA, such as PHB, PHBV, and PHBHHx, produced by wild-type microorganisms, are still facing problems of high cost and poor properties, and scientists are developing novel methods to lower the cost of PHA or discover novel PHA with high value-added applications or better properties using synthetic biology and metabolic engineering [17]. In many cases, precursors such as fatty acids, alcohols, or functional monomers are expensive, and new pathways are being established to synthesize PHA monomers in vivo from low cost glucose [16, 18]. This approach is very important if the PHA is to be produced on an industrial scale [19]. Recent advances in systems biology have improved the amount of information that can be collected, and synthetic biology tools are developing modeling and molecular implementation methods, promising to move microbial engineering from the iterative approach to a design-oriented paradigm [20].

2 Metabolic Pathways of PHA Synthesis

Many bacteria have been found to produce various polyhydroxyalkanoate (PHA) biopolyesters [8]. For example, *Ralstonia eutropha* was mostly studied in producing PHB and PHBV [21] and *Pseudomonas putida* is well-known for synthesizing mcl-PHA [22, 23]. The specificity of a PHA synthase (PhaC) is the most important

element determining PHA monomer compositions in different microorganisms [24–26]. PhaC from *Ralstonia eutropha* has been known to be able to polymerize PHA monomers consisting of three (C3) to five (C5) carbon chain lengths termed short-chain-length PHA or scl PHA [27], including poly(3-hydroxypropionate) (P3HP) [28, 29], poly(3-hydroxybutyrate) (PHB) [30], poly(4-hydroxybutyrate) (P4HB) [31, 32], poly(3-hydroxyvalerate) (PHV) [33], and copolymers of 3-hydroxypropionate and 4-hydroxybutyrate (P3HP4HB) [11], as well as similar copolymers of P3HB4HB [18], P3HP3HB [34], and PHBV [5, 8]. Many *Pseudomonas* spp. contain PhaCs that can polymerize monomers of six (C6) to fourteen (C14) carbon-chain-length to form medium-chain-length PHA (or mcl PHA) [35]. Very few bacteria were found to have PhaCs that can polymerize C4 to C14 to form scl-mcl copolymers [36, 37]. Wild-type *Ralstonia eutropha* H16 can only produce scl PHA, when introducing PHA synthase gene $phaC2_{Ps}$ from *Pseudomonas stutzeri* strain 1317 into PHA synthase gene $phbC_{Re}$ negative mutant *R. eutropha* PHB-4, the recombinant *R. eutropha* having the ability to synthesize mcl PHA. During the cultivation on gluconate, the presence of $phaC2_{Ps}$ in *R. eutropha* PHB-4 led to the accumulation of PHB homopolymer at 40.9 wt% in dry cells. When using fatty acids as carbon sources, the recombinant successfully produced PHA copolyesters containing both scl PHA and mcl PHA of 4–12 carbon atoms in length. When cultivated on a mixture of gluconate and a fatty acid, the monomer composition of accumulated PHA was strongly affected and the monomer content was easily regulated by the addition of fatty acids in the cultivation medium [36]. A series of optimization strategies were reported on the PHA synthase $PhaC2_{Ps}$ in *E. coli*, codon optimization of the gene and mRNA stabilization with a hairpin structure were conducted, and the function of the optimized PHA synthase was tested in *E. coli*. The transcript was more stable after the hairpin structure was introduced, both codon optimization and hairpin introduction increasing the protein expression level compared with the wild-type $PhaC2_{Ps.}$ The optimized $PhaC2_{Ps}$ increased PHB production by approximately 16-fold to 30% of the cell dry weight. When grown on dodecanoate, the recombinant *E. coli* harboring the optimized gene $phaC2_{Ps}O$ with a hairpin structure in the 5′ untranslated region was able to synthesize fourfold more PHA, consisting of 3HB and mcl 3HA, compared to the recombinant harboring the wild-type $phaC2_{Ps}$ [38].

The authors' group summarized a metabolic pathways map leading to PHA formation (Fig. 3). The most studied PHA synthesis pathways are discussed in the following. Pathway I, starting from sugar to scl PHA, especially PHB, glucose was used as carbon source to produce acetyl-CoA first, followed by metabolism to acetoacetyl-CoA and 3-hydroxybutyryl-CoA, entering the polymerization process to form PHB. The recombinant *E. coli* also showed high productivity of PHA after introducing the *phaCAB* operon from *Ralstonia eutropha*. Based on this pathway, more synthetic pathways were developed to produce more PHA with other structures [18, 34]. Pathway II begins from fatty acid(s) as substrate to enter the β-oxidation cycle, leading to formation of *R*-3-hydroxyacyl-CoA monomers for mostly mcl PHA synthesis [39]. Pathway III directs acetyl-CoA to malonyl-CoA to

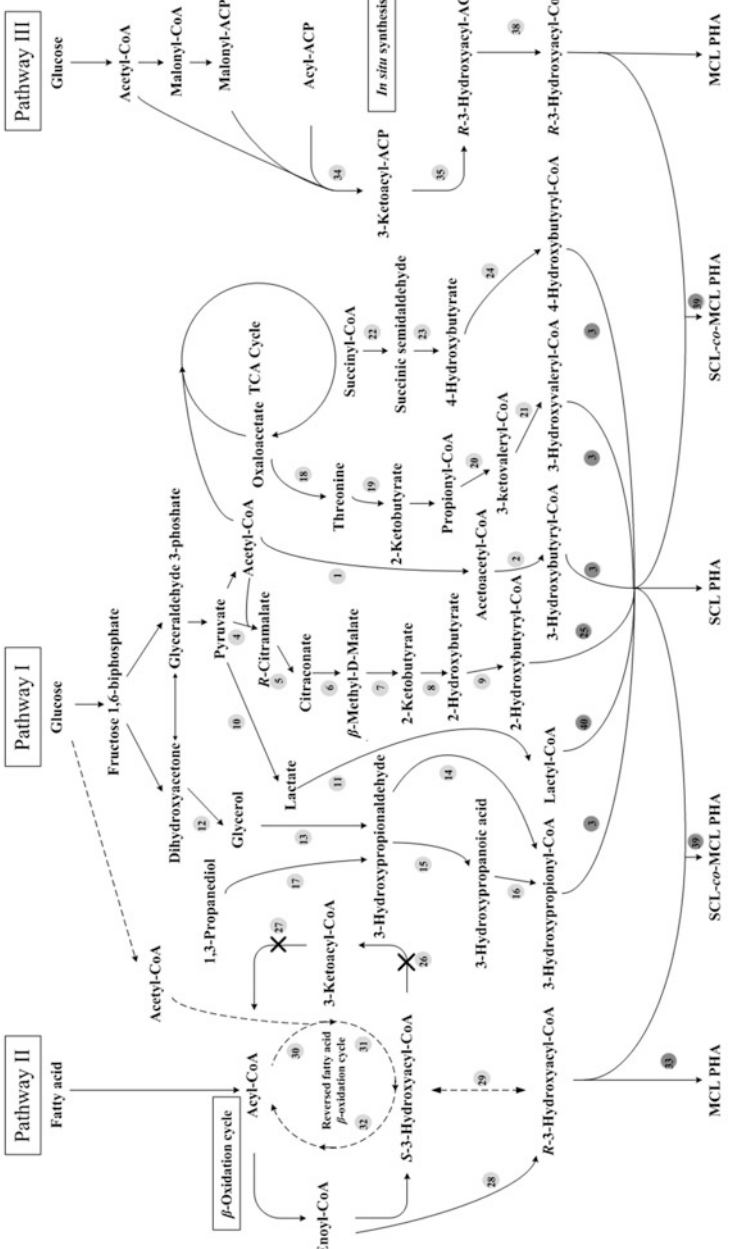

Fig. 3 Metabolic pathways for PHA synthesis [9]. *1*: PhaA, β-ketothiolase; *2*: PhaB, NADP dependent acetoacetyl-CoA reductase; *3*: PhaC, PHA synthase; *4*: CimA, citramalate synthase; *5*: LeuCD, 3-isopropylmalate isomerase; *6*: 3-isopropylmalate dehydrogenase; *7*: LeuB, 3-isopropylmalate dehydrogenase; *8*: PanE, 2HB dehydrogenase; *9*: Pct5404: propionyl-CoA transferase; *10*: IdhA, lactate dehydrogenase; *11*: Pct $_{Cp}$: propionate CoA transferase; *12*: GPD1, glycerol-3-P dehydrogenase, GPP2, glycerol-3-P phosphatase; *13*: DhaB, glycerol dehydratase; *14*: PduP, propionaldehyde dehydrogenase; *15*: AldD, aldehyde dehydrogenase; *16*: PCS', propanoyl-CoA synthetase; *17*: DhaT, 1,3-propanediol dehydrogenase; *18*: ThrA, aspartokinase I, ThrB, homoserine kinase;

3-ketoacyl-ACP for forming *R*-3-hydroxyacyl-CoA monomers [40, 41]. Glucose was also used as carbon source to produce novel PHA with high value-added products, such as P3HP, which is discussed later [34]. The types of PHA formed depend not only on monomer supply pathways, but also on specificity of PHA synthases. Generally, a low specificity of a PhaC allows formation of diverse PHA structures [36]. As the properties of copolymer of scl-PHA and mcl-PHA are drawing more attention, a lot of work is focusing on the production of scl-*co*-mcl PHA using a low specificity of a PhaC [25, 42, 43].

3 Diversity of PHA

Diversity of PHA has been focused not only on monomer variations but also on the composition of PHA, especially on PHA main chain structures (Table 2). PHA was first discovered in the form of poly-3-hydroxybutyrate (PHB) in the last century [13]. New monomers 3-hydroxyvalerate (3HV) and 3-hydroxyhexanoate (3HHx) were detected as components of PHA in bacteria in activated sewage sludge in the 1970s [44]. Then, 10–15 years afterward, *Pseudomonas oleovorans* was found to be able to produce a series of PHA containing 3-hydroxyhexanoate (3HHx), 3-hydroxyoctanoate (3HO), 3-hydroxydecanoate (3HD), and 3-hydroxydodecanoate (3HDD) when grown on different alkanes or fatty acids as substrates [16, 33]. In the following years, scientists started to modify the PHA pathways or introduce the PHA pathways into a better host. For example, the *phaCAB* operon for PHB production was transformed into *E. coli*, and the non-PHA producing bacteria also showed high PHB productivity with the heterologous PHB pathway [30]. An increasing number of novel PHA were synthesized using mostly structure-related substrates and, in 1995, 91 different hydroxyalkanoic acids were reported as monomers in PHA [8]. PHA diversity was further increased by producing functional PHA, grafted with other chemicals and polymers [45–47]. From then on, diversity of PHA was further expanded to include PHA polymer

Fig. 3 (continued) ThrAC, threonine synthase; *19*: IlvA, threonine deaminase; *20*: PhaA, β-ketothiolase; *21*: PhaB, NADP-dependent acetoacetyl-CoA reductase; *22*: SucD: succinic semialdehyde dehydrogenase; *23*: 4hbD, 4-hydroxybutyrate dehydrogenase; *24*: OrfZ, 4-hydroxybutyrate-CoA transferase; *25*: $PhaC1_{Ps6\text{-}19}$, PHA synthase from *Pseudomonas* sp. MBEL 6-19; *26*: fadB, *S*-3-hydroxyacyl-CoA dehydrogenase; *27*: fadA, 3-ketothiolase; *28*: PhaJ, enoyl-CoA hydratase; *29*: epimerase; *30*: YqeF/FadA, thiolase; *31*: FadB, hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase; *32*: YdiO, enoyl-CoA reductase, Ter, trans-2-enoyl-CoA reductase from *Treponema denticola*; *33*: PhaC, type II PHA synthase; *34*: β-ketoacyl-ACP synthase; *35*: β-ketoacyl-ACP reductase; *36*: β-hydroxyacyl-ACP dehydrase; *37*: enoyl-ACP reductase; *38*: PhaG, 3-hydroxyacyl-acyl carrier protein-coenzyme A transferase; *39*: PhaC1 (STQK), PHA synthase derived from *Pseudomonas* sp. 61-3 PHA synthase; *40*: engineered $PhaC1_{Ps6\text{-}19}$, PHA synthase

Table 2 Diversity of PHA [9]

Types	Polymer structures
Homopolymers	PHB, P3HP, P4HB, PHV, PTE, PLA, P3HHx, P3HHp, P3HO, P3HD, P3HDD, P3HPhV, P3HPE, PHU, P3H6PHx
Random copolymers	P(3HB-*co*-3HV), P(3HB-*co*-4HB), P(3HB-*co*-3HHx), P(3HP-*co*-4HB), P(3HB-*co*-3HP), P(3HB-*co*-mcl 3HA), P(3HHx-*co*-3HO-3HD-3HDD), P(3HB-*co*-LA)
Block copolymers	P3HB-*b*-P3HBV, P3HB-*b*-4HB, P3HP-*b*-4HB, P3HB-*b*-3HHx, P3HB-*b*-3HP, P3HHx-*b*-P(3HD3HDD)
Graft polymers	PS-*g*-PHA, PMMA-*g*-PHA, PHA-*g*-PAA, PHA-*g*-AA-CS, PHA-*g*-AA-COS, PHA-*g*-Cellulose, PEG-*g*-PHA, PEGMA-*g*-PHO, PLA-*g*-PHA, VI-*g*-PHO, GDD-*g*-PHO, PHOU-*g*-Jeffamine, PHOU-*g*-POSS, PHBV-*g*-PVK, PHBV-*g*-PA

3HB 3-hydroxybutyrate, *3HP* 3-hydroxypropionate, *4HB* 4-hydroxybutyrate, *3HV* 3-hydroxyvalcrate, *PTE* polythioester, *PLA* polylactic acid, *3HHx* 3-hydroxyhexanoate, *3HHp* 3-hydroxyheptanoate, *3HO* 3-hydroxyoctanoate, *3HD* 3-hydroxydecanoate, *3HDD* 3-hydroxydodecanoate, *3HPhV* 3-hydroxy-5-phenylvalerate, *3HPE* 3-hydroxy-4-pentenoic acid, *PHU* polyhydroxyundecenoate, *3H6PHx* 3-hydroxy-6-phenylhexanoate, *PS-g-PHA* poly(styrene peroxide)-*g*-PHA, *PMMA-g-PHA* poly(methyl methacrylate peroxide)-*g*-PHA, *PHA-g-PAA* PHA-*g*-poly(acrylic acid), *PHB-g-AA/starch* acrylic acid grafted poly(3-hydroxybutyric acid)/starch, *PHA-g-AA-CS* PHA-*g*-AA-chitosan, *PHA-g-AA-COS* PHA-*g*-AA-chitooligosaccharide, *PEG-g-PHA* poly(ethylene glycol)-*g*-PHA, *PEGMA-g-PHO* monoacrylate-poly(ethylene glycol)-*g*-PHO, *PLA-g-PHA* poly(lactic acid)-*g*-PHA, *VI-g-PHO* vinylimidazole-grafted poly (3-hydroxyoctanoate), *GDD-g-PHO* glycerol 1,3-diglycerol diacrylate-*g*-PHO, *PHOU-g-Jeffamine* PHOU-*g*-α-amino-ω–methoxy poly(oxyethylene-*co*-oxypropylene), (Jeffamine®)-*g*-PHOU, *PHOU-g-POSS* PHOU-*g*-polyhedral oligomeric silsesquioxane, *PHBV-g-PVK* PHBV-*g*-poly(phenyl vinyl ketone), *PHBV-g-PA* PHBV-*g*-poly(acrylamide)

chains with various microstructures, such as homopolymers, random copolymers, block copolymers, block-random copolymers, functional polymers, graft polymers, and thiopolyesters, as well as their various combinations [10, 11, 45]. Among the diverse PHA, grafted PHA polymers can be most easily extended to a wider diversity, and this is a topic that requires further elucidation [46, 48]. However, only very few PHA are commercially available for application developments, including PHB, PHBV, P3HB4HB, and PHBHHx. All other PHA have been prepared by individual laboratories across the world in very small amounts out of academic curiosity. How to accelerate the pace of discovery and deployment of advanced PHA materials has been a central question for all PHA researchers and stakeholders. All these depend on the availability of the diverse PHA in sufficient quantities for studies of their thermal and mechanical properties, as well as other application potentials. It should be a global effort to establish platforms to supply diverse PHA in sufficient quantities for various developments.

3.1 Homopolymers

So far, only limited homopolymers have been reported, including scl PHA: PHB [49], P3HP [34], P4HB [32], microbial polylactic acid (PLA) [42], and PHV [33], mcl PHA: P3HHx, P3HHp (poly3-hydroxyheptanoate) [50], PHO, P3HD, P3HDD, and P3HTD or poly(3-hydroxytetradecanoate) [35, 51], as well as functional PHA: poly(3-hydroxy-5-phenylvalerate) or P(3HPhV) [47], poly(3-hydroxy-4-pentenoate) [46], poly(3-hydroxy-10-undecenoate) [52], and poly(3-hydroxy-6-phenylhexanoate) [48]. With the success of engineering the β-oxidation pathway, more and more homopolymers can be synthesized. *Pseudomonas putida* KT2442 often produces mcl PHA consisting of 3HHx, 3HO, 3HD, 3HDD, and 3HTD, and when it was knocked out with its β-oxidation related genes *fadA*, *fadB*, *fadB2x*, *fadAx*, and *phaG*, the mutant *P. putida* KTQQ20 synthesized homopolymer poly-3-hydroxydecanoate (PHD) when grown on decanoic acid [35]. Mcl PHA producer *Pseudomonas entomophila* L48 was also studied for homopolymer production, when genes encoding 3-hydroxyacyl-CoA dehydrogenase, 3-ketoacyl-CoA thiolase, and acetyl-CoA acetyltransferase in the β-oxidation pathway were knocked out. The mutant *P. entomophila* LAC26 accumulated over 90 wt% PHA consisting of 99 mol% 3HDD using dodecanoic acid as a carbon source. The β-oxidation-inhibited mutant of *P. entomophila* was also studied to produce benzene containing PHA, poly(3-hydroxy-5-phenylvalerate) using 5-phenylvaleric acid as carbon source and homopolymer P(3-hydroxy-9-decenoate) using 9-decenol as carbon source [45, 47]. Synthetic biology also makes it possible to create novel PHA with designed structures and compositions.

3.2 Random Copolymers

Most of the commercially produced PHA are random copolymers, including P(3HB-*co*-3HV) or PHBV, P(3HB-*co*-4HB) or P3HB4HB, and P(3HB-*co*-3HHx) or PHBHHx, which have been produced on an industrial scale [2]. Copolymers of mcl PHA termed P(3HHx-*co*-3HO-*co*-3HD-*co*-3HDD) are commonly synthesized by many *Pseudomonads* belonging to the rRNA homology group I [39], but it is too soft for any application [53]. Recently, random copolymers of P(3HP-*co*-4HB) [11], P(3HB-*co*-3HP) [34], poly(3HB-*co*-3MP) [54], and P(3HB-*co*-LA) [42] were found to be accumulated by recombinant *E. coli,* and these copolymers demonstrated improved properties over the existing ones. However, the yield of PHA production needs to be improved for further industrial scale production.

3.3 Block Copolymers

Pederson et al. [10] reported the first PHA block copolymer of PHB-*b*-PHBV, and the material was found to have anti-ageing property. Block copolymerization is a method of controlling the thermodynamic nature of a polymer, and it is able to withstand the ageing effect that leads to the brittleness of a polymer material [10]. Starting in 2011, the authors' lab and other groups have succeeded in making a series of diblock copolymers, including PHB-*b*-P3HVHHp [55], PHB-*b*-P4HB [56], PHB-*b*-PHHx [57], P3HB-*b*-P3HP [58], P3HP-*b*-P4HB [59], and P3HHx-*b*-P(3HD-*co*-3HDD) [60]. The sequential feeding of two or more structurally related carbon substrates led to biosynthesis of block copolymers. For example, by first feeding 1,3-propanediol and late addition of 1,4-butanediol to cultures, the engineered *E. coli* synthesized block copolymers of P3HP-*b*-P4HB [59]. All the diblock copolymers were found to have one or more improved properties over their two relative homopolymers, random copolymers or blend polymers. Compositions of diblock copolymers can be adjusted based on monomer substrate ratios in the feeds, leading to adjustable polymer properties. Although multiple-block PHA are still difficult to synthesize, with the development of synthetic biology it should become possible to realize the accurate control of monomer composition and then production of block PHA with diverse structures on a larger scale.

3.4 Graft Polymers

As it is possible to introduce functional groups into PHA chains, such as double or triple bonds, epoxy, carbonyl, cyano, phenyl, and halogen [46], graft PHA polymers can be formed by inserting small molecules or larger polymers into the PHA side chains, leading to dramatic changes PHA properties. So far, successful PHA graft polymers include poly(styrene peroxide)-*g*-PHA or PS-*g*-PHA [61], poly(methyl methacrylate peroxide)-*g*-PHA or PMMA-*g*-PHA [62], PHA-*g*-poly(acrylic acid) or PHA-*g*-PAA [63], PHA-*g*-AA-chitooligosaccharide or PHA-*g*-AA-COS [64], PHA-*g*-Cellulose [65], poly(ethylene glycol)-*g*-PHA or PEG-*g*-PHA [66], monoacrylate-poly(ethylene glycol)-*g*-PHO or PEGMA-*g*-PHO [67], poly(lactic acid)-*g*-PHA or PLA-*g*-PHA [68], vinylimidazole-*g*-PHO or VI-*g*-PHO [69], glycerol-1,3-diglycerol diacrylate-*g*-PHO or GDD-*g*-PHO [70], (Jeffamine®)-*g*-PHOU or PHOU-*g*-Jeffamine, PHOU-*g*-α-amino-ω-methoxy poly(oxyethylene-*co*-oxypropylene) [71], PHOU-*g*-polyhedral oligomeric silsesquioxane or PHOU-*g*-POSS [72], PHBV-*g*-poly(phenyl vinyl ketone) or PHBV-*g*-PVK [73], and PHBV-*g*-poly(acrylamide) or PHBV-*g*-PA [74]. Graft copolymers were mostly synthesized by chemical modification. For example, side carboxylic groups of the PHA were coupled with end hydroxyl groups of methoxy-poly(ethylene glycol) (MePEG) or methoxy-poly(lactic acid) (MePLA) in the presence of *N*,*N*′-dicylohexylcarbodiimide (DCC) [68]. There are endless possibilities to create new graft PHA homo- or copolymers.

4 Engineering Pathways for Controlling PHA Biosynthesis

4.1 Pathways for scl PHA

Microbial metabolic engineering has been exploited as a powerful approach for enhanced production of novel polyesters. A designed pathway assembled using a synthetic biology approach could also precisely control the PHA composition. The use of recombinant *E. coli* enabled an efficient production of poly (4-hydroxybutyrate) or P4HB using glucose as a sole carbon source when a pathway was established containing genes encoding succinic semialdehyde dehydrogenase of *Clostridium kluyveri* and PHB synthase of *Ralstonia eutropha* combined with inactivation of native succinate semialdehyde dehydrogenase genes *sad* and *gabD* to enhance the carbon flux toward P4HB biosynthesis [32]. When the PHB accumulation pathway of *Ralstonia eutropha* was co-expressed with the P4HB synthesis pathway, the recombinant *E. coli* produces P(3HB-*co*-4HB) from glucose [18].

Aeromonas hydrophila 4AK4 normally produces copolyesters PHBHHx. Recombinant *A. hydrophila* 4AK4 expressing *vgb* and *fadD* genes encoding *Vitreoscilla* hemoglobin and *E. coli* acyl-CoA synthase, respectively, was found to produce homopolymer poly(3-hydroxyvalerate) (PHV) (C5) using undecanoic acid as a solo carbon source [75]. At the same time, 3-hydroxyvalerate monomer can also be supplied via the threonine degradation pathway. Recently, it became possible to produce PHA containing 2-hydroxybutyrate [76] or lactate [42]. In addition, P3HP can be produced from 1,3-propandiol [29], glycerol alone [77], and glucose as sole carbon source [34].

PHA synthesis genes *phbC* and *orfZ* cloned from *Ralstonia eutropha* H16 and *Clostridium kluyveri*, respectively, were transformed into a β-oxidation weakened *Pseudomonas putida* KTOY08ΔGC, a mutant of *P. putida* KT2442, and the resulting mutant termed KTHH06 was able to produce P3HB-*b*-P4HB diblock copolymer [56].

4.2 Synthesis of Poly(3-hydroxypropionate-co-4-hydroxybutyrate) with Fully Controllable Structures by Recombinant Escherichia coli *Containing an Engineered Pathway*

Recently, microbial copolyesters containing 3HP have become increasingly interesting because of the ultrahigh strength brought about by 3HP, and these include P (3HB-*co*-3HP), P(3HP-*co*-3HB-*co*-3HH-*co*-3HO), P(4HB-*co*-3HP-*co*-Lactate), P (4HB-*co*-3HP-co-2HP), P(3HB-*co*-3HP-4HB-*co*-Lactate), and P(3HB-*co*-3HP-*co*-4HB-*co*-2HP) [78]. Natural bacteria are unable to produce 3-hydroxyproionate

(3HP) and 4-hydroxybutyrate (4HB) as building blocks for PHA synthase to make the unnatural biopolyester P(3HP-*co*-4HB) [11]. However, precursors of 3HP and 4HB can come from 1,3-propanediol (PDO) [29] and 1,4-butanediol (BDO) [79], respectively. Copolyesters of 3-hydroxypropionate (3HP) and 4-hydroxybutyrate (4HB), abbreviated as P(3HP-*co*-4HB), were synthesized by *E. coli* harboring a synthetic pathway consisting of five heterologous genes including *orfZ* encoding 4-hydroxybutyrate-coenzyme A transferase from *Clostridium kluyveri* [80, 81], *pcs'* encoding the ACS domain of tri-functional propionyl-CoA ligase (PCS) from *Chloroflexus aurantiacus* [82], *dhaT* and *aldD* encoding dehydratase and aldehyde dehydrogenase from *Pseudomonas putida* KT2442 [83], and *phaC1* encoding PHA synthase from *Ralstonia eutropha* (Fig. 4) [11, 29]. When grown on mixtures of 1,3-propanediol (PDO) and 1,4-butanediol (BDO), compositions of 4HB in microbial P(3HP-*co*-4HB) were controllable ranging from 12 mol% to 82 mol% depending on PDO:BDO ratios. Their mechanical and thermal properties showed obvious changes depending on the monomer ratios (Table 3). Morphologically, P(3HP-*co*-4HB) films only became fully transparent when monomer 4HB content was around 67 mol% (Fig. 5) [11].

Several key enzymes were considered as important for making P(3HP-*co*-4HB) copolymers with flexible 4HB content: propionyl-CoA synthetase (PCS') from the 3-hydroxypropionate cycle of phototrophic green non-sulfur eubacterium *Chloroflexus aurantiacus* is very likely to convert 3HP to 3HP-CoA, and 4HB-coenzyme, a transferase gene *orfz* from *Clostridium kluyveri*, was found to turn 4HB into 4HB-CoA effectively [82]. Genes *dhaT* and *aldD* were found to turn 1,4-butanediol (BDO) or/and 1,3-propanediol (PDO) into 4HB or/and 3HP, respectively. The enzyme encoded by *dhaT* was mostly active with substrates containing two primary alcohol groups separated by one or two carbon atoms such as 1,3-propanediol or 1,4-butanediol, and 3HP or/and 4HB yield were affected by expression levels of these two genes [29, 79, 83]. Promoter of PHA synthesis genes *phaCAB* operon from *Ralstonia eutropha* (P_{Re}) was demonstrated to be more active than *lac* promoter or T7 promoter transcriptionally in *E. coli*. Finally, PHA synthase

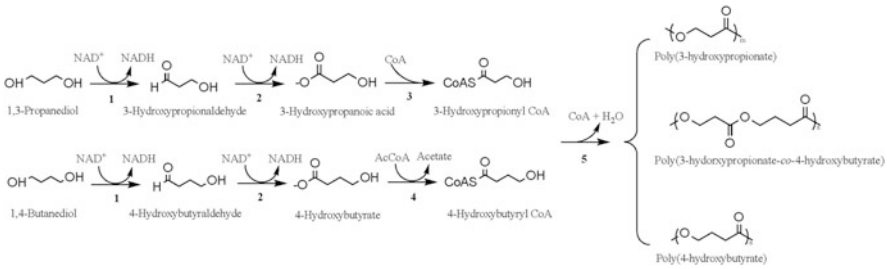

Fig. 4 Construction of P(3HP-*co*-4HB) biosynthetic pathways in recombinant *Escherichia coli* [11]. Enzymes for each numbered step are as follows: (*1*) 1,3-propanediol dehydrogenase; (*2*) aldehyde dehydrogenase; (*3*) propanoyl-CoA synthetase; (*4*) 4-hydroxybutyrate coenzyme A transferase; (*5*) PHA synthase

Table 3 Physical characterization of microbial P(3HP-*co*-4HB) [11]

	Thermal properties		Mechanical properties			
P(3HP-*co*-4HB) 4HB (mol%)	T_m (°C)	T_g (°C)	σ_y (MPa)	σ_{mt} (MPa)	ε_b (%)	E (MPa)
0	78.13	−17.85	33.83 ± 0.76	21.54 ± 1.10	497.6 ± 6.2	2889.3 ± 698.8
11.86	61.68	−24.42	12.98 ± 0.30	48.82 ± 1.73	1248.3 ± 63.4	3.9 ± 1.4
25.48	62.74	−31.28	1.70 ± 0.18	6.36 ± 0.76	962.9 ± 20.5	14.5 ± 0.4
37.89	63.46	−36.14	0.92 ± 0.03	0.54 ± 0.04	1611.0 ± 19.3	4.4 ± 0.1
67.00	64.77	−41.87	0.64 ± 0.03	0.34 ± 0.01	429.2 ± 161.3	1.8 ± 0.1
81.84	35.65	−29.48	2.61 ± 0.05	6.33 ± 1.76	594.7 ± 126.2	18.5 ± 0.5
100	61	−47	13.77 ± 1.41	34.66 ± 0.98	696.6 ± 43.6	180.9 ± 59.2
PHB	171.8	3.1	–	18.0 ± 0.7	3.0 ± 0.4	1,470 ± 78
PHBV	114	−5	–	26	27	1,900
$PHBHH_X$	125	0	–	7.0 ± 0.5	400 ± 36	283.2 ± 21

T_m melting temperature, T_g glass transition temperature, σ_y yield strength, σ_{mt} maximum tension strength, ε_b elongation at break, E Young's modulus

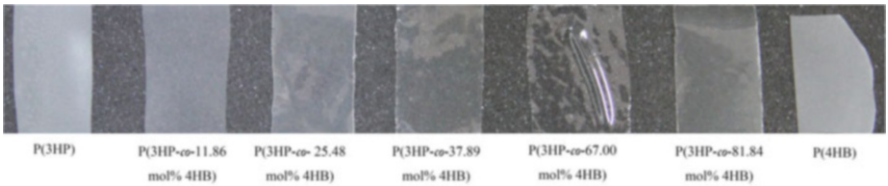

Fig. 5 Transparency of P(3HP-*co*-4HB) consisting of different monomer compositions [11]. From *left* to *right*: P(3HP), P(3HP-*co*-12 mol% 4HB), P(3HP-*co*-25 mol% 4HB), P (3HP-*co*-38 mol% 4HB), P(3HP-*co*-67 mol% 4HB), P(3HP-*co*-82 mol% 4HB), P(4HB)

PhaC1 of *R. eutropha* has sufficient activity for polymerizing SCL PHA monomers [11, 29].

A mixture of PDO and BDO in cultures of the recombinant *E. coli* S17-1 resulted in formation of copolyesters P(3HP-*co*-4HB) consisting of 3HP and 4HB. Compositions of the 3HP and 4HB in P(3HP-*co*-4HB) could be adjusted by changing the ratios of PDO to BDO. For example, 63 wt% P(17 mol% 3HP-*co*-83 mol% 4HB) was accumulated when the PDO:BDO ratio was 1/10; whereas a ratio of 1:1 led to the formation of P(70 mol% 3HP-*co*-30 mol% 4HB). When PDO/BDO was equal to 10/15 (or 2/3), only 2.3 wt% P(88 mol% 3HP-*co*-12 mol% 4HB) was synthesized, indicating the toxicity of high BDO or PDO concentration. Especially when the total concentration of BDO and PDO were over 20 g/L, the toxicity became very obvious, as indicated by significant reduction on CDW and PHA production. Obviously, a copolymer consisting of a defined 3HP:4HB ratio can be produced by adjusting the ratios of PDO:BDO. In this study, P(3HP-*co*-4HB) consisting of 17 mol% 3HP–88 mol% 3HP were obtained. Interestingly, the transparency of P (3HP-*co*-4HB) was also found to be dependent on monomer compositions. Only P (3HP-*co*-67 mol% 4HB) was a totally transparent material, whereas other PHA including P(3HP), P(3HP-*co*-12 mol% 4HB), P(3HP-*co*-25 mol% 4HB), P(3HP-*co*-38 mol% 4HB), P(3HP-*co*-82 mol% 4HB), and P(4HB) were observed to be less transparent [11].

The addition of 4HB monomer into P3HP led to the formation of P(3HP-*co*-4HB) which clearly lowered the P3HP melting temperatures (T_m) and the glass transition temperature (T_g) from 78°C and −18°C to 61–65°C and −24°C to –41°C with the 4HB ratio increased from 12 mol% to 67 mol% (Table 3). Interestingly, P (3HP-*co*-82 mol% 4HB) was revealed to have a much lower T_m of 36°C and a higher T_g of −29°C compared to other copolymers. T_m seemed to stabilize at around 63°C in copolymers consisting of 12–67 mol% 4HB. T_g decreased from −24°C to −42°C with 4HB content increasing from 12 mol% to 67 mol%. Homopolymer P4HB had the lowest T_g of −47°C with a T_m of 61°C (Table 3).

Copolymerization reduced yield strengths and Young's modulus of both P3HP and P4HB (Table 3). However, the elongation at breaks showed an improvement for P(3HP-*co*-4HB) consisting of 12–38 mol% 4HB over P3HP and P4HB. On the

other hand, only P(3HP-*co*-12 mol% 4HB) had an increase on maximum tension strength over other homo- and copolymers. In terms of thermal and mechanical properties, P(3HP-*co*-4HB) seems to be unique in combined properties compared with commercial PHA such as PHB, PHBV, and PHBHHx.

As PDO and BDO can be respectively biosynthesized from glucose [84, 85], it becomes possible to establish an engineering pathway for production of P(3HP-*co*-4HB). Block copolymers of P3HB-*b*-P3HP could also be produced [59]. The two pathways supplied 3HP and 4HB monomers independently, leading to the formation of homopolymer P3HP in the absence of 4HB, of P4HB in the absence of 3HP, or to random copolymers of P(3HP-*co*-4HB) when 3HP and 4HB were both available.

4.3 *Poly(3-hydroxybutyrate*-co-*3-hydroxypropionate) from Glucose by Engineering* Escherichia coli

Poly(3-hydroxypropionate) (P3HP), an scl-PHA containing three carbon atoms without side chain, shows the best combined mechanical properties, including an elongation at break of more than 600%, and a Young's modulus of 3 GPa [11]. P3HP therefore stands out as a PHA member that holds great promise. No microorganism has been known to synthesize homopolymer P3HP so far. Thus, recombinants have been developed to produce P3HP. Andreessen et al. [28] first reported bacterial synthesis of P3HP using glycerol as carbon source in a two-step fed-batch fermentation. Wang et al. [77] modified the process by replacing the strict anaerobic glycerol dehydratase from *Clostridium butyricum* with the vitamin B12-dependent glycerol dehydratase DhaB123 from *Klebsiella pneumonia*. Zhou et al. [29] used 1,3-propanediol as a precursor to produce over 90% P3HP in *E. coli* cell dry weight (CDW). There were attempts to synthesize P3HP from an unrelated carbon source starting with acetyl-CoA [86, 87]. The related pathway involves carboxylation of acetyl-CoA to malonyl-CoA, reduction of malonyl-CoA to 3HP, its coupling to CoA, and their following polymerization. This recombinant pathway led to only 1.32 g/L CDW containing 0.98% P3HP [87].

The authors' lab reported that multiple genes from various sources were assembled into a new pathway for the production of P3HP from glucose as a sole carbon source, including *gpd1* (glycerol-3-P dehydrogenase) and *gpp2* (glycerol-3-P phosphatase) from *Saccharomyces cerevisiae* [88, 89], *dhaB1-3* (glycerol dehydratase) and *gdrAB* (glycerol dehydratase reactivating factor) from *Klebsiella pneumonia* [90, 91], *pduP* (propionaldehyde dehydrogenase) from *Salmonella typhimurium* [92, 93], *phaC* (PHA synthase) from *Ralstonia eutropha* [26, 94], *aldD* (aldehyde dehydrogenase) and *dhaT* (1,3-propanediol dehydrogenase) from *Pseudomonas putida* KT2442 [79, 83], and *pcs'* (propanoyl-CoA synthetase) from *Chloroflexus aurantiacus* [29]. When the plasmid containing the above multiple genes was

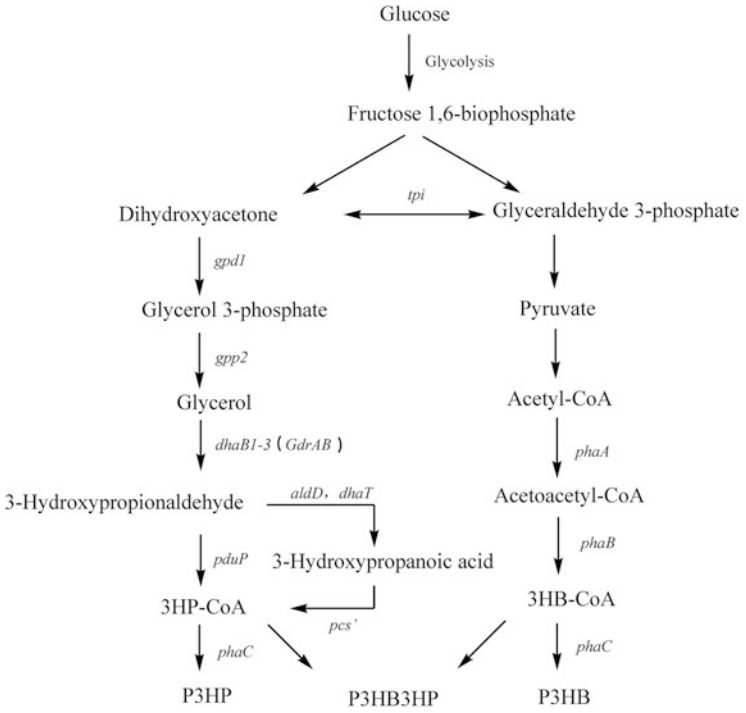

Fig. 6 Construction of P3HP and P(3HB-*co*-3HP) biosynthetic pathways from glucose as a sole carbon source in recombinant *Escherichia coli* [34]. Enzymes encoded by each gene are described below: *gpd1* glycerol-3-P dehydrogenase (*Saccharomyces cerevisiae*), *gpp2* glycerol-3-P phosphatase (*Saccharomyces cerevisiae*), *dhaB1-3* glycerol dehydratase (*Klebsiella pneumoniae*), *gdrAB* glycerol dehydratase reactivating factors (*Klebsiella pneumoniae*), *pduP* propionaldehyde dehydrogenase (*Salmonella typhimurium*), *phaC* polyhydroxyalkanoate synthase (*Ralstonia eutropha*), *aldD* aldehyde dehydrogenase (*Pseudomonas putida*), *dhaT* 1,3-propanediol dehydrogenase (*Pseudomonas putida*), *pcs'* propanoyl-CoA synthetase (*Chloroflexus aurantiacus*), *phaA* β-ketothiolase (*Ralstonia eutropha*), *phaB* NADPH-dependent acetoacetyl-CoA reductase (*Ralstonia eutropha*)

transformed into *E. coli*, up to 18.4% P3HP homopolymer was produced from glucose (Fig. 6) [34].

The expression of the two genes *gpd1* and *gpp2* allows dihydroxyacetone from glucose glycolysis to form glycerol-3-phosphate, which is further hydrolyzed to generate glycerol [95]. Glycerol is converted to 3-hydroxypropionaldehyde by glycerol dehydratase (DhaB1-3) from *Klebsiella pneumonia,* which is an important intermediate for P3HP, and 3-hydroxypropionaldehyde is converted to 3-hydroxypropionate (3HP) by aldehyde dehydrogenase (AldD) cloned from *Pseudomonas putida* KT2442. Propionyl-CoA synthetase (PCS') from *Chloroflexus aurantiacus* should be able to change 3HP to 3HP-CoA. At the same time, 3-hydroxypropionaldehyde can also be directly turned into 3HP-CoA by

propionaldehyde dehydrogenase (PduP) from *Salmonella typhimurium*. To increase the activity of glycerol dehydratase, *gdrAB*, a reactivation factor for glycerol dehydratase was inserted into the above-mentioned pathway. When gene *pudP* was used to replace *aldD* and *dhaT*, the resulting plasmid pDC02 became the only plasmid containing the entire pathway from glucose to P3HP. Recombinant *E. coli* Trans1-T1 (pDC02) produced over 18% P3HP in over 5 g/L CDW when grown in glucose LB medium whereas in the glucose mineral medium, 12% P3HP was accumulated in 3 g/L CDW. More P3HP accumulation from glucose is expected when the metabolic flux is further optimized [34].

When a P3HB synthesis pathway containing the P3HB synthesis operon *phaCAB* from *Ralstonia eutropha* was added to the P3HP synthesis pathway, the recombinant harboring the P3HB and P3HP pathways started to produce random copolymers of P3HB3HP from glucose as the sole carbon source. This study demonstrated that ultra-strong polyhydroxyalkanoates (PHA), mainly P3HP and P3HB3HP, can be synthesized from low cost glucose using synthetic biology approaches.

The two plasmids p15apCAB and pDC02, which harbor three genes and nine genes from different microorganisms responsible for P3HP and P3HB syntheses from glucose, respectively, can be regarded as bio-devices or bio-bricks that are assembled to perform their functions (Fig. 7). This study can serve as a typical synthetic biology example that uses bio-bricks or bio-devices to achieve biological functions. In this case, it was the synthetic biology for production of novel bio-polyesters. In total, 11 heterogeneous genes were cloned from other microorganisms and were assembled to become new pathways to meet our new demands.

In the future, the two polyester synthesis pathways could be transformed into other microbial hosts after codon optimization to enhance P3HB3HP production by some industrial microbial hosts [34].

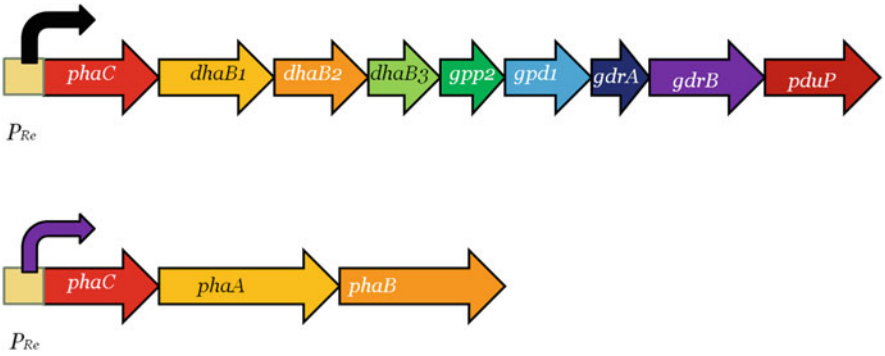

Fig. 7 Orders of gene arrangements on plasmids pDC02 and p15apCAB, respectively [34]

4.4 Engineering the β-Oxidation Pathway on the Chromosome for mcl PHA Synthesis

Many *Pseudomonas* spp. are able to utilize fatty acids via their β-oxidation to obtain both energy and substrates for cell growth. The β-oxidation pathway shortens the fatty acid chain lengths in each cycle by two carbon atoms, generating several PHA monomers of different lengths, which can result in the formation of random PHA copolymers (Fig. 8).

Recently, the authors' lab succeeded in engineering the β-oxidation pathway encoded on the chromosomes of *Pseudomonas putida* and *Pseudomonas entomophiles*, resulting in controllable PHA composition, including formation of PHA homopolymers and composition-adjustable random copolymers and block copolymers [35, 51, 60]. To avoid the changing of fatty acid substrate structures, chromosomal genes related to β-oxidation were selectively deleted to weaken β-oxidation in *Pseudomonas* spp., so that fatty acids can maintain their structures when used as PHA monomer precursors.

Mutant *Pseudomonas putida* KTQQ20, a derivative of *P. putida* KT2442, deleted key fatty acid degradation enzymes encoded by genes *fadB*, *fadA*, *fadB2x*,

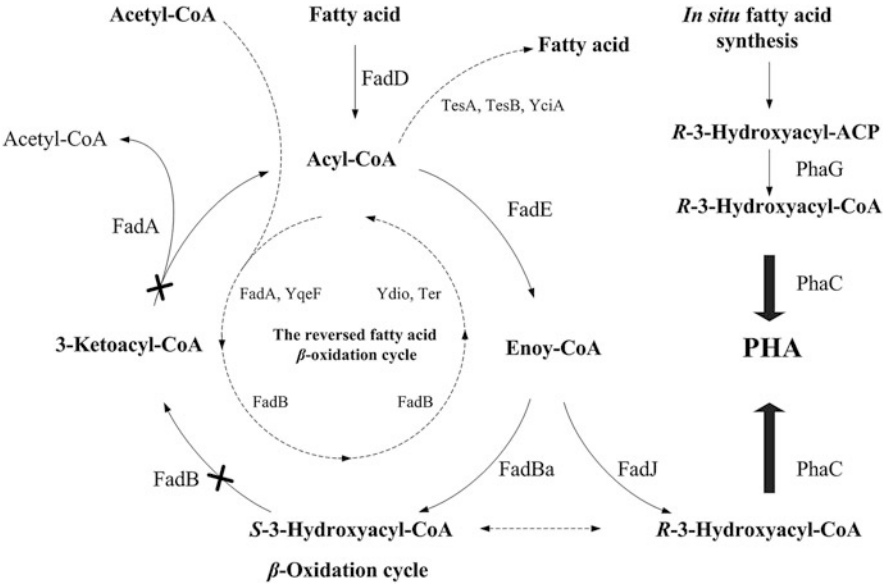

Fig. 8 The weakened beta-oxidation cycle, reversed fatty acid beta-oxidation cycle and in situ fatty acid synthesis. Enzymes in β-oxidation cycle: FadD fatty acid-CoA ligase, FadE acyl-CoA dehydrogenase, FadBa S-enoyl-CoA hydratase, FadB 3-hydroxyacyl-CoA dehydrogenase, FadA acetyl-CoA acetyltransferase, PhaJ R-enoyl-CoA hydratase, PhaC PHA synthase, PhaG 3-hydroxyacyl-CoA-acyl carrier protein transferase. Genes in the reversed fatty acid β-oxidation cycle: *yqeF*/*fadA* thiolase, *fadB* hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase, *ydiO* enoyl-CoA reductase, *ter* trans-2-enoyl-CoA reductase, *tesA*/*tesB*/*yciA* thioesterase

and *fadAx*, as well as *PP2047* and *PP2048* encoding 3-hydroxyacyl-CoA dehydrogenase and acyl-CoA dehydrogenase, respectively, combined with the deletion of *phaG* encoding 3-hydroxyacyl-CoA-acyl carrier protein transferase, becomes defective in fatty acid β-oxidation activity. The strain was now able to synthesize homopolymer poly(3-hydroxydecanoate) or PHD and P(3HD-*co*-84 mol% 3HDD) when grown on decanoic acid or dodecanoic acid, respectively [35]. When grown on mixtures of the sodium salt of hexanoate (C6) and decanoate (C10), it produced random copolymers of P(3HHx-*co*-3HD) with monomer compositions easily regulated by varying the C6:C10 ratio. *P. putida* KTQQ20 also produced diblock copolymer P3HHx-*b*-P(3HD-*co*-3HDD) when sodium salts of hexanoate (C6) and decanoate (C10) were fed to its culture one after another [60].

Pseudomonas entomophila strain L48, a strong fatty acid utilizer, was also investigated for microbial production of mcl PHA. A total of 70.2% of *P. entomophila* genes have orthologs with the *P. putida* genome, of which >96% are found in synteny. The β-oxidation activity of *P. entomophila* was weakened by deleting similar genes on its chromosome as in *P. putida*. The resulting *P. entomophila* LAC26 accumulated over 90 wt% PHA consisting of 99 mol% 3HDD. Homopolymers of C6–C14 were all accumulated, respectively, when an equal chain length of a fatty acid was fed to the mutant for related PHA homopolymer production [51].

4.5 Pathways for scl and mcl PHA Copolymers

P. putida KTOYO6 is a fatty acid β-oxidation impaired mutant in which genes of 3-ketoacyl-CoA thiolase (*fadA*) and 3-hydroxyacyl-CoA dehydrogenase (*fadB*) were deleted to a maximum level to improve fatty acid utilization for PHA synthesis [53]. When its mcl PHA synthase (C6–C14) was replaced by a less specific synthase operon $phaPCJ_{Ac}$ which could synthesize both scl and mcl monomers (C3–C7) from *Aeromonas caviae,* recombinant *P. putida* KTOYO6ΔC ($phaPCJ_{Ac}$) was able to produce a diblock copolymer of PHB-*b*-PHVHHp by controlling the sequential feeding time of sodium butyrate and sodium heptanoate. When cultivated on mixtures of sodium salts of butyrate (C4) and hexanoate (C6), random copolymers of P(3HB-*co*-3HHx) were accumulated with monomer contents adjustable by C4:C6 ratios [55].

5 Functional PHA

When cultures of engineered strains, such as *P. putida* KTQQ20 or *P. entomophila* LAC23, were fed with fatty acids containing functional groups such as double or triple bonds, epoxy, carbonyl, cyano, phenyl and halogen group, respectively [46],

the resulting PHA contains the functional groups on the side chains, allowing further chemical modifications (grafting) on the side chains.

Homopolymers with 100 mol% content of aromatic moieties, random copolymers, or a blend of both have been produced [47]. Hydrophilic PHA bearing alkoxy, acetoxy, or hydroxyl groups are also of great interest, as they show enhanced solubility and biocompatibility [46].

The β-oxidation weakened *P. entomophila* LAC23 was found able to accumulate PHA containing phenyl groups on the side chains. When cultured in 5-phenylvaleric acid, only homopolymer poly(3-hydroxy-5-phenylvalerate) was synthesized. Copolyesters of 3-hydroxy-5-phenylvalerate (3HPhV) and 3-hydroxydodecanoate (3HDD) were also successfully produced by *P. entomophila* LAC23 when grown on mixtures of phenylvaleric acid and dodecanoic acid. Compositions of 3HPhV in P(3HPhV-*co*-3HDD) were controllable, ranging from 3% to 32%, depending on dodecanoic acid:5-phenylvaleric acid ratios [47]. Although the production of PHA with functional groups is still facing high costs and low productivity, the toxicity of substrates also affect the growth of microorganisms, and PHA with functional groups needs be produced from unrelated carbon sources in future studies.

6 Engineering the Bacterial PHA Synthesis Using CRISPRi

Clustered regularly interspaced short palindromic repeats interference (CRISPRi) is a powerful technology used to regulate eukaryotic genomes [96]. CRISPRi has also been reported to control PHA biosynthesis pathway flux and to adjust PHA composition. First, an *E. coli* strain was engineered by introducing a pathway for the production of P3HB4HB from glucose [18]. The native gene *sad*, encoding succinate semi-aldehyde dehydrogenase, was regulated by CRISPRi using five specially designed single guide RNAs (sgRNAs) for controlling carbon flux toward 4-hydroxybutyrate (4HB) biosynthesis in *E. coli*. The system allowed formation of P3HB4HB consisting of 1–9 mol% 4HB. Additionally, succinate, generated by succinyl-CoA synthetase and succinate dehydrogenase (respectively encoded by genes *sucC*, *sucD*, *sdhA*, and *sdhB*) was channeled preferentially to the 4HB precursor using selected sgRNAs such as *sucC2*, *sucD2*, *sdhB2*, and *sdhA1* via CRISPRi. The resulting 4HB content in P3HB4HB could be adjusted from 1.4 mol% to 18.4 mol% depending on the expression levels of down-regulated genes (Fig. 9). The results show that CRISPRi is a feasible approach to simultaneously manipulate multiple genes and control metabolic flux in *E. coli* [97].

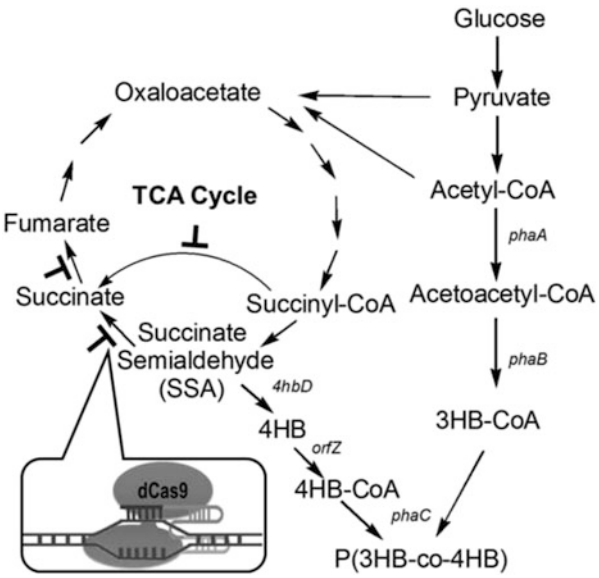

Fig. 9 CRISPRi as a tool to control P3HB4HB biosynthesis pathway flux and to adjust 3HB/4HB composition [97]. Engineered pathways for P3HB4HB synthesis by recombinant *Escherichia coli*. The CRISPRi system was used to repress gene transcription initiation and elongation in the related pathways. To obtain P3HB4HB consisting of various 4HB ratios, several genes can be manipulated simultaneously, including following genes: *phaA* beta-ketothiolase, *phaB* NADPH-dependent acetoacetyl-CoA reductase, *phaC* PHA synthase, *sucD* succinate semi-aldehyde dehydrogenase, *4hbD* 4-hydroxybutyrate dehydrogenase, *orfZ* 4-hydroxybutyrate CoA transferase

7 Engineering the Bacterial Shapes for Enhanced Polyhydroxyalkanoates Accumulation

Most bacteria have a small size ranging from 0.5 μm to 2 μm, preventing the bacterial cells from accumulating large amounts of inclusion bodies intracellularly, even though the bacteria are able to grow very fast. To overcome the size limitation, it is important to make bacterial cells larger. That is to say, a larger intracellular space is needed for more inclusion body accumulation. Various approaches were taken to increase the bacterial cell sizes, including deletion on actin-like protein gene *mreB*, weak expression of *mreB* in *mreB* deletion mutant, and weak expression of *mreB* in *mreB* deletion mutant under inducible expression of *sulA*, the inhibitor of division ring protein gene *ftsZ*. All of the methods resulted in different levels of increases in bacterial sizes and PHB granules accumulation [98].

MreB, the actin-like bacterial cytoskeletons, which also affects bacterial morphology, was considered a suitable engineering target for expanding the cell volumes [99]. When *mreB* was deleted, *E. coli* changed from rods to spherical

shapes, and some cells even increased their sizes to diameters of around 10 μm. More PHB granules were accumulated in the large *E. coli* JM109SG (Δ*mreB*) cells. However, *E. coli* JM109SG (Δ*mreB*) also appeared to be fragile and a fraction of cells ruptured during the growth stage. This phenomenon showed that MreB may provide critical support for maintaining the cell shape. Ectopic expression of *MreB* in a wild-type bacterium was found to interfere with normal MreB cytoskeleton formation, resulting in a larger cell size compared with that of a wild type. To increase the cell size further, the *mreB* gene was compensated by constitutively expressing *mreB* in a weaker manner in *MreB* deleted *E. coli* JM109SG together with an arabinose-inducible *sulA* gene encoding an inhibitory protein for the formation of the cell division ring (FtsZ ring), the overexpression of which leads to elongated cells. Remarkably, an increase of over 100% PHB accumulation was observed in recombinant *E. coli* overexpressing *mreB* in an *mreB* deletion mutant under inducible expression of gene *ftsZ* inhibiting protein SulA (Fig. 10). The molecular mechanism of enlarged bacterial size was found to be directly related to the weakened cytoskeleton, which was the result of broken skeleton helix [98]. The larger *E. coli* cells make it possible to produce more PHA.

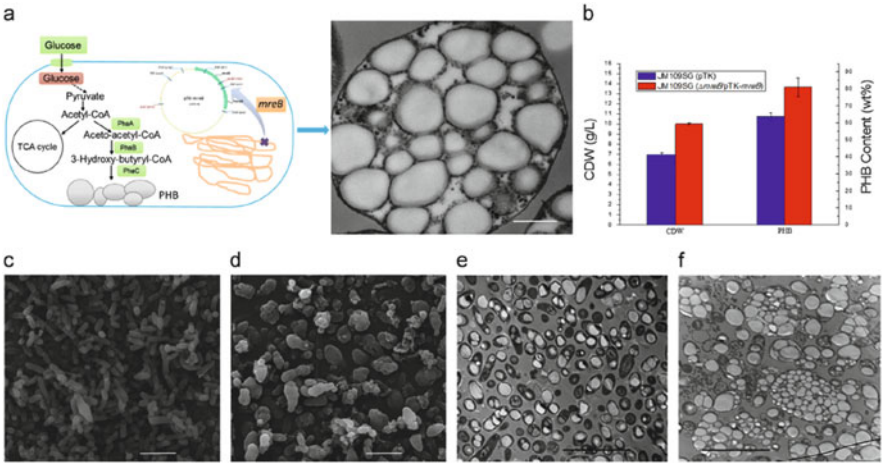

Fig. 10 Electron microscopy studies on morphology and PHB production by *E. coli* JM109SG (Δ*mreB*) overexpressing *mreB* [98]. (**a**) Schematic of PHB accumulation in *E. coli* JM109SG (Δ*mreB*) overexpressing *mreB*. *Scale bar*: 0.5 μm; (**b**) Growth and PHB accumulation by recombinants harboring pBHR68 cultivated in minimal medium at 30°C for 10 h followed by addition of 20 g/L glucose and another 40 h of growth. *Error bars*: s.d. ($n = 3$). *E. coli* JM109SG (pTK/pBHR68) (**c**) and *E. coli* JM109SG(ΔmreB/pTK-mreB/pBHR68) (**d**) were grown in LB medium at 30°C for 10 h followed by addition of 20 g/L glucose and another 40 h of growth. TEM on sections of cells of control *E. coli* JM109SG (pTK/pBHR68) (**e**) and *E. coli* JM109SG (ΔmreB/pTK-mreB/pBHR68) (**f**) cultivated in the LB medium at 30°C for 10 h, followed by addition of 20 g/L glucose and another 40 h of growth. *Scale bar*: 5 μm

8 Conclusion

The application of PHA as a low-cost biodegradable plastic has been hampered by its higher production cost and the difficulty to control precisely their structures and properties. Global efforts have been made to develop technology for lowering the PHA production cost. With the successful construction of β-oxidation weakened *Pseudomonas* spp. as PHA production platforms, it is possible to control the formation of homopolymers and random- and block copolymers including monomer structures and ratios, and this allows us to obtain PHA with consistent properties. At the same time, it is possible to introduce various functional groups into the PHA side chains in a quantitative way, which provides more opportunities for side-chain grafting. Functional PHA together with endless possibilities for grafting have provided us with limitless ways of making new PHA, possibly with some high value-added functionalities. With the development of synthetic biology, it also becomes possible to construct unnatural pathways to produce novel PHA with strong value-added properties. It is widely held that within 5 or 10 years, many novel properties including environmental responsiveness, shape memory ability, controllable biodegradability, and mechanical ultra-strength will be developed from the diverse PHA materials. Thus, with the development of synthetic biology, we open a new PHA golden era.

Acknowledgements This research was financially supported by 973 Basic Research Fund (Grant No. 2012CB725201), National High Tech 863 Grants (Grant No. 2012AA023102), and a Grant from National Natural Science Foundation of China (Grant No. 31270146).

References

1. Brandl H, Gross RA, Lenz RW, Fuller RC (1990) Plastics from bacteria and for bacteria: poly (beta-hydroxyalkanoates) as natural, biocompatible, and biodegradable polyesters. Adv Biochem Eng Biotechnol 41:77–93
2. Chen GQ (2009) A microbial polyhydroxyalkanoates (PHA) based bio- and materials industry. Chem Soc Rev 38:2434–2446
3. Poirier Y, Nawrath C, Somerville C (1995) Production of polyhydroxyalkanoates, a family of biodegradable plastics and elastomers, in bacteria and plants. Biotechnology 13:142–150
4. Chen GQ, Wu Q (2005) The application of polyhydroxyalkanoates as tissue engineering materials. Biomaterials 26:6565–6578
5. Gao X, Chen JC, Wu Q, Chen GQ (2011) Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels. Curr Opin Biotechnol 22:768–774
6. Volova TG, Gladyshev MI, Trusova MY, Zhila NO (2006) Degradation of polyhydroxyalkanoates and the composition of microbial destructors under natural conditions. Microbiology 75:593–598
7. Steinbüchel A, Lütke-Eversloh T (2003) Metabolic engineering and pathway construction for biotechnological production of relevant polyhydroxyalkanoates in microorganisms. Biochem Eng J 16:81–96
8. Steinbüchel A, Valentin HE (1995) Diversity of bacterial polyhydroxyalkanoic acids. FEMS Microbiol Lett 128:219–228

9. Meng DC, Shen R, Yao H, Chen JC, Wu Q, Chen GQ (2014) Engineering the diversity of polyesters. Curr Opin Biotechnol 29:24–33
10. Pederson EN, McChalicher CW, Srienc F (2006) Bacterial synthesis of PHA block copolymers. Biomacromolecules 7:1904–1911
11. Meng DC, Shi ZY, Wu LP, Zhou Q, Wu Q, Chen JC, Chen GQ (2012) Production and characterization of poly(3-hydroxypropionate-*co*-4-hydroxybutyrate) with fully controllable structures by recombinant *Escherichia coli* containing an engineered pathway. Metab Eng 14:317–324
12. Doi Y (1989) Production, properties, and biodegradation of microbial copolyesters of 3-hydroxybutyrate and 4-hydroxybutyrate. Abstr Pap Am Chem S 197:45-Btec
13. Lemoigne M (1926) Products of dehydration and of polymerization of β-hydroxybutyric acid. Bull Soc Chem Biol 8:770–782
14. Savenkova L, Gercberga Z, Nikolaeva V, Dzene A, Bibers I, Kalnin M (2000) Mechanical properties and biodegradation characteristics of PHB-based films. Process Biochem 35:573–579
15. Fiedler S, Steinbuchel A, Rehm BH (2002) The role of the fatty acid beta-oxidation multienzyme complex from *Pseudomonas oleovorans* in polyhydroxyalkanoate biosynthesis: molecular characterization of the *fadBA* operon from *P. oleovorans* and of the enoyl-CoA hydratase genes *phaJ* from *P. oleovorans* and *Pseudomonas putida*. Arch Microbiol 178:149–160
16. Timm A, Steinbüchel A (1990) Formation of polyesters consisting of medium-chain-length 3-hydroxyalkanoic acids from gluconate by *Pseudomonas aeruginosa* and other fluorescent pseudomonads. Appl Environ Microbiol 56:3360–3367
17. Chanprateep S (2010) Current trends in biodegradable polyhydroxyalkanoates. J Biosci Bioeng 110:621–632
18. Li ZJ, Shi ZY, Jian J, Guo YY, Wu Q, Chen GQ (2010) Production of poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) from unrelated carbon sources by metabolically engineered *Escherichia coli*. Metab Eng 12:352–359
19. Chen GQ, Patel MK (2012) Plastics derived from biological sources: present and future: a technical and environmental review. Chem Rev 112:2082–2099
20. Tyo KE, Kocharin K, Nielsen J (2010) Toward design-based engineering of industrial microbes. Curr Opin Microbiol 13:255–262
21. Kichise T, Fukui T, Yoshida Y, Doi Y (1999) Biosynthesis of polyhydroxyalkanoates (PHA) by recombinant *Ralstonia eutropha* and effects of PHA synthase activity on in vivo PHA biosynthesis. Int J Biol Macromol 25:69–77
22. Lee SY, Wong HH, Choi JI, Lee SH, Lee SC, Han CS (2000) Production of medium-chain-length polyhydroxyalkanoates by high-cell-density cultivation of *Pseudomonas putida* under phosphorus limitation. Biotechnol Bioeng 68:466–470
23. Tan IKP, Kumar KS, Theanmalar M, Gan SN, Gordon B (1997) Saponified palm kernel oil and its major free fatty acids as carbon substrates for the production of polyhydroxyalkanoates in *Pseudomonas putida* PGA1. Appl Microbiol Biotechnol 47:207–211
24. Hyakutake M, Saito Y, Tomizawa S, Mizuno K, Tsuge T (2011) Polyhydroxyalkanoate (PHA) synthesis by class IV PHA synthases employing *Ralstonia eutropha* PHB-4 as host strain. Biosci Biotechnol Biochem 75:1615–1617
25. Rehm BHA, Steinbüchel A (1999) Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis. Int J Biol Macromol 25:3–19
26. Yuan W, Jia Y, Tian J, Snell KD, Muh U, Sinskey AJ, Lambalot RH, Walsh CT, Stubbe J (2001) Class I and III polyhydroxyalkanoate synthases from *Ralstonia eutropha* and *Allochromatium vinosum*: characterization and substrate specificity studies. Arch Biochem Biophys 394:87–98
27. Sudesh K, Abe H, Doi Y (2000) Synthesis, structure and properties of polyhydroxyalkanoates: biological polyesters. Prog Polym Sci 25:1503–1555
28. Andreessen B, Lange AB, Robenek H, Steinbüchel A (2010) Conversion of glycerol to poly (3-hydroxypropionate) in recombinant *Escherichia coli*. Appl Environ Microbiol 76:622–626

29. Zhou Q, Shi ZY, Meng DC, Wu Q, Chen JC, Chen GQ (2011) Production of 3-hydroxypropionate homopolymer and poly(3-hydroxypropionate-*co*-4-hydroxybutyrate) copolymer by recombinant *Escherichia coli*. Metab Eng 13:777–785
30. Slater SC, Voige WH, Dennis DE (1988) Cloning and expression in *Escherichia coli* of the *Alcaligenes eutrophus* H16 poly-beta-hydroxybutyrate biosynthetic pathway. J Bacteriol 170:4431–4436
31. Steinbüchel A, Valentin H, Schönebaum A (1994) Application of recombinant gene technology for production of polyhydroxyalkanoic acids: biosynthesis of poly(4-hydroxybutyric acid) homopolyester. J Environ Polym Degrad 2:67–74
32. Zhou XY, Yuan XX, Shi ZY, Meng DC, Jiang WJ, Wu LP, Chen JC, Chen GQ (2012) Hyperproduction of poly(4-hydroxybutyrate) from glucose by recombinant *Escherichia coli*. Microb Cell Fact 11:54
33. Steinbüchel A, Debzi E-M, Marchessault R, Timm A (1993) Synthesis and production of poly (3-hydroxyvaleric acid) homopolyester by *Chromobacterium violaceum*. Appl Microbiol Biotechnol 39:443–449
34. Meng DC, Wang Y, Wu LP, Shen R, Chen JC, Wu Q, Chen GQ (2015) Production of poly (3-hydroxypropionate) and poly(3-hydroxybutyrate-*co*-3-hydroxypropionate) from glucose by engineering *Escherichia coli*. Metab Eng 29:189–195
35. Liu Q, Luo G, Zhou XR, Chen GQ (2011) Biosynthesis of poly(3-hydroxydecanoate) and 3-hydroxydodecanoate dominating polyhydroxyalkanoates by beta-oxidation pathway inhibited *Pseudomonas putida*. Metab Eng 13:11–17
36. Chen JY, Song G, Chen GQ (2006) A lower specificity PhaC2 synthase from *Pseudomonas stutzeri* catalyses the production of copolyesters consisting of short-chain-length and medium-chain-length 3-hydroxyalkanoates. Anton Leeuw Int J G 89:157–167
37. Loo CY, Lee WH, Tsuge T, Doi Y, Sudesh K (2005) Biosynthesis and characterization of poly (3-hydroxybutyrate-*co*-3-hydroxyhexanoate) from palm oil products in a *Wautersia eutropha* mutant. Biotechnol Lett 27:1405–1410
38. Gao X, Yuan XX, Shi ZY, Guo YY, Shen XW, Chen JC, Wu Q, Chen GQ (2012) Production of copolyesters of 3-hydroxybutyrate and medium-chain-length 3-hydroxyalkanoates by *E. coli* containing an optimized PHA synthase gene. Microb Cell Fact 11:130
39. Huisman GW, de Leeuw O, Eggink G, Witholt B (1989) Synthesis of poly-3-hydroxyalkanoates is a common feature of fluorescent pseudomonads. Appl Environ Microbiol 55:1949–1954
40. Klinke S, Ren Q, Witholt B, Kessler B (1999) Production of medium-chain-length poly (3-hydroxyalkanoates) from gluconate by recombinant *Escherichia coli*. Appl Environ Microbiol 65:540–548
41. Verwoert II, Verbree EC, van der Linden KH, Nijkamp HJ, Stuitje AR (1992) Cloning, nucleotide sequence, and expression of the *Escherichia coli* fabD gene, encoding malonyl coenzyme A-acyl carrier protein transacylase. J Bacteriol 174:2851–2857
42. Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol Bioeng 105:161–171
43. Zou XH, Chen GQ (2007) Metabolic engineering for microbial production and applications of copolyesters consisting of 3-hydroxybutyrate and medium-chain-length 3-hydroxyalkanoates. Macromol Biosci 7:174–182
44. Wallen LL, Rohwedde WK (1974) Poly-beta-hydroxyalkanoate from activated-sludge. Environ Sci Technol 8:576–579
45. Li SJ, Cai LW, Wu LP, Zeng GD, Chen JC, Wu Q, Chen GQ (2014) Microbial synthesis of functional homo-, random, and block polyhydroxyalkanoates by beta-oxidation deleted *Pseudomonas entomophila*. Biomacromolecules 15:2310–2319
46. Olivera E, Arcos M, Naharro G, Luengo J (2010) Unusual PHA biosynthesis. In: Chen GG-Q (ed) Plastics from bacteria. Microbiol Monographs, vol 14. Springer, Berlin, Heidelberg, pp. 133–186

47. Shen R, Cai L, Meng D, Wu L, Guo K, Dong G, Liu L, Chen J, Wu Q, Chen G (2014) Benzene containing polyhydroxyalkanoates homo- and copolymers synthesized by genome edited *Pseudomonas entomophila*. Sci China Life Sci 57:4–10
48. Abraham GA, Gallardo A, San Roman J, Olivera ER, Jodra R, Garcia B, Minambres B, Garcia JL, Luengo JM (2001) Microbial synthesis of poly(beta-hydroxyalkanoates) bearing phenyl groups from *Pseudomonas putida*: chemical structure and characterization. Biomacromolecules 2:562–567
49. Slater S, Houmiel KL, Tran M, Mitsky TA, Taylor NB, Padgette SR, Gruys KJ (1998) Multiple beta-ketothiolases mediate poly(beta-hydroxyalkanoate) copolymer synthesis in *Ralstonia eutropha*. J Bacteriol 180:1979–1987
50. Wang HH, Li XT, Chen GQ (2009) Production and characterization of homopolymer polyhydroxyheptanoate (P3HHp) by a *fadBA* knockout mutant *Pseudomonas putida* KTOY06 derived from *P. putida* KT2442. Process Biochem 44:106–111
51. Chung AL, Jin HL, Huang LJ, Ye HM, Chen JC, Wu Q, Chen GQ (2011) Biosynthesis and characterization of poly(3-hydroxydodecanoate) by beta-oxidation inhibited mutant of *Pseudomonas entomophila* L48. Biomacromolecules 12:3559–3566
52. Jagoda A, Ketikidis P, Zinn M, Meier W, Kita-Tokarczyk K (2011) Interactions of biodegradable poly([*R*]-3-hydroxy-10-undecenoate) with 1,2-dioleoyl-sn-glycero-3-phosphocholine lipid: a monolayer study. Langmuir 27:10878–10885
53. Ouyang SP, Luo RC, Chen SS, Liu Q, Chung A, Wu Q, Chen GQ (2007) Production of polyhydroxyalkanoates with high 3-hydroxydodecanoate monomer content by *fadB* and *fadA* knockout mutant of *Pseudomonas putida* KT2442. Biomacromolecules 8:2504–2511
54. Lütke-Eversloh T, Bergander K, Luftmann H, Steinbüchel A (2001) Biosynthesis of poly (3-hydroxybutyrate-*co*-3-mercaptobutyrate) as a sulfur analogue to poly(3-hydroxybutyrate) (PHB). Biomacromolecules 2:1061–1065
55. Li SY, Dong CL, Wang SY, Ye HM, Chen GQ (2011) Microbial production of polyhydroxyalkanoate block copolymer by recombinant *Pseudomonas putida*. Appl Microbiol Biotechnol 90:659–669
56. Hu D, Chung AL, Wu LP, Zhang X, Wu Q, Chen JC, Chen GQ (2011) Biosynthesis and characterization of polyhydroxyalkanoate block copolymer P3HB-*b*-P4HB. Biomacromolecules 12:3166–3173
57. Tripathi L, Wu LP, Chen J, Chen GQ (2012) Synthesis of diblock copolymer poly-3-hydroxybutyrate-*block*-poly-3-hydroxyhexanoate [PHB-*b*-PHHx] by a beta-oxidation weakened *Pseudomonas putida* KT2442. Microb Cell Fact 11:44
58. Wang Q, Yang P, Xian M, Liu H, Cao Y, Yang Y, Zhao G (2013) Production of block copolymer poly(3-hydroxybutyrate)-*block*-poly(3-hydroxypropionate) with adjustable structure from an inexpensive carbon source. ACS Macro Lett 2:996–1000
59. Tripathi L, Wu LP, Meng D, Chen J, Chen GQ (2013) Biosynthesis and characterization of diblock copolymer of p(3-hydroxypropionate)-*block*-p(4-hydroxybutyrate) from recombinant *Escherichia coli*. Biomacromolecules 14:862–870
60. Tripathi L, Wu LP, Meng DC, Chen JC, Wu Q, Chen GQ (2013) *Pseudomonas putida* KT2442 as a platform for the biosynthesis of polyhydroxyalkanoates with adjustable monomer contents and compositions. Bioresour Technol 142:225–231
61. Hazer B (1994) Preparation of polystyrene poly(β-hydroxy nonanoate) graft copolymers. Polym Bull 33:431–438
62. Cakmakli B, Hazer B, Borcakli M (2001) Poly(styrene peroxide) and poly(methyl methacrylate peroxide) for grafting on unsaturated bacterial polyesters. Macromol Biosci 1:348–354
63. Zhang J, Kasuya K, Takemura A, Isogai A, Iwata T (2013) Properties and enzymatic degradation of poly(acrylic acid) grafted polyhydroxyalkanoate films by plasma-initiated polymerization. Polym Degrad Stab 98:1458–1464
64. Hu SG, Jou CH, Yang MC (2003) Antibacterial and biodegradable properties of polyhydroxyalkanoates grafted with chitosan and chitooligosaccharides via ozone treatment. J Appl Polym Sci 88:2797–2803

65. Yalpani M, Marchessault RH, Morin FG, Monasterios CJ (1991) Synthesis of poly (3-hydroxyalkanoate) (PHA) conjugates: PHA-carbohydrate and PHA-synthetic polymer conjugates. Macromolecules 24:6046–6049
66. Domenek S, Langlois V, Renard E (2007) Bacterial polyesters grafted with poly(ethylene glycol): behaviour in aqueous media. Polym Degrad Stab 92:1384–1392
67. Kim HW, Chung CW, Hwang SJ, Rhee YH (2005) Drug release from and hydrolytic degradation of a poly(ethylene glycol) grafted poly(3-hydroxyoctanoate). Int J Biol Macromol 36:84–89
68. Renard E, Tanguy PY, Samain E, Guerin P (2003) Synthesis of novel graft polyhydroxyalkanoates. Macromol Symp 197:11–18
69. Chung MG, Kim HW, Kim BR, Kim YB, Rhee YH (2012) Biocompatibility and antimicrobial activity of poly(3-hydroxyoctanoate) grafted with vinylimidazole. Int J Biol Macromol 50:310–316
70. Kim HW, Chung MG, Kim YB, Rhee YH (2008) Graft copolymerization of glycerol 1,3-diglycerolate diacrylate onto poly(3-hydroxyoctanoate) to improve physical properties and biocompatibility. Int J Biol Macromol 43:307–313
71. Le Fer G, Babinot J, Versace D-L, Langlois V, Renard E (2012) An efficient thiol-ene chemistry for the preparation of amphiphilic PHA-based graft copolymers. Macromol Rapid Commun 33:2041–2045
72. Ishida K, Hortensius R, Luo X, Mather PT (2012) Soft bacterial polyester-based shape memory nanocomposites featuring reconfigurable nanostructure. J Polym Sci B Polym Phys 50:387–393
73. Park J, Park JG, Choi WM, Ha CS, Cho WJ (1999) Synthesis and photo- and biodegradabilities of poly[(hydroxybutyrate-*co*-hydroxyvalerate)-*g*-phenyl vinyl ketone]. J Appl Polym Sci 74:1432–1439
74. Lee HS, Lee TY (1997) Graft polymerization of acrylamide onto poly(hydroxybutyrate-*co*-hydroxyvalerate) films. Polymer 38:4505–4511
75. Shen XW, Yang Y, Jian J, Wu Q, Chen GQ (2009) Production and characterization of homopolymer poly(3-hydroxyvalerate) (PHV) accumulated by wild type and recombinant *Aeromonas hydrophila* strain 4AK4. Bioresour Technol 100:4296–4299
76. Park SJ, Lee TW, Lim SC, Kim TW, Lee H, Kim MK, Lee SH, Song BK, Lee SY (2012) Biosynthesis of polyhydroxyalkanoates containing 2-hydroxybutyrate from unrelated carbon source by metabolically engineered *Escherichia coli*. Appl Microbiol Biotechnol 93:273–283
77. Wang Q, Yang P, Liu CS, Xue YC, Xian M, Zhao G (2013) Biosynthesis of poly (3-hydroxypropionate) from glycerol by recombinant *Escherichia coli*. Bioresour Technol 131:548–551
78. Andreessen B, Steinbüchel A (2010) Biosynthesis and biodegradation of 3-hydroxypropionate-containing polyesters. Appl Environ Microbiol 76:4919–4925
79. Zhang L, Shi ZY, Wu Q, Chen GQ (2009) Microbial production of 4-hydroxybutyrate, poly-4-hydroxybutyrate, and poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) by recombinant microorganisms. Appl Microbiol Biotechnol 84:909–916
80. Hein S, Sohling B, Gottschalk G, Steinbuchel A (1997) Biosynthesis of poly(4-hydroxybutyric acid) by recombinant strains of *Escherichia coli*. FEMS Microbiol Lett 153:411–418
81. Sohling B, Gottschalk G (1996) Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri*. J Bacteriol 178:871–880
82. Alber BE, Fuchs G (2002) Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. J Biol Chem 277:12137–12143
83. Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, dos Santos VAPM, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Lee PC, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen J, Timmis KN, Dusterhoft A,

Tummler B, Fraser CM (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4:799–808
84. Celinska E (2010) Debottlenecking the 1,3-propanediol pathway by metabolic engineering. Biotechnol Adv 28:519–530
85. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7:445–452
86. Alber B, Olinger M, Rieder A, Kockelkorn D, Jobst B, Hugler M, Fuchs G (2006) Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. J Bacteriol 188:8551–8559
87. Wang Q, Liu CS, Xian M, Zhang YG, Zhao G (2012) Biosynthetic pathway for poly (3-hydroxypropionate) in recombinant *Escherichia coli*. J Microbiol 50:693–697
88. Albertyn J, Vantonder A, Prior BA (1992) Purification and characterization of glycerol-3-phosphate dehydrogenase of *Saccharomyces cerevisiae*. FEBS Lett 308:130–132
89. Norbeck J, Pahlman AK, Akhtar N, Blomberg A, Adler L (1996) Purification and characterization of two isoenzymes of DL-glycerol-3-phosphatase from *Saccharomyces cerevisiae* – identification of the corresponding *GPP1* and *GPP2* genes and evidence for osmotic regulation of *Gpp2p* expression by the osmosensing mitogen-activated protein kinase signal transduction pathway. J Biol Chem 271:13875–13881
90. Forage RG, Foster MA (1982) Glycerol fermentation in *Klebsiella pneumoniae* – functions of the coenzyme-B12-dependent glycerol and diol dehydratases. J Bacteriol 149:413–419
91. Liang QF, Zhang HJ, Li SN, Qi QS (2011) Construction of stress-induced metabolic pathway from glucose to 1,3-propanediol in *Escherichia coli*. Appl Microbiol Biotechnol 89:57–62
92. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC (1999) The propanediol utilization (*pdu*) operon of *Salmonella enterica serovar typhimurium* LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B-12-dependent 1,2-propanediol degradation. J Bacteriol 181:5967–5975
93. Leal NA, Havemann GD, Bobik TA (2003) PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica serovar Typhimurium* LT2. Arch Microbiol 180:353–361
94. Hiramitsu M, Doi Y (1993) Microbial synthesis and characterization of poly (3-hydroxybutyrate-*co*-3-hydroxypropionate). Polymer 34:4782–4786
95. Salles IM, Forchhammer N, Croux C, Girbal L, Soucaille P (2007) Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. Metab Eng 9:152–159
96. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183
97. Lv L, Ren YL, Chen JC, Wu Q, Chen GQ (2015) Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: controllable P(3HB-*co*-4HB) biosynthesis. Metab Eng 29:160–168
98. Jiang XR, Wang H, Shen R, Chen GQ (2015) Engineering the bacterial shapes for enhanced inclusion bodies accumulation. Metab Eng 29:227–237
99. van den Ent F, Amos LA, Lowe J (2001) Prokaryotic origin of the actin cytoskeleton. Nature 413:39–44

Adv Biochem Eng Biotechnol (2018) 162: 175–216
DOI: 10.1007/10_2016_22

Published online: 3 December 2017

Engineering and Evolution of *Saccharomyces cerevisiae* to Produce Biofuels and Chemicals

Timothy L. Turner, Heejin Kim, In Iok Kong, Jing-Jing Liu, Guo-Chang Zhang, and Yong-Su Jin

Abstract To mitigate global climate change caused partly by the use of fossil fuels, the production of fuels and chemicals from renewable biomass has been attempted. The conversion of various sugars from renewable biomass into biofuels by engineered baker's yeast (*Saccharomyces cerevisiae*) is one major direction which has grown dramatically in recent years. As well as shifting away from fossil fuels, the production of commodity chemicals by engineered *S. cerevisiae* has also increased significantly. The traditional approaches of biochemical and metabolic engineering to develop economic bioconversion processes in laboratory and industrial settings have been accelerated by rapid advancements in the areas of yeast genomics, synthetic biology, and systems biology. Together, these innovations have resulted in rapid and efficient manipulation of *S. cerevisiae* to expand fermentable substrates and diversify value-added products. Here, we discuss recent and major advances in rational (relying on prior experimentally-derived knowledge) and combinatorial (relying on high-throughput screening and genomics) approaches to engineer *S. cerevisiae* for producing ethanol, butanol, 2,3-butanediol, fatty acid ethyl esters, isoprenoids, organic acids, rare sugars, antioxidants, and sugar alcohols from glucose, xylose, cellobiose, galactose, acetate, alginate, mannitol, arabinose, and lactose.

T.L. Turner and G.-C. Zhang
Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: tlturne2@illinois.edu; zgc1984@illinois.edu

H. Kim, I.I. Kong, J.-J. Liu, and Y.-S. Jin (✉)
Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: hkim248@illinois.edu; ikong2@illinois.edu; jil263@illinois.edu; ysjin@illinois.edu

Keywords Biofuels, Metabolic engineering, Renewable chemicals, *Saccharomyces cerevisiae*, Yeast

Contents

1 Introduction

As human society has grown and developed, our demand for fuels and commodity chemicals has accelerated. This demand has manifested as many different outputs for both fuels and chemicals. For fuels, we have two major categories: transportation fuels and non-transportation fuels. Here we mainly discuss transportation fuels, which are currently primarily derived from non-renewable fossil fuels. These hydrocarbons, such as coal, petroleum, or natural gas, are processed into gasoline, ethanol, jet fuel, or other specialized products [1]. Approximately 80% of energy use by humans is derived from fossil fuels, with up to 58% consumed for transportation [2, 3]. Because the rate of natural production of fossil fuels has for decades been increasingly outpaced by humanity's usage, renewable alternatives for transportation fuels are considered a societal necessity [1].

As with fuels, many non-fuel chemicals are produced using non-renewable fossil fuel feedstocks. This petrochemical-based system is non-renewable and, as with fuels, an alternative method of production is needed to allow for continued advancement of human society. In particular, the petrochemical industry produces chemicals used in nearly every industry on Earth. Many bulk chemicals, such as ethylene and propylene, are produced in the 1–100 million annual tons range [4]. The specific uses of these chemicals can vary greatly: in some cases, such as artemisinic acid, only one major use is currently considered (as a precursor to an antimalarial drug) [5], whereas other chemicals, such as lactic acid, have numerous uses, including as a plastic precursor or as a food preservative [6]. Collectively, reliable and sustainable industrial production of chemicals is a necessity for ongoing human progress.

The finite supply of fossil fuels [7, 8], the risks associated with harvesting hard-to-obtain fossil fuels [9–11], and the concerns about manmade climate change related to fossil fuel use [12–15] have collectively pushed researchers and governments toward producing fuels and chemicals from renewable biomass by engineered microbes [16, 17]. Although many microbes have been studied for the production of renewable fuels and chemicals, yeasts, *Saccharomyces cerevisiae* in particular, have served as major platform microbes for many of these studies.

S. cerevisiae, also known as brewer's yeast, is a well-studied microorganism, even beyond its traditional use for the production of beer and other fermented foods and beverages [18]. Extensive tools exist for the manipulation and engineering of yeasts [19–22]. These tools have allowed for harnessing the native ability of *S. cerevisiae* to grow in minimal medium, their generally recognized as safe (GRAS) designation, and their tolerance to low pH and acidic conditions [23, 24]. With these tools and inherent physiological advantages, scientific advances for the production of fuels and chemicals from biomass by *S. cerevisiae* have improved dramatically in recent years. In this review we discuss these recent developments as they relate to feedstock utilization as well as production of fuels and chemicals with additional insight on the future economic outlook of these processes.

2 Yeast Fermentation Technologies

With modern metabolic engineering techniques improvements following their advent in the 1970s and the more recent development of synthetic biology procedures, yeast engineering technologies have grown dramatically [16]. Many yeast engineering approaches follow a scheme known as the "Design, Build, Test, and Learn" cycle [25, 26]. This scheme first requires a target outcome or goal. For example, a target goal could be to produce ethanol from the pentose sugar xylose by engineered *S. cerevisiae*, which natively are unable to ferment xylose.

Once the desired outcome is determined, a parental yeast strain, often a wild-type strain, is selected as the target organism to be engineered. The steps for engineering the parental strain are as follow: (1) *Designing* the specific yeast engineering steps, including plasmids and transformation protocols, (2) *Building* the engineered strain by introduction of target genetic perturbations, (3) *Testing* the newly-developed strain, often involving fermentation and sampling, and (4) *Learning* from the new strain (Fig. 1). The new knowledge obtained from this process can then be factored into the design of the next strain and the cycle can repeat until the target outcome is reached. This systematic approach has led to significant advances in the development of engineered *S. cerevisiae* capable of fermenting novel substrates for the production of fuels and chemicals. Although not all studies explicitly state this four-step process, the general concept is applicable in many cases.

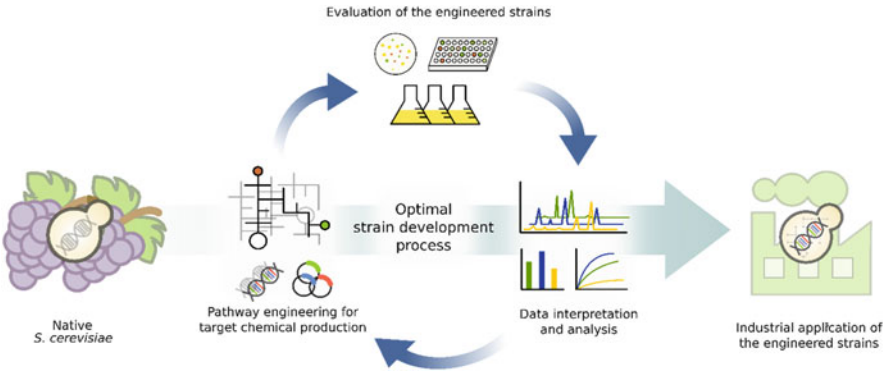

Fig. 1 Schematic demonstrating the step-by-step process for the Design, Build, Test, and Learn metabolic engineering/synthetic biology cycle used to develop engineered *Saccharomyces cerevisiae* for industrial-scale production of renewable fuels and chemicals

2.1 Major Objectives and Feedstocks for Yeast Fermentations

The first-generation biofuels from cornstarch or sugarcane juice have been industrialized for decades; however the food-vs-fuel conflict has limited its further expansion [27]. Ethanol is considered the major and most highly produced of the first-generation biofuels. Despite the food-vs-fuel concerns, 23.8 billion gallons of ethanol are produced annually, primarily from cornstarch or sugarcane juice [28]. Transitioning from first-generation biofuel feedstocks (cornstarch and sugarcane juice) to second-generation feedstocks (lignocellulosic biomass) is a key objective of modern yeast fermentation research.

The second-generation biofuels from non-food lignocellulosic biomass, which is a renewable carbon source, has offered an excellent opportunity to address the food-vs-fuel issue [29, 30]. Lignocellulosic hydrolysates obtained from corn stover [31], bagasse [32, 33], sorghum biomass [34], and marine plants [35, 36] after pretreatment and hydrolysis contain substantial amounts of hexoses (six-carbon, C6 sugar) and pentoses (five-carbon, C5 sugar) which can be used as renewable carbon sources for the production of bioethanol and other value-added products (Fig. 2). Lignocellulosic hydrolysates are commonly composed of ~70% cellodextrins and glucose and 30% xylose [37], although this can vary by biomass source and processing protocol. Marine hydrolysate sugar compositions can vary wildly: as a percent of total solids, red algae hydrolysates can be composed of ~18% glucose, ~30% total of galactose/xylose/arabinose, and ~8% mannose; green algae hydrolysates can consist of ~8% glucose, 6% total of galactose/xylose/arabinose, and 5% mannose; finally, brown algae can be composed of 6–7% glucose and between 2% (*Sargassum fulvellum*) and 30% (*Laminaria japonica*) mannitol [38]. However, natively, the yeast *S. cerevisiae* cannot use pentoses, such as xylose, and cannot efficiently ferment all hexoses. Therefore, another major objective of yeast

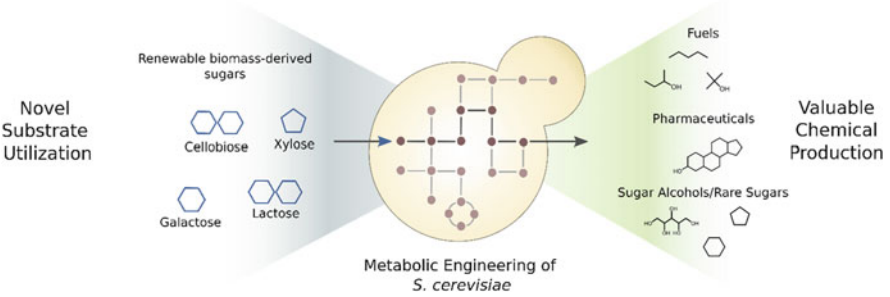

Fig. 2 A selection of major sugar substrates (inputs) which are processed by engineered yeast to generate target products (outputs)

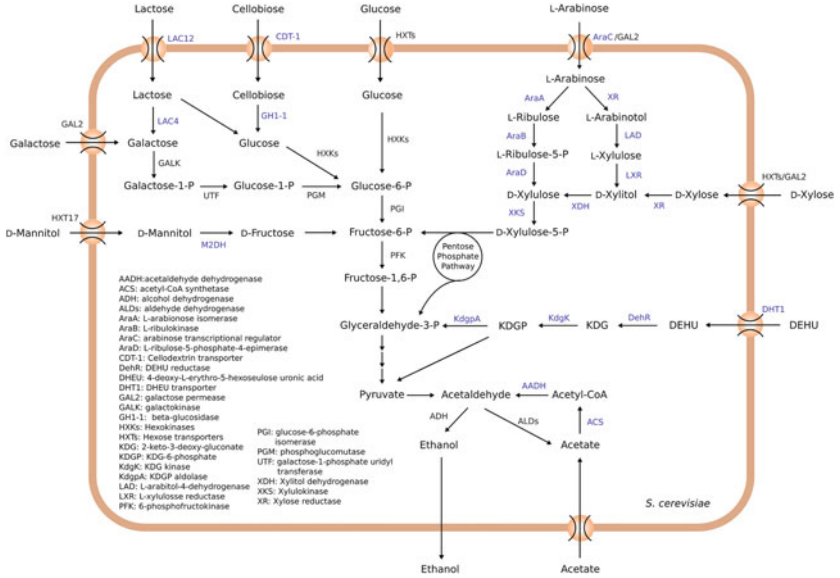

Fig. 3 A diagram of substrates which are fermentable by *Saccharomyces cerevisiae* via native or heterologous (*blue text*) pathways

fermentation research is to improve the selection of sugars capable of being fermented by *S. cerevisiae* for the purpose of industrial fermentation (Fig. 3). In Sect. 2.2 we discuss the currently available substrates for native and engineered *S. cerevisiae* strains.

2.2 *Native and Non-Native Substrate Utilization by* Saccharomyces cerevisiae

Glucose

Glucose is the most preferred carbon source for *S. cerevisiae* [39] and can be fermented more rapidly than any other sugar. To date, no other carbon source has been found to be consumed more rapidly or efficiently than glucose in any wild-type or engineered *S. cerevisiae*. The major industrial source of glucose is from cornstarch, although wheat is also sometimes used. The first generation of biofuels is based on the hydrolysis of cornstarch and very high gravity (VHG) fermentations have been conducted to decrease the process costs [32, 40, 41]. Several studies have focused on enhancing the fitness of *S. cerevisiae* in the presence of high concentrations of glucose. For example, Guadalupe-Medina et al. created a *GPD1*- and *GPD2*-negative *S. cerevisiae* that anaerobically produced ethanol at a high yield from glucose [42]. However, this strain became sensitive to high concentrations of ethanol, but the problem was alleviated by employing a laboratory evolution strategy with serial subculturing of the *GPD1*/*GPD2*-deleted strain on ethanol [42]. Because glucose fermentations by *S. cerevisiae* are very rapid and efficient, further improvements for glucose fermentations by engineered *S. cerevisiae* are likely to focus on improving strain tolerance to harsh fermentation media conditions, especially those found in cellulosic hydrolysates.

Xylose

Harvested terrestrial biomass is processed into a sludge-like product known as a hydrolysate. In terrestrial biomass, hydrolysates contain both C6 and C5 sugars. However, the most widely-used fermenting microorganism, *S. cerevisiae*, cannot metabolize pentose sugars such as xylose and arabinose which are abundant in cellulosic hydrolysates. Therefore, numerous studies have attempted to construct metabolically engineered *S. cerevisiae* capable of fermenting pentose as rapidly as glucose [43–45]. Xylose metabolism can be introduced into *S. cerevisiae* using a bacterial or fungal metabolic route for xylose assimilation [46, 47]. The bacterial pathway uses only one enzyme, xylose isomerase (XI), for converting xylose into xylulose [44, 45]. Xylulose is later phosphorylated by xylulose kinase (XK) into xylulose-5-phosphate (X5P) and then enters the non-oxidative pentose phosphate pathway (PPP) for further metabolism toward pyruvate. Using the XI pathway, an ethanol yield from xylose as high as 0.45 g/g has been achieved [44]. Another study by Lee et al. engineered an *S. cerevisiae* to harbor a bacterial xylose pathway to express a mutant xylose isomerase (*xylA3**) from *Piromyces sp.* with aldose reductase (*GRE3*) and *PHO13* deletions coupled with overexpression of the *S. cerevisiae* native xylulokinase (*XKS1*) and *S. stipitis* transaldolase (*TAL1*) [44]. Zhou et al. also overexpressed the *Piromyces sp.* xylose isomerase gene

(*XYLA*), *S. stipitis* xylulose kinase (*XYL3*), and genes of the non-oxidative pentose phosphate pathway [45].

The fungal xylose assimilation pathway consists of two oxidoreductases, NADPH-linked xylose reductase (XR) and NAD-linked xylitol dehydrogenase (XDH) [43]. Several researchers developed platforms for consuming these specific substrates, such as introducing xylose-metabolizing enzymes into *S. cerevisiae* to produce a rapid and efficient xylose-fermenting strain [47–50]. For example, Kim et al. introduced the fungal pathway by strong and balanced expression of genes from *Scheffersomyces stipitis* consisting of xylose reductase (XR, encoded by *XYL1*), xylitol dehydrogenase (XDH, encoded by *XYL2*), and xylulose kinase (XK, encoded by *XYL3*) with the addition of the genetic disruption of alkaline phosphatase (*PHO13*) and acetaldehyde dehydrogenase (*ALD6*) [43]. This series of genetic manipulations using the fungal XR/XDH/XK pathway resulted in an ethanol yield of 0.35 g/g from xylose [43].

Collectively, these studies have developed numerous xylose-fermenting *S. cerevisiae* capable of rapid and efficient xylose fermentation. Despite these advances, even the fastest xylose fermentations by engineered yeasts are still slower than the fastest glucose fermentations, and so further studies to improve xylose fermentation rates and yields by *S. cerevisiae* are ongoing.

Arabinose

Similar to xylose metabolism, different L-arabinose metabolizing pathways have been identified in bacteria [51] and fungi [52, 53]. The bacterial pathway for L-arabinose utilization converts L-arabinose into X5P via L-ribulose-5-phosphate (L5P) using three enzymes (an isomerase, a kinase, and an epimerase). When L-arabinose isomerase (*araA*), L-ribulokinase (*araB*), and L-ribulose-5-phosphate 4-epimerase (*araD*) from *Lactobacillus plantarum* were expressed in *S. cerevisiae*, L-arabinose fermentation was observed [51]. The fungal L-arabinose utilization pathway converts L-arabinose into L-arabinitol by aldose reductase (*GRE3* from *S. cerevisiae* or *XYL1* from *Scheffersomyces stipitis*), L-xylulose by L-arabinitol 4-dehydrogenase (*LAD* from *Trichoderma reesei*), xylitol by L-xylulose reductase (*LXR* from *T. reesei*), D-xylulose by (*XDH* from *S. stipitis*), and lastly X5P by xylulokinase (*XYL3*) [52, 53]. As X5P is a gateway metabolite in the PPP, it can be converted to pyruvate and ethanol. Recently, researchers have used codon optimized bacterial pathways for L-arabinose fermentation in *S. cerevisiae* because of the inefficient L-arabinose utilization and high byproduct (L-arabinitol) yield of fungal pathways caused by severe redox imbalance [54].

Other than introducing xylose and arabinose pathways into *S. cerevisiae*, known hexose transporters, as potential xylose and arabinose transporters, have been investigated. Several hexose transporters were proven to be responsible for the uptake of pentose sugars. For instance, Hxt7p, Hxt5p, and Gal2p improve xylose uptake [55] and Gal2p also facilitates the transport of L-arabinose [56]. However, these hexose transporters exhibited very low affinity to pentoses and preferred

D-glucose. Therefore, for the improvement of xylose and L-arabinose fermentations, efforts were made to find high-affinity xylose or L-arabinose specific transporters. Heterologous transporters were discovered with higher affinity for xylose over glucose, such as Gxs1p from *Candida intermedia* [57], Xut3p from *S. stipitis* [58], and Mgt05196p from *Meyerozyma guilliermondii* [59]. Nonetheless, it is still challenging to have both the specificity and efficiency of xylose transport, and further evolutionary adaptation and protein engineering are required [59, 60]. Heterologous overexpression of *STP2* from *Arabidopsis thaliana* and *ARAT* from *S. stipitis* in *S. cerevisiae* also led to improved anaerobic L-arabinose fermentation, especially at low L-arabinose concentrations, although these two transporters still are inhibited in the presence of glucose [61]. Recently, Wang et al. have engineered an *S. cerevisiae* strain capable of producing an ethanol yield of 0.43 g/g from arabinose, one of the highest reported yields to date [62].

Cellobiose

Another major sugar of interest is cellobiose, a β(1,4)-linked dimer of D-glucose, which is readily released from larger cellodextrins from cellulose by cellulases after acidic treatment of terrestrial biomass [63]. However, *S. cerevisiae* cannot naturally metabolize cellobiose because of the lack of a cellobiose transporter and intracellular β-glucosidase. A high-affinity cellodextrin transporter (*cdt-1* or *cdt-2*) and intracellular β-glucosidase (*gh1-1*) were identified from the cellulolytic fungus *Neurospora crassa* [64]. The cellobiose transporters and the intracellular β-glucosidase promote efficient cellobiose fermentation and ethanol production when expressed in *S. cerevisiae* [64]. The intracellular β-glucosidase can be replaced by cellobiose phosphorylase, which produces glucose and glucose-1-phosphate from cellobiose. Efficient cellobiose fermentation by engineered yeast expressing a cellobiose transporter and a bacterial cellobiose phosphorylase has also been demonstrated [65]. Because cellobiose does not induce glucose inhibition on other carbon sources, simultaneous cofermentation of cellobiose and xylose [66, 67] as well as cellobiose and galactose [68] has been achieved. Simultaneous cofermentation is necessary for efficient and rapid industrial-scale fermentation of hydrolysates.

Alginate and Mannitol

Another type of sustainable non-lignocellulosic biomass is marine biomass, such as macroalgae or seaweed. The most abundant sugars in brown macroalgae are alginate, mannitol, and glucan (presented as laminarin or cellulose). However, industrial microbes are unable to metabolize the alginate, which represents 30–60% of total sugars in brown macroalgae. Alginate is a linear block copolymer of two uronates, β-D-mannuronate (M) and α-L-guluronate (G), arranged in varying sequences [69]. Some microbes can metabolize alginate natively by

depolymerization of alginate into oligomers by alginate lyases (Alys). These oligomers are further degraded into unsaturated monomers by oligoalginate lyase (Oal) and the monomers are rearranged spontaneously into 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH). DEH is then converted into 2-keto-3-deoxyl-glucontate (KDG) by DEH reductase (DehR), and KDG is a common metabolite that can enter into the Entner–Doudoroff (ED) pathway and yield pyruvate and glyceraldehydrate-3-phosphate via KDG kinase (KdgK) and KDG-6-aldolase (Eda).

However, these natural microbes, such as *Sphingomonas sp*., lack the robustness necessary for industrial fermentation conditions and have limited availability of genetic and metabolic engineering tools. Therefore, researchers have introduced and expressed the genes responsible for the alginate degradation, transport, and metabolism into the well-characterized microorganism *Escherichia coli*, which is naturally capable of utilizing mannitol and D-glucose. A 36-kb pair DNA fragment from *Vibrio splendidus* encoding enzymes necessary for alginate degradation, transport, and metabolism was discovered. After introducing the alginate metabolism, the heterologous homoethanol pathway consisting of *Zymomonas mobilis* pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase B (*adhB*) were also introduced for efficient ethanol production [70], which later demonstrated the feasibility of utilizing macroalgae as a microbial host for ethanol production.

Although engineered *E. coli* provided the proof of concept for metabolizing alginate, mannitol, and D-glucose, *S. cerevisiae* is a more amenable host for industrial-scale ethanol production. Therefore, Enquist-Newman et al. attempted to re-engineer the alginate and mannitol catabolic pathways into *S. cerevisiae* [71]. They discovered an alginate monomer (DEHU) transporter from the alginolytic eukaryote *Asteromyces cruciatus*. Through the genome integration and overexpression of this transporter and with the necessary bacterial alginate degradation genes and essential genes for mannitol consumption, including an NAD^+-dependent mannitol-2-dehydrogenase (M2DH) and a mannitol transporter, the engineered *S. cerevisiae* was able to metabolize DEHU and mannitol [71]. As a result, the engineered *S. cerevisiae* strain produced ethanol from mannitol and DEHU at 83% of the maximum theoretical yield.

Galactose

S. cerevisiae are naturally capable of fermenting galactose, a C6 monosaccharide, into ethanol through the Leloir pathway. In the Leloir pathway, galactose is converted to UDP-glucose and then glucose-1-phosphate. Phosphoglucomutase converts glucose 1-phosphate to glucose 6-phosphate. Whereas the rest of the metabolic pathway is identical, the ethanol yield and productivity from galactose by *S. cerevisiae* is significantly lower than from glucose [72]. Through overexpression of a truncated *TUP1* gene, which codes for a general transcription repressor, Lee et al. were able to improve galactose consumption rate and ethanol productivity by 250% compared to a control *S. cerevisiae* [72]. By combining

enhanced galactose metabolism with a heterologous cellobiose pathway, an engineered *S. cerevisiae* could be employed for fermenting red seaweed hydrolysates. The major components of red seaweed (*Gelidium amansii*), cellulose and galactan, can be hydrolyzed to produce a mixture of cellobiose and galactose [36, 68]. Cellobiose and galactose can be cofermented by engineered yeast [68] because the two sugars are transported with high affinity by independent transporters (CDT-1 and Gal2). Recently, one research group has focused on using high concentrations of galactose as an adaptation pressure on yeast to improve galactose consumption rates and ethanol productivity [73, 74]. Further improvements for producing ethanol from galactose and red seaweed are necessary for industrial-scale ethanol production, especially as demand for second-generation biofuels continues to grow.

Acetate

Acetate is one of the major inhibitors present in lignocellulosic hydrolysates which can hamper *S. cerevisiae* fermentation capabilities. In addition, acetate is also produced as a major component from the pyrolysis of lignin [75, 76]. Recently, an interesting solution was developed to convert acetate from a fermentation component or inhibitor into a valuable product. By coupling the consumption of acetate and xylose, the redox imbalance of xylose fermentation by the heterologous xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulokinase (XK) pathway can be alleviated and the inhibitor (acetate) can be detoxified [77]. As a major result, the entire bioethanol fermentation process was improved compared to the control, increasing the ethanol yield by 6% (to 0.414 g/g) and reducing byproduct formation by 11% [77]. This process was further advanced by generating an engineered *S. cerevisiae* which expresses a cellobiose-utilizing pathway in addition to the aforementioned acetate and xylose pathways, allowing for efficient fermentation of multiple lignocellulosic sugars (xylose and cellobiose) and fermentation inhibitors (acetate) [78]. Finally, a peak ethanol yield of 0.463 g ethanol/g xylose was achieved by an XR/XDH-expressing *S. cerevisiae* through upregulation of acetylating acetaldehyde dehydrogenase (AADH) and acetyl-CoA synthetase (ACS) [79]. Compared to the control strain, the engineered strain was able to produce 18.4% more ethanol, 41.3% less glycerol, and consume 4.1 g/L of acetate from a cellulosic hydrolysate [79]. Collectively, these acetate-utilization studies are a significant breakthrough for the in situ detoxification of acetate by *S. cerevisiae* for ethanol production. Additional improvements could convert this fermentation into an industrial-scale ready process.

Lactose

Lactose is a disaccharide consisting of the monomers glucose and galactose. The primary source of lactose is from milk or fermented dairy products. Annually,

millions of tons of lactose are produced by the dairy industry. As a result of the acid whey fermentation process, a significant amount of lactose is trapped in the harsh and acidic acid whey slurry. Many studies have been conducted to find efficient uses for this trapped lactose.

Several studies have attempted to create a lactose-consuming *S. cerevisiae* by introducing *LAC4* and *LAC12* from *Kluyveromyces marxianus* and *Kluyveromyces lactis* into *S. cerevisiae* [80–84]. These studies resulted in the development of engineered *S. cerevisiae* capable of fermenting lactose. By expressing the *LAC4* and *LAC12* genes into the *MIG1* and *NTH1* gene-encoding regions in *S. cerevisiae*, respectively, Zou et al. engineered a strain capable of producing 63.3 g/L of ethanol from approximately 150 g/L lactose in 120 h from concentrated cheese whey [84]. By disrupting the function of the *MIG1* and *NTH1* genes, the engineered strain had highly reduced glucose repression. Although *Kluyveromyces spp.* are yeasts which can natively ferment lactose, their genetics are not as well-understood as that of *S. cerevisiae*, suggesting that improvements of *S. cerevisiae* for lactose fermentation may be ideal.

3 Biofuel Production by Engineered or Evolved Yeast

S. cerevisiae offers many advantages for producing sustainable and economically viable biofuels from renewable feedstocks. It has been widely used as an important eukaryotic model for fundamental molecular biology research with numerous synthetic biology tools developed as compared to most other microorganisms, perhaps second only to *E. coli*. Recently developed yeast engineering tools include the use of zinc-finger nucleases [85], yeast oligo-mediated genome engineering [21], and most notably the clustered regularly-interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas9) system) [22]. However, *S. cerevisiae* is considered to be more robust than *E. coli*, with *S. cerevisiae* possessing a higher tolerance to low pH/high acid conditions, resulting in preference for the eukaryote for fermentation of biomass hydrolysate. *S. cerevisiae* has been used extensively as a platform cell factory for first-generation, industrial-scale bioethanol production [86]. Because of its unique robustness toward harsh fermentation conditions and the substantial availability of yeast engineering tools, introducing new metabolic engineering pathways into *S. cerevisiae* has been used for producing alternative products beyond bioethanol (Fig. 4 and Table 1).

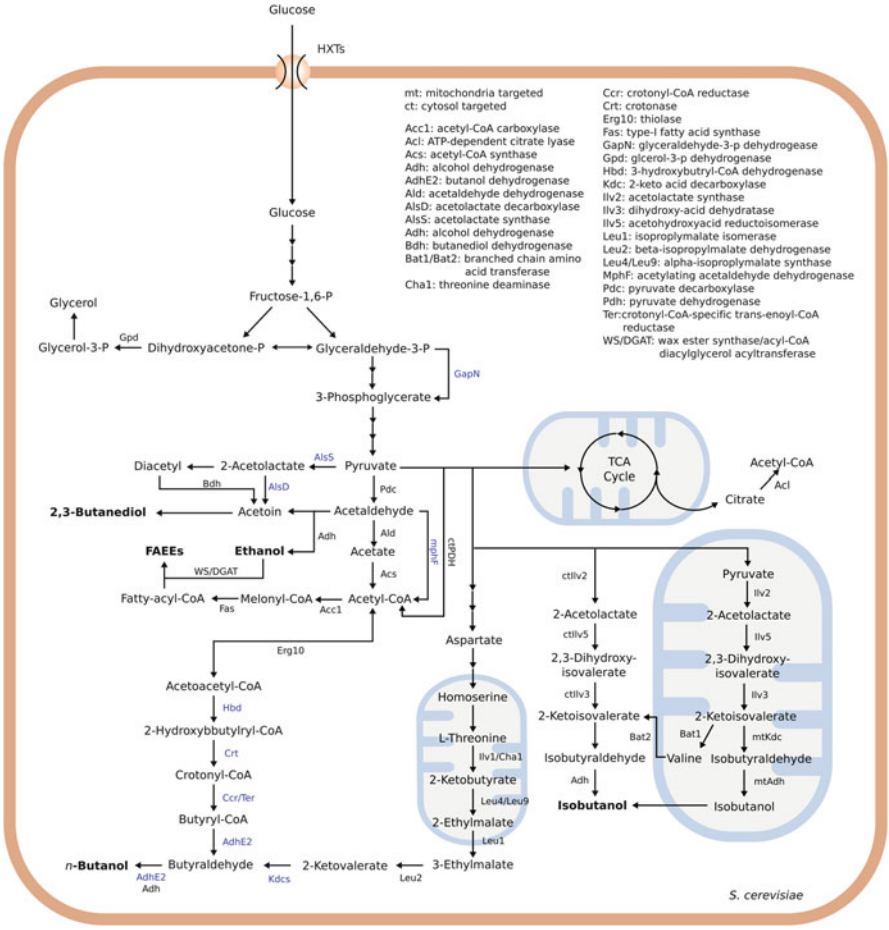

Fig. 4 A diagram of fuels that can be produced by *Saccharomyces cerevisiae* via native or heterologous (*blue text*) pathways

3.1 Biofuel Production by S. cerevisiae

Ethanol

First-generation biofuel production focused almost entirely on producing bioethanol from corn or sugarcane juice. Although many research directions were investigated to improve bioethanol production, one major direction focused on glycerol, a common byproduct of ethanol fermentations. During anaerobic yeast fermentations, the biosynthesis of proteins, nucleic acids, and lipids from biomass production generate excess cytosolic reduced redox cofactors such as NADH. Formation of glycerol serves as an essential electron sink for oxidizing NADH into NAD^+ in the cytosol. Tremendous research efforts focused on minimizing the formation of the unwanted glycerol byproduct generated during the bioethanol

Table 1 Biobased fuels from *Saccharomyces cerevisiae*

Product	Substrate	Result	Genetic modification(s)	Reference
Ethanol	Glucose	97.8% tm	*FPS1ΔGPD2Δ* and 80-bp 3′ truncation of *GPD1* native promoter	[39]
Ethanol	Xylose	0.35 g/g xylose	*S. stipitis XYL1*, *XYL2*, and *XYL3* balanced expression and *PHO13ΔALD6Δ*	[43]
Ethanol	Arabinose	0.43 g/g arabinose	*L. plantarum araA*, *araB*, and *araD* expression and overexpression of *TAL1*, *TKL1*, *RPE1*, *RKI1*, and *GAL2* with adaptive evolution	[62]
Ethanol	Cellobiose	86.3% tm	*N. crassa cdt-1* and *gh1-1* integration	[64]
Ethanol	Xylose and cellobiose	0.39 g/g xylose and cellobiose	*S. stipitis XYL1*, *XYL2*, and *XYL3* and *N. crassa cdt-1* and *gh1-1* balanced expression	[66]
Ethanol	Cellobiose and galactose	0.36 g/g galactose and cellobiose	*N. crassa cdt-1* and *gh1-1* integration	[68]
Ethanol	Mannitol and DEHU	83% tm	*A. cruciatus* YEL070W/YNR073C, *HXT13*, *HXT17*, and *YNR071C* expression	[71]
Ethanol	Galactose	0.46 g/g galactose	Laboratory evolution on galactose	[73]
Ethanol	Acetate and xylose	6% improved yield and 11% reduced byproduct formation	*E. coli adhE* integration with *S. stipitis XYL1*, *XYL2*, and *XYL3* balanced expression and *PHO13ΔALD6Δ*	[77]
Ethanol	Acetate, xylose, and cellobiose	~9% improved yield	*E. coli adhE* and *N. crassa cdt-1* and *gh1-1* integration with *S. stipitis XYL1*, *XYL2*, and *XYL3* balanced expression and *PHO13ΔALD6Δ*	[78]
Ethanol	Glucose	10% improved yield	*GLN1* and *GLT1* overexpression and *GDH1Δ*	[90]
Ethanol	Glucose	97.4% tm	*B. cereus gapN*, *E. coli frdA*, and *mhpF* expression	[94]
Ethanol	Glucose	10% improved yield	*FPS1Δ* reducing glycerol production	[95]
Ethanol	Glucose	Tolerance up to 90 g/L EtOH in wheat liquefact SSF	Native *GPD1* and *GPD2* promoters replaced with lower-strength *TEF1* promoter mutants in *GPD1Δ* or *GPD2Δ* strains	[96]
1-Butanol	Galactose	Tenfold increase	*C. beijerinckii adhe2*, *hbd*, *crt*, with *S. cerevisiae ERG10*, and *S. collinus ccr* expression	[98]
1-Butanol	Glucose	16.3 mg/L titer	*T. denticola Ter* and *S. enterica ACS2* expression, *ADH2* and *ALD6* overexpression, and *MLS1ΔCIT2Δ*	[99]

(continued)

Table 1 (continued)

Product	Substrate	Result	Genetic modification(s)	Reference
1-Butanol	Glucose	120 mg/L titer	*E. coli* PDH genes and acetyl-CoA synthetase gene expression with *ADH1ΔADH4ΔGPD1ΔGPD2Δ*	[100]
1-Butanol	Glucose	242.8 mg/L titer	Leucine biosynthesis pathway overexpression and *ILV2ΔADH1Δ*	[101]
Isobutanol	Glucose	4.12 mg/g	*ILV2*, *ILV3*, *ILV5*, and *BAT2* overexpression	[102]
Isobutanol	Glucose	6.40 mg/g	Located isobutanol pathway into the mitochondria	[103]
Isobutanol	Glucose	15 mg/g	δ-Integration used to assemble isobutanol pathway genes into the yeast chromosome	[104]
Isobutanol	Glucose	1.62 g/L titer	PDH complex activity reduction *LPD1Δ* and transhydrogenase-like shunt expression	[106]
FAEE	Glucose	6.3 mg/L titer	*M. hydrocarbonoclasticus* wax ester synthase expression	[109]
FAEE	Glucose	6.3-fold increase	*M. hydrocarbonoclasticus* wax ester synthase, *S. cerevisiae FAA1*, and *B. ammoniagenes bafas* and *ppt* expression	[112]

tm theoretical maximum

production process because carbon directed toward glycerol reduced carbon availability for ethanol synthesis. Two structural genes encoding cytosolic NADH-dependent glycerol-3-phosphate dehydrogenases, *GPD1* and *GPD2*, play important roles in redox balance and osmoregulation. These genes are also both induced under high osmotic conditions and during anaerobic fermentation. Glycerol formation can be reduced by deleting one or both genes [87]. However, yeast cells with this double deletion of *GPD1* and *GPD2* become unable to grow anaerobically because of the lack of alternative pathways to oxidize NADH. The single deletion of *GPD2* showed improved ethanol yields by decreasing glycerol production, but the deletion also hindered cell growth and ethanol productivity [88]. Reduced glycerol production also increased the osmosensitivity and diminished the general robustness of the engineered yeast [89].

Other studies have focused on the metabolic engineering of the cellular redox metabolism. An ammonium assimilation pathway that consumes NADH and ATP was utilized. Deletion of the NADPH-dependent glutamate dehydrogenase gene *GDH1* and respectively co-overexpression of the glutamate synthase gene *GLT1* and the glutamine synthetase gene *GLN1* showed a significant reduction in glycerol by 38% and improved ethanol yield by 10% [90]. Another alternative way of redox engineering would be the reduction of the surplus cytosolic NADH with lower ATP

production by replacing the natural glyceraldehyde-3-phosphate dehydrogenase with a non-phosphorylating, $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase (*gapN*) from *Streptococcus sp*. mutants or *Bacillus cereus* [91, 92]. One interesting demonstration of cofactor metabolism is that overexpression of *E. coli mhpF* was able to restore the anaerobic growth of a *GPD1*- and *GPD2*-deleted mutant under the presence of acetate by re-oxidizing the NADH through the reduction of acetic acid to ethanol [93]. By combining these genetic modifications, overexpression of the NAD^+-dependent fumarate reductase *frdA* or NAD^+-dependent acetaldehyde dehydrogenase *mhpF* from *E. coli* with *gapN* can improve the ethanol yield to above 97% of the maximum theoretical yield compared to wild-type yeast [94]. Furthermore, *gapN* expression with the combination of *TPS1* and *TPS2* overexpression showed reduced glycerol production and improved ethanol yield [92]. By blocking the export of glycerol through deletion of *FPS1* encoding a glycerol facilitator, yet another method to reduce glycerol production and improve ethanol yield was uncovered [95].

Promoter engineering has been used as an alternative approach to modulate the expression of *GPD1* and *GPD2*. For example, *S. cerevisiae* mutants with the lower-strength *TEF1* promoter replacing the native *GPD1* and/or *GPD2* promoters produced less glycerol and more ethanol without reducing the robustness of the host strain toward osmotic stress [96]. With the *FPS1*- and *GPD2*-deleted yeast strain background (KAM15 strain), the mutants with 3′ truncation of the *GPD1* promoter by 20-, 60-, or 80-bp displayed varied expression strength of *GPD1* and had an unaffected osmotic response. The glycerol production by the engineered yeast was also reduced by 16% and 31% in mutants with 60- and 80-bp truncated promoters, respectively, in high-gravity (VHG) fermentations. The ethanol yield reached 0.499 g/g in the mutant with an 80-bp truncated promoter [40].

1-Butanol

Higher-chain alcohols provide higher energy density and are considered as potential next-generation gasoline substitutes. One of the primary target alcohols is 1-butanol. Although 1-butanol was traditionally produced from *Clostridium* species through the acetone-butanol-ethanol (ABE) fermentation process, or by engineered *E. coli* with a titer up to 30 g/L [97], there are several advantages of using *S. cerevisiae* for 1-butanol production. In addition to the general robustness of *S. cerevisiae* toward fermentation inhibitors and low pH, *S. cerevisiae* also does not have phage contamination issues and has better resistance to high 1-butanol concentrations. However, only a low concentration of 1-butanol was produced from the native 1-butanol metabolic pathway in *S. cerevisiae*, which prompted several labs to look for heterologous pathways to improve 1-butanol production.

Steen et al. introduced and expressed in *S. cerevisiae* several isozymes from different organisms to create a biosynthetic 1-butanol pathway with a peak 1-butanol titer of 2.5 mg/L [98]. This pathway consisted of converting acetyl-

CoA into acetoacetyl-CoA, which was reduced to 3-hydroxybutyryl-CoA, and later crotonyl-CoA. Butyryl-CoA is the reduced form of crotonyl-CoA which is later further reduced into butyraldehyde and finally reduced into 1-butanol. This pathway consisted of overexpression of a thiolase (*ERG10*) from *S. cerevisiae*, an NADH-dependent 3-hydroxybutyryl-CoA dehydrogenase (*hbd*) and crotonase (*crt*) from *Clostridium beijerinckii*, an NADH-dependent crotonyl-CoA reductase (*ccr*) from *Streptomyces collinus* and butanol dehydrogenase (*adhe2*) from *C. beijerinckii* [98].

Krivoruchko et al. initially increased 1-butanol titers up to 6.6 mg/L by engineering yeast with higher flux toward cytosolic acetyl-CoA, which is the precursor for 1-butanol biosynthesis in addition to the overexpression of the heterologous enzymes for the 1-butanol biosynthetic pathway as follow [99]. First, NADH-dependent crotonyl-CoA-specific *trans*-enoyl-CoA reductase (*Ter*) from *Treponema denticola* replaced the *ccr* to avoid the reverse oxidation of butyryl-CoA to crotonyl-CoA. Second, to increase the cytosolic acetyl-CoA supply, a pyruvate dehydrogenase (*PDH*) bypass was created by overexpression of endogenous alcohol dehydrogenase (*ADH2*), NADP-dependent aldehyde dehydrogenase (*ALD6*), codon-optimized acetyl-CoA synthetase (*ACS2*) from *Salmonella enterica*, and acetyl-CoA acetyltransferase (*ERG10*). Lastly, deletion of malate synthase (*MLS1*) or citrate synthase (*CIT2*) reduced the drainage of acetyl-CoA through the glyoxylate pathway, and the 1-butanol titer increased to 16.3 mg/L [99].

Therefore, intracellular availability of cytosolic acetyl-CoA is considered an important factor for 1-butanol production in yeast. NADH availability could also be a strong driving force toward 1-butanol production. Therefore, NADH-dependent alcohol dehydrogenase (*ADH*) and glycerol-3-phosphate dehydrogenase (*GPD*) can be deleted to increase the NADH availability and reduce the unwanted byproducts such as ethanol and glycerol. Lian et al. produced up to 120 mg/L 1-butanol by inactivating *ADH* and *GPD*, introducing the butanol biosynthesis pathway genes, and, most importantly, introducing a PDH-bypass pathway, cytosolic localized PDH, and ATP-dependent citrate lyase (*ACL*) [100].

Despite the limited accumulation of 1-butanol from the native *S. cerevisiae* pathway, some researchers have focused on improving the native pathway by focusing on threonine catabolism. Si et al. utilized genes from leucine biosynthesis (*LEU1*, *LEU2*, *LEU4*, and *LEU9*), together with threonine deaminase genes (*ILV1*/*CHA1*), 2-keto acid decarboxylases (*KDC*s) from *Lactococcus lactis*, and alcohol dehydrogenases (*ADH*s) from *S. cerevisiae* [101]. The pathway consists of many steps, starting with L-threonine to 2-ketobutyrate to 2-ketovalerate, and so forth, eventually ending at 1-butanol. Deletion of *ADH* allowed the engineered *S. cerevisiae* to produce more than 120 mg/L of 1-butanol from glucose in a complex yeast-peptone medium. By amplifying the leucine biosynthesis pathway via overexpression of several key genes and eliminating the competing pathways, the highest reported 1-butanol titer of 242.8 mg/L in *S. cerevisiae* with *ADH1*- and *ILV2*-deletions was achieved [101].

Isobutanol

Isobutanol is another example of a target alcohol which has a higher energy density than ethanol. The isobutanol biosynthesis pathway is closely linked to the biosynthesis of branched-chain amino acids via the Ehrlich pathway. 2-Ketoisovalerate (KIV), an intermediate of valine biosynthesis, is decarboxylated to isobutyraldehyde by 2-ketoacid decarboxylase (*KDC*) and later reduced into isobutanol by alcohol dehydrogenase (*ADH*). However, the protein synthesis of KIV occurs in the yeast mitochondria whereas the other two enzymes, Kdc and Adh, are found in the yeast cytosol. For isobutanol synthesis in *S. cerevisiae*, pyruvate must transfer into mitochondria and then KIV must be transported into the cytosol.

The first report for isobutanol overproduction in yeast utilized simultaneous overexpression of endogenous genes (*ILV2*, *ILV3*, and *ILV5*) of the mitochondrial valine biosynthesis pathway. The resulting strain produced isobutanol with a yield up to 0.97 mg isobutanol/g glucose in minimal medium [102]. Additional overexpression of the cytosolic branched-chain amino acid aminotransferase (*BAT2*) increased the isobutanol yield up to 3.86 mg/g glucose [102]. Finally, a yield of 4.12 mg/g glucose was achieved by the engineered yeast in an aerobic condition with complex yeast-peptone medium [102]. Avalos et al. demonstrated that locating the complete isobutanol pathway into the mitochondria resulted in substantial increases in isobutanol as compared with the native pathway which is split between the cytosol and the mitochondria. *KDC*s and *ADH*s were overexpressed in the cytosol or imported into mitochondria by fusing them with an N-terminal targeting signal, and the isobutanol yield reached up to 6.40 mg/g glucose with a titer up to 0.635 g/L [103]. This study suggested that the availability of the KIV intermediate and the increased local enzyme concentration would be beneficial for isobutanol production. Another research group, Yuan and Ching, developed a similar approach with a δ-integration system to assemble the genes into the yeast chromosomes with the resulting isobutanol yield up to 15 mg/g glucose [104].

The opposite strategy is to relocate the pathway into the cytosol. By re-localization and codon-optimization of the mitochondrial valine synthesis enzymes together, along with the overexpression of decarboxylase (*ARO10*) and alcohol dehydrogenase (*ADH2*) genes, isobutanol production was improved to the highest titer of 0.63 g/L and a yield of approximately 15 mg/g glucose [105]. Isobutanol production was further improved in engineered *S. cerevisiae* by two strategies. First, the elimination of competing pathways by deletion of a pyruvate dehydrogenase complex component (*LPD1*) to avoid competing with acetyl-CoA biosynthesis in the mitochondria. Second, resolving cofactor imbalance by the implementation of the transhydrogenase-like shunt, which pyruvate cyclically converted into oxaloacetate, malate, and back to pyruvate causing simultaneous conversion of NADH to NADPH. The final isobutanol titer reached 1.62 g/L and a yield of 16 mg/g glucose [106]. However, even this heightened result is still

considerably below that of engineered *E. coli*, reported to generate isobutanol titers up to several grams per liter [107]. These results suggest that considerable improvements are necessary before yeast-based isobutanol production can be competitive on an industrial scale.

Fatty Acids

Fatty acids (FAs) and lipids are also valuable chemicals for numerous industrial applications. Lipids are condensed from a glycerol-3-phosphate backbone with the completed FA synthesized from acetyl-CoA. Fatty acid ethyl esters (FAEEs) can be used for diesel or jet fuel production. FAEEs can be formed by esterification of fatty acyl-CoAs and ethanol. Kalscheuer et al. first studied FAEE production in yeast [108] by heterologous expression of an unspecific bacterial acyltransferase, a wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase (WS/DGAT), from *Acinetobacter calcoaceticus ADP1*. Later, Shi et al. screened five different wax ester synthases in *S. cerevisiae* and found the wax ester synthase from *Marinobacter hydrocarbonoclasticus* performed best with the highest titer of FAEE at 6.3 mg/L [109]. Overexpression of acetyl-coA carboxylase (*ACC1*) led to an increase of FAEE titer to 8.2 mg/L [109]. de Jong et al. continued the study by increasing the acyl-CoA synthesis which later enhanced the production of FAEE by increasing the NADPH and acetyl-CoA pools in two ways [110]. First, overexpression of alcohol dehydrogenase (*ADH2*), acetaldehyde dehydrogenase (*ALD6*), and a heterologous acetyl-CoA synthase variant from *Salmonella enterica* (acs_{SE}^{L641P}) was conducted to re-channel the carbon flow for acetyl-CoA with the ethanol degradation pathway. Wax ester synthase from *M. hydrocarbonoclasticus* was also overexpressed. Second, a phosphoketolase pathway was established by overexpression of *xpkA* and *ack* from *Aspergillus nidulans* for the conversion of xylulose-5-phosphate to acetyl-phosphate and glyceraldehyde-3-phosphate and acetyl phosphate to acetate. The resulting engineered *S. cerevisiae* strain proved to have a 1.7-fold improvement for FAEE production compared to the control strain, with 5.1 mg/g dry cell weight [110].

In the same year, Valle-Rodrigez et al. eliminated the non-essential fatty acid utilization pathway such as steryl esters (SEs) and triacylglycerols (TAGs) by deletion of *DGA1*, *LRO1*, *ARE1*, and *ARE2* [111]. The researchers also deleted *POX1* to avoid degradation of FAs and overexpressed wax ester synthase (WS) from *M. hydrocarbonoclasticus* DSM 8798 which generated a final FAEE titer of up to 17.2 mg/L [111]. Recently, Eriksen et al. investigated the heterologous expression of Type-I fatty acid synthase (FAS) from *Brevibacterium ammoniagenes* coupled with WS/DGAT [112]. They found the strain harboring the orthologous FAS yielded a 6.3-fold increased FAEE titer compared to strains without FAS. The FAEE titer was 10.498 mg/g DCW with the overexpression of Type-I fatty acid synthase (*bafas* and *ppt1*) from *Brevibacterium ammoniagenes*, *FAA1* from *S. cerevisiae*, and wax ester synthase from *M. hydrocarbonoclasticus* [112]. However, additional studies and demonstrations must be conducted for

further improvement of the titers for FAEE, because the above-mentioned titers from engineered *S. cerevisiae* are still relatively low for industrial applications.

4 Chemical Production by Engineered or Evolved Yeast

There has been an intensive effort for the engineering of *S. cerevisiae* to produce non-fuel, value-added chemicals. Historically, *S. cerevisiae* has been used for ethanol production by the food or fuel industries, but scientific advances for the purpose of ethanol production by yeast can often easily be applied to non-fuel production. As mentioned in previous sections of this review, *S. cerevisiae* has GRAS status and their genetic system has been studied heavily. Thus, many genetic tools are available [21, 22, 85] which ease the engineering of this host organism to produce nonconventional target products. These products include food additives, pharmaceuticals, advanced biofuels, and valuable chemicals for industrial applications.

Natively, *S. cerevisiae* produces numerous minor and major intermediates and metabolites, especially those throughout the glycolytic pathway, the pentose phosphate pathway, and the tricarboxylic acid pathway [113]. However, to accumulate a significant concentration of these intermediates (or other, non-native compounds) for industrial purposes, considerable engineering or evolution of *S. cerevisiae* is often necessary. Methods, such as the Design, Build, Test, and Learn approach (Fig. 1) or tools such as CRISPR/Cas9 [22] have been largely applied for the purpose of producing ethanol by yeast fermentations, but can be and have been easily re-tooled for constructing yeast capable of producing many other chemicals. These chemicals cover many broad categories including isoprenoids, fatty acids, organic acids, rare sugars, sugar alcohols, and others. A recent tour de force of *S. cerevisiae* engineering came from Galanie et al., in which the group required 23 enzymes from bacteria, mammals, plants, and yeast to produce a tiny amount of opioids, albeit at roughly five orders of magnitude below what would be necessary for industrial scale-up [114]. However, this demonstrates a future for yeast biotechnology in which a single biosynthetic pathway can create downstream products that may otherwise take multiple chemical catalysis steps (Fig. 5 and Table 2).

4.1 Chemical Production by S. cerevisiae

2,3-Butanediol

2,3-Butanediol (2,3-BD) is an increasingly popular target chemical because of its wide applications for synthesizing diverse products such as pharmaceuticals, cosmetics, and industrial solvents. As 2,3-BD is mostly produced by pathogenic bacteria, it is difficult to apply the bacteria to industrial fermentations.

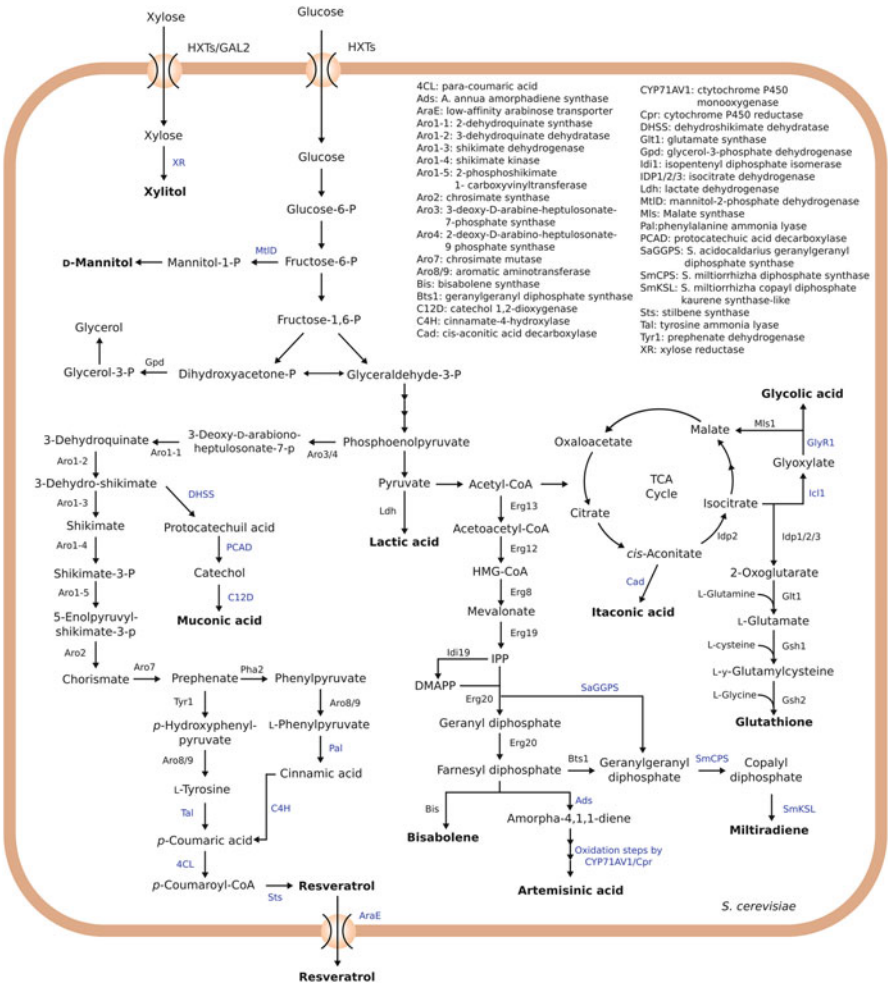

Fig. 5 A diagram of non-fuel chemicals that can be produced by *Saccharomyces cerevisiae* via native or heterologous (*blue text*) pathways

S. cerevisiae can produce 2,3-BD naturally, but at a very low concentration, because of ethanol serving as the major fermentative end product. Therefore, researchers have engineered *S. cerevisiae* to generate a higher titer of 2,3-BD by the elimination of ethanol production through the disruption of alcohol dehydrogenases (*ADH1*, *ADH3*, and *ADH5*). Ng et al. achieved a titer of 2.29 g 2,3-BD/L with a yield of 0.113 g/g glucose [115]. Kim et al. further eliminated the competing pathways by deleting all three pyruvate decarboxylase genes (*PDC1*, *PDC5*, and *PDC6*) and generated a Pdc-deficient mutant to improve the 2,3-BD titer [116]. However, Pdc-deficient mutants had defects such as slow growth, and they required acetate or ethanol supplementation as a carbon source. The Pdc-deficient mutants also suffered from redox imbalance because of glucose repression. The

Table 2 Biobased chemicals from *Saccharomyces cerevisiae*

Product	Substrate	Result	Genetic modification(s)	Reference
2,3-BDO	Glucose	2.29 g/L titer	*ADH1ΔADH3ΔADH5Δ*	[115]
2,3-BDO	Glucose	72.9 g/L titer	*B. subtilis AlsS* and *AlsD*, *L. lactis NoxE*, and *S. cerevisiae* overexpression with *ADH1ΔADH2ΔADH3ΔADH4ΔADH5Δ GPD1ΔGPD2Δ*	[118]
Hydrocodone	Glucose	~0.3 μg/L titer	Expression of 23 genes encoding for various enzymes, overexpression of two native genes, and inactivation of one native gene	[114]
Geraniol	Glucose	5 mg/L titer	*ERG20* mutation and *O. basilicum* monoterpene synthase expression	[122]
Cineole	Galactose	1 g/L titer	Overexpression of *HMG2*, *ERG20*, and *IDI1* with expression of two genes encoding for terpene synthases from *S. fruticosa* and *S. pomifera*	[125]
Bisabolene	Glucose and galactose	>900 mg/L titer	Overexpression of *ERG10*, *IDI*, *ERG20*, *tHMGR*, and Upc2-1 with *A. grandis* BIS expression	[126]
Bisabolene	Glucose or galactose	5.2 g/L titer	Deletion of YJL062W and YPL064W	[127]
Taxadiene	Glucose	8.7 mg/L titer	Expression of codon-optimized *T. chinensis* TDS, *S. acidocaldarious* GGPPS, mUpc2-1, and truncated HMG-CoA reductase isoenzyme 1	[128]
Miltiradiene	Glucose	488 mg/L titer	Expression of copalyl diphosphate synthase, overexpression of a truncated HMG-CoA reductase and a mUpc2-1, and overexpression of a fusion gene of *ERG20* and *BTS1* together with *S. acidocaldarius* GGPS	[130]
Artemisinic acid	Glucose and galactose	2.5 g/L titer	Multiple mevalonate pathway modification, galactose as an inducer, and Pmet3 promoter controlling *ERG9*	[5]
Amorpha-4,11-diene	Glucose	40 g/L titer	Overexpression of every mevalonate pathway enzyme through *ERG20* and an optimized fermentation process	[136]
Lactic acid	Glucose	81.5% tm	Bovine *LDH* and *PDC1ΔPDC5Δ*	[139]
Lactic acid	Glucose and xylose	69% tm	*R. oryzae ldhA* with *S. stipitis XYL1*, *XYL2*, and *XYL3* balanced expression and *PHO13ΔALD6Δ*	[140]
Lactic acid	Glucose, xylose, and cellobiose	66% tm	*R. oryzae ldhA* with *S. stipitis XYL1*, *XYL2*, *XYL3*, and *N. crassa cdt-1* and *gh1-1* balanced expression with *PHO13ΔALD6Δ*	[141]
Itaconic acid	Glucose	168 mg/L titer	*A. terreus* CAD with GPD promoter, *ADE3ΔBNA2ΔTES1Δ*	[148]

(continued)

Table 2 (continued)

Product	Substrate	Result	Genetic modification(s)	Reference
Succinic acid	Glucose	12.97 g/L titer	Cytosolic retargeting of *MDH3*, *FRDS1*, and *E. coli FumC* with *PYC2* overexpression and *GPD1ΔFUM1Δ*	[152]
Succinic acid	Glucose	43-fold increase	*SDH3ΔSER3ΔSER33Δ* and directed evolution	[154]
Glycolic acid	Xylose and ethanol	~1 g/L titer	*A. thaliana GLYR1* and *MLS1ΔIDP2Δ* with *ICL1* and XR/XDK/XK xylose utilization pathway expression	[156]
Xylitol	Xylose and cellobiose	~100% tm	*S. stipitis XYL1*, *N. crassa cdt-1* and *gh1-1* expression with *ALD6*, *IDP2*, and *ZWF1* overexpression	[166]
Xylitol	Glucose and xylose	~100% tm	Two *XYL1* genes, *ZWF1*, and *ACS1* expression with fed-batch optimization	[167]

tm theoretical maximum

researchers identified point mutation A81P in the transcription regulator Mth1 involved in glucose sensing, which is necessary for glucose tolerance. They also introduced a bacterial 2,3-BD pathway by converting pyruvate into α-acetolactate and then acetoin, respectively, by acetolactate synthase (*alsS*) and acetolactate decarboxylase (*alsD*), and then acetoin is reduced into 2,3-BD by butanediol dehydrogenase (*BDH1*) from *Bacillus subtilis*. Finally, the engineered *S. cerevisiae* produced a titer up to 96.2 g/L under a fed-batch fermentation with a yield of 0.28 g/g glucose [116].

Recently, Kim et al. attempted to minimize the glycerol byproduct formation by decreasing the intracellular NADH/NAD^+ from the expression of NADH oxidase (*noxE*) from *L. lactis*, and the resulting engineered yeast strain was able to produce 2,3-BD with a yield of 0.359 g/g glucose [117]. With a similar approach, Kim and Hahn tried to minimize glycerol production in engineered *S. cerevisiae* with the additional deletion of glycerol-3-phosphate dehydrogenase (*GPD1* and *GPD2*), creating a strain which could produce a 2,3-BD titer of up to 72.9 g/L in a fed-batch fermentation and with a yield of up to 0.41 g/g glucose [118].

Isoprenoids

Isoprenoids, also known as terpenes, are a diverse group of chemical compounds typically utilized as medicines, cosmetics, nutritional supplements, food additives, or even as a potential future biofuels [119]. *S. cerevisiae* harbor natural metabolic pathways to produce certain isoprenoids, although yields and productivities are very poor [120]. Despite the poor natural production, isoprenoids are of great interest because of their diverse structures and wide range of potential uses. Monoterpenes (C10) and sesquiterpenes (C15) are two of the main candidates for jet fuel and biodiesel alternatives because of their low freezing temperature and high

ignition stability properties. To produce isoprenoids, acetyl-CoA production is of a high importance because all isoprenoids share the mevalonate metabolic pathway starting from acetyl-CoA [121–123]. Either the bacterial 1-deoxyl-D-xylulose 5-phosphate (DXP) pathway or the eukaryote/archaea mevalonate (MVA) pathway is essential for the biosynthesis of isoprenoids. Both pathways end with the formation of five-carbon monomers dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP). DMAPP and IPP are then condensed and modified by prenyltransferases to form isoprenoid precursors such as geranyl pyrophosphate (GPP, C10) and farnesyl pyrophosphate (FPP, C15) [124].

Monoterpenes (C10) are derived from GPP by monoterpene synthases. Fischer et al. is the first group able to produce geraniol, a monoterpene and alcohol, with a titer of up to 5 mg/L in *S. cerevisiae* by a mutation of *ERG20* (farnesyl pyrophosphate synthase) and the overexpression of heterologous geraniol synthase (monoterpene synthases) from *Ocimum basilicum* [122]. To improve the monoterpene biosynthesis, Ignea et al. used the yeast sterol biosynthesis pathway genes *HMG2*, *ERG20*, and *IDI1* and co-expression of two terpene synthase enzymes (cineole synthase) from *Salvia fruticosa* and *Salvia pomifera.* The final titer of cineole was up to 1 g/L [125].

Sesquiterpenes (C15) are another isoprenoid-derived potential fuel source which has recently gained interest for several industrial applications. Bisabolene, a precursor of bisabolane, was produced at a titer of over 900 mg/L in engineered *S. cerevisiae* by Peralta-Yahya et al. [126]. The yeast was first engineered by overexpression of acetyl-CoA acetyltransferase (*ERG10*), isoprenyl diphosphate isomerase (*IDI1*), and farnesyl pyrophosphate synthase (*ERG20*), truncated HMG-CoA reductase (*tHMGR*), and the transcriptional regulator of the sterol pathway (Upc2-1). Then researchers examined six different bisabolene synthases isolated from *Arabidopsis thaliana*, *Picea abies*, *Pseudotsuga menziesii*, and *Abies grandis*. Finally they developed the highest titer with the codon-optimized bisabolene synthase (BIS) from *A. grandis* [126]. Recently, Özaydın et al. screened the *S. cerevisiae* deletion collection for carotenoid production and constructed a strain producing the highest titer of up to 5.2 g/L of bisabolene through double deletion of YJL064W and YPL062W [127].

Several diterpenes (C20) have also been produced by engineered yeast. In 2008, a titer of 8.7 mg/L of taxadiene was achieved from engineered *S. cerevisiae* [128]. This feat was achieved through two general metabolic modifications: (1) coexpression of a codon-optimized *Taxus chinensis* taxadiene synthase and a *Sulfolobus acidocaldarius* geranylgeranyl diphosphate synthase and (2) expression of a truncated 3-hydroxyl-3-methylglutaryl-CoA reductase isoenzyme and a mutant regulatory protein UPC2-1 allowing for steroid uptake in anerobic conditions. In 2012, miltiradiene, another diterpene, was overproduced through metabolic engineering of *S. cerevisiae*. Zhou et al. achieved a peak titer of miltiradiene of 365 mg/L [129] and Dai et al. obtained 488 mg/L through a fed-batch fermentation [130]. The 488 mg/L titer was achieved through multiple *S. cerevisiae* metabolic engineering and fermentation technology steps: (1) overexpression of a mutated global regulatory factor (*upc2.1*) and a truncated 3-hydroxyl-3-methylglutaryl-CoA

reductase (*tHMGR*), (2) copalyl diphosphate synthase was first expressed, (3) overexpression of a fusion gene of farnesyl diphosphate synthase (*ERG20*) and an endogenous geranylgeranyl diphosphate (*BTS1*) together with a geranylgeranyl diphosphate synthase from *Sulfolobus acidocaldarius* (*SaGGPS*), and (4) use of a fed-batch fermentation [130].

Artemisinin is a sesquiterpene lactone which has received notoriety as an antimalarial drug following its discovery by You-You Tu in the 1970s [131, 132]. Unfortunately, the natural isolation and industrial production process for artemisinin is not always reliable, and shortages of this vital drug have been reported [133]. Production of artemisinin through a reliable and sustainable microbial cell factory could be a viable alternative. Several labs have worked to construct such a process. An important precursor for artemisinin production, amorpha-4,11-diene, was produced by Lindahl et al. in 2006 [134]. This result was achieved by subcloning the amorpha-4,11-diene synthase from *Artemisia annua* into a galactose-inducible, high-copy number pYeDP60 plasmid and subsequent transformation of the plasmid into an *S. cerevisiae* strain. Although further optimizations are needed before industrial-scale applications, the final titer, 600 μg/L, served as an important step toward microbial production of artemisinin.

Within a year of the report of amorpha-4,11-diene the process of producing artemisinic acid from engineered yeast was published. Artemisinic acid serves as the immediate precursor of artemisinin and can undergo further chemical synthesis to produce artemisinin. In their report, Ro et al. achieved a peak titer of ~100 mg/L of artemisinic acid [135]. A multitude of engineering steps were necessary to achieve this production in an engineered *S. cerevisiae*, broadly including increasing farnesyl pyrophosphate (FPP) production and reducing its use for sterols, expressing the amorphadiene synthase gene from *A. annua* into the improved FPP-producing strain, and cloning a novel cytochrome P450 to provide a three-step oxidation pathway from amorphadiene to artemisinic acid.

More recently, significant boosts in the production of both amorpha-4,11-diene and artemisinic acid from engineered *S. cerevisiae* have been reported. Lenihan et al. produced a titer of 2.5 g/L of artemisinic acid from an engineered *S. cerevisiae* by using a defined medium containing galactose as a carbon source and inducer in a fed-batch process which utilized a precise agitation and feed pump rate [5]. A Pmet3 promoter was used to control *ERG9*, which improved precursor availability for artemisinic acid synthesis by limiting sterol synthesis. Later, Westfall et al. hypothesized that high titers of artemisinic acid may be unachievable without improvement to the production of necessary precursors [136], such as the previously discussed amorpha-4,11-diene. Through overexpression of every mevalonate pathway enzyme through *ERG20* and fermentation optimization resulted in a considerably titer of 40 g/L amorpha-4,11-diene [136].

Organic Acids

Organic acids are widely used for many applications including usage as food additives. However, organic acids also serve as building blocks of many larger polymers by undergoing several steps of chemical catalysis. For example, lactic acid is produced by engineered *S. cerevisiae* by introducing lactate dehydrogenase (*ldh*). Through catalysis, polylactic acid (also known as polylactide; PLA) can be produced [137]. PLA is a renewable and biodegradable polyester used for many purposes including as a filament for 3D printing, for producing medical screws/implants, and for producing plastic dinnerware. Numerous studies have been conducted for producing lactic acid from engineered *S. cerevisiae* from a variety of feedstocks including glucose [138, 139], xylose [140], and cellobiose [141]. Currently, no study using engineered yeast has been able to achieve the theoretical maximum of lactic acid production from glucose, xylose, cellobiose, or a mixture of these carbon sources, so work is ongoing to improve these fermentation processes. Of the studies which have generated lactic acid-producing *S. cerevisiae*, a variety of *ldh* sources have been used, including bovine materials [142, 143], *Rhizopus oryzae* [140, 141, 144], *Bifidobacterium longum* [142], and *Lactobacillus plantarum* [145, 146]. Moving forward, expression of *ldh* from yet-unstudied sources into *S. cerevisiae* may prove useful for producing specific ratios of L- or D-lactic acid, which can be beneficial for specific industrial applications.

Because itaconic acid has many industrial uses, including serving as a copolymer for producing plastics and rubbers [147], this compound is another interesting organic acid which has recently been produced at a laboratory-scale in engineered *S. cerevisiae*. To achieve a peak titer of 168 mg/L of itaconic acid from *S. cerevisiae*, several metabolic engineering steps were implemented [148]. First, the *cis*-aconitic acid decarboxylase encoding gene (CAD) from *Aspergillus terreus* was expressed in an *S. cerevisiae* strain under the control of a strong "Enhanced" GPD promoter. Second, several gene targets including *ADE3*, *BNA2*, and *TES1* were identified by a genome-wide stoichiometric model, deleted, and assessed for itaconic acid production improvements. Finally, the triple deletion strain expressing the *A. terreus* CAD was grown in optimized fermentation conditions including a high cell density to provide the peak titer of 168 mg/L itaconic acid. However, scale-up to a cost-effective and efficient industrial-scale process require further optimization, as a titer of more than 80 g/L of itaconic acid is considered necessary [148].

As with itaconic acid, muconic acid is another platform chemical which can act as a precursor for the production of many useful products, including various renewable plastics [149]. The first reported instance of muconic acid production by engineered *S. cerevisiae* was in 2012, resulting in a peak titer of approximately 1.56 mg/L muconic acid [150]. However, by 2013, several metabolic engineering improvements allowed for production of 141 mg/L muconic acid from an engineered *S. cerevisiae* [151]. Several metabolic engineering steps were needed to produce this result. First, *Candida albicans* catechol 1,2-dioxygenase,

Enterobacter cloacae protocatechuic acid decarboxylase, and *Podospora anserine* dehydroshikimate dehydratase were expressed in an *S. cerevisiae* strain. Then *ARO3* was deleted and a feedback-resistant mutant *ARO4* was expressed to reduce shikimate pathway feedback inhibition. Next, *ZWF1* was deleted and *TKL1* was overexpressed to increase precursor flux into the target pathway. Finally, several heterologous enzyme levels were balanced, resulting in the final titer of 141 mg/L muconic acid [151].

Succinic acid is a value-added organic acid which can be overproduced by engineered yeast [152–154]. Similar to lactic acid, succinic acid can be used as a precursor to several polyesters [155]. Furthermore, succinic acid is designated as GRAS by the U.S. Food and Drug Administration, which has allowed its use in the food industry as an acidity regulator. As an intermediate of the citric acid cycle (or tricarboxylic acid cycle), yeast natively produces succinic acid if provided with an aerobic environment, but overproduction of succinic acid requires multiple genetic perturbations. For example, Otero et al. constructed an engineered *S. cerevisiae* with deletions of *SDH3*, *SER3*, and *SER33* to reduce primary succinate-consuming reactions and to interrupt glycolysis-derived serine [154]. The resulting engineered yeast displayed a 30-fold improvement in succinic acid titer and a 43-fold improvement in succinic acid yield as compared to the control strain.

Beyond succinic acid, glycolic acid, a C2 hydroxy acid, has gained attention in recent years. The global glycolic acid production in 2011 was approximately 40,000,000 kg with this expected to more than double by 2018 [156]. Glycolic acid is often used as a building block of a polyglycolate. The polyglycolate polymer is used as a packaging material because of its high gas permeability and mechanical strength. However, most glycolic acid is produced in a chemical process which relies on non-renewable fossil resources [156]. As an alternative, a biological route for the production of glycolic acid exists which involves converting glyoxylate through glyoxylate reductase into glycolic acid. To overproduce glycolic acid successfully, efficient glyoxylate reductase activity in an engineered *S. cerevisiae* is required. A further improvement, up to approximately 1 g/L glycolic acid, can be achieved by deletions of the malate synthase (*MLS1*) and the cytosolic form of isocitrate dehydronase (*IDP2*) genes [156]. As the current generation of organic acids produced by *S. cerevisiae* continues to improve and develop, it is likely that new, rare, or hard-to-obtain organic acids can be produced in laboratories by engineered *S. cerevisiae* strains.

Rare Sugars, Sugar Alcohols, and Antioxidants

Sugars such as L-ribose, D-allose, D-tagatose, and D-psicose are classified as rare sugars. As the name implies, these sugars are rarely found in nature, but they have beneficial health properties. L-Ribose, for example, is considered a very important intermediate to produce chemicals for pharmaceutical and food products [157, 158]. Although D-ribose is very common in nature, L-ribose is not found in

nature based on current knowledge. The driving demand for L-ribose production is its potential as a building block for L-nucleoside-based pharmaceutical compounds. L-Nucleoside-based compounds or analogs play an important role in treating viral infections and cancers [159]. Currently, research regarding rare sugar production by engineered yeast is very limited.

Sugar alcohols such as erythritol, xylitol, or sorbitol have a high demand in the food industry because of their sweetening properties without causing dental caries [160]. Although generally difficult, one positive aspect of sugar alcohol production is that, in general, sugar alcohols are not fermentable by *S. cerevisiae*, which limits reuptake by engineered yeast designed to overproduce target sugar alcohols. The interest in producing sugar alcohols dates back more than 50 years, with at least one study investigating D-arabitol production in S*accharomyces spp*. [161]. More recently, a minute titer of 44 μg/mL mannitol was produced by expression of multiple copies of the *E. coli* mannitol-1-phosphate dehydrogenase gene (*mtlD*) into *S. cerevisiae* [162]. This titer was later improved upon by Costenoble et al. by producing a titer of nearly 400 mg/L of mannitol in an engineered *S. cerevisiae* in anaerobic conditions [163]. Primarily, this was achieved by expression of the *E. coli mtlD* into an *S. cerevisiae* strain and deletion of *GPD1* and *GPD2* followed by an oxygen-sparged fermentation which was switched to nitrogen-sparging during the exponential growth phase.

As one primary example of a well-known sugar alcohol, xylitol shares similar sweetening power with sucrose, but it does not contribute to dental caries and has a cooling effect when eaten. A chemical hydrogenation process to produce xylitol has existed for decades [164] but, more recently, several groups have produced high xylitol titers and yields from biological, engineered yeast systems [165–167]. Oh et al. were able to produce xylitol rapidly and efficiently using an engineered *S. cerevisiae* expressing xylose reductase (*XYL1*), a cellodextrin transporter (*cdt-1*), and an intracellular β-glucosidase (*gh1-1*) via simultaneous utilization of xylose and cellobiose [166]. As a result, the engineered *S. cerevisiae* was able to produce xylitol at the maximum theoretical yield by co-utilization of xylose and cellobiose.

Because antioxidants have been considered potentially beneficial as supplements to the human diet, there has been increased interest in efficiently producing these compounds from a consistently obtainable source rather than depending on extraction from seasonally-available produce. Resveratrol is one of these compounds of interest, as it is a common component of grape skins and wines made from these skins [168]. Many studies discussing the engineering of *S. cerevisiae* and other microbes for the microbe-based production of resveratrol have been published in recent years [169–172]. In one example, an engineered *S. cerevisiae* expressing a codon-optimized bacterial tyrosine ammonia lyase and an *E. coli* high-capacity, low-affinity arabinose transporter (*araE*) were able to produce a peak of 3.44 mg/L at 48 h in a laboratory-scale grape juice fermentation [172]. This result is an important step from an industrial standpoint, as it represented a method to increase the resveratrol concentration in white wine, which in most cases has a significantly lower resveratrol concentration than red wine.

As with resveratrol production, glutathione is another antioxidant which has been extensively studied for production by engineered *S. cerevisiae* [173–178]. - Microbe-based production of glutathione is currently the primary industrial process for glutathione synthesis, although it can also be produced by chemical synthesis [179]. Because the microbe-based process is the major method of industrial-scale production, many varied processes to improve the titer, yield, and productivity have been explored. Recently, a titer of 320 mg/L of glutathione was achieved by a laboratory-evolved *S. cerevisiae* strain in an acrolein-containing medium [178]. Acrolein is an aldehyde which is toxic to yeast cells [180], although glutathione has been shown to act as a defense against acrolein toxicity, suggesting that cells which have increased resistance to acrolein may be overproducing glutathione [181]. Based on this knowledge, several *S. cerevisiae* strains were evolved over 250 generations on increasing concentrations of acrolein. Finally, *S. cerevisiae* strain A4-19 was isolated, which displayed glucose consumption rates, growth rates, and ethanol production rates similar to the parental A4 strain, yet had increased acrolein resistance and a glutathione titer of 320 mg/L, approximately twofold larger than the parental strain [178].

5 Current Scope and Future Outlook of Industrial Fuel and Chemical Production by Yeast

As discussed in Sects. 3 and 4, many advances have been made in recent years in yeast metabolic engineering and synthetic biology for the purpose of biofuel and renewable chemical production. Collectively, these new technologies have resulted in *S. cerevisiae* strains capable of fermenting a variety of substrates, such as xylose and cellobiose, with improved target product yields and productivities. Only a fraction of these laboratory developments have seen implementation at an industrial scale because of prohibitive costs, difficulty in scale-up, or low yields and productivities. For industrial-scale biofuel production, *S. cerevisiae* is the primary yeast species seeing usage, although lab-scale biofuel production by non-*S. cerevisiae* yeast, such as *Yarrowia lipolytica* and *Schizosaccharomyces pombe*, has seen growth in recent years [182, 183]. However, several non-*S. cerevisiae* microbes are used for industrial chemical production because of the wide range of target chemicals produced by the biobased chemical industry. Although *S. cerevisiae* is extremely hardy and can be easily engineered, there are instances where other microbes are preferred for a target product. Perhaps the most notable example is the use of engineered *E. coli* for the production of recombinant insulin [184], and over 150 recombinant therapeutics have been approved by the European Medicines Agency [185]. However, only approximately one-third of approved therapeutics utilize engineered *E. coli*, with *S. cerevisiae* and other yeasts also accounting for a significant portion of industrial therapeutics, fuels, and chemicals [185].

At the industrial scale, ethanol is the major biofuel target, especially by engineered *S. cerevisiae* [186, 187]. Ethanol is commonly used as a fuel additive for the creation of gasoline-ethanol blends. The use of ethanol blends in the United States has grown from less than 5 vol% to over 10 vol% in the past decade [188]. This growth is at least partially attributed to the United States Environmental Protection Agency's Renewable Fuel Standard, which requires up to 17.4 billion gallons of renewable fuel production by 2016, of which 0.21 billion gallons must be cellulosic biofuel [189]. The total production requirement for renewable fuels can increase to 36 billion gallons by 2020 [190].

To achieve the renewable fuel standards set by the United States and other governments, industrial fuel producers have used *S. cerevisiae* as their platform microbial strain of choice. As of 2014, approximately 23.8 billion gallons of ethanol are produced on an annual basis worldwide, almost entirely from fermentation by *S. cerevisiae* [28]. The United States and Brazil are responsible for the vast majority of global bioethanol production, annually producing 14.3 billion gallons and 6.2 billion gallons, respectively [28]. In the United States, corn serves as the primary feedstock, whereas in Brazil, sugarcane is the major feedstock for the purpose of bioethanol production [191, 192].

As the two major bioethanol-producing countries, both nations have considerable motivation for the success of their respective ethanol industries. In Brazil, ethanol serves as a transportation fuel at nearly a 1:1 ratio with gasoline [193]. In the United States, roughly 40% of corn produced is used for the purpose of producing ethanol [194]. Both nations provide protection to their bioethanol industries in the form of tax breaks, subsidies, or increased tariffs toward imported ethanol. Moving forward, it is expected that these economic benefits are likely to shift away from first-generation biofuels (using corn and sugarcane juice as the feedstock) toward second-generation biofuels (using corn stover, switchgrass, and miscanthus). As government and environmental protection groups provide further incentives for renewable biofuel production by engineered yeast, scientific advances developed for producing fuel can be modified and applied to the production of non-fuel chemicals by engineered yeast. However, despite legislation in the United States and elsewhere to encourage biofuel production, no equivalent guidelines exist to provide incentive specifically for the purpose of biobased, non-fuel chemical production. A global effort to limit average global Earth surface temperatures to increasing by no more than 2°C relative to temperatures in the late nineteenth century by reducing greenhouse gas emissions has provided a minor incentive for renewable chemical production [195]. The influence this legislation has on biobased chemicals is small because of less than 10% of total fossil fuels being employed for chemical catalysis, with the vast majority going toward the energy and transportation fuel industries [196, 197].

Since early 2014, global oil prices have fallen rapidly and dramatically [198]. Unsurprisingly, as fossil fuel costs decrease, the economic production of biofuels and renewable chemicals becomes increasingly less viable. Not only are second-generation (lignocellulosic) biofuels at economic risk, but even the currently more cost-effective first-generation biofuels become difficult to produce in a

cost-effective manner. Roboredo et al. suggest that "huge state subsidies" would be needed to maintain viable biofuel production amidst the crashing oil prices [199]. Although the short-term outlook on biofuel and renewable chemical production is uncertain, it is anticipated that the continuing volatility of oil prices is likely to encourage further research for efficient, economical, and renewable biofuel and renewable chemical production.

Although there are many companies which produce renewable fuels or fuel additives, there also exist many companies worldwide which employ microbial fermentation for the production of non-fuel, renewable chemicals. In many cases, the exact specifications of the species of microbe used or the precise metabolic pathway engineering protocol are not entirely disclosed. However, some of the more notable companies using a yeast-based fermentation platform include DSM, Verdezyne, BioAmber, Amyris, and NatureWorks, which produce, respectively, succinic acid [200], adipic acid [201], 1,4-butanediol [202], farnesene [203], and lactic acid [204].

6 Conclusion

Equipped with rapid advances in metabolic engineering, synthetic biology, and genomics, the production of fuels and non-fuel chemicals by engineered *S. cerevisiae* has developed tremendously. Several of these advances have transitioned to industrial-scale fermentation processes, allowing for the sustainable production of many valuable chemicals from renewable biomass. Despite these advances and growing numbers of industrial examples, many barriers still exist, which can hinder the further adoption of *S. cerevisiae* industrial fermentations.

Currently, global oil prices have reached the lowest levels in approximately a decade [199]. Low oil prices are a major detriment not only to the cost-effective production of renewable fuels and chemicals but also to consumer and government sentiment regarding the short-term importance of developing a renewable chemical industry infrastructure. Furthermore, reduced oil prices significantly lower the cost of petroleum-based chemicals, which places additional pressure on renewable, fermentation-based biochemical production. Despite these pressures, many industrial biobased processes, such as succinic acid production (from *E. coli*) [205] and bioethanol production (from *S. cerevisiae*) [186, 187], are still considered to be feasible or even preferential to petrochemical production.

Moving forward, newer and more complex industrial-scale fuels and chemicals can be produced by engineered *S. cerevisiae* as volatile oil prices and depletion of finite fossil fuels encourage investment in biobased alternatives. Nearly all industrial-scale *S. cerevisiae* fermentations start as laboratory-scale studies following the "Design, Build, Test, and Learn" cycle (Fig. 1), but simpler single-step metabolic pathways, such as producing lactic acid by a heterologous lactate dehydrogenase [140], can give way to complex, multi-step pathways, such as producing

opioids [114]. Collectively, the impact of engineered *S. cerevisiae* on the biobased fuel and chemical industries is likely to expand in the near future.

References

1. Hubbert MK (1956) Nuclear energy and the fossil fuel. Drilling Prod Pract 36
2. Nigam PS, Singh A (2011) Production of liquid biofuels from renewable resources. Prog Energy Combust Sci 37:52–68
3. Escobar JC, Lora ES, Venturini OJ, Yáñez EE, Castillo EF, Almazan O (2009) Biofuels: environment, technology and food security. Renew Sustain Energy Rev 13:1275–1287
4. Zeng A, Biebl H (2002) Bulk chemicals from biotechnology: the case of 1,3-propanediol production and the new trends. In: Tools and applications of biochemical engineering science. Springer, Berlin, pp 239–259
5. Lenihan JR, Tsuruta H, Diola D, Renninger NS, Regentin R (2008) Developing an industrial artemisinic acid fermentation process to support the cost-effective production of antimalarial artemisinin-based combination therapies. Biotechnol Prog 24:1026–1032
6. Martinez FAC, Balciunas EM, Salgado JM, González JMD, Converti A, de Souza Oliveira RP (2013) Lactic acid properties, applications and production: a review. Trends Food Sci Technol 30:70–83
7. Berg P, Boland A (2014) Analysis of ultimate fossil fuel reserves and associated CO_2 emissions in IPCC scenarios. Nat Resour Res 23:141–158
8. Shafiee S, Topal E (2009) When will fossil fuel reserves be diminished? Energy Policy 37:181–189
9. Keranen KM, Weingarten M, Abers GA, Bekins BA, Ge S (2014) Induced earthquakes. Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. Science 345:448–451
10. Throupe R, Simons R, Mao X (2013) A review of hydro "fracking" and its potential effects on real estate. J Real Estate Lit 21:205–232
11. Jackson RB, Vengosh A, Carey JW, Davies RJ, Darrah TH, O'Sullivan F, Pétron G (2014) The environmental costs and benefits of fracking. Annu Rev Env Resour 39:327–362
12. Somerville RC, Hassol SJ (2011) The science of climate change. Phys Today 64:48
13. Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. Science 328:1523–1528
14. Friedlingstein P, Andrew R, Rogelj J, Peters G, Canadell J, Knutti R, Luderer G, Raupach M, Schaeffer M, van Vuuren D (2014) Persistent growth of CO_2 emissions and implications for reaching climate targets. Nat Geosci 7:709–715
15. Montzka SA, Dlugokencky EJ, Butler JH (2011) Non-CO_2 greenhouse gases and climate change. Nature 476:43–50
16. Stephanopoulos G (2012) Synthetic biology and metabolic engineering. ACS Synth Biol 1:514–525
17. Jang Y, Kim B, Shin JH, Choi YJ, Choi S, Song CW, Lee J, Park HG, Lee SY (2012) Bio-based production of C2–C6 platform chemicals. Biotechnol Bioeng 109:2437–2459
18. Ferreira I, Pinho O, Vieira E, Tavarela J (2010) Brewer's *Saccharomyces* yeast biomass: characteristics and potential applications. Trends Food Sci Technol 21:77–84
19. Krivoruchko A, Siewers V, Nielsen J (2011) Opportunities for yeast metabolic engineering: lessons from synthetic biology. Biotechnol J 6:262–276
20. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012)

Saccharomyces genome database: the genomics resource of budding yeast. Nucleic Acids Res 40:D700–D705
21. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM (2013) Yeast oligo-mediated genome engineering (YOGE). ACS Synth Biol 2:741–749
22. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343
23. Abbott DA, Zelle RM, Pronk JT, Van Maris AJ (2009) Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. FEMS Yeast Res 9:1123–1136
24. Benjaphokee S, Hasegawa D, Yokota D, Asvarak T, Auesukaree C, Sugiyama M, Kaneko Y, Boonchird C, Harashima S (2012) Highly efficient bioethanol production by a *Saccharomyces cerevisiae* strain with multiple stress tolerance to high temperature, acid and ethanol. New Biotechnol 29:379–386
25. Crook NC, Schmitz AC, Alper HS (2013) Optimization of a yeast RNA interference system for controlling gene expression and enabling rapid metabolic engineering. ACS Synth Biol 3:307–313
26. Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol 12:355–367
27. Panagiotopoulos I, Bakker R, de Vrije T, Claassen P, Koukios E (2013) Integration of first and second generation biofuels: fermentative hydrogen production from wheat grain and straw. Bioresour Technol 128:345–350
28. Renewable Fuels Association (2014) World Fuel Ethanol Production. In: http://ethanolrfa.org/pages/World-Fuel-Ethanol-Production. Accessed 21 Sept 2015
29. Scott F, Conejeros R, Aroca G (2013) Attainable region analysis for continuous production of second generation bioethanol. Biotechnol Biofuels 6:171-6834-6-171
30. Wang P, Dudareva N, Morgan JA, Chapple C (2015) Genetic manipulation of lignocellulosic biomass for bioenergy. Curr Opin Chem Biol 29:32–39
31. Parreiras LS, Breuer RJ, Narasimhan RA, Higbee AJ, La Reau A, Tremaine M, Qin L, Willis LB, Bice BD, Bonfert BL (2014) Engineering and two-stage evolution of a lignocellulosic hydrolysate-tolerant *Saccharomyces cerevisiae* strain for anaerobic fermentation of xylose from AFEX pretreated corn stover. PLoS One 9:e107499
32. Pereira FB, Guimarães PM, Teixeira JA, Domingues L (2010) Optimization of low-cost medium for very high gravity ethanol fermentations by *Saccharomyces cerevisiae* using statistical experimental designs. Bioresour Technol 101:7856–7863
33. Macrelli S, Galbe M, Wallberg O (2014) Effects of production and market factors on ethanol profitability for an integrated first and second generation ethanol plant using the whole sugarcane as feedstock. Biotechnol Biofuels 7:26–41
34. Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, Anderson R, Olson SN, Rooney W (2014) Energy sorghum—a genetic model for the design of C4 grass bioenergy crops. J Exp Bot 65:3479–3489
35. Behera S, Singh R, Arora R, Sharma NK, Shukla M, Kumar S (2015) Scope of algae as third generation biofuels. Front Bioeng Biotechnol 2:90
36. Wei N, Quarterman J, Jin Y (2013) Marine macroalgae: an untapped resource for producing fuels and chemicals. Trends Biotechnol 31:70–77
37. Carroll A, Somerville C (2009) Cellulosic biofuels. Annu Rev Plant Biol 60:165–182
38. Kim N, Li H, Jung K, Chang HN, Lee PC (2011) Ethanol production from marine algal hydrolysates using *Escherichia coli* KO11. Bioresour Technol 102:7466–7469
39. Busti S, Coccetti P, Alberghina L, Vanoni M (2010) Glucose signaling-mediated coordination of cell growth and cell cycle in *Saccharomyces cerevisiae*. Sensors 10:6195–6240
40. Ding WT, Zhang GC, Liu JJ (2013) 3′ Truncation of the GPD1 promoter in *Saccharomyces cerevisiae* for improved ethanol yield and productivity. Appl Environ Microbiol 79:3273–3281

41. Puligundla P, Smogrovicova D, Obulam VSR, Ko S (2011) Very high gravity (VHG) ethanolic brewing and fermentation: a research update. J Ind Microbiol Biotechnol 38:1133–1144
42. Guadalupe-Medina V, Metz B, Oud B, Der Graaf CM, Mans R, Pronk JT, Maris AJ (2014) Evolutionary engineering of a glycerol-3-phosphate dehydrogenase-negative, acetate-reducing *Saccharomyces cerevisiae* strain enables anaerobic growth at high glucose concentrations. Microb Biotechnol 7:44–53
43. Kim SR, Skerker JM, Kang W, Lesmana A, Wei N, Arkin AP, Jin Y (2013) Rational and evolutionary engineering approaches uncover a small set of genetic changes efficient for rapid xylose fermentation in *Saccharomyces cerevisiae*. PLoS One 8, e57048
44. Lee SM, Jellison T, Alper HS (2014) Systematic and evolutionary engineering of a xylose isomerase-based pathway in *Saccharomyces cerevisiae* for efficient conversion yields. Biotechnol Biofuels 7:122-014-0122-x. eCollection 2014
45. Zhou H, Cheng J, Wang BL, Fink GR, Stephanopoulos G (2012) Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. Metab Eng 14:611–622
46. Sarthy AV, McConaughy BL, Lobo Z, Sundstrom JA, Furlong CE, Hall BD (1987) Expression of the *Escherichia coli* xylose isomerase gene in *Saccharomyces cerevisiae*. Appl Environ Microbiol 53:1996–2000
47. Kötter P, Amore R, Hollenberg CP, Ciriacy M (1990) Isolation and characterization of the *Pichia stipitis* xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant. Curr Genet 18:493–500
48. Walfridsson M, Hallborn J, Penttila M, Keranen S, Hahn-Hagerdal B (1995) Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase. Appl Environ Microbiol 61:4184–4190
49. Ho NW, Chen Z, Brainard AP (1998) Genetically engineered *Saccharomyces* yeast capable of effective cofermentation of glucose and xylose. Appl Environ Microbiol 64:1852–1859
50. Chu BC, Lee H (2007) Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. Biotechnol Adv 25:425–441
51. Wisselink HW, Toirkens MJ, del Rosario Franco Berriel M, Winkler AA, van Dijken JP, Pronk JT, van Maris AJ (2007) Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose. Appl Environ Microbiol 73:4881–4891
52. Richard P, Londesborough J, Putkonen M, Kalkkinen N, Penttila M (2001) Cloning and expression of a fungal L-arabinitol 4-dehydrogenase gene. J Biol Chem 276:40631–40637
53. Richard P, Putkonen M, Väänänen R, Londesborough J, Penttilä M (2002) The missing link in the fungal L-arabinose catabolic pathway, identification of the L-xylulose reductase gene. Biochemistry (NY) 41:6432–6437
54. Bettiga M, Bengtsson O, Hahn-Hagerdal B, Gorwa-Grauslund MF (2009) Arabinose and xylose fermentation by recombinant *Saccharomyces cerevisiae* expressing a fungal pentose utilization pathway. Microb Cell Fact 8:40-2859-8-40
55. Hamacher T, Becker J, Gardonyi M, Hahn-Hagerdal B, Boles E (2002) Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology 148:2783–2788
56. Becker J, Boles E (2003) A modified *Saccharomyces cerevisiae* strain that consumes L-arabinose and produces ethanol. Appl Environ Microbiol 69:4144–4150
57. Leandro M, Gonçalves P, Spencer-Martins I (2006) Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H symporter. Biochem J 395:543–549
58. Young EM, Comer AD, Huang H, Alper HS (2012) A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. Metab Eng 14:401–411

59. Wang C, Bao X, Li Y, Jiao C, Hou J, Zhang Q, Zhang W, Liu W, Shen Y (2015) Cloning and characterization of heterologous transporters in *Saccharomyces cerevisiae* and identification of important amino acids for xylose utilization. Metab Eng 30:79–88
60. Young EM, Tong A, Bui H, Spofford C, Alper HS (2014) Rewiring yeast sugar transporter preference through modifying a conserved protein motif. Proc Natl Acad Sci U S A 111:131–136
61. Subtil T, Boles E (2011) Improving L-arabinose utilization of pentose fermenting *Saccharomyces cerevisiae* cells by heterologous expression of L-arabinose transporting sugar transporters. Biotechnol Biofuels 4:38-6834-4-38
62. Wang C, Shen Y, Zhang Y, Suo F, Hou J, Bao X (2013) Improvement of L-arabinose fermentation by modifying the metabolic pathway and transport in *Saccharomyces cerevisiae*. BioMed Res Int 2013:1–9
63. Hamelinck CN, Van Hooijdonk G, Faaij AP (2005) Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. Biomass Bioenergy 28:384–410
64. Galazka JM, Tian C, Beeson WT, Martinez B, Glass NL, Cate JH (2010) Cellodextrin transport in yeast for improved biofuel production. Science 330:84–86
65. Ha S, Galazka JM, Oh EJ, Kordić V, Kim H, Jin Y, Cate JH (2013) Energetic benefits and rapid cellobiose fermentation by *Saccharomyces cerevisiae* expressing cellobiose phosphorylase and mutant cellodextrin transporters. Metab Eng 15:134–143
66. Ha SJ, Galazka JM, Kim SR, Choi JH, Yang X, Seo JH, Glass NL, Cate JH, Jin YS (2011) Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. Proc Natl Acad Sci U S A 108:504–509
67. Li S, Du J, Sun J, Galazka JM, Glass NL, Cate JH, Yang X, Zhao H (2010) Overcoming glucose repression in mixed sugar fermentation by co-expressing a cellobiose transporter and a β-glucosidase in *Saccharomyces cerevisiae*. Mol BioSyst 6:2129–2132
68. Ha SJ, Wei Q, Kim SR, Galazka JM, Cate JH, Jin YS (2011) Cofermentation of cellobiose and galactose by an engineered *Saccharomyces cerevisiae* strain. Appl Environ Microbiol 77:5822–5825
69. Pawar SN, Edgar KJ (2012) Alginate derivatization: a review of chemistry, properties and applications. Biomaterials 33:3279–3305
70. Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CN, Kim PB, Cooper SR, Raisner RM, Herman A, Sivitz AB, Lakshmanaswamy A, Kashiyama Y, Baker D, Yoshikuni Y (2012) An engineered microbial platform for direct biofuel production from brown macroalgae. Science 335:308–313
71. Enquist-Newman M, Faust AME, Bravo DD, Santos CNS, Raisner RM, Hanel A, Sarvabhowman P, Le C, Regitsky DD, Cooper SR (2014) Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform. Nature 505:239–243
72. Lee K, Hong M, Jung S, Ha S, Yu BJ, Koo HM, Park SM, Seo J, Kweon D, Park JC (2011) Improved galactose fermentation of *Saccharomyces cerevisiae* through inverse metabolic engineering. Biotechnol Bioeng 108:621–631
73. Ra CH, Kim YJ, Lee SY, Jeong G, Kim S (2015) Effects of galactose adaptation in yeast for ethanol fermentation from red seaweed, *Gracilaria verrucosa*. Bioprocess Biosyst Eng :1–8
74. Kim H, Ra CH, Kim S (2013) Ethanol production from seaweed (*Undaria pinnatifida*) using yeast acclimated to specific sugars. Biotechnol Bioprocess Eng 18:533–537
75. Patwardhan PR, Brown RC, Shanks BH (2011) Understanding the fast pyrolysis of lignin. ChemSusChem 4:1629–1636
76. Wang K, Kim KH, Brown RC (2014) Catalytic pyrolysis of individual components of lignocellulosic biomass. Green Chem 16:727–735
77. Wei N, Quarterman J, Kim SR, Cate JH, Jin Y (2013) Enhanced biofuel production through coupled acetic acid and xylose consumption by engineered yeast. Nat Commun 4:2580

78. Wei N, Oh EJ, Million G, Cate JH, Jin Y (2015) Simultaneous utilization of cellobiose, xylose, and acetic acid from lignocellulosic biomass for biofuel production by an engineered yeast platform. ACS Synth Biol 4:707–713
79. Zhang G, Kong II, Wei N, Peng D, Turner TL, Sung BH, Sohn J, Jin Y (2016) Optimization of an acetate reduction pathway for producing cellulosic ethanol by engineered yeast. Biotechnol Bioeng 113:2587–2596
80. Domingues L, Lima N, Teixeira JA (2001) Alcohol production from cheese whey permeate using genetically modified flocculent yeast cells. Biotechnol Bioeng 72:507–514
81. Guimaraes PM, Francois J, Parrou JL, Teixeira JA, Domingues L (2008) Adaptive evolution of a lactose-consuming *Saccharomyces cerevisiae* recombinant. Appl Environ Microbiol 74:1748–1756
82. Guimarães PM, Teixeira JA, Domingues L (2008) Fermentation of high concentrations of lactose to ethanol by engineered flocculent *Saccharomyces cerevisiae*. Biotechnol Lett 30:1953–1958
83. Sreekrishna K, Dickson RC (1985) Construction of strains of *Saccharomyces cerevisiae* that grow on lactose. Proc Natl Acad Sci U S A 82:7909–7913
84. Zou J, Guo X, Shen T, Dong J, Zhang C, Xiao D (2013) Construction of lactose-consuming *Saccharomyces cerevisiae* for lactose fermentation into ethanol fuel. J Ind Microbiol Biotechnol 40:353–363
85. Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8:74–79
86. Mosier NS, Ileleji KE (2014) How fuel ethanol is made from corn. Bioenerg Biomass Biofuel 379–384
87. Ansell R, Granath K, Hohmann S, Thevelein JM, Adler L (1997) The two isoenzymes for yeast NAD^+-dependent glycerol 3-phosphate dehydrogenase encoded by GPD1 and GPD2 have distinct roles in osmoadaptation and redox regulation. EMBO J 16:2179–2187
88. Valadi H, Larsson C, Gustafsson L (1998) Improved ethanol production by glycerol-3-phosphate dehydrogenase mutants of *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 50:434–439
89. Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev 66:300–372
90. Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J (2000) Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. Metab Eng 2:69–77
91. Bro C, Regenberg B, Förster J, Nielsen J (2006) In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab Eng 8:102–111
92. Guo Z, Zhang L, Ding Z, Shi G (2011) Minimization of glycerol synthesis in industrial ethanol yeast without influencing its fermentation performance. Metab Eng 13:49–59
93. Guadalupe Medina V, Almering MJ, van Maris AJ, Pronk JT (2010) Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. Appl Environ Microbiol 76:190–195
94. Zhang L, Tang Y, Guo Z, Ding Z, Shi G (2011) Improving the ethanol yield by reducing glycerol formation using cofactor regulation in *Saccharomyces cerevisiae*. Biotechnol Lett 33:1375–1380
95. Zhang A, Kong Q, Cao L, Chen X (2007) Effect of FPS1 deletion on the fermentation properties of *Saccharomyces cerevisiae*. Lett Appl Microbiol 44:212–217
96. Hubmann G, Guillouet S, Nevoigt E (2011) Gpd1 and Gpd2 fine-tuning for sustainable reduction of glycerol formation in *Saccharomyces cerevisiae*. Appl Environ Microbiol 77:5857–5867
97. Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, Liao JC (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. Appl Environ Microbiol 77:2905–2915

98. Steen EJ, Chan R, Prasad N, Myers S, Petzold CJ, Redding A, Ouellet M, Keasling JD (2008) Metabolic engineering of *Saccharomyces cerevisiae* for the production of *n*-butanol. Microb Cell Fact 7:36
99. Krivoruchko A, Serrano-Amatriain C, Chen Y, Siewers V, Nielsen J (2013) Improving biobutanol production in engineered *Saccharomyces cerevisiae* by manipulation of acetyl-CoA metabolism. J Ind Microbiol Biotechnol 40:1051–1056
100. Lian J, Si T, Nair NU, Zhao H (2014) Design and construction of acetyl-CoA overproducing *Saccharomyces cerevisiae* strains. Metab Eng 24:139–149
101. Si T, Luo Y, Xiao H, Zhao H (2014) Utilizing an endogenous pathway for 1-butanol production in *Saccharomyces cerevisiae*. Metab Eng 22:60–68
102. Chen X, Nielsen KF, Borodina I, Kielland-Brandt MC, Karhumaa K (2011) Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. Biotechnol Biofuels 4:2089–2090
103. Avalos JL, Fink GR, Stephanopoulos G (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. Nat Biotechnol 31:335–341
104. Yuan J, Ching CB (2014) Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds. ACS Synth Biol 4:23–31
105. Brat D, Weber C, Lorenzen W, Bode HB, Boles E (2012) Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast *Saccharomyces cerevisiae*. Biotechnol Biofuels 5:65
106. Matsuda F, Ishii J, Kondo T, Ida K, Tezuka H, Kondo A (2013) Increased isobutanol production in *Saccharomyces cerevisiae* by eliminating competing pathways and resolving cofactor imbalance. Microb Cell Fact 12:119
107. Huo Y, Cho KM, Rivera JGL, Monte E, Shen CR, Yan Y, Liao JC (2011) Conversion of proteins into biofuels by engineering nitrogen flux. Nat Biotechnol 29:346–351
108. Kalscheuer R, Luftmann H, Steinbuchel A (2004) Synthesis of novel lipids in *Saccharomyces cerevisiae* by heterologous expression of an unspecific bacterial acyltransferase. Appl Environ Microbiol 70:7119–7125
109. Shi S, Valle-Rodríguez JO, Khoomrung S, Siewers V, Nielsen J (2012) Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. Biotechnol Biofuels 5:10.1186
110. de Jong BW, Shi S, Siewers V, Nielsen J (2014) Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. Microb Cell Fact 13:39
111. Valle-Rodríguez JO, Shi S, Siewers V, Nielsen J (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways. Appl Energy 115:226–232
112. Eriksen DT, HamediRad M, Yuan Y, Zhao H (2015) Orthogonal fatty acid biosynthetic pathway improves fatty acid ethyl ester production in *Saccharomyces cerevisiae*. ACS Synth Biol 4:808–814
113. Costenoble R, Picotti P, Reiter L, Stallmach R, Heinemann M, Sauer U, Aebersold R (2011) Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. Mol Syst Biol 7:464
114. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD (2015) Complete biosynthesis of opioids in yeast. Science 349:1095–1100
115. Ng CY, Jung M, Lee J, Oh M (2012) Production of 2,3-butanediol in *Saccharomyces cerevisiae* by *in silico* aided metabolic engineering. Microb Cell Fact 11:68
116. Kim S, Seo S, Jin Y, Seo J (2013) Production of 2,3-butanediol by engineered *Saccharomyces cerevisiae*. Bioresour Technol 146:274–281

117. Kim J, Seo S, Zhang G, Jin Y, Seo J (2015) Expression of Lactococcus lactis NADH oxidase increases 2,3-butanediol production in Pdc-deficient *Saccharomyces cerevisiae*. Bioresour Technol 191:512–519
118. Kim S, Hahn J (2015) Efficient production of 2,3-butanediol in *Saccharomyces cerevisiae* by eliminating ethanol and glycerol production and redox rebalancing. Metab Eng 31:94–101
119. George KW, Alonso-Gutierrez J, Keasling JD, Lee TS (2015) Isoprenoid drugs, biofuels, and chemicals—artemisinin, farnesene, and beyond. In: Biotechnology of isoprenoids. Springer, Cham, pp 355–389
120. Tippmann S, Chen Y, Siewers V, Nielsen J (2013) From flavors and pharmaceuticals to advanced biofuels: production of isoprenoids in *Saccharomyces cerevisiae*. Biotechnol J 8:1435–1444
121. Brennan TC, Turner CD, Krömer JO, Nielsen LK (2012) Alleviating monoterpene toxicity using a two-phase extractive fermentation for the bioproduction of jet fuel mixtures in *Saccharomyces cerevisiae*. Biotechnol Bioeng 109:2513–2522
122. Fischer MJ, Meyer S, Claudel P, Bergdoll M, Karst F (2011) Metabolic engineering of monoterpene synthesis in yeast. Biotechnol Bioeng 108:1883–1892
123. Albertsen L, Chen Y, Bach LS, Rattleff S, Maury J, Brix S, Nielsen J, Mortensen UH (2011) Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. Appl Environ Microbiol 77:1033–1040
124. d'Espaux L, Mendez-Perez D, Li R, Keasling JD (2015) Synthetic biology for microbial production of lipid-based biofuels. Curr Opin Chem Biol 29:58–65
125. Ignea C, Cvetkovic I, Loupassaki S, Kefalas P, Johnson CB, Kampranis SC, Makris AM (2011) Improving yeast strains using recyclable integration cassettes, for the production of plant terpenoids. Microb Cell Fact 10:1–18
126. Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483
127. Özaydın B, Burd H, Lee TS, Keasling JD (2013) Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. Metab Eng 15:174–183
128. Engels B, Dahm P, Jennewein S (2008) Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 10:201–206
129. Zhou YJ, Gao W, Rong Q, Jin G, Chu H, Liu W, Yang W, Zhu Z, Li G, Zhu G (2012) Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. J Am Chem Soc 134:3234–3241
130. Dai Z, Liu Y, Huang L, Zhang X (2012) Production of miltiradiene by metabolically engineered *Saccharomyces cerevisiae*. Biotechnol Bioeng 109:2845–2853
131. White NJ (1997) Assessment of the pharmacodynamic properties of antimalarial drugs in vivo. Antimicrob Agents Chemother 41:1413–1422
132. Tu Y (2011) The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat Med 17:1217–1220
133. Kindermans J, Pilloy J, Olliaro P, Gomes M (2007) Ensuring sustained ACT production and reliable artemisinin supply. Malar J 6:125
134. Lindahl A, Olsson ME, Mercke P, Tollbom Ö, Schelin J, Brodelius M, Brodelius PE (2006) Production of the artemisinin precursor amorpha-4,11-diene by engineered *Saccharomyces cerevisiae*. Biotechnol Lett 28:571–580
135. Ro D, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440:940–943
136. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production of amorphadiene in yeast, and its

conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109:E111–E118
137. Auras RA, Lim L, Selke SE, Tsuji H (2011) Poly(lactic acid): synthesis, structures, properties, processing, and applications. Wiley
138. John RP, Nampoothiri KM, Pandey A (2007) Fermentative production of lactic acid from biomass: an overview on process developments and future perspectives. Appl Microbiol Biotechnol 74:524–534
139. Ishida N, Saitoh S, Onishi T, Tokuhiro K, Nagamori E, Kitamoto K, Takahashi H (2006) The effect of pyruvate decarboxylase gene knockout in *Saccharomyces cerevisiae* on L-lactic acid production. Biosci Biotechnol Biochem 70:1148–1153
140. Turner TL, Zhang G, Kim SR, Subramaniam V, Steffen D, Skory CD, Jang JY, Yu BJ, Jin Y (2015) Lactic acid production from xylose by engineered *Saccharomyces cerevisiae* without PDC or ADH deletion. Appl Microbiol Biotechnol 99(19):8023–8033
141. Turner TL, Zhang G, Oh EJ, Subramaniam V, Adiputra A, Subramaniam V, Skory CD, Jang JY, Yu BJ, Park I (2015) Lactic acid production from cellobiose and xylose by engineered *Saccharomyces cerevisiae*. Biotechnol Bioeng 113:1075–1083
142. Ishida N, Saitoh S, Tokuhiro K, Nagamori E, Matsuyama T, Kitamoto K, Takahashi H (2005) Efficient production of L-lactic acid by metabolically engineered *Saccharomyces cerevisiae* with a genome-integrated L-lactate dehydrogenase gene. Appl Environ Microbiol 71:1964–1970
143. Saitoh S, Ishida N, Onishi T, Tokuhiro K, Nagamori E, Kitamoto K, Takahashi H (2005) Genetically engineered wine yeast produces a high concentration of L-lactic acid of extremely high optical purity. Appl Environ Microbiol 71:2789–2792
144. Skory CD (2003) Lactic acid production by *Saccharomyces cerevisiae* expressing a *Rhizopus oryzae* lactate dehydrogenase gene. J Ind Microbiol Biotechnol 30:22–27
145. Colombié S, Dequin S, Sablayrolles JM (2003) Control of lactate production by *Saccharomyces cerevisiae* expressing a bacterial LDH gene. Enzyme Microb Technol 33:38–46
146. Colombié S, Sablayrolles J (2004) Nicotinic acid controls lactate production by K1-LDH: a *Saccharomyces cerevisiae* strain expressing a bacterial LDH gene. J Ind Microbiol Biotechnol 31:209–215
147. Tate BE (1967) Polymerization of itaconic acid and derivatives. In: Fortschritte der Hochpolymeren-Forschung. Springer, Berlin, pp 214–232
148. Blazeck J, Miller J, Pan A, Gengler J, Holden C, Jamoussi M, Alper HS (2014) Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production. Appl Microbiol Biotechnol 98:8155–8164
149. Xie N, Liang H, Huang R, Xu P (2014) Biotechnological production of muconic acid: current status and future prospects. Biotechnol Adv 32:615–622
150. Weber C, Bruckner C, Weinreb S, Lehr C, Essl C, Boles E (2012) Biosynthesis of *cis*,*cis*-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*. Appl Environ Microbiol 78:8421–8430
151. Curran KA, Leavitt JM, Karim AS, Alper HS (2013) Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng 15:55–66
152. Yan D, Wang C, Zhou J, Liu Y, Yang M, Xing J (2014) Construction of reductive pathway in *Saccharomyces cerevisiae* for effective succinic acid fermentation at low pH value. Bioresour Technol 156:232–239
153. Agren R, Otero JM, Nielsen J (2013) Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. J Ind Microbiol Biotechnol 40:735–747
154. Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, Nielsen J (2013) Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. PLoS One 8, e54144

155. Goerz O, Ritter H (2013) Polymers with shape memory effect from renewable resources: crosslinking of polyesters based on isosorbide, itaconic acid and succinic acid. Polym Int 62:709–712
156. Koivistoinen OM, Kuivanen J, Barth D, Turkia H, Pitkänen J, Penttilä M, Richard P (2013) Glycolic acid production in the engineered yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Microb Cell Fact 12:1
157. Beerens K, Desmet T, Soetaert W (2012) Enzymes for the biocatalytic production of rare sugars. J Ind Microbiol Biotechnol 39:823–834
158. Hu C, Li L, Zheng Y, Rui L, Hu C (2011) Perspectives of biotechnological production of L-ribose and its purification. Appl Microbiol Biotechnol 92:449–455
159. Okano K (2009) Synthesis and pharmaceutical application of L-ribose. Tetrahedron 65:1937–1949
160. Milgrom P, Ly KA (2012) The role of sugar alcohols, xylitol, and chewing gum in preventing dental diseases. Comprehensive preventive dentistry. Wiley, West Sussex, pp 146–158
161. Ingram JM, Wood WA (1965) Enzymatic basis for D-arbitol production by *Saccharomyces rouxii*. J Bacteriol 89:1186–1194
162. Chaturvedi V, Bartiss A, Wong B (1997) Expression of bacterial *mtlD* in *Saccharomyces cerevisiae* results in mannitol synthesis and protects a glycerol-defective mutant from high-salt and oxidative stress. J Bacteriol 179:157–162
163. Costenoble R, Adler L, Niklasson C, Liden G (2003) Engineering of the metabolism of *Saccharomyces cerevisiae* for anaerobic production of mannitol. FEMS Yeast Res 3:17–25
164. Wisniak J, Hershkowitz M, Leibowitz R, Stein S (1974) Hydrogenation of xylose to xylitol. Ind Eng Chem Prod Res Dev 13:75–79
165. Oh E, Bae Y, Kim K, Park Y, Seo J (2012) Effects of overexpression of acetaldehyde dehydrogenase 6 and acetyl-CoA synthetase 1 on xylitol production in recombinant *Saccharomyces cerevisiae*. Biocatal Agric Biotechnol 1:15–19
166. Oh EJ, Ha S, Rin Kim S, Lee W, Galazka JM, Cate JH, Jin Y (2013) Enhanced xylitol production through simultaneous co-utilization of cellobiose and xylose by engineered *Saccharomyces cerevisiae*. Metab Eng 15:226–234
167. Jo J, Oh S, Lee H, Park Y, Seo J (2015) Dual utilization of NADPH and NADH cofactors enhances xylitol production in engineered *Saccharomyces cerevisiae*. Biotechnol J 10:1935–1943
168. Baur JA, Pearson KJ, Price NL, Jamieson HA, Lerin C, Kalra A, Prabhu VV, Allard JS, Lopez-Lluch G, Lewis K (2006) Resveratrol improves health and survival of mice on a high-calorie diet. Nature 444:337–342
169. Becker JV, Armstrong GO, van der Merwe MJ, Lambrechts MG, Vivier MA, Pretorius IS (2003) Metabolic engineering of *Saccharomyces cerevisiae* for the synthesis of the wine-related antioxidant resveratrol. FEMS Yeast Res 4:79–85
170. Beekwilder J, Wolswinkel R, Jonker H, Hall R, de Vos CH, Bovy A (2006) Production of resveratrol in recombinant microorganisms. Appl Environ Microbiol 72:5670–5672
171. Sydor T, Schaffer S, Boles E (2010) Considerable increase in resveratrol production by recombinant industrial yeast strains with use of rich medium. Appl Environ Microbiol 76:3361–3363
172. Wang Y, Halls C, Zhang J, Matsuno M, Zhang Y, Yu O (2011) Stepwise increase of resveratrol biosynthesis in yeast *Saccharomyces cerevisiae* by metabolic engineering. Metab Eng 13:455–463
173. Hara KY, Kiriyama K, Inagaki A, Nakayama H, Kondo A (2012) Improvement of glutathione production by metabolic engineering the sulfate assimilation pathway of *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 94:1313–1319
174. Taskin M (2013) A new strategy for improved glutathione production from *Saccharomyces cerevisiae*: use of cysteine-and glycine-rich chicken feather protein hydrolysate as a new cheap substrate. J Sci Food Agric 93:535–541

175. Zhao Y, Bian X, You X, Shao F, Xiang X, Deng X, Zhao G, Xu J (2013) Nystatin-enhanced glutathione production by *Saccharomyces cerevisiae* depends on γ-glutamylcysteine synthase activity and K. Eng Life Sci 13:156–162
176. Mezzetti F, De Vero L, Giudici P (2014) Evolved *Saccharomyces cerevisiae* wine strains with enhanced glutathione production obtained by an evolution-based strategy. FEMS Yeast Res 14:977–987
177. Oraby MM, Allababidy M, Ramadan E (2014) Enhancement of antioxidant glutathione production by *Saccharomyces cerevisiae* growing under stressful condition. Int J 5:160–165
178. Patzschke A, Steiger MG, Holz C, Lang C, Mattanovich D, Sauer M (2015) Enhanced glutathione production by evolutionary engineering of *Saccharomyces cerevisiae* strains. Biotechnol J 10:1719–1726
179. Li Y, Wei G, Chen J (2004) Glutathione: a review on biotechnological production. Appl Microbiol Biotechnol 66:233–242
180. Kwolek-Mirek M, Bednarska S, Bartosz G, Biliński T (2009) Acrolein toxicity involves oxidative stress caused by glutathione depletion in the yeast *Saccharomyces cerevisiae*. Cell Biol Toxicol 25:363–378
181. Vollenweider S, Evers S, Zurbriggen K, Lacroix C (2010) Unraveling the hydroxypropionaldehyde (HPA) system: an active antimicrobial agent against human pathogens. J Agric Food Chem 58:10315–10322
182. Bakhiet SE, Mahmoud MA (2015) Production of bio-ethanol from molasses by *Schizosaccharomyces* species. Annu Res Rev Biol 7:45–53
183. Tai M, Stephanopoulos G (2013) Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. Metab Eng 15:1–9
184. Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc Natl Acad Sci U S A 76:106–110
185. Huang C, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. J Ind Microbiol Biotechnol 39:383–399
186. Geddes CC, Nieves IU, Ingram LO (2011) Advances in ethanol production. Curr Opin Biotechnol 22:312–319
187. Jang Y, Park JM, Choi S, Choi YJ, Cho JH, Lee SY (2012) Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. Biotechnol Adv 30:989–1000
188. Anderson J, DiCicco D, Ginder J, Kramer U, Leone T, Raney-Pablo H, Wallington T (2012) High octane number ethanol–gasoline blends: quantifying the potential benefits in the United States. Fuel 97:585–594
189. United States Environmental Protection Agency (2015) EPA Proposes Renewable Fuel Standards for 2014, 2015, and 2016, and the Biomass-Based Diesel Volume for 2017. In: http://www3.epa.gov/otaq/fuels/renewablefuels/documents/420f15028.pdf. Accessed 21 Sept 2015
190. Sissine F (2007) Energy independence and security cct of 2007: a summary of major provisions. Library of Congress, Congressional Research Service, Washington DC
191. de Souza Dias MO, Maciel Filho R, Mantelatto PE, Cavalett O, Rossell CEV, Bonomi A, Leal MRLV (2015) Sugarcane processing for ethanol and sugar in Brazil. Environ Dev 15:35–51
192. Mumm RH, Goldsmith PD, Rausch KD, Stein HH (2014) Land usage attributed to corn ethanol production in the United States: sensitivity to technological advances in corn grain yield, ethanol conversion, and co-product utilization. Biotechnol Biofuels 7:61-6834-7-61. eCollection 2014
193. Goldemberg J (2013) Sugarcane ethanol: strategies to a successful program in Brazil. In: Advanced biofuels and bioproducts. Springer, New York, pp 13–20

194. Wallington T, Anderson J, Mueller S, Kolinski Morris E, Winkler S, Ginder J, Nielsen OJ (2012) Corn ethanol production, food exports, and indirect land use change. Environ Sci Technol 46:6379–6384
195. IPCC (2014) Climate Change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. In: Field CB, Barros VR, Dokken DJ, Mach KJ, Mastrandrea MD, Bilir TE, Chatterjee M, Ebi KL, Estrada YO, Genova RC, Girma B, Kissel ES, Levy AN, MacCracken S, Mastrandrea PR, White LL (eds) Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Canbridge/New York, p 1132
196. Kircher M (2012) The transition to a bio-economy: emerging from the oil age. Biofuels Bioprod Biorefin 6:369–375
197. Dale BE, Anderson JE, Brown RC, Csonka S, Dale VH, Herwick G, Jackson RD, Jordan N, Kaffka S, Kline KL (2014) Take a closer look: biofuels can support environmental, economic and social goals. Environ Sci Technol 48:7200–7203
198. Wang Q, Li R (2016) Impact of cheaper oil on economic system and climate change: a SWOT analysis. Renew Sustain Energy Rev 54:925–931
199. Reboredo FH, Lidon F, Pessoa F, Ramalho JC (2016) The fall of oil prices and the effects on biofuels. Trends Biotechnol 34:3–6
200. Verwaal R, Wu L, Damveld RA, Sagt CMJ (2009) Succinic acid production in a eukaryotic cell. BMC Biotechnol 9:48
201. Picataggio S, Beardslee T (2013) Biological methods for preparing adipic acid. US Patent No. 8,343,752
202. Fruchey OS, Manzer LE, Dunuwila D, Keen BT, Albin BA, Clinton NA, Dombek BD (2011) Processes for producing butanediol (BDO), diaminobutane (DAB), succinic dinitrile (SDN) and succinamide (DAM). US Patent Application No. 14/117,141
203. Gardner TS, Hawkins KM, Meadows AL, Tsong AE, Tsegaye Y (2013) Production of acetyl-coenzyme a derived isoprenoids. US Patent No. 8,603,800
204. Miller M, Suominen P, Aristidou A, Hause BM, Van Hoek P, Dundon CA (2012) Lactic acid-producing yeast cells having nonfunctional L- or D-lactate:ferricytochrome C oxidoreductase cells. US Patent No. 8,137,953
205. Pinazo JM, Domine ME, Parvulescu V, Petru F (2015) Sustainability metrics for succinic acid production: a comparison between biomass-based and petrochemical routes. Catal Today 239:17–24

Adv Biochem Eng Biotechnol (2018) 162: 217–264
DOI: 10.1007/10_2016_21

Published online: 22 November 2016

Corynebacterium glutamicum for Sustainable Bioproduction: From Metabolic Physiology to Systems Metabolic Engineering

Judith Becker, Gideon Gießelmann, Sarah Lisa Hoffmann, and Christoph Wittmann

Abstract Since its discovery 60 years ago, *Corynebacterium glutamicum* has evolved into a workhorse for industrial biotechnology. Traditionally well known for its remarkable capacity to produce amino acids, this Gram-positive soil bacterium, has become a flexible, efficient production platform for various bulk and fine chemicals, materials, and biofuels. The central turnstile of all these achievements is our excellent understanding of its metabolism and physiology. This knowledge base, together with innovative systems metabolic engineering concepts, which integrate systems and synthetic biology into strain engineering, has upgraded *C. glutamicum* into one of the most successful industrial microorganisms in the world.

Keywords Diaminopentane, Industrial biotechnology, Industrial raw material, Lysine, Metabolic network, Pathway engineering, Rational design

Contents

J. Becker, G. Gießelmann, S.L. Hoffmann, and C. Wittmann (✉)
Institute of Systems Biotechnology, Saarland University, 66123 Saarbrücken, Germany
e-mail: christoph.wittmann@uni-saarland.de

1 Introduction

The Gram-positive soil bacterium *Corynebacterium glutamicum* belongs to the veterans of industrial biotechnology. Its natural capability to produce and secrete glutamate in high amounts originally led to its discovery about 60 years ago [1]. From early on, its versatile metabolism, nutritional flexibility, and process robustness were major drivers to establish and develop industrial strains and processes from scratch. Among them, amino acid production has evolved most rapidly, today being a multi-billion dollar business [2, 3]. Over the past decades, global competition among leading companies in the field steadily demanded innovation to improve key performance indicators: yield, titer, and productivity. For this reason, *C. glutamicum* has become one of the best characterized microorganisms worldwide with regard to substrate spectrum and nutrient requirement [4], catabolic and anabolic pathways and their regulation [5], the underlying biochemistry [6], and response to environmental conditions [7]. On entering the era of genetic engineering, this provided a detailed knowledge base for targeted modification of enzymes and pathways to optimize established fermentation processes. More recently, powerful molecular tools for genome-based engineering together with technologies to analyze genome, transcriptome, proteome, metabolome, and fluxome have enabled the next level of strain engineering: tailored optimization on a systems wide level. Successful expression of heterologous genes in *C. glutamicum* even allowed crossing natural boundaries and paving the way to non-natural products. Systems and synthetic metabolic engineering has enabled *C. glutamicum* to produce a wide portfolio of products: biofuels, bulk and fine chemicals, polymer building blocks, polymers, feed additives, and products for nutrition and health care [4, 8, 9]. The central turnstile of all achievements is metabolism and physiology. Core carbon metabolism has to function properly in a successful cell factory. For maximal conversion of external carbon sources into desired products, metabolism has to keep producing cells alive and simultaneously provide energy, carbon building blocks, and redox power for biosynthesis. Without doubt, metabolism and physiology shape the basis for the industrial success of *C. glutamicum* and deserve a close and detailed view. Accordingly, this chapter summarizes our current knowledge in this area. In addition, prominent strategies

and showcases highlight the upgrade of *C. glutamicum* into one of the most important cell factories in white biotechnology.

2 Metabolism: Pathway Principles and Engineering Strategies

C. glutamicum is a soil-dwelling microorganism. It belongs to the high GC content Gram-positive bacteria, the Actinobacteria. Cells are small, non-motile, and non-spore forming. Their shape is typically club-like, explaining the name "coryne-form" (club-shaped). The type strain, *C. glutamicum* ATCC 13032, possesses a circular chromosome of 3.3 Mb and a plasmid of 0.5 Mb [10]. From early on, the industrial relevance of *C. glutamicum* has driven the investigation of its biochemistry, with a strong focus on the pathways of core metabolism and their regulation, synthesizing products of interest from substrates of interest. This provides a highly detailed portrayal of *C. glutamicum*, which substantially guides metabolic engineering approaches.

2.1 Carbon Core Metabolism

2.1.1 Substrate Uptake

C. glutamicum is able to use a variety of carbon sources as growth and energy substrates, including sugars [11, 12], sugar alcohols [13], and organic acids [14–18].

Hexoses

Sugar uptake in *C. glutamicum* is mediated by phosphotransferase systems (PTS), first described by Mori and Shiio [19]. During transport, the sugar is phosphorylated at the expense of phosphoenolpyruvate (PEP) (Fig. 1). For *C. glutamicum,* four PTS variants have been reported, being specific for glucose, fructose, sucrose, and mannose [20]. All systems consist of three distinct proteins: enzyme I (EI), histidine protein (HPr), and enzyme II (EII). EII mediates substrate specificity and the corresponding protein variants are encoded by *ptsG, ptsF*, and *ptsS* for the uptake of glucose, fructose, and sucrose, respectively [21, 25]. The two general components EI and HPr are encoded by the genes *ptsI* and *ptsH*. A remaining glucose assimilation activity, observed in *ptsG* null mutants, has shown the existence of an additional PTS-independent uptake system for glucose. Such an uptake would require intracellular phosphorylation of glucose. Indeed, a glucose kinase (*glk*) is present in *C. glutamicum* [25, 26] and might contribute up to 15% of total glucose

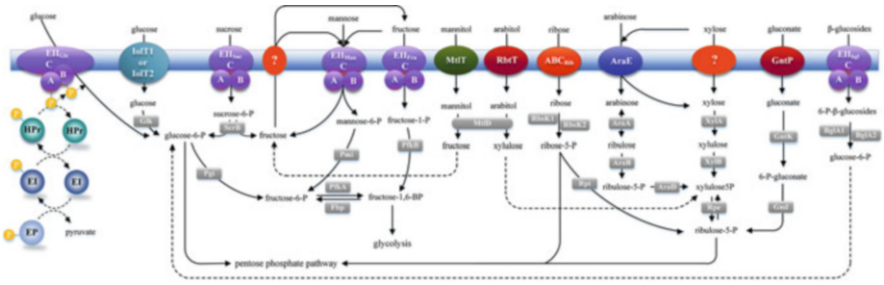

Fig. 1 Overview on transporters and metabolic reactions in C. glutamicum for uptake and conversion of different industrially relevant sugars and sugar alcohols, accessible from renewable biomass [6, 13, 20–24]. The transporter set comprises phosphotransferase systems for uptake of glucose, fructose, sucrose, mannose, and β-glucosides, an ABC transporter for ribose, an H^+-symporter for arabinose, permeases for gluconate and arabitol, a MFS-type transporter for mannitol, and a transporter of the major facilitator superfamily for non-PTS-mediated glucose uptake. Transporters for fructose secretion and xylose uptake have not been identified so far. ABC_{Rib} ATP-binding cassette transporter for ribose import, AraA arabinose isomerase, AraB ribulokinase, AraE arabinose H^+-symporter, AraD ribulose 5-phosphate 4-epimerase, BglA phospho-β-glucosidases BglA1 and BglA2, EI general PTS-component enzyme I, EII_{bgl} β-glucosides-specific PTS component, EII_{Fru} fructose-specific PTS component, EII_{Glc} glucose-specific PTS component, EII_{Man} mannose-specific PTS component, EII_{Suc} sucrose-specific PTS component, Fbp fructose-1,6-bisphosphatase, Glk glucokinase, Gnd 6-phosphogluconate dehydrogenase, GntK gluconate kinase, GntP gluconate permease, HPr histidine protein, IolT myo-inositol transporter 1 and 2, MtlD mannitol 2-dehydrogenase, MtlT mannitol transporter, PEP phosphoenolpyruvate, PfkA 6-phosphofructokinase, PfkB fructose 1-phosphate kinase, Pgi phosphoglucoisomerase, Pmi phosphomannoseisomerase, RbsK ribokinase 1 and 2, RbtT ribitol transporter, Rpe ribulose 5-phosphate epimerase, Rpi ribose 5-phosphate epimerase, ScrB sucrose 6-phosphate hydrolase, XylA xylose isomerase, XylB xylulokinase

uptake [27]. Recently, two transporters for *myo*-inositol, (*iolT1* and *iolT2*) were identified, both of which mediate glucose uptake in *C. glutamicum* [28, 29]. Metabolic engineering of *C. glutamicum* toward utilization of a PTS-independent glucose uptake was beneficially applied to improve lysine [30] and succinate production [31]. Similarly, an additional uptake system was suggested for fructose produced by residual growth of a *ptsF* null mutant of *C. glutamicum* on this sugar [32, 33]. In this case, the mannose PTS has been identified to also transport fructose (Fig. 1). Metabolic flux analysis revealed that the mannose PTS is responsible for a relative fructose uptake flux of 8%, whereas 92% of fructose enters the cell via the fructose-specific PTS at the level of fructose 1,6-bisphosphate [34]. This finding led to the identification of fructose 1,6-bisphosphatase activity as bottleneck for fructose- and sucrose-based lysine production [34, 35]. This was overcome by targeted overexpression of the encoding *fbp* gene, which resulted in a substantially improved production performance [36]. Though fructose can also be taken up by the two *myo*-inositol transporters [37], the lack of fructokinase activity avoids further metabolization [38].

Pentoses

Ribose is the only five carbon sugar that is naturally utilized by a variety of *C. glutamicum* strains. As in most bacteria, the uptake of ribose occurs through an ATP-binding cassette (ABC) transporter [39]. The genes, encoding the ribose-specific ABC transporter and its corresponding transcriptional regulatory protein (RbsR), are organized in an operon [40]. Subsequent to uptake, ribose is phosphorylated by one of the two ribokinases RbsK1 and RbsK2, which yields the pentose phosphate (PP) pathway intermediate ribose 5-phosphate (Fig. 1). Double-deletion of the two encoding genes *rbsk1* and *rbsk2* results in the inability to grow on ribose as sole carbon source [40]. Utilization of the pentose arabinose is a rare feature for *C. glutamicum*. The required enzymes L-arabinose isomerase (AraA), L-ribulokinase (AraB), and L-ribulose 5-phosphate 4-epimerase (AraD) are missing in most strains. As an exception, *C. glutamicum* ATCC 31831 possesses an *araBDA* operon and is able to grow on L-arabinose [41]. The upstream region of the gene cluster contains genes for a negative LacI-type transcriptional regulator (*araR*) and a high-affinity arabinose-inducible H^+-symporter (*araE*). Deletion of the latter strongly impairs growth at low arabinose concentration, whereas high substrate concentration supports normal growth, indicating the presence of a so far unidentified additional transporter. Metabolic engineering strategies for utilizing arabinose for growth and production rely on heterologous expression of the arabinose gene cluster of *Escherichia coli*, making arabinose bioavailable for *C. glutamicum* type strain ATCC 13032 [42, 43]. Natural xylose users have not been described so far. However, *C. glutamicum* can take up xylose from the environment and harbors a functional *xylB* gene, encoding xylulokinase activity [44]. Type strains, however, lack xylose isomerase activity, required as the essential link to channel xylose into central carbon metabolism [44]. Related to the relevance of xylose as renewable feedstock, the xylose-assimilation pathway has been reconstructed in *C. glutamicum* to allow growth [44] and production of organic acids [44], proteinogenic and non-proteinogenic amino acids [45, 46], and diamines [46–48].

Gluconate and β-Glucosides

Gluconate enters the cell via a specific permease (GntP) and is subsequently phosphorylated into 6-phosphogluconate, an intermediate of the oxidative PP pathway [49]. Two GntR-type regulators, GntR1 and GntR2, control gluconate metabolism and PTS-mediated glucose uptake. In the absence of gluconate, genes involved in gluconate metabolism (e.g., *gntP* and *gntK*) are repressed, whereby transcription of *ptsG* and *ptsS*, responsible for PTS-dependent glucose and sucrose uptake, is enhanced [49]. The metabolism of β-glucosides is, similarly to arabinose, a strain-specific feature. The strain *C. glutamicum* R, for instance, possesses two gene clusters (*bglF-bglA-bglG* and *blfF2-bglA2-bglG2*) for uptake and degradation

of β-glucosides such as salicin, arbutin, and methyl-β-glucoside, whereas such gene clusters are not present in the genome of C. *glutamicum* ATCC 13032 [6, 50].

Sugar-Alcohols

Arabitol can be used by *C. glutamicum* as sole carbon and energy source [13]. In the presence of arabitol, the catabolic operon, comprising the genes *xylB*, *rbtT*, *mtlD*, and *sixA*, is induced via a regulator, that is, AtlR [13]. Arabitol is taken up via the permease RbtT, and is then oxidized into xylulose by NAD-dependent arabitol dehydrogenase, encoded by *mtlD* (Fig. 1). Subsequent phosphorylation into xylulose 5-phosphate represents an overlap to xylose metabolism and relies on *xylB*-encoded xylulokinase. In addition, *C. glutamicum* carries a mannitol catabolic operon, but the presence of the auto-regulator protein AtlT (MtlR) prevents mannitol utilization. Deletion of the *mtlR* gene abolishes repression and enables transcription, likely producing polycistronic mRNA of the two structural genes *mtlT* and *mtlD*, encoding an MFS-type transporter and NAD-dependent mannitol 2-dehydrogenase (Fig. 1), respectively [22]. As *C. glutamicum* lacks fructokinase activity, fructose, the product of mannitol oxidation, cannot be phosphorylated within the cell. Further metabolization involves fructose efflux by a so far unassigned transporter and re-uptake by the fructose-specific PTS [25, 32, 33].

2.1.2 Embden–Meyerhof–Parnas Pathway

The Embden–Meyerhof–Parnas (EMP) pathway is a major route for catabolic breakdown of sugars and sugar alcohols. Pathway control mainly occurs by metabolic regulation of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and pyruvate kinase (PYK), which are sensitive to the redox and energy state of the cell [10]. The redox state is hereby sensed as NADH/NAD ratio [32], whereas the energy level is sensed as absolute concentration of ATP and AMP, respectively [51]. Interestingly, the enzyme 6-phosphofructokinase does not show a classical regulation pattern, which would be activation by low energy metabolites and inhibition by high energy metabolites, but it is inhibited by ADP instead [10]. Gluconeogenesis, the antagonist pathway, is transcriptionally induced by selected carbon sources such as pyruvate, lactate, glutamate, and acetate [17, 18, 52]. Additional control is taken at the level of fructose 1,6-bisphosphatase (FBPase), which is strongly sensitive to metabolic regulation by AMP, PEP, and its own substrate [10]. During growth on glucose, *C. glutamicum* ATCC 13032 channels roughly 50% of carbon through the EMP pathway [53]. The relative contribution of the pathway to glucose degradation changes in response to cellular requirements. Systematic investigation of different lysine-producing strains reveals a reduced flux through the EMP pathway with increasing production performance [36, 53–58]. Interestingly, when the carbon source is fructose, the EMP pathway becomes the major catabolic route and carries more than 90% of the total flux [34]. This

relates to the entry point of fructose at the level of F16BP combined with a lack of in vivo FBPase activity, which forces the fructose carbon downstream into the EMP pathway [34]. Utilization of the glucose-fructose disaccharide sucrose results in an intermediate flux pattern of the two hexoses [35]. As mentioned above, this observation led to the identification of FBPase as a bottleneck for lysine production [34, 35] and corresponding overexpression of the encoding gene toward improvement [36]. Other EMP pathway-related engineering strategies toward improved lysine production comprise deletion of *pgi*, encoding phosphoglucoisomerase [59], deletion of *pyk*, encoding pyruvate kinase [60], and co-factor engineering of glyceraldehyde 3-phosphate dehydrogenase [61–63] for improved supply of NADPH. For products such as lactate and alanine, high glycolytic fluxes are favorable. Overexpression of glycolytic enzymes was hereby successfully applied for improving production [64–66]. Ornithine and arginine production also profited from overexpression of the glycolytic gene *pgk*, encoding phosphoglycerate kinase [67].

2.1.3 Pentose Phosphate Pathway

The pentose phosphate (PP) pathway represents an alternative glycolytic route in *C. glutamicum*. The oxidative part is comprised of glucose 6-phosphate (G6P) dehydrogenase (*zwf-opcA* genes), 6-phosphogluconolactonase (*devB* gene), and 6-phosphogluconate (6PG) dehydrogenase (*gnd* gene). It is most relevant for the supply of redox power [51, 68]. The regenerative or non-oxidative route provides building blocks and also recycles excess carbon back into the EMP pathway. It involves transketolase (*tkt* gene) and transaldolase (*tal* gene) activity [51]. Transcriptional regulation has been little studied, although GntR-like regulators have been discovered as repressors of the PP pathway genes *tkt*, *tal*, *zwf*, *opcA*, and *devB*, respectively [69]. Quantification of metabolite pools and elucidation of kinetic properties identified the enzymes G6P and 6PG dehydrogenase as major control points for carbon flux, mainly through sensing of the NADPH/NADP ratio [68]. The PP pathway is crucial for amino acid overproduction. As an example, increased lysine production requires an increased flux into the pathway [58, 70]. This observation stimulated the design of strains with increased PP pathway flux to improve lysine production. Successful examples demonstrate overexpression and modification of *zwf*, encoding G6P dehydrogenase [54], implementation of a point mutation into the *gnd* gene [71], deletion of *pgi*, encoding phosphoglucoisomerase [59], start of codon engineering of *zwf* and *pgi* [72], and overexpression of the full *tkt*-operon [55]. The findings from flux analysis during growth on fructose [34] and sucrose [35] further unraveled FBPase as an additional target to enhance PP pathway flux, the overexpression of which is beneficial for lysine production [36]. Similar engineering strategies were applied to improve other NADPH-demanding production processes including those of diaminopentane [47, 73], L-isoleucine [74], L-valine [75], L-arginine [76], and L-ornithine [77], underlining the high importance of the PP pathway for biotechnological production

in *C. glutamicum*. Beyond its essential role for NADPH supply, the PP pathway provides carbon building blocks for, for example, biosynthesis of aromatic compounds. The reactions of transketolase and transaldolase, comprising the non-oxidative PP pathway, were of high importance. Overexpression of the *tkt* gene was successfully applied for targeted improvement of the production of L-phenylalanine [78] and L-tryptophan [79].

2.1.4 Tricarboxylic Acid Cycle and Glyoxylate Shunt

The tricarboxylic acid (TCA) cycle is a key metabolic pathway of aerobic microorganisms such as *C. glutamicum*. It supplies biosynthetic precursors and energy: ATP (or GTP), and NADH and FADH for subsequent ATP generation via the respiratory chain and ATP synthase. Several nodes are under sophisticated metabolic and transcriptional control to modulate the carbon flux through the TCA cycle. The flux partitioning between the TCA cycle and the glyoxylate (Glx) shunt is adjusted at the level of isocitrate through metabolic control of the TCA cycle enzyme isocitrate dehydrogenase [10, 80], whereas isocitrate lyase, the entry enzyme into the Glx shunt, is controlled on the transcriptional level [17, 81] and inhibited by several metabolites including 3-phosphoglycerate, 6-phospogluconate, PEP, F16BP, succinate, and glyoxylate [82]. Further control of the TCA cycle occurs at the level of the 2-oxoglutarate dehydrogenase complex (ODHC) [83]. ODHC is activated and inhibited by several effector molecules [84] and also regulated by the ODHC repressor protein OdhI [85]. In the improvement of *C. glutamicum* for L-glutamate production, the alteration of control of ODHC has proven valuable [85–89]. In line with this, production of L-glutamate-derived γ-amino butyrate could be improved by deletion of *odhA* [90]. For other added-value products, such as L-lysine and its daughter product diaminopentane, the TCA cycle displays a competing pathway. Here, approaches for improving production efficiency intentionally reduced the flux through the TCA cycle, whereby citrate synthase [91] and isocitrate dehydrogenase [47, 92] were selected as engineering targets. An innovative strategy coupled lysine formation to the TCA cycle flux through the elimination of succinyl-CoA synthase [93]. Engineering of itaconic acid overproduction considered targeted downregulation of isocitrate dehydrogenase [94] by using rare translational start codons [72, 92]. Combined with the deletion of malate synthase, this strategy was similarly applied for glycolate production [95].

2.1.5 Pyruvate Metabolism

Pyruvate and PEP represent a central switch-point in metabolism. They function as highly connected hubs between the EMP pathway and the TCA cycle, take part in PTS-dependent substrate uptake, are the starting point for overflow metabolism, and serve as building blocks for anabolism. *C. glutamicum* possesses a rich

enzymatic set around the PEP-pyruvate node: pyruvate carboxylase (PCx), PEP carboxylase (PEPCx), pyruvate kinase (PK), pyruvate dehydrogenase (PDHC), pyruvate:quinone oxidoreductase (PQO), PEP carboxykinase (PEPCk), malic enzyme (MalE), and a putative oxaloacetate decarboxylase (Odx) [96, 97]. Related to the concerted action of multiple carboxylation and decarboxylation reactions in vivo, the metabolism of *C. glutamicum* is highly flexible, important in order to respond rapidly to altering conditions [97, 98]. Metabolic regulation of the different enzymes seems significant. PCx is the major anaplerotic enzyme, contributing to 90% of total flux in vivo, although in vitro activity of PEPCx is substantially higher [98–100]. The relevance of this metabolic switch-point entailed substantial engineering strategies toward production of different industrial goods (Table 1).

2.2 Anabolism

Cells of *C. glutamicum* are mainly composed of five macromolecules, namely protein, DNA, RNA, lipids, and cell wall carbohydrates (Fig. 2). As do almost all *Corynebacterium* species, *C. glutamicum* exhibits a complex cell wall architecture: a peptidoglycan layer covers the plasma membrane, which itself is bound to arabinogalactan, a complex hetero-polysaccharide meshwork [122]. The plasma membrane mainly contains oleic acid (18:1) and palmitic acid (16:1) [123]. The cell wall is rather unique, as it contains diaminopimelic acid and an outer membrane with mycolic acids [124]. The anabolic pathways in *C. glutamicum* are well-established. Biomass building blocks, such as amino acids, nucleotides, fatty acids, and carbohydrates, are synthesized from glucose 6-phosphate, fructose 6-phosphate, ribose 5-phosphate, erythrose 4-phosphate, glyceraldehyde 3-phosphate, 3-phosphoglycerate, pyruvate, phosphoenolpyruvate, acetyl-CoA, 2-oxoglutarate, succinyl-CoA, and oxaloacetate [58], and they are then assembled to the corresponding macromolecules and also occur as free intracellular pools [58, 93, 121]. Beside carbon precursors, anabolism also demands reducing equivalents and energy. Because of the huge interest in *C. glutamicum*, its cellular composition has been precisely determined and the specific demand for certain precursors as well as redox power and energy is well known [58] and can be used, for example, to infer metabolic fluxes [125, 126]. The synthesis of 1 g cell dry mass requires about 16,400 μmol of NADPH [58]. Interestingly, the ATP demand of 6,779 μmol g^{-1} is mainly because of polymerization of cell protein, but not precursor biosynthesis.

2.3 Regulation and Control of Metabolism

In the past few years, transcriptional regulation of *C. glutamicum* has been studied extensively, largely driven by powerful whole-genome transcriptome profiling

Table 1 Metabolic engineering of pyruvate metabolism in *C. glutamicum* toward improved production of industrially relevant goods

Enzyme	Modification	Product	Effect	References
Pyruvate dehydrogenase	Δ*aceE*	L-Valine	+	[101]
	aceE^{A1G}	L-Lysine	+	[102]
	aceE A16[a]	Isobutanol	−	[103]
		L-Lysine	+	[72]
		L-Valine	+	[104]
		L-Lysine	+	[104]
		2-Oxoisovalerate	+	[104]
Pyruvate kinase	Δ*pyk*	L-Glutamate	+	[105]
		L-Lysine	±	[106]
		L-Lysine	−	[60]
Pyruvate carboxylase	pEKE3x-*pyc*	Putrescine	+	[107]
	pyc^{P485S} [b]	L-Glutamate	+	[100]
	P_{sod}*pyc*^{P485S}	L-Threonine	+	[100]
	Δ*pyc*	L-Lysine	+	[100]
		L-Lysine	+	[108]
		L-Lysine	+	[55]
		L-Glutamate	+	[109]
		Succinate	±	[110]
		Lactate	±	[110]
		Isobutanol	+	[103]
PEP carboxylase	Δ*ppc*	Ethanol	+	[111]
	ppc_{mut}[c]	Succinate	−	[110]
	pAJ43-*ppc*	L-Glutamate	−	[109]
	pECt-*ppc*	Isobutanol	±	[103]
	ppc^{N917G} [d]	L-Tryptophane	+	[112]
	ppc^{D299N} [d]	L-Threonine	+	[113]
		L-Proline	+	[113]
		L-Glutamate	+	[109]
		L-Lysine	+	[114]
		L-Glutamate	+	[115]
PEP carboxykinase	Δ*pck*	L-Lysine	+	[116, 117]
	pEK-*pck*	L-Lysine	−	[116]
Malic enzyme	pVWEx1-*malE*	L-Lysine	±	[118]
	Δ*malE*	Isobutanol	−	[119]

The individual effects of genetic changes on production are given as follows: improved (+), decreased (−), not changed (±)

[a]Replacement of the native promoter by *dapA* promoter variant A16 [120]

[b]Pyruvate carboxylase variant with reduced sensitivity for inhibition

[c]Mutated PEP carboxylase isolate with 75% reduced activity

[d]PEP carboxylase variant with reduced sensitivity for inhibition

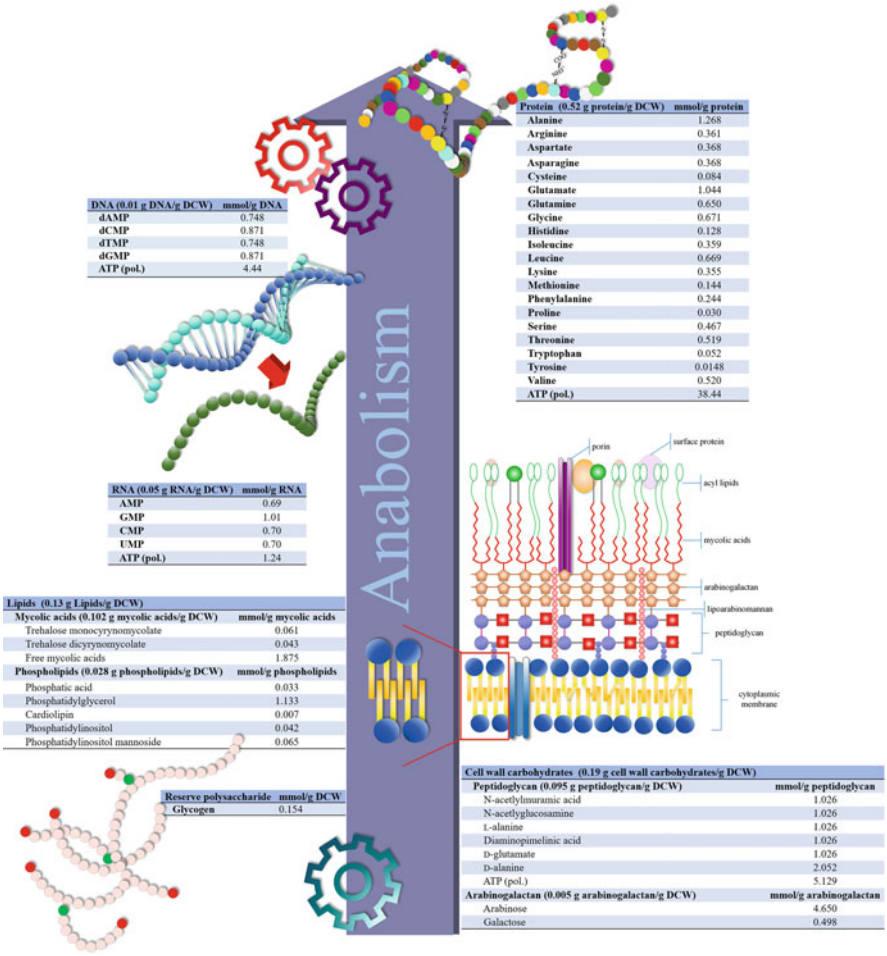

Fig. 2 Cellular composition of *C. glutamicum*. The anabolic demand for synthesis of the macromolecules was taken from previous work [58, 121]

technologies such as DNA microarrays and RNA sequencing. This has unraveled a highly complex transcriptional regulatory network (TRN), consisting of about 160 interacting genes for DNA-binding transcription regulators, various sigma (σ) factors, and additional regulator proteins [127]. Today, the TRN of *C. glutamicum* is available from interactive databases, such as the web-based platform "CoryneRegNet," which are continuously updated with novel findings [128]. Knowledge of pathway and expression regulation has substantially supported metabolic engineering. Here we give an overview of the general features complemented with applications in the field of metabolic engineering.

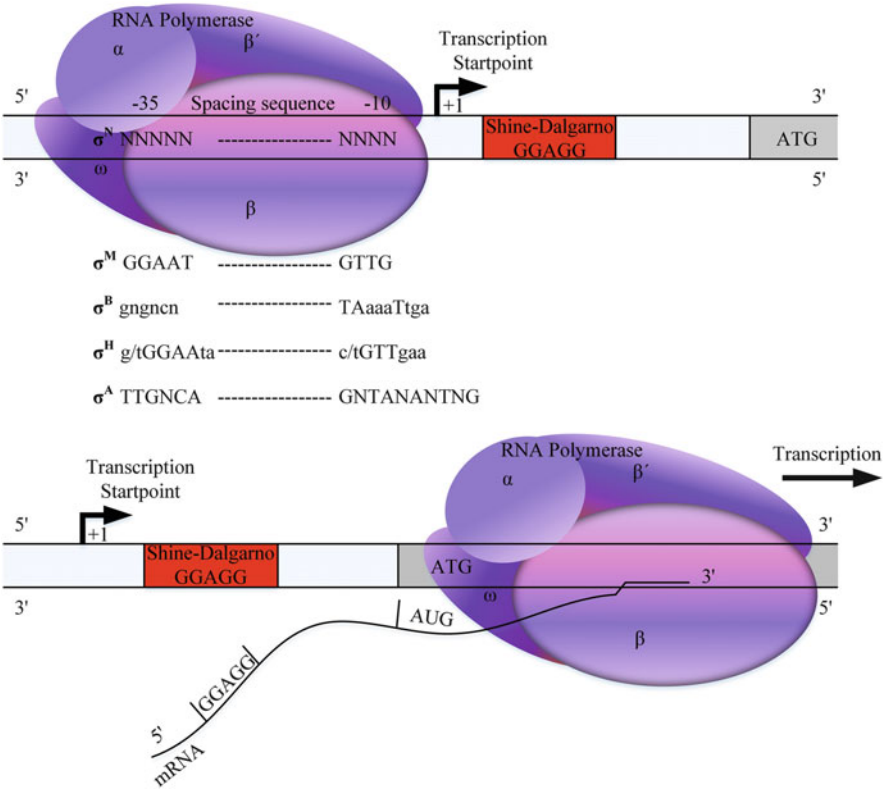

Fig. 3 Molecular structure of the transcriptional initiation site in *C. glutamicum*, mediating RNA polymerase binding and start of transcription with α, β, β´, and ω indicating the RNA polymerase subunits. Consensus sequences at −35 and −10 upstream of the transcription start point are given for the sigma factors σ^A, σ^B, σ^M, and σ^H. *Upper case letters* indicate sequence conservation of over 80%, *lower case letters* of over 40% [129]

2.3.1 Sigma Factors

Promoter sequences upstream of structural genes display the core piece for the regulation of gene expression in *C. glutamicum*. In short, the promoter region usually consists of 40–50 base pairs and binds the transcription machinery, the RNA polymerase complex. Hereby, a regulator protein, called sigma (σ) factor, interacts with the partially melted double-stranded DNA and affects the binding capacity of the complex and transcription initiation. In addition, promoter consensus sequences, 35 and 10 nucleotides upstream, of the transcriptional start point at position +1 (Fig. 3) and the spacing sequence between the consensus sequences, influence the initiation of transcription [130, 131]. *C. glutamicum* possesses different sigma factors, which all belong to the σ^{70} family. Cells use varied expression of the individual sigma factors to regulate their gene expression, amongst other mechanisms. Overall, seven sigma factors – σ^A, σ^B, σ^C, σ^D, σ^E, σ^H, and σ^M –

have been discovered in *C. glutamicum* [130, 132, 133]. The expression of housekeeping genes is mainly controlled by σ^A, whereas stress related expression is primarily under the control of σ^B, σ^H, and σ^M [134–136]. Mutants lacking σ^B are more sensitive to heat, cold, salt, acid, and alcohol stress [137]. If oxygen supply is limited, σ^B positively promotes genes of glucose uptake and several genes of the EMP pathway and the TCA cycle [138].

2.3.2 Transcriptional Regulators

The basic level of the TRN consists of local regulators, each of which control a small subset of only a few genes, related to a rather specific function. An example is fructose assimilation, regulated by the local regulator FruR that controls the expression of fructose specific genes [5, 139, 140]. Such functionally related genes are usually located in an operon- or divergon-like structure, the latter being a pair of divergently transcribed operons. Master regulators with superimposed function take a higher level of control within the TRN. They orchestrate complex cellular programs, related to carbon metabolism (e.g., RamAB, SugR, SigB), nitrogen metabolism (AmtR), phosphor metabolism (PhoR), sulfur (e.g., McbR, CysR) and iron (DtxR) homeostasis, respiration/anaerobiosis (ArnR), and stress responses (e.g., LexA, SigH) for cell survival [141, 142]. As an example, AmtR inhibits transcription of *amtA* (amino-methyl transferase), *amtB* (ammonium transporter), *glnA* (glutamine synthetase), *gltBD* (glutamate synthase), and *dapD* (tetrahydrodipicolinate succinylase). It also controls linked pathways of creatinine and urea metabolism: *codA* (creatinine deaminase), *crnT* (creatinine transporter), *urtABCDE* (ABC-type urea transporter), and *ureABCEFGD* (urease). Hereby, AmtR activity itself senses the ammonium level, which is mediated by a signal cascade of UTase (uridylyltransferase) and the regulatory protein GlnK [143, 144]. Deletion of AmtR results in deregulation of the ammonium uptake system in *C. glutamicum* [145]. Detailed knowledge of nitrogen metabolism appears especially valuable for amino acid and diamine production processes. The ammonium level thereby not only defines the assimilation route and thus the "energetic cost" of uptake [146] but might also influence pathway usage as demonstrated for lysine production [146, 147]. These findings guided metabolic design and engineering of *C. glutamicum* for the production of L-lysine [55], diaminopentane [73], and ectoine [148].

The McbR regulator has been at the focus of researchers as it takes substantial control in the biosynthesis of the feed amino acid methionine [149]. Its deletion entailed oxidative stress [150] and imbalances in the metabolism of sulfur-containing amino acids, resulting in the accumulation of pathway intermediates and the activation of normally silent pathways [151–153].

The major response regulator in *C. glutamicum* for heat and oxidative stress is SigH (σ^H) [7], a sigma factor responsible for the transcription of *sigA*, *sigM*, and *sigB* [7, 154]. When cells are exposed to heat stress, σ^H activates transcription of *clpC* (Clp ATPase subunit), *clpP1P2* (Clp protease subunits), and *clgR* (ClgR,

positive regulator of *clpP1P2*). In parallel, σ^H controls ClgR via modulation of stability and transcription [155, 156]. During heat shock response, σ^H additionally controls the expression of other regulators: ClgR, SufR, WhcA, and WhcE [157]. Depending on the imposed temperature, the regulatory system shows an intensity-dependent response for HrcA/CIRCE regulated genes, but not for genes regulated by HspR/HAIR [158]. The heat stress response cascade results in the activation of molecular chaperones which stabilize the cellular proteins. During exposition to oxidative stress, σ^H activates the expression of *whcE* and *whcA* with *whcE* being a repressor of *whcA* and *whcB* under normal growth conditions [159]. As phenotypic results, increased expression of *whcB* and downregulation or deletion of *whcE* improves growth. Besides heat shock proteins, molecular chaperones and ATP-dependent proteases are upregulated [7]. Upon stress, *C. glutamicum* also changes the expression of genes of core carbon metabolism [160]. At increased temperature, citrate synthase gene *gltA* is expressed less, whereas the expression of *malE*, encoding malic enzyme, is increased [161]. These natural metabolic responses can be harnessed for bio-based production. Examples include improved lysine [161] and ectoine [148] production at higher temperature. Under hyperosmotic stress, most substrate is used for ATP generation and is directed toward glycolysis and TCA to satisfy higher demand for maintenance [162].

The different modules of the TRN in *C. glutamicum* are not fully autonomous, but are interconnected via regulators that function as a kind of interface [142, 163]. Negative autoregulation of the master regulators enables a fine-tuned gene expression together with a rapid response to imposed environmental changes [164, 165]. Regarding global regulators at the very top of the TRN cascade, the only protein identified so far in *C. glutamicum* is GlxR (Fig. 4). It mediates the cellular response to altering levels of the signal molecule cAMP [142, 167, 168]. The numerous GlxR-specific DNA-binding sites in the genome of *C. glutamicum* suggest global regulation of up to 14% of *C. glutamicum* genes and transcription regulators by this global regulator.

2.3.3 Small RNAs

Small RNAs are short, non-coding RNA molecules which control the stability and the translation efficiency of mRNA. These regulatory elements dynamically change gene expression. Small RNAs seem to be the most abundant posttranscriptional regulators in bacterial cells. Transcriptome sequencing revealed a variety of these regulatory elements to be present in *C. glutamicum* [169]. For higher organisms, such as fission yeast and the multicellular model *Caenorhabditis elegans*, the role of small RNAs has been extensively investigated [170]. For *C. glutamicum*, however, we know less at present, but first insights into transcriptional regulation via small regulatory RNAs reveal that they may also play an important role in this bacterium [171]. Analysis of RNA-seq data from the sRNA cDNA library of *C. glutamicum* verified short transcripts in the known transcriptional attenuators

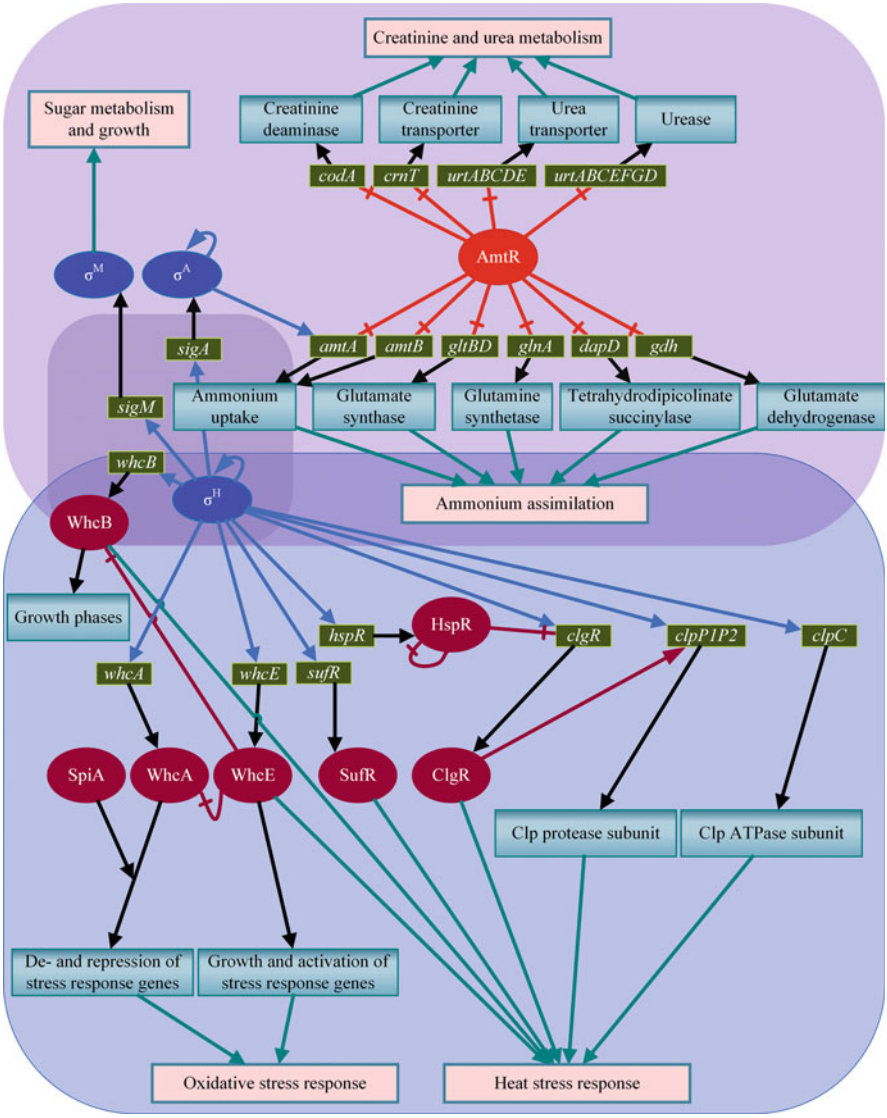

Fig. 4 Transcriptional regulatory network of *C. glutamicum*, involving the σ factors σ^A, σ^M, and σ^H and the master regulators AmtR, WchB, and HspR, the latter being involved in regulation of ammonium assimilation, oxidative stress, and heat shock response, respectively. Repression is indicated by *T-bars*, activation is given by *arrows*. AmtR regulatory protein for ammonium assimilation, *amtA* amino-methyltransferase, *amtB* ammonium transporter, *glnA* glutamine synthetase, *gltBD* glutamate synthase, *dapD* tetrahydrodipicolinate succinylase, *codA* creatinine deaminase, *crnT* creatinine transporter, *urtABCDE* ABC-type urea transporter, *ureABCEFGD* urease, *gdh* glutamine synthetase expressing gene, *clpC*Clp ATPase subunit, *clpP1P2* Clp protease subunits, *clgR* positive regulator of *clpP1P2, sigA*, *sigM*, *sigH* respective σ-factors, *whcB* WhcB regulator of growth phase transition, *whcA* WhcA regulator with SpiA regulation of stress response genes, *whcE* WhcE regulator of growth and activation of stress response, *hspR* HspR repressor of *clgR* transcription, *ClgR* regulation of *clpP1P2* and heat stress response, *sufR* SufR regulator of heat stress response [7, 133, 143–145, 155, 157–159, 166]

sites of the *trp* operon, the *ilvBNC* operon, and the *leuA* gene [172]. Further elucidation promises advances for the production of related products such as aromatic and branched-chain amino acids and biofuels.

3 Molecular Tools for Genetic Engineering

When approaching systems and synthetic metabolic engineering, the availability of molecular tools of DNA manipulation is essential. First steps toward genetic engineering of *C. glutamicum* were initiated in the 1980s by the discovery and isolation of natural plasmids [173, 174] and the invention of DNA transfer methods [175]. Meanwhile, episomal and genome-based DNA manipulation are routine techniques because of the availability of the genome sequence [176, 177], the development and advancement of episomal [129, 178] and integrative plasmids [179, 180], optimized transformation methods [181, 182], and cloning and expression procedures [183, 184].

3.1 Plasmids

For genetic manipulation of *C. glutamicum*, different types of plasmids have been developed. Autonomously replicating plasmids are mainly based on the naturally occurring cryptic set of *C. glutamicum* plasmids [129]. For amplification, maintenance, and propagation, plasmids are designed as *C. glutamicum*/*E. coli* shuttle vectors and are equipped with selection markers conferring antibiotic resistance [129, 184, 185]. This enables application as cloning, promoter probe, and expression vectors [10]. Modification of the chromosome of *C. glutamicum* became possible by the application of the DNA vectors lacking replicon elements. Chromosomal integration is commonly permitted via homologous recombination [186, 187] through site-specific insertion sequences of IS-elements and phage sequences [185, 186, 188–190]. Discovery and application of the *Bacillus subtilis* levansucrase (*sacB*) as counter-selectable marker was a major breakthrough for genome-based manipulation of *C. glutamicum* [179]. This conditionally lethal marker system displays the most convenient system genetic engineering of *C. glutamicum* [36, 101, 148, 187, 191, 192]. The survival rate of plasmids, subsequent to transformation, remains a critical factor as it directly correlates to the success rate of genetic manipulation. In this regard, circumvention of the natural defense system of *C. glutamicum* to degrade foreign DNA enzymatically substantially improved genetic engineering [10]. Successful strategies include exposure to heat, solvent, or pH stress [193–195], plasmid transfer through *C. glutamicum* compatible hosts [196], recruitment of intermediate cloning host for adding the *C. glutamicum* specific DNA-methylation pattern [197], or the use of synthetic [198, 199] and non-methylated DNA [182, 200].

3.2 Promoters and Codon Usage

The σ-factor binding region influences the strength of a promoter, because the nucleotide sequence defines the binding efficiency of the RNA polymerase and thus the efficiency of transcription initiation. Consequently, promoter modification bears plenty of optimization possibilities for metabolic engineering of *C. glutamicum*. In this regard, the constitutive and strong promoter sequence of superoxide dismutase (Sod), elongation factor TU (Eftu), and the chaperone GroEL have proved to be valuable for targeted increase of gene expression in *C. glutamicum* [14, 36, 48, 55, 76, 77, 197]. Beyond the natural set of promoters, synthetic promoter libraries have been developed through site-directed mutagenesis and randomization of promoter length [120, 201, 202]. The variety of weak and strong promoters obtained confers higher flexibility for gradually decreasing or increasing gene expression and thus fine-tuning of enzyme and pathway activities. Successful applications include the production of L-valine [201] and L-lysine [91] and high-level expression of endo-xylanase [203]. The set of synthetic promoters also includes inducible promoters relying on IPTG [202, 203] or on carbon sources such as gluconate and maltose, allowing substrate-dependent pathway modification [204]. With the advent of synthetic metabolic engineering, the recruitment of genes from heterologous donor strains became more and more convenient. This was often hampered by the different genetic peculiarity of donor and host regarding GC content and codon usage. Here, substantial benefit was achieved by using synthetic genes which were codon-optimized for *C. glutamicum* as being done for the *E. coli*-derived lysine decarboxylase gene *ldcC* to improve diaminopentane production [197].

3.3 Small RNAs

Small RNAs dynamically control gene expression, which is attractive for strain engineering. *Trans*-encoded sRNA (Fig. 5c–e), transcribed from regions separate from their target genes, have the ability to inhibit and to promote translation of the target mRNA. Pairing with the 5′UTR and the ribosome-binding site blocks translation. The formation of sRNA–mRNA complexes leads to degradation by RNAses. Activation of translation is triggered if *trans*-encoded sRNAs prevent the formation of inhibitory structures around the ribosome-binding site [205–208]. A *cis*-antisense sRNA (Fig. 5a, b), transcribed from the opposite strand of the target DNA shows high complementarity and acts rather specific in three different ways. By binding to the ribosome-binding site of the target mRNA, translation is inhibited and RNA degradation is activated. A *cis*-antisense sRNA that binds in the intergenic region between two genes of an operon can cause the cleavage into two mRNAs. Such sRNAs function as transcriptional terminators as well [205, 209–211]. The length of sRNA molecules varies between 50 and 300 nt [169, 212]. Most small RNAs discovered in *C. glutamicum* are structured antisense,

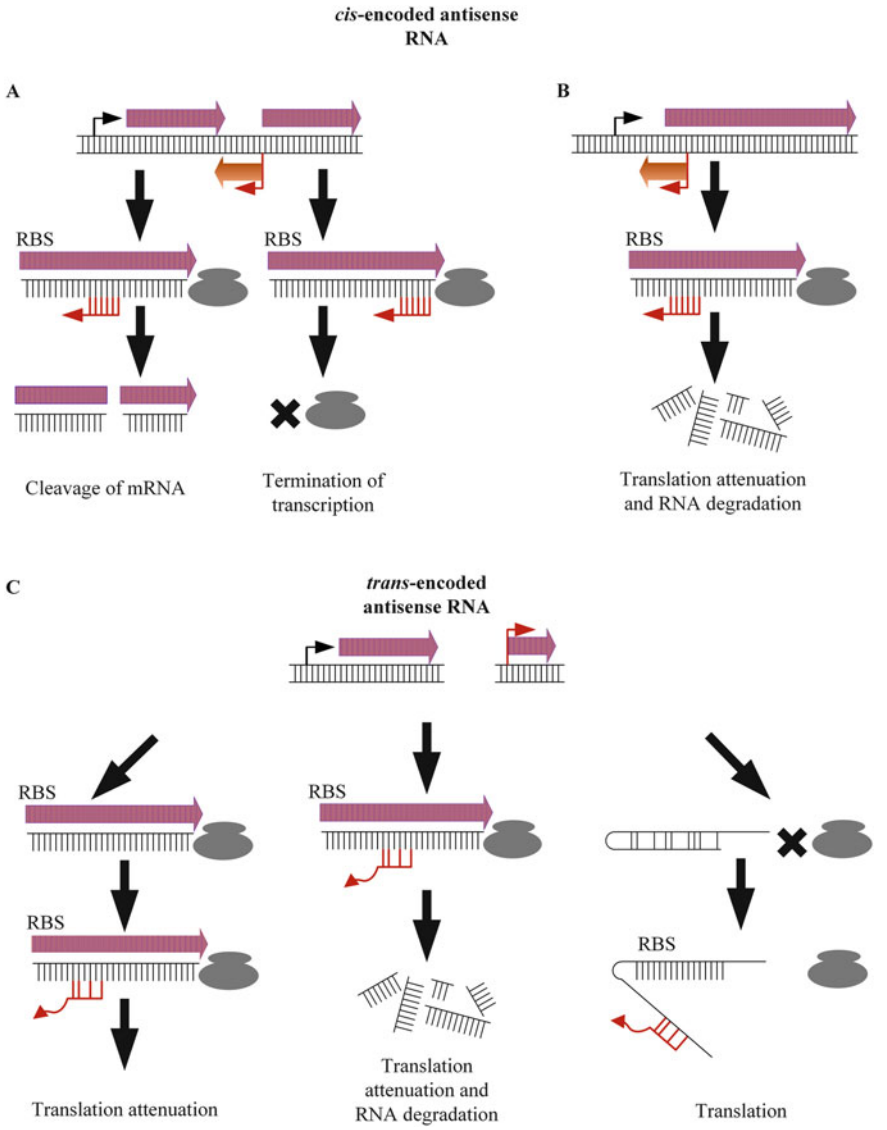

Fig. 5 Transcriptional and translational regulation via *cis*- and *trans*-encoded antisense RNA. *Cis*-encoded antisense RNA attenuates translation and induces mRNA degradation. *Trans*-encoded antisense RNA are transcribed from a gene sequence distant from its target mRNA. (**a**) *Cis*-encoded antisense RNA (*orange*) is transcribed between two genes (*green*) and is highly complementary to the target mRNA. The sRNA (*red*) is able to function as transcriptional terminator or cleaves the mRNA into two mRNA fragments in order to alter translation. (**b**) Pairing with bases near the ribosome binding site. (**c**) Imperfect base pairing, translation of mRNA is suppressed. (**d**) RNAse degradation can be triggered by binding of the *trans*-antisense small RNA. (**e**) The translation is initiated by inhibiting the formation of RBS-blocking structures (adapted from [205])

whereas *cis*-antisense, *trans*-encoded RNAs, and other regulatory elements, such as riboswitches, have been found as well. The exact functions of most of these elements have not been completely understood so far. The non-coding 6S RNA is non-existent for *C. glutamicum*, unlike most bacteria [169, 213, 214]. Recently, bioinformatics tools have been developed to predict translation initiation rate of mRNA regulated by small RNAs. Such approaches deliver insight into the huge potential of directed regulation with small RNAs for metabolic engineering [215]. For *Escherichia coli*, the regulatory function of short transcripts has been exploited for metabolic engineering to silence genes and fine-tune gene expression. Computational approaches were used to predict targets and binding efficiency [216–219]. In contrast to *E. coli*, *C. glutamicum* lacks a protein similar to the RNA chaperone *hfq*, which was used as supporting protein for metabolic engineering in *E. coli* [169, 217, 220].

3.4 Riboswitches

Riboswitches are sequences at the 5′ end of mRNAs. Their conformation can vary, depending on cellular conditions. Riboswitches consist of two units, the aptamer and the expression platform, respectively. The aptamer displays the binding part of the riboswitch. By direct binding of a specific metabolite as a ligand, the structure of the expression platform is modulated, resulting in either transcription termination or initiation, for example by forming a hairpin structure [221–224]. Recent studies report the modulation of *C. glutamicum* metabolism by *Bacillus subtilis* and *E. coli* lysine riboswitches. The constructed strains carried lysine riboswitch between the promoter and the start codon of the *gltA* gene, encoding citrate synthase. Lysine binds as ligand to the riboswitch and promotes transcription termination. In this way, the TCA cycle activity was down-regulated [221]. The lysine "Off" riboswitch of *Escherichia coli* was further engineered to function as the lysine "On" riboswitch in *C. glutamicum* by randomizing the genetic sequence between the aptamer and the ribosome-binding site, which caused an upregulation of target genes [225].

3.5 Translation Efficiency

The translation initiation rate (TIR) is a key parameter in promoter-regulated gene expression [131]. Translation usually starts by binding of the ribosome to the Shine–Dalgarno sequence at the 5′-end of the mRNA. Different Shine–Dalgarno regions are highly conserved and differ only slightly from the consensus sequence GGAGG. Recognition occurs by a complementary anti-Shine–Dalgarno sequence, contained in the ribosomal 16S rRNA. The ribosomal binding site has a strong impact on the protein expression level and is consequently a valuable molecular

tool for strain engineering [226, 227]. In addition, the translational start codon as well as the genetic context and the mRNA structure are crucial factors for translation initiation [135, 228]. Start codon design is straightforward for the metabolic engineering of *C. glutamicum*, whereby the most abundant start codon ATG results in higher expression levels, as compared to the rare variants GTG and TTG [72, 92, 229]. The translational efficiency is also influenced by the 5'-untranslated region of the mRNA (UTR). Modification of the UTR is highly attractive for rational strain engineering, as translation efficiency can alternatively be reduced or increased to attenuate competing and to stimulate supporting pathways [230–232]. The mode of action is a change of mRNA conformation that influences transcript stability via sensitivity for endonuclease cleavage [233]. In *C. glutamicum*, the effect of UTR manipulation was nicely demonstrated for GFP expression [203, 226]. Beyond modification of the expression strength, appropriate secretory signal peptides can be used to navigate proteins to the secretory apparatus for efficient protein secretion [226]. With all these molecular engineering systems to hand, today's researcher are well-equipped for targeted engineering of the *C. glutamicum* metabolism to improve already existing production processes or to establish new ones.

4 Advanced Strain Engineering for Industrial Bioproduction

The power of system metabolic engineering strategies together with a strong need for bio-based and eco-efficient production of chemicals, materials, and fuels are major drivers for the development of novel production processes from renewables [103, 197, 234, 235]. For *C. glutamicum* an extensive product spectrum is meanwhile obtainable through fermentation of diverse biomass components – some fully established at the market, some still at the periphery (Fig. 6).

The early years of strain development were strongly involved with random mutagenesis and selection, which, on the one hand allowed fast progress and improvement without a detailed knowledge of metabolism and physiology and, on the other hand, entailed a vast metabolic burden, manifested as growth deficiency, auxotrophy, and low robustness and vitality [3]. With ever-increasing knowledge of metabolic pathways and the development of molecular tools for DNA manipulation, classical methods were complemented by targeted pathway design and engineering for local manipulation of production-relevant key enzymes and reactions [36, 54, 72, 79, 86, 92, 100, 118, 242, 243]. However, a systems perspective was still missing until the establishment of advanced tools for global analysis of the cell on the various layers, namely genome, transcriptome, proteome, metabolome, and fluxome [244]. This enabled the next level of rational strain engineering – systems and synthetic metabolic engineering [12, 245, 246]. Systems biology data combined with computational platforms now provided an expert basis for global conceptual design and substantially promoted cellular engineering. This

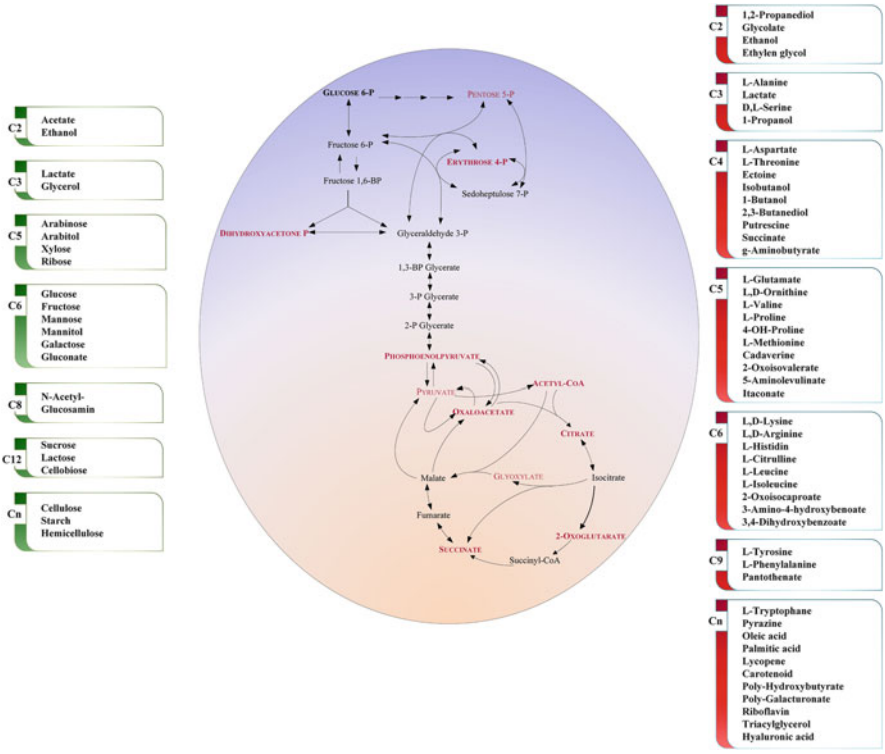

Fig. 6 Product portfolio of metabolically engineered *C. glutamicum* through fermentation from renewable feedstocks [3, 9, 12, 94, 137, 236–241]. Products comprise commercial goods for diverse application sectors including feed and food, health and hygiene, energy and transportation, textiles, packaging and housing, and agricultural and technical application

was a milestone toward generating tailored cell factories of *C. glutamicum* for production of L-lysine [55], diaminopentane [73], and L-arginine [76]. When targeting defined metabolic features, enzymes and biosynthetic pathways, that is, systems metabolic engineering with all its aspects, is most valuable for defining the best strategy. Proceeding further to multi-target tolerance issues involving more complex mechanisms, this approach becomes limiting and was hence recently complemented by evolutionary engineering [245]. Grown in a stress-imposed environment, *C. glutamicum* was successively evolved to become more tolerant to oxidative [247], thermal [248], solvent [248] and methanol [249] stress, followed by systems biology analysis for unravelling the underlying cellular features which are conferring tolerance. Combining evolution with biosensor-coupled product detection [250] or targeted metabolic re-engineering promises strains with improved production performance and greater robustness.

4.1 Human and Animal Health and Nutrition

Products for health and nutrition have the longest history in biotechnology, with *C. glutamicum* being one of the major producers [9, 12, 251, 252]. Among all products obtainable by *C. glutamicum*, L-amino acids have the longest tradition and hold the largest market share. Meanwhile, processes for other products including non-proteinogenic amino acids [253–255], vitamins [236, 256], flavors and fragrances [257], and other nutrients and health care products [148, 258, 259] are also on the rise.

4.1.1 L-Lysine

Large-scale production for the feed amino acid L-lysine was established in the 1950s, using mutants from iterative rounds of random mutagenesis. Almost exclusively produced by *C. glutamicum* [260], L-lysine belongs to the world's top-selling amino acids with an annual production volume of around 2.5 million tons per year. Throughout decades of research and development, pioneering discoveries disclosed a set of genetic targets for local metabolic engineering. The modifications can be categorized according to their generic function for biosynthesis [242, 261] and export [262], supply of carbon building blocks [100, 108, 113, 116] and redox power [36, 54, 72], and competing reactions [91, 92, 102, 108]. Despite a given benefit for production, local engineering approaches failed to generate producer strains for competitive industrial application, which, however, was more recently overcome by systems metabolic engineering. With a titer of 120 g/L, a yield of 55%, and productivity of 4.0 g/L/h, respectively, the current benchmark of L-lysine production has been achieved by comprehensive systems metabolic engineering. Based on comparative in vivo and in silico flux studies, 12 genetic traits were predicted to upgrade a non-producing wild type into a tailored hyper-producer [55]. Beyond the obvious benefit of this design-based strategy for industrial lysine production, this concept is highly promising for promoting strain and process development to bring novel products on market.

4.1.2 L-Arginine and L-Ornithine

L-Arginine and L-ornithine are intermediates of the urea cycle and thus metabolically closely related. L-Arginine is a semi-essential amino acid with both anabolic and regulatory function as one of the proteinogenic amino acids and has a pronounced vasodilatory effect. To L-ornithine, a positive effect with regard to treatment of liver diseases and strengthening of the heart is ascribed [263]. Both amino acids are natural products of *C. glutamicum* and first reports on fermentative production date back to the early years of amino acid fermentation [264–267]. As compared to L-lysine, the regulation of biosynthesis is much more complex and

involves feedback inhibition of *N*-acetyl-glutamate kinase (ArgB) by arginine, and transcriptional control imposed by the arginine repressor ArgR [234, 268, 269]. The strategy for generating a genetically defined arginine producer thus involved genome breeding, that is, a comparative sequence analysis of wild type and classical producers, to identify potential regulatory mutations [108, 234]. Another strategy, combining systems metabolic engineering and mutagenesis recently bore a strain capable of producing 93 g L^{-1} arginine with a yield of 0.4 g g^{-1} glucose [76]. Random mutagenesis thereby aimed at increased tolerance toward L-arginine analogues, being equivalent to addressing pathway regulation. This was supported by additional removal of repressors of the arginine operon. Further systems metabolic engineering involved optimization of NADPH levels through promoter and start codon engineering, disruption of L-glutamate exporter to increase L-arginine precursor, and flux optimization of rate-limiting L-arginine biosynthetic reactions [76].

As with the L-arginine strategy, removal of pathway regulation was a key issue for L-ornithine production [77, 263, 270]. Starting from basic producers, alternative engineering strategies generated strains with good production performance. One strategy thereby combined evolutionary engineering with subsequent transcriptional profiling. This revealed the upregulation of *pgi* (encoding glucose-6-phosphate isomerase), *pfkA* (encoding 6-phosphofructokinase), *gap* (encoding glyceraldehyde-3-phosphate dehydrogenase), *pyk* (encoding pyruvate kinase), *pyc* (encoding pyruvate carboxylase), *gltA* (encoding citrate synthase), *gdh* (encoding glutamate dehydrogenase), *argB* (encoding acetylglutamate kinase), and *argJ* (encoding the bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase) in sum tunneling carbon from glucose via glutamate toward ornithine [263]. Expressional changes of enzymes involved in redox metabolism pointed to the relevance of NADPH supply [263, 271]. Overall, the combination of metabolic and evolutionary engineering yielded a production of 24 g L^{-1} ornithine [263]. The final titer was more than doubled (51.5 g L^{-1}) by another study completely relying on rational design and engineering [77]. Subsequent to removal of pathway regulation and elimination of competing reactions, the biosynthetic gene cluster *argCJBD* was overexpressed on plasmid and efficient NADPH supply was assured through pentose phosphate pathway engineering [77].

4.1.3 Ectoine

In recent years, the pharmaceutical and cosmetics industry started to exploit the stabilizing and function-preserving effects of ectoines for health and hygiene products [272]. Related to the natural function of these chemical chaperones as protecting agents against high osmolarity or temperature, their biosynthesis in natural producers is, in general, a stress response [273–275]. Current production processes are accordingly dependent on provoking high salinity [272, 273, 276], which requires expensive process equipment because of corrosive effects. To overcome this drawback, lysine-producing *C. glutamicum* was genetically modified

for salinity-decoupled production of ectoines. This was achieved by genome-based integration of the codon-optimized *Pseudomonas stutzeri* gene cluster *ectABCD*, encoding 2,4-diaminobutyrate acetyltransferase (*ectA*), L-2,4-diaminobutyrate transaminase (*ectB*), ectoine synthase (*ectC*), and ectoine hydroxylase (*ectD*), respectively, in the *ddh* gene locus of *C. glutamicum*, whereby expressional control was taken by the strong and constitutive promoter of elongation factor Tu [148]. Elimination of by-product formation through deletion of the lysine exporter and subsequent bioprocess development allowed the production of 4.5 g L^{-1} ectoine with an estimable productivity of 6.7 g L^{-1} day^{-1} [148]. Proceeding further to systems-wide engineering and integration of evolutionary strategies (here previously described thermo-tolerance [248] appears most promising in light of the better production performance at increased temperature [148]) can certainly generate improved producers, making ectoine production with *C. glutamicum* more and more attractive.

4.1.4 Terpenoids

Many high-value products including the anti-cancer drug taxadiene, the anti-malaria drug artemisinin, and the colorful carotenoids belong to the substance class of terpenoids [12]. For all, the biosynthesis originates from the common pathway intermediate isopentenyl pyrophospate (IPP), which is further metabolized to the respective product. Metabolic engineering of *C. glutamicum* for terpenoid production so far mainly focused on carotenoids. Wild type strains already possess native gene clusters, which were modified by deletion of the *crtEb* gene, encoding lycopene elongase, and overexpression of *crtE*, *crtB*, and *crtI*, encoding prenyl transferase, phytoene synthase, and phytoene desaturase to establish lycopene overproduction with a yield of 2.4 mg/g cell dry mass [258]. Elongation of the pathway and introduction of glycosyltransferases from different donor strains allowed production of beta-carotene and zeaxanthin as well as glycosylated derivatives thereof [259]. Moreover, *C. glutamicum* was engineered to produce (+)-valencene. Production relied on heterologous expression of (+)-valencene synthase from the sweet orange *Citrus sinensis* or from *Nootka cypress*, whereby improved supply of the precursor farnesyl pyrophosphate (FPP) by additional overexpression of the FPP synthase from *E. coli* or *S. cerevisiae* was crucial [277].

4.2 Platform Chemicals and Materials

In our post-industrial and petrochemically coined era, it is hardly possible to imagine life without plastics. Driven by the need and desire for replacing fossil raw materials to achieve sustainability, bio-plastics are currently experiencing a renaissance [12, 268, 278, 279]. For *C. glutamicum* this includes fermentative supply of chemical building blocks such as succinate [110, 280, 281],

diaminopentane (cadaverine) [197, 282], diaminobutane (putrescine) [107, 235], lactate [280, 283], and propanediol [284, 285], as well as direct polymer production, mainly polyesters [286–288]. Here, we focus on engineering strategies for diaminopentane and succinate, representing building blocks for high-value polyamide production.

4.2.1 Succinate

The relevance and attractiveness of bio-succinate becomes obvious from the great effort of world-leading (bio)chemical companies including Myriant (with ThyssenKrupp Uhde), BioAmber (joint venture with Mitsui & Co.) and Succinity GmbH (joint venture of BASF SE & Corbion Purac) to establish industrial scale fermentation processes [12, 280]. Despite not yet in the focus for production, *C. glutamicum* has been engineered for high-level succinate production [289]. Key modifications comprise overexpression [290] and feedback deregulation [291] of pyruvate carboxylase to enhance anaplerotic carboxylation, which was additionally stimulated by increasing the CO_2 level via bicarbonate supplementation [110, 290]. Formation of by-products was eliminated through deletion of lactate [290, 291] and acetate [291] formation routes, and further improvement was achieved by overexpression of glyceraldehyde 3-phosphate dehydrogenase [291] and manipulation of redox metabolism [110, 291, 292]. Aerobic [293, 294] and micro-aerobic [295] production processes are currently evaluated, whereby elimination of by-product formation, disruption of the TCA cycle downstream of succinate, and overexpression of anaplerotic carboxylation seem most relevant for production [294]. Further improvement was achieved by acetate recycling and increased flux of the oxidative TCA cycle by amplified expression of citrate synthase [293]. The resulting process is biphasic and comprises an aerobic growth and an anaerobic production phase. The latter leads to titers up to 146 g L^{-1} [290] and yields even surpassing 1.0 g g^{-1} [291]. However, the additional time and raw material needed to provide the cells through aerobic growth reduces the overall performance significantly, and still below that of natural anaerobic producers [296, 297].

4.2.2 Diaminopentane

The high research interest in diaminopentane (DAP) is because the diamine serves as building block for polyamides, top-level industrial polymers with advanced material properties and a current market volume of several million tons per year [73, 278]. DAP is a naturally occurring degradation product of lysine, putting the spotlight for process development on the industrial lysine-producer *C. glutamicum* [197, 268, 282]. As *C. glutamicum* does not possess natural lysine-degradation pathways, DAP production in *C. glutamicum* was approached by heterologous expression of the CadA [282] and the LdcC [197] lysine decarboxylase variant of

E. coli. The host cell was pre-designed by feedback deregulation of lysine biosynthesis. Though both enzyme variants enabled production, the CadA strategy suffered from incomplete lysine conversion and seems inferior [282]. A further benefit for production was achieved by debottlenecking the carbon flux through the terminal biosynthesis, enhanced supply of the central building block oxaloacetate, and expression of a codon-optimized *ldcC* gene under control of the strong *tuf* promoter [197]. Metabolome analysis then revealed substantial secretion of N-acetylated-DAP as a major by-product. This involved the discovery and deletion of a so far unknown *trans*-acetylase, catalyzing the undesired cross-reaction with the non-natural metabolite DAP [298]. Additional engineering of product export [299], combined with metabolic manipulation on a systems biology level comprising attenuation of competing pathways and enhanced supply of redox power [73], created a streamlined cell factory that converts more than 40% of the consumed glucose into diaminopentane during exponential growth in batch culture [73]. - Fed-batch process implementation leveled production to a molar yield of 50% with a maximum titer of 88 g L^{-1} and a space-time yield of 2.2 g L^{-1} h^{-1} [73]. The generation of the advanced DAP producer represents a breakthrough for providing fully bio-based polyamides such as PA5.4 (copolymerization with succinic acid) and PA5.10 (copolymerization with sebacic acid from castor oil). The latter was successfully manufactured as polymer in pure and glass fiber reinforced form [73]. The excellent material properties of the novel bio-based PA5.10, surpassing that of conventional petrochemical nylons PA6.6 and PA6, are encouraging.

4.3 Biofuels

Though *C. glutamicum* is not a designated example for biofuels, some attempts have been made to establish biofuel production, mainly processes for ethanol [111] and isopropanol [103, 300].

4.3.1 Ethanol

In *C. glutamicum*, ethanol fermentation strictly relies on heterologous genes. Obtaining pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*) from *Zymomonas mobilis* under expressional control of the promoter of the lactate dehydrogenase gene (*ldhA*), a basic producer was generated [111]. Subsequent deletion of the genes *ldhA* and *ppc*, encoding phosphoenolpyruvate carboxylase, avoided formation of the major by-products lactate and succinate. Under oxygen limitation, a yield of 0.53 g g^{-1} was achieved, though at low product concentration, likely related to tolerance issues [12, 111]. The ethanol production rate was, however, substantially higher than that of many other bacteria reported so far [8]. At a high cell density of 60 g L^{-1} cell dry weight, growth arrested cells produced ethanol at a rate of 30 g L^{-1} h^{-1} [111]. This appears to be a promising

starting point for further improvement. With regard to the antiseptic activity of ethanol, tolerance is one of the key issues. Recent laboratory evolution experiments proved valuable in conferring tolerance to methanol [249] and accompanying tolerance to thermal and solvent stress [248]. Identification of the assignable cause and subsequent re-engineering [249] seems to be a reasonable approach for next level ethanol producers.

4.3.2 Isobutanol

Strain engineering for efficient isobutanol production has profited substantially from the knowledge gained from branched-chain amino acid fermentation [3, 103, 119, 301]. The isobutanol strategy [103] thus relied on a previous strain design for valine-producing *C. glutamicum* [101, 302, 303]. Complete pathway design comprised *alsS* (acetohydroxy acid synthase from *B. subtilis*) and *ilvCD* (acetohydroxyacid isomeroreductase and dihydroxyacid dehydratase from *C. glutamicum*) along with downstream genes for the subsequent decarboxylation (*kivd*, encoding ketoacid decarboxylase from *L. lactis*) and reduction (*adhA*, encoding alcohol dehydrogenase from C. *glutamicum*) of 2-ketoisovalerate to isobutanol [103]. In a Δ*pyc* Δ*ldh* background, this enabled production of 4.9 g L^{-1} isobutanol. Higher level production relied on a 2-ketovalerate production strain that was additionally modified by inactivation of lactate and malate dehydrogenases, implementation of ketoacid decarboxylase from *Lactococcus lactis*, alcohol dehydrogenase (ADH2) from *S. cerevisiae*, and expression of the *pntAB* transhydrogenase gene from *E. coli* [300]. The highest titer so far of 73 g L^{-1} was achieved by an approach that, in addition to metabolic engineering, also considered tolerance issues, which was addressed by process operation via continuous solvent extraction during fermentation [304].

4.4 Toward Non-food Substrates

Strategies to achieve sustainability are strongly driving new biosynthetic chemistry processes from renewable feedstocks [4]. Traditionally, however, biotechnology builds on glucose and starch as fermentation substrates – raw materials that are equally serving for human nutrition. Beyond these feedstocks, there are vast amounts of other, so far unused, bio-based substrates, including lignocellulosic biomass, natural oils, or waste streams from different industries [289]. To make them bioavailable for *C. glutamicum*, some assimilation routes were implemented to establish processes from diverse sugars [42, 44, 48, 305–308], oligo- and polymers thereof [309–312], alcohols [313], sugar alcohols [314, 315], organic acids [14], and green juices [15, 237]. As one of the major novel raw materials, pentoses are discussed below. In addition, interesting developments toward direct use of polymeric raw materials in one-step consolidated bioprocesses are described.

4.4.1 Pentoses

Being a major constituent of hemicellulose, pentose sugar is highly abundant on Earth and thus of value and relevance when talking about alternative fermentation feedstocks [4]. For enabling growth on xylose in *C. glutamicum,* heterologous expression of a single gene – *xylA*, encoding xylose isomerase – is sufficient, though additional overexpression of *xylB*, encoding xylulokinase, supports xylose assimilation [44]. Using this strategy, diaminopentane can be produced by modified *C. glutamicum* strains from xylose as single carbon source and, beyond its pure form, from xylose-containing sugar mixtures obtained from hemicellulose hydrolyzates [48]. Integrated analysis of the physiological response to xylose fermentation on the level of transcriptome and in vivo fluxes provided new insights into xylose metabolism and unraveled further optimization targets for systems-wide engineering [47]. A superior strain created from these findings efficiently converted xylose into diaminopentane with a yield of 32% and a titer of 60.1 g/L [47]. Arabinose-based production was established using the recombinant arabinose-operon of *E. coli* [42, 43]. In additional to extension of the substrate spectrum, a major issue is tolerance against toxic substances, such as furfural, hydroxymethylfurfural (HMF), and phenol, typically present in lignocellulosic feedstocks after pre-treatment [4, 316, 317]. *C. glutamicum* has a natural capacity for detoxification of furfural [318] which was recently associated with the *fudC* gene, conveying the ability for reduction of furfural to furfuryl alcohol [319]. Enhanced robustness can also be conferred by overexpression of mycothiol glycosyltransferase *mshA* [320]. For the future, further discovery of tolerance mechanisms and their manipulation remain essential for realizing lignocellulose-based biorefineries with *C. glutamicum*.

4.4.2 Sugar Polymers

The commercially used fermentation substrates are naturally bound in homo- or heteropolymers with specific and polymer-dependent composition and branching, in general requiring pretreatment before use [321]. This was overcome in recombinant *C. glutamicum* strains that express polymer-degrading pathways from diverse donor strains for direct utilization of starch [309], cellobiose [322–324], cellulose [325, 326], and lignocellulose [327]. As most polymers are not transported into the cell, the hydrolyzing enzymes need to get to the extracellular environment. This was addressed by different approaches involving either secretory systems for releasing soluble enzymes into the medium [310, 325] or cell-surface display of the degrading enzymes [311, 322, 327, 328]. The enzymatic set mainly comprises hydrolase such as α-amylase [309–311, 328, 329], β-glucosidase [322, 323], cellulases [327], and endoglucanase [325, 326]. However, hydrolysis rates are usually higher at higher temperature and lower pH as compared to the optimal growth environment of *C. glutamicum* [48, 330], favoring strains with increased tolerance through metabolic and evolutionary engineering

[248, 331]. Recent and current studies also aim at substrate co-utilization toward implementation of consolidated bioprocesses [332, 333].

5 Conclusions

We are currently standing at a turning point from a petroleum-based industry to a bio-based economy. Riboflavin production is an impressive and promising example, where petrochemical production was completely replaced by bio-based production with optimized strains of *Bacillus subtilis* and *Ashbya gossypii* within only 30 years [230]. The strong foundation for this sustainable development has been built since the discovery and establishment of microbial fermentation processes [10]. Throughout the last few decades, innovations in strain engineering, automation, and mechanization have provided invaluable tools to face upcoming challenges. Most encouraging, systems and synthetic metabolic engineering has provided novel and alternative routes for producing chemicals, materials, nutrients, and health care products from renewable raw materials [12, 280, 334]. Success stories from engineering *C. glutamicum* toward production of lysine [55], arginine [76], diaminopentane [73], and valine [335] give substantial encouragement. Some of these products have already become established in the market, some with decades-lasting tradition, whereas other innovative products need to prove their benefit and value to advance from the state of feasibility to the state of commercialization. What remains open at this point beyond pathway engineering comes from large-scale set-ups of industrial production. These fermentations typically have their own peculiarities such as nutrient, pH, oxygen and temperature gradients related to mixing issues, as well as complex, partly toxic raw materials. Strain robustness and vitality are therefore key targets to be addressed more precisely in the future. Evolutionary approaches to address the required multi-target cellular responses to overcome process stress have recently been found to be valuable for improving the tolerance of *C. glutamicum* [159, 248, 249] and production performance [250, 271]. Upon integrating evolutionary adaptation and re-engineering within the concept of systems and synthetic metabolic engineering, we can expect a new level of strain engineering to move forward to a bio-based economy.

References

1. Kinoshita S, Udaka S, Shimono M (1957) Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. J Gen Appl Microbiol 3 (3):193–205
2. Leuchtenberger W, Huthmacher K, Drauz K (2005) Biotechnological production of amino acids and derivatives: current status and prospects. Appl Microbiol Biotechnol 69(1):1–8

3. Becker J, Wittmann C (2012) Systems and synthetic metabolic engineering for amino acid production – the heartbeat of industrial strain development. Curr Opin Biotechnol 23 (5):718–726
4. Buschke N, Schäfer R, Becker J, Wittmann C (2013) Metabolic engineering of industrial platform microorganisms for biorefinery applications--optimization of substrate spectrum and process robustness by rational and evolutive strategies. Bioresour Technol 135:544–554
5. Soliman S, Tang Y (2015) Natural and engineered production of taxadiene with taxadiene synthase. Biotechnol Bioeng 112(2):229–235
6. Blombach B, Seibold GM (2010) Carbohydrate metabolism in *Corynebacterium glutamicum* and applications for the metabolic engineering of L-lysine production strains. Appl Microbiol Biotechnol 86(5):1313–1322
7. Ehira S, Teramoto H, Inui M, Yukawa H (2009) Regulation of *Corynebacterium glutamicum* heat shock response by the extracytoplasmic-function sigma factor SigH and transcriptional regulators HspR and HrcA. J Bacteriol 191(9):2964–2972
8. Becker J, Kind S, Wittmann C (2012) Systems metabolic engineering of *Corynebacterium glutamicum* for biobased production of chemicals, materials and fuels. In: Wittmann C, Lee SY (eds) Systems metabolic engineering. Springer, Dordrecht, Heidelberg, New York, London, pp 152–191
9. Becker J, Wittmann C (2012) Bio-based production of chemicals, materials and fuels – *Corynebacterium glutamicum* as versatile cell factory. Curr Opin Biotechnol 23(4):631–640
10. Becker J, Wittmann C (2016) Industrial microorganisms: *Corynebacterium glutamicum*. In: Wittmann C, Liao JC (eds) Industrial iotechnology. Advanced biotechnology. Wiley-VCH, Weinheim, pp 183–222
11. Sugimoto S, Shiio I (1989) Fructose metabolism and regulation of 1-phosphofructokinase and 6-phosphofructokinase in *Brevibacterium flavum*. Agric Biol Chem 53:1261–1268
12. Becker J, Wittmann C (2015) Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. Angew Chem Int Ed Engl 54:3328–3350
13. Laslo T, von Zaluskowski P, Gabris C, Lodd E, Rückert C, Dangel P, Kalinowski J, Auchter M, Seibold G, Eikmanns BJ (2012) Arabitol metabolism of *Corynebacterium glutamicum* and its regulation by AtlR. J Bacteriol 194(5):941–955
14. Neuner A, Heinzle E (2011) Mixed glucose and lactate uptake by *Corynebacterium glutamicum* through metabolic engineering. Biotechnol J 6(3):318–329
15. Neuner A, Wagner I, Sieker T, Ulber R, Schneider K, Peifer S, Heinzle E (2013) Production of L-lysine on different silage juices using genetically engineered *Corynebacterium glutamicum*. J Biotechnol 163(2):217–224
16. Gerstmeir R, Wendisch VF, Schnicke S, Ruan H, Farwick M, Reinscheid D, Eikmanns BJ (2003) Acetate metabolism and its regulation in *Corynebacterium glutamicum*. J Biotechnol 104(1-3):99–122
17. Hayashi M, Mizoguchi H, Shiraishi N, Obayashi M, Nakagawa S, Imai J, Watanabe S, Ota T, Ikeda M (2002) Transcriptome analysis of acetate metabolism in *Corynebacterium glutamicum* using a newly developed metabolic array. Biosci Biotechnol Biochem 66 (6):1337–1344
18. Rittmann D, Schaffer S, Wendisch VF, Sahm H (2003) Fructose-1,6-bisphosphatase from *Corynebacterium glutamicum*: expression and deletion of the *fbp* gene and biochemical characterization of the enzyme. Arch Microbiol 180(4):285–292
19. Mori M, Shiio I (1987) Pyruvate formation and sugar metabolism in an amino acid-producing bacterium, *Brevibacterium flavum*. Agric Biol Chem 51(1):129–138
20. Parche S, Burkovski A, Sprenger GA, Weil B, Krämer R, Titgemeyer F (2001) *Corynebacterium glutamicum*: a dissection of the PTS. J Mol Microbiol Biotechnol 3(3):423–428
21. Ikeda M (2012) Sugar transport systems in *Corynebacterium glutamicum*: features and applications to strain development. Appl Microbiol Biotechnol 96(5):1191–1200

22. Peng X, Okai N, Vertes AA, Inatomi K, Inui M, Yukawa H (2011) Characterization of the mannitol catabolic operon of *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 91 (5):1375–1387
23. Sasaki M, Teramoto H, Inui M, Yukawa H (2011) Identification of mannose uptake and catabolism genes in *Corynebacterium glutamicum* and genetic engineering for simultaneous utilization of mannose and glucose. Appl Microbiol Biotechnol 89(6):1905–1916
24. Xu J, Han M, Zhang J, Guo Y, Zhang W (2014) Metabolic engineering *Corynebacterium glutamicum* for the L-lysine production by increasing the flux into L-lysine biosynthetic pathway. Amino Acids 46(9):2165–2175
25. Moon MW, Park SY, Choi SK, Lee JK (2007) The phosphotransferase system of *Corynebacterium glutamicum*: features of sugar transport and carbon regulation. J Mol Microbiol Biotechnol 12(1-2):43–50
26. Park S-Y, Kim H-K, Yoo S-K, Oh T-K, Lee J-K (2000) Characterization of glk, a gene coding for glucose kinase of *Corynebacterium glutamicum*. FEMS Microbiol Lett 188(2):209–215
27. Cocaign-Bousquet M, Guyonvarch A, Lindley ND (1996) Growth rate-dependent modulation of carbon flux through central metabolism and the kinetic consequences for glucose-limited chemostat cultures of *Corynebacterium glutamicum*. Appl Environ Microbiol 62 (2):429–436
28. Ikeda M, Mizuno Y, Awane S, Hayashi M, Mitsuhashi S, Takeno S (2011) Identification and application of a different glucose uptake system that functions as an alternative to the phosphotransferase system in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90 (4):1443–1451
29. Lindner SN, Seibold GM, Henrich A, Krämer R, Wendisch VF (2011) Phosphotransferase system-independent glucose utilization in *Corynebacterium glutamicum* by inositol permeases and glucokinases. Appl Environ Microbiol 77(11):3571–3581
30. Lindner SN, Seibold GM, Krämer R, Wendisch VF (2011) Impact of a new glucose utilization pathway in amino acid-producing *Corynebacterium glutamicum*. Bioeng Bugs 2 (5):291–295
31. Zhou Z, Wang C, Xu H, Chen Z, Cai H (2015) Increasing succinic acid production using the PTS-independent glucose transport system in a *Corynebacterium glutamicum* PTS-defective mutant. J Ind Microbiol Biotechnol 42(7):1073–1082
32. Dominguez H, Rollin C, Guyonvarch A, Guerquin-Kern JL, Cocaign-Bousquet M, Lindley ND (1998) Carbon-flux distribution in the central metabolic pathways of *Corynebacterium glutamicum* during growth on fructose. Eur J Biochem 254(1):96–102
33. Moon M-W, Kim H-J, Oh T-K, Shin C-S, Lee J-S, Kim S-J, Lee J-K (2005) Analyses of enzyme II gene mutants for sugar transport and heterologous expression of fructokinase gene in *Corynebacterium glutamicum* ATCC 13032. FEMS Microbiol Lett 244(2):259–266
34. Kiefer P, Heinzle E, Zelder O, Wittmann C (2004) Comparative metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* cultured on glucose or fructose. Appl Environ Microbiol 70(1):229–239
35. Wittmann C, Kiefer P, Zelder O (2004) Metabolic fluxes in *Corynebacterium glutamicum* during lysine production with sucrose as carbon source. Appl Environ Microbiol 70 (12):7277–7287
36. Becker J, Klopprogge C, Zelder O, Heinzle E, Wittmann C (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(12):8587–8596
37. Chen Y, Zhou YJ, Siewers V, Nielsen J (2015) Enabling technologies to advance microbial isoprenoid production. Adv Biochem Eng Biotechnol 148:143–160
38. Dominguez H, Lindley ND (1996) Complete sucrose metabolism requires fructose phosphotransferase activity in *Corynebacterium glutamicum* to ensure phosphorylation of liberated fructose. Appl Environ Microbiol 62(10):3878–3880

39. Higgins CF (2001) ABC transporters: physiology, structure and mechanism – an overview. Res Microbiol 152(3–4):205–210
40. Nentwich SS, Brinkrolf K, Gaigalat L, Hüser AT, Rey DA, Mohrbach T, Marin K, Pühler A, Tauch A, Kalinowski J (2009) Characterization of the LacI-type transcriptional repressor RbsR controlling ribose transport in *Corynebacterium glutamicum* ATCC 13032. Microbiology 155(Pt 1):150–164
41. Kawaguchi H, Sasaki M, Vertes AA, Inui M, Yukawa H (2009) Identification and functional analysis of the gene cluster for L-arabinose utilization in *Corynebacterium glutamicum*. Appl Environ Microbiol 75(11):3419–3429
42. Kawaguchi H, Sasaki M, Vertes AA, Inui M, Yukawa H (2008) Engineering of an L-arabinose metabolic pathway in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 77(5):1053–1062
43. Schneider J, Niermann K, Wendisch VF (2011) Production of the amino acids L-glutamate, L-lysine, L-ornithine and L-arginine from arabinose by recombinant *Corynebacterium glutamicum*. J Biotechnol 154(2-3):191–198
44. Kawaguchi H, Vertes AA, Okino S, Inui M, Yukawa H (2006) Engineering of a xylose metabolic pathway in *Corynebacterium glutamicum*. Appl Environ Microbiol 72 (5):3418–3428
45. Eberhardt D, Jensen JV, Wendisch VF (2014) L-Citrulline production by metabolically engineered *Corynebacterium glutamicum* from glucose and alternative carbon sources. AMB Express 4(1):85
46. Meiswinkel TM, Gopinath V, Lindner SN, Nampoothiri KM, Wendisch VF (2013) Accelerated pentose utilization by *Corynebacterium glutamicum* for accelerated production of lysine, glutamate, ornithine and putrescine. Microb Biotechnol 6(2):131–140
47. Buschke N, Becker J, Schäfer R, Kiefer P, Biedendieck R, Wittmann C (2013) Systems metabolic engineering of xylose-utilizing *Corynebacterium glutamicum* for production of 1,5-diaminopentane. Biotechnol J 8(5):557–570
48. Buschke N, Schröder H, Wittmann C (2011) Metabolic engineering of *Corynebacterium glutamicum* for production of 1,5-diaminopentane from hemicellulose. Biotechnol J 6 (3):306–317
49. Frunzke J, Engels V, Hasenbein S, Gätgens C, Bott M (2008) Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. Mol Microbiol 67(2):305–322
50. Yin H, Zhuang YB, Li EE, Bi HP, Zhou W, Liu T (2015) Heterologous biosynthesis of costunolide in *Escherichia coli* and yield improvement. Biotechnol Lett 37(6):1249–1255
51. Yokota A, Lindley ND (2005) Central metabolism: sugar uptake and conversion. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC, Boca Raton, pp 215–240
52. Han SO, Inui M, Yukawa H (2007) Expression of *Corynebacterium glutamicum* glycolytic genes varies with carbon source and growth phase. Microbiology 153(Pt 7):2190–2202
53. Wittmann C, Heinzle E (2002) Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing Corynebacteria. Appl Environ Microbiol 68(12):5843–5859
54. Becker J, Klopprogge C, Herold A, Zelder O, Bolten CJ, Wittmann C (2007) Metabolic flux engineering of L-lysine production in *Corynebacterium glutamicum*--over expression and modification of G6P dehydrogenase. J Biotechnol 132(2):99–109
55. Becker J, Zelder O, Haefner S, Schröder H, Wittmann C (2011) From zero to hero–design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. Metab Eng 13(2):159–168
56. Krömer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C (2004) In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J Bacteriol 186(6):1769–1784

57. Marx A, Striegel K, de Graaf AA, Sahm H, Eggeling L (1997) Response of the central metabolism of *Corynebacterium glutamicum* to different flux burdens. Biotechnol Bioeng 56 (2):168–180
58. Wittmann C, De Graaf AA (2005) Metabolic flux analysis in *Corynebacterium glutamicum*. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC, Boca Raton, pp 277–304
59. Marx A, Hans S, Möckel B, Bathe B, de Graaf AA, McCormack AC, Stapleton C, Burke K, O'Donohue M, Dunican LK (2003) Metabolic phenotype of phosphoglucose isomerase mutants of *Corynebacterium glutamicum*. J Biotechnol 104(1-3):185–197
60. Gubler M, Jetten M, Lee SH, Sinskey AJ (1994) Cloning of the pyruvate kinase gene (*pyk*) of *Corynebacterium glutamicum* and site-specific inactivation of *pyk* in a lysine-producing *Corynebacterium lactofermentum* strain. Appl Environ Microbiol 60(7):2494–2500
61. Bommareddy RR, Chen Z, Rappert S, Zeng AP (2014) A de novo NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. Metab Eng 25:30–37
62. Takeno S, Hori K, Ohtani S, Mimura A, Mitsuhashi S, Ikeda M (2016) L-Lysine production independent of the oxidative pentose phosphate pathway by *Corynebacterium glutamicum* with the *Streptococcus mutans gapN* gene. Metab Eng 37:1–10
63. Takeno S, Murata R, Kobayashi R, Mitsuhashi S, Ikeda M (2010) Engineering of *Corynebacterium glutamicum* with an NADPH-generating glycolytic pathway for L-lysine production. Appl Environ Microbiol 76(21):7154–7160
64. Tsuge Y, Yamamoto S, Kato N, Suda M, Vertes AA, Yukawa H, Inui M (2015) Overexpression of the phosphofructokinase encoding gene is crucial for achieving high production of D-lactate in *Corynebacterium glutamicum* under oxygen deprivation. Appl Microbiol Biotechnol 99(11):4679–4689
65. Tsuge Y, Yamamoto S, Suda M, Inui M, Yukawa H (2013) Reactions upstream of glycerate-1,3-bisphosphate drive *Corynebacterium glutamicum* (D)-lactate productivity under oxygen deprivation. Appl Microbiol Biotechnol 97(15):6693–6703
66. Yamamoto S, Gunji W, Suzuki H, Toda H, Suda M, Jojima T, Inui M, Yukawa H (2012) Overexpression of genes encoding glycolytic enzymes in *Corynebacterium glutamicum* enhances glucose metabolism and alanine production under oxygen deprivation conditions. Appl Environ Microbiol 78(12):4447–4457
67. Reddy GK, Wendisch VF (2014) Characterization of 3-phosphoglycerate kinase from *Corynebacterium glutamicum* and its impact on amino acid production. BMC Microbiol 14:54
68. Moritz B, Striegel K, De Graaf AA, Sahm H (2000) Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux in vivo. Eur J Biochem 267 (12):3442–3452
69. Teramoto H, Inui M (2013) Regulation of sugar uptake, glycolysis and the pentose phosphate pathway in *Corynebacterium glutamicum*. In: Yukawa H, Inui M (eds) *Corynebacterium glutamicum* – biology and biotechnology, vol 23, Microbiology monographs. Springer, Berlin-Heidelberg, pp 263–279
70. Becker J, Wittmann C (2013) Pathways at work: metabolic flux analysis of the industrial cell factory *Corynebacterium glutamicum*. In: Yukawa H, Inui M (eds) *Corynebacterium glutamicum* – biology and biotechnology, vol 23, Microbiology monographs. Springer, Berlin-Heidelberg, pp 217–237
71. Ohnishi J, Katahira R, Mitsuhashi S, Kakita S, Ikeda M (2005) A novel *gnd* mutation leading to increased L-lysine production in *Corynebacterium glutamicum*. FEMS Microbiol Lett 242 (2):265–274
72. Becker J, Buschke N, Bücker R, Wittmann C (2010) Systems level engineering of *Corynebacterium glutamicum* – reprogramming translational efficiency for superior production. Eng Life Sci 10:430–438

73. Kind S, Neubauer S, Becker J, Yamamoto M, Völkert M, Abendroth GV, Zelder O, Wittmann C (2014) From zero to hero – production of bio-based nylon from renewable resources using engineered *Corynebacterium glutamicum*. Metab Eng 25:113–123
74. Shi F, Li K, Huan X, Wang X (2013) Expression of NAD(H) kinase and glucose-6-phosphate dehydrogenase improve NADPH supply and L-isoleucine biosynthesis in *Corynebacterium glutamicum* ssp. *lactofermentum*. Appl Biochem Biotechnol 171(2):504–521
75. Bartek T, Blombach B, Zonnchen E, Makus P, Lang S, Eikmanns BJ, Oldiges M (2010) Importance of NADPH supply for improved L-valine formation in *Corynebacterium glutamicum*. Biotechnol Prog 26(2):361–371
76. Park SH, Kim HU, Kim TY, Park JS, Kim SS, Lee SY (2014) Metabolic engineering of *Corynebacterium glutamicum* for L-arginine production. Nat Commun 5:4618
77. Kim SY, Lee J, Lee SY (2015) Metabolic engineering of *Corynebacterium glutamicum* for the production of L-ornithine. Biotechnol Bioeng 112(2):416–421
78. Zhang C, Zhang J, Kang Z, Du G, Chen J (2015) Rational engineering of multiple module pathways for the production of L-phenylalanine in *Corynebacterium glutamicum*. J Ind Microbiol Biotechnol 42(5):787–797
79. Ikeda M, Katsumata R (1999) Hyperproduction of tryptophan by *Corynebacterium glutamicum* with the modified pentose phosphate pathway. Appl Environ Microbiol 65 (6):2497–2502
80. Eikmanns BJ, Rittmann D, Sahm H (1995) Cloning, sequence analysis, expression, and inactivation of the *Corynebacterium glutamicum icd* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. J Bacteriol 177(3):774–782
81. Wendisch VF, Spies M, Reinscheid DJ, Schnicke S, Sahm H, Eikmanns BJ (1997) Regulation of acetate metabolism in *Corynebacterium glutamicum*: transcriptional control of the isocitrate lyase and malate synthase genes. Arch Microbiol 168(4):262–269
82. Ozaki H, Shiio I (1968) Regulation of the TCA and glyoxylate cycles in *Brevibacterium flavum*. I. Inhibition of isocitrate lyase and isocitrate dehydrogenase by organic acids related to the TCA and glyoxylate cycles. J Biochem 64(3):355–363
83. Shirai T, Nakato A, Izutani N, Nagahisa K, Shioya S, Kimura E, Kawarabayasi Y, Yamagishi A, Gojobori T, Shimizu H (2005) Comparative study of flux redistribution of metabolic pathway in glutamate production by two coryneform bacteria. Metab Eng 7 (2):59–69
84. Shiio I, Ujigawa-Takeda K (1980) Presence and regulation of α-ketoglutarate dehydrogenase complex in a glutamate-producing bacterium, *Brevibacterium flavum*. Agric Biol Chem 44 (8):1897–1904
85. Kim J, Fukuda H, Hirasawa T, Nagahisa K, Nagai K, Wachi M, Shimizu H (2010) Requirement of de novo synthesis of the OdhI protein in penicillin-induced glutamate production by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 86(3):911–920
86. Asakura Y, Kimura E, Usuda Y, Kawahara Y, Matsui K, Osumi T, Nakamatsu T (2007) Altered metabolic flux due to deletion of *odhA* causes L-glutamate overproduction in *Corynebacterium glutamicum*. Appl Environ Microbiol 73(4):1308–1319
87. Kim J, Hirasawa T, Sato Y, Nagahisa K, Furusawa C, Shimizu H (2009) Effect of *odhA* overexpression and *odhA* antisense RNA expression on Tween-40-triggered glutamate production by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 81(6):1097–1106
88. Niebisch A, Kabus A, Schultz C, Weil B, Bott M (2006) Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. J Biol Chem 281(18):12300–12307
89. Schultz C, Niebisch A, Gebel L, Bott M (2007) Glutamate production by *Corynebacterium glutamicum*: dependence on the oxoglutarate dehydrogenase inhibitor protein OdhI and protein kinase PknG. Appl Microbiol Biotechnol 76(3):691–700
90. Wang N, Ni Y, Shi F (2015) Deletion of *odhA* or *pyc* improves production of gamma-aminobutyric acid and its precursor L-glutamate in recombinant *Corynebacterium glutamicum*. Biotechnol Lett 37(7):1473–1481

91. van Ooyen J, Noack S, Bott M, Reth A, Eggeling L (2012) Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity. Biotechnol Bioeng 109(8):2070–2081
92. Becker J, Klopprogge C, Schröder H, Wittmann C (2009) Metabolic engineering of the tricarboxylic acid cycle for improved lysine production by *Corynebacterium glutamicum*. Appl Environ Microbiol 75(24):7866–7869
93. Kind S, Becker J, Wittmann C (2013) Increased lysine production by flux coupling of the tricarboxylic acid cycle and the lysine biosynthetic pathway--metabolic engineering of the availability of succinyl-CoA in *Corynebacterium glutamicum*. Metab Eng 15:184–195
94. Otten A, Brocker M, Bott M (2015) Metabolic engineering of *Corynebacterium glutamicum* for the production of itaconate. Metab Eng 30:156–165
95. Zahoor A, Otten A, Wendisch VF (2014) Metabolic engineering of *Corynebacterium glutamicum* for glycolate production. J Biotechnol. 192:366–375
96. Sauer U, Eikmanns BJ (2005) The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS Microbiol Rev 29(4):765–794
97. Eikmanns BJ (2005) Central metabolism: tricarboxylic acid cycle and anaplerotic reactions. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC, Boca Raton, pp 241–276
98. Petersen S, de Graaf AA, Eggeling L, Mollney M, Wiechert W, Sahm H (2000) In vivo quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*. J Biol Chem 275(46):35932–35941
99. Peters-Wendisch PG, Kreutzer C, Kalinowski J, Patek M, Sahm H, Eikmanns BJ (1998) Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. Microbiology 144(Pt 4):915–927
100. Peters-Wendisch PG, Schiel B, Wendisch VF, Katsoulidis E, Möckel B, Sahm H, Eikmanns BJ (2001) Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. J Mol Microbiol Biotechnol 3(2):295–300
101. Blombach B, Schreiner ME, Holatko J, Bartek T, Oldiges M, Eikmanns BJ (2007) L-Valine production with pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum*. Appl Environ Microbiol 73(7):2079–2084
102. Blombach B, Schreiner ME, Moch M, Oldiges M, Eikmanns BJ (2007) Effect of pyruvate dehydrogenase complex deficiency on L-lysine production with *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 76(3):615–623
103. Smith KM, Cho KM, Liao JC (2010) Engineering *Corynebacterium glutamicum* for isobutanol production. Appl Microbiol Biotechnol 87(3):1045–1055
104. Buchholz J, Schwentner A, Brunnenkan B, Gabris C, Grimm S, Gerstmeir R, Takors R, Eikmanns BJ, Blombach B (2013) Platform engineering of *Corynebacterium glutamicum* with reduced pyruvate dehydrogenase complex activity for improved production of L-lysine, L-valine, and 2-ketoisovalerate. Appl Environ Microbiol 79(18):5566–5575
105. Sawada K, Zen-in S, Wada M, Yokota A (2010) Metabolic changes in a pyruvate kinase gene deletion mutant of *Corynebacterium glutamicum* ATCC 13032. Metab Eng 12(4):401–407
106. Becker J, Klopprogge C, Wittmann C (2008) Metabolic responses to pyruvate kinase deletion in lysine producing *Corynebacterium glutamicum*. Microb Cell Fact 7:8
107. Nguyen AQ, Schneider J, Reddy GK, Wendisch VF (2015) Fermentative production of the diamine putrescine: system metabolic engineering of *Corynebacterium glutamicum*. Metabolites 5(2):211–231
108. Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K, Ikeda M (2002) A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. Appl Microbiol Biotechnol 58(2):217–223
109. Sato H, Orishimo K, Shirai T, Hirasawa T, Nagahisa K, Shimizu H, Wachi M (2008) Distinct roles of two anaplerotic pathways in glutamate production induced by biotin limitation in *Corynebacterium glutamicum*. J Biosci Bioeng 106(1):51–58

110. Inui M, Murakami S, Okino S, Kawaguchi H, Vertes AA, Yukawa H (2004) Metabolic analysis of *Corynebacterium glutamicum* during lactate and succinate productions under oxygen deprivation conditions. J Mol Microbiol Biotechnol 7(4):182–196
111. Inui M, Kawaguchi H, Murakami S, Vertes AA, Yukawa H (2004) Metabolic engineering of *Corynebacterium glutamicum* for fuel ethanol production under oxygen-deprivation conditions. J Mol Microbiol Biotechnol 8(4):243–254
112. Ikeda M (2005) L-tryptophan production. In: Eggeling L, Bott M (eds) Handbook of *Corynebacterium glutamicum*. CRC, Boca Raton, pp 489–509
113. Sano K, Ito K, Miwa K, Nakamori S (1987) Amplification of the phosphoenol pyruvate carboxylase gene of *Brevibacterium lactofermentum* to improve amino acid production. Agric Biol Chem 51(2):597–599
114. Chen Z, Bommareddy RR, Frank D, Rappert S, Zeng AP (2014) Deregulation of feedback inhibition of phosphoenolpyruvate carboxylase for improved lysine production in *Corynebacterium glutamicum*. Appl Environ Microbiol 80(4):1388–1393
115. Wada M, Sawada K, Ogura K, Shimono Y, Hagiwara T, Sugimoto M, Onuki A, Yokota A (2015) Effects of phosphoenolpyruvate carboxylase desensitization on glutamic acid production in *Corynebacterium glutamicum* ATCC 13032. J Biosci Bioeng 121:172–177
116. Petersen S, Mack C, de Graaf AA, Riedel C, Eikmanns BJ, Sahm H (2001) Metabolic consequences of altered phosphoenolpyruvate carboxykinase activity in *Corynebacterium glutamicum* reveal anaplerotic regulation mechanisms in vivo. Metab Eng 3(4):344–361
117. Riedel C, Rittmann D, Dangel P, Möckel B, Petersen S, Sahm H, Eikmanns BJ (2001) Characterization of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the enzyme for growth and amino acid production. J Mol Microbiol Biotechnol 3(4):573–583
118. Georgi T, Rittmann D, Wendisch VF (2005) Lysine and glutamate production by *Corynebacterium glutamicum* on glucose, fructose and sucrose: roles of malic enzyme and fructose-1,6-bisphosphatase. Metab Eng 7(4):291–301
119. Blombach B, Eikmanns BJ (2011) Current knowledge on isobutanol production with *Escherichia coli, Bacillus subtilis* and *Corynebacterium glutamicum*. Bioeng Bugs 2 (6):346–350
120. Vasicova P, Patek M, Nesvera J, Sahm H, Eikmanns B (1999) Analysis of the *Corynebacterium glutamicum dapA* promoter. J Bacteriol 181(19):6188–6191
121. Kjeldsen KR, Nielsen J (2009) In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. Biotechnol Bioeng 102(2):583–597
122. Burkovski A (2013) Cell envelope of corynebacteria: structure and influence on pathogenicity. ISRN Microbiol 2013:935736
123. Yuzawa S, Eng CH, Katz L, Keasling JD (2014) Enzyme analysis of the polyketide synthase leads to the discovery of a novel analog of the antibiotic alpha-lipomycin. J Antibiot (Tokyo) 67(2):199–201
124. Puech V, Chami M, Lemassu A, Lanéelle M-A, Schiffler B, Gounon P, Bayan N, Benz R, Daffé M (2001) Structure of the cell envelope of corynebacteria: importance of the non-covalently bound lipids in the formation of the cell wall permeability barrier and fracture plane. Microbiology 147(5):1365–1382
125. Wittmann C (2007) Fluxome analysis using GC-MS. Microb Cell Fact 6:6
126. Wittmann C, Kim HM, Heinzle E (2004) Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. Biotechnol Bioeng 87(1):1–6
127. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14(3):283–291
128. Pauling J, Röttger R, Tauch A, Azevedo V, Baumbach J (2012) CoryneRegNet 6.0—updated database content, new analysis methods and novel features focusing on community demands. Nucleic Acids Res 40(D1):D610–D614

129. Patek M, Nesvera J (2013) Promoters and plasmid vectors of *Corynebacterium glutamicum*. In: Yukawa H, Inui M (eds) *Corynebacterium glutamicum* – biology and biotechnology, vol 23, Microbiology monographs. Springer, Berlin-Heidelberg, pp 51–88
130. Lemuth K, Steuer K, Albermann C (2011) Engineering of a plasmid-free *Escherichia coli* strain for improved in vivo biosynthesis of astaxanthin. Microb Cell Fact 10:29
131. Howat S, Park B, Oh IS, Jin YW, Lee EK, Loake GJ (2014) Paclitaxel: biosynthesis, production and future prospects. N Biotechnol 31(3):242–245
132. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57(1):441–466
133. Restaino OF, Bhaskar U, Paul P, Li L, De Rosa M, Dordick JS, Linhardt RJ (2013) High cell density cultivation of a recombinant *E. coli* strain expressing a key enzyme in bioengineered heparin production. Appl Microbiol Biotechnol 97(9):3893–3900
134. Larisch C, Nakunst D, Huser AT, Tauch A, Kalinowski J (2007) The alternative sigma factor SigB of *Corynebacterium glutamicum* modulates global gene expression during transition from exponential growth to stationary phase. BMC Genomics 8:4
135. Taniguchi H, Wendisch VF (2015) Exploring the role of sigma factor gene expression on production by *Corynebacterium glutamicum*: sigma factor H and FMN as example. Front Microbiol 6:740
136. Nakunst D, Larisch C, Hüser AT, Tauch A, Pühler A, Kalinowski J (2007) The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. J Bacteriol 189 (13):4696–4707
137. Plassmeier J, Li Y, Rueckert C, Sinskey AJ (2016) Metabolic engineering *Corynebacterium glutamicum* to produce triacylglycerols. Metab Eng 33:86–97
138. Rados D, Carvalho AL, Wieschalka S, Neves AR, Blombach B, Eikmanns BJ, Santos H (2015) Engineering *Corynebacterium glutamicum* for the production of 2,3-butanediol. Microb Cell Fact 14(1):171
139. Yu H, Luscombe NM, Qian J, Gerstein M (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends in Genetics 19(8):422–427
140. Wang J, Guleria S, Koffas MA, Yan Y (2015) Microbial production of value-added nutraceuticals. Curr Opin Biotechnol 37:97–104
141. Nishimura T, Teramoto H, Inui M, Yukawa H (2011) Gene expression profiling of *Corynebacterium glutamicum* during anaerobic nitrate respiration: induction of the SOS response for cell survival. J Bacteriol 193(6):1327–1333
142. Kohl TA, Tauch A (2009) The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: detection of the corynebacterial core regulon and integration into the transcriptional regulatory network model. J Biotechnol 143(4):239–246
143. Fang MY, Zhang C, Yang S, Cui JY, Jiang PX, Lou K, Wachi M, Xing XH (2015) High crude violacein production from glucose by *Escherichia coli* engineered with interactive control of tryptophan pathway and violacein biosynthetic pathway. Microb Cell Fact 14:8
144. Walter B, Hanssler E, Kalinowski J, Burkovski A (2007) Nitrogen metabolism and nitrogen control in corynebacteria: variations of a common theme. J Mol Microbiol Biotechnol 12 (1-2):131–138
145. Jakoby M, Nolden L, Meier-Wagner J, Krämer R, Burkovski A (2000) AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. Mol Microbiol 37(4):964–977
146. Silberbach M, Burkovski A (2006) Application of global analysis techniques to *Corynebacterium glutamicum*: new insights into nitrogen regulation. J Biotechnol 126(1):101–110
147. Sonntag K, Eggeling L, De Graaf AA, Sahm H (1993) Flux partitioning in the split pathway of lysine synthesis in *Corynebacterium glutamicum*. Quantification by ^{13}C- and ^{1}H-NMR spectroscopy. Eur J Biochem 213(3):1325–1331

148. Becker J, Schäfer R, Kohlstedt M, Harder BJ, Borchert NS, Stöveken N, Bremer E, Wittmann C (2013) Systems metabolic engineering of *Corynebacterium glutamicum* for production of the chemical chaperone ectoine. Microb Cell Fact 12:110
149. Rey DA, Pühler A, Kalinowski J (2003) The putative transcriptional repressor McbR, member of the TetR-family, is involved in the regulation of the metabolic network directing the synthesis of sulfur containing amino acids in *Corynebacterium glutamicum*. J Biotechnol 103(1):51–65
150. Krömer JO, Bolten CJ, Heinzle E, Schröder H, Wittmann C (2008) Physiological response of *Corynebacterium glutamicum* to oxidative stress induced by deletion of the transcriptional repressor McbR. Microbiology 154(Pt 12):3917–3930
151. Krömer JO, Fritz M, Heinzle E, Wittmann C (2005) In vivo quantification of intracellular amino acids and intermediates of the methionine pathway in *Corynebacterium glutamicum*. Anal Biochem 340(1):171–173
152. Krömer JO, Heinzle E, Schröder H, Wittmann C (2006) Accumulation of homolanthionine and activation of a novel pathway for isoleucine biosynthesis in *Corynebacterium glutamicum* McbR deletion strains. J Bacteriol 188(2):609–618
153. Krömer JO, Heinzle E, Wittmann C (2006) Quantification of S-adenosyl methionine in microbial cell extracts. Biotechnol Lett 28(2):69–71
154. Toyoda K, Inui M (2015) Regulons of global transcription factors in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 100:45–60
155. Engels S, Schweitzer JE, Ludwig C, Bott M, Schaffer S (2004) clpC and clpP1P2 gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor σH. Mol Microbiol 52(1):285–302
156. Toyoda K, Teramoto H, Yukawa H, Inui M (2015) Expanding the regulatory network governed by the extracytoplasmic function sigma factor sigmaH in *Corynebacterium glutamicum*. J Bacteriol 197(3):483–496
157. Busche T, Silar R, Picmanova M, Patek M, Kalinowski J (2012) Transcriptional regulation of the operon encoding stress-responsive ECF sigma factor SigH and its anti-sigma factor RshA, and control of its regulatory network in *Corynebacterium glutamicum*. BMC Genomics 13:445
158. Osman A, Tzortzis G, Rastall RA, Charalampopoulos D (2013) High yield production of a soluble bifidobacterial beta-galactosidase (BbgIV) in *E. coli* DH5alpha with improved catalytic efficiency for the synthesis of prebiotic galactooligosaccharides. J Agric Food Chem 61(9):2213–2223
159. Lee JY, Kim HJ, Kim ES, Kim P, Kim Y, Lee HS (2013) Regulatory interaction of the *Corynebacterium glutamicum whc* genes in oxidative stress responses. J Biotechnol 168 (2):149–154
160. Inui M, Suda M, Okino S, Nonaka H, Puskás LG, Vertès AA, Yukawa H (2007) Transcriptional profiling of *Corynebacterium glutamicum* metabolism during organic acid production under oxygen deprivation conditions. Microbiology 153(8):2491–2504
161. Ohnishi J, Hayashi M, Mitsuhashi S, Ikeda M (2003) Efficient 40 degrees C fermentation of L-lysine by a new *Corynebacterium glutamicum* mutant developed by genome breeding. Appl Microbiol Biotechnol 62(1):69–75
162. Varela C, Agosin E, Baez M, Klapa M, Stephanopoulos G (2003) Metabolic flux redistribution in *Corynebacterium glutamicum* in response to osmotic stress. Appl Microbiol Biotechnol 60(5):547–555
163. Kohl TA, Baumbach J, Jungwirth B, Pühler A, Tauch A (2008) The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: in silico and in vitro detection of DNA binding sites of a global transcription regulator. J Biotechnol 135(4):340–350
164. Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. J Mol Biol 323(5):785–793

165. Schiraldi C, Alfano A, Cimini D, Rosa MD, Panariello A, Restaino OF (2012) Application of a 22L scale membrane bioreactor and cross-flow ultrafiltration to obtain purified chondroitin. Biotechnol Prog 28(4):1012–1018
166. Kim TH, Park JS, Kim HJ, Kim Y, Kim P, Lee HS (2005) The *whcE* gene of *Corynebacterium glutamicum* is important for survival following heat and oxidative stress. Biochem Biophys Res Commun 337(3):757–764
167. Kim H-J, Kim T-H, Kim Y, Lee H-S (2004) Identification and characterization of glxR, a gene involved in regulation of glyoxylate bypass in *Corynebacterium glutamicum*. J Bacteriol 186(11):3453–3460
168. Polen T, Schluesener D, Poetsch A, Bott M, Wendisch VF (2007) Characterization of citrate utilization in *Corynebacterium glutamicum* by transcriptome and proteome analysis. FEMS Microbiol Lett 273(1):109–119
169. Mentz A, Neshat A, Pfeifer-Sancar K, Pühler A, Rückert C, Kalinowski J (2013) Comprehensive discovery and characterization of small RNAs in *Corynebacterium glutamicum* ATCC 13032. BMC Genomics 14(1):714
170. Cimini D, De Rosa M, Carlino E, Ruggiero A, Schiraldi C (2013) Homologous overexpression of RfaH in *E. coli* K4 improves the production of chondroitin-like capsular polysaccharide. Microb Cell Fact 12:46
171. Pfeifer-Sancar K, Mentz A, Rückert C, Kalinowski J (2013) Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. BMC Genomics 14:888
172. Neshat A, Mentz A, Ruckert C, Kalinowski J (2014) Transcriptome sequencing revealed the transcriptional organization at ribosome-mediated attenuation sites in *Corynebacterium glutamicum* and identified a novel attenuator involved in aromatic amino acid biosynthesis. J Biotechnol 190:55–63
173. Santamaria R, Gil J, Mesas J, Martin J (1984) Characterization of an endogenous plasmid and development of cloning vectors and a transformation system in *Brevibacterium lactofermentum*. J Gen Microbiol 130:2237–2246
174. Miwa K, Matsui H, Terabe M, Nakamori S, Sano K, Momose H (1984) Cryptic plasmids in glutamic acid producing bacteria. Agric Biol Chem 48(11):2901–2903
175. Katsumata R, Ozaki A, Oka T, Furuya A (1984) Protoplast transformation of glutamate-producing bacteria with plasmid DNA. J Bacteriol 159(1):306–311
176. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfefferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J Biotechnol 104(1–3):5–25
177. Ikeda M, Nakagawa S (2003) The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. Appl Microbiol Biotechnol 62(2-3):99–109
178. Eikmanns BJ, Kleinertz E, Liebl W, Sahm H (1991) A family of *Corynebacterium glutamicum*/*Escherichia coli* shuttle vectors for cloning, controlled gene expression, and promoter probing. Gene 102(1):93–98
179. Jäger W, Schäfer A, Pühler A, Labes G, Wohlleben W (1992) Expression of the *Bacillus subtilis sacB* gene leads to sucrose sensitivity in the Gram-positive bacterium *Corynebacterium glutamicum* but not in *Streptomyces lividans*. J Bacteriol 174(16):5462–5465
180. Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, Pühler A (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145 (1):69–73
181. Liebl W, Bayerl A, Schein B, Stillner U, Schleifer KH (1989) High efficiency electroporation of intact *Corynebacterium glutamicum* cells. FEMS Microbiol Lett 65(3):299–303

182. Bonamy C, Guyonvarch A, Reyes O, David F, Leblon G (1990) Interspecies electrotransformation in Corynebacteria. FEMS Microbiol Lett 54(1-3):263–269
183. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6 (5):343–345
184. Nesvera J, Patek M (2011) Tools for genetic manipulations in *Corynebacterium glutamicum* and their applications. Appl Microbiol Biotechnol 90(5):1641–1654
185. Kirchner O, Tauch A (2003) Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. J Biotechnol 104(1-3):287–299
186. Vertes AA, Inui M, Yukawa H (2005) Manipulating corynebacteria, from individual genes to chromosomes. Appl Environ Microbiol 71(12):7633–7642
187. Schäfer A, Schwarzer A, Kalinowski J, Pühler A (1994) Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from *Corynebacterium glutamicum* ATCC 13032 and analysis of its role in intergeneric conjugation with *Escherichia coli*. J Bacteriol 176(23):7309–7319
188. Vertes AA, Asai Y, Inui M, Kobayashi M, Kurusu Y, Yukawa H (1994) Transposon mutagenesis of coryneform bacteria. Mol Gen Genet 245(4):397–405
189. Bonamy C, Labarre J, Cazaubon L, Jacob C, Le Bohec F, Reyes O, Leblon G (2003) The mobile element IS1207 of *Brevibacterium lactofermentum* ATCC21086: isolation and use in the construction of Tn5531, a versatile transposon for insertional mutagenesis of *Corynebacterium glutamicum*. J Biotechnol 104(1–3):301–309
190. Moreau S, Blanco C, Trautwetter A (1999) Site-specific integration of corynephage phi16: construction of an integration vector. Microbiology 145(Pt 3):539–548
191. Tan Y, Xu D, Li Y, Wang X (2012) Construction of a novel *sacB*-based system for marker-free gene deletion in *Corynebacterium glutamicum*. Plasmid 67(1):44–52
192. Okibe N, Suzuki N, Inui M, Yukawa H (2011) Efficient markerless gene replacement in *Corynebacterium glutamicum* using a new temperature-sensitive plasmid. J Microbiol Methods 85(2):155–163
193. Schäfer A, Kalinowski J, Pühler A (1994) Increased fertility of *Corynebacterium glutamicum* recipients in intergeneric matings with *Escherichia coli* after stress exposure. Appl Environ Microbiol 60(2):756–759
194. van der Rest ME, Lange C, Molenaar D (1999) A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. Appl Microbiol Biotechnol 52(4):541–545
195. Schäfer A, Kalinowski J, Simon R, Seep-Feldhaus AH, Pühler A (1990) High-frequency conjugal plasmid transfer from Gram-negative *Escherichia coli* to various Gram-positive coryneform bacteria. J Bacteriol 172(3):1663–1666
196. Reyes O, Guyonvarch A, Bonamy C, Salti V, David F, Leblon G (1991) 'Integron'-bearing vectors: a method suitable for stable chromosomal integration in highly restrictive corynebacteria. Gene 107(1):61–68
197. Kind S, Jeong WK, Schröder H, Wittmann C (2010) Systems-wide metabolic pathway engineering in *Corynebacterium glutamicum* for bio-based production of diaminopentane. Metab Eng 12(4):341–351
198. Horton RM (1995) PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. Mol Biotechnol 3(2):93–99
199. Ankri S, Reyes O, Leblon G (1996) Electrotransformation of highly DNA-restrictive corynebacteria with synthetic DNA. Plasmid 35(1):62–66
200. Vertes A, Hatakeyama K, Inui M, Kobayashi M, Kurusu Y, Yukawa H (1993) Replacement recombination in coryneform bacteria: high efficiency integration requirement for non-methylated plasmid DNA. Biosci Biotechnol Biochem 57:2036–2038
201. Cao W, Ma W, Zhang B, Wang X, Chen K, Li Y, Ouyang P (2016) Improved pinocembrin production in *Escherichia coli* by engineering fatty acid synthesis. J Ind Microbiol Biotechnol 43:557–566

202. Rytter JV, Helmark S, Chen J, Lezyk MJ, Solem C, Jensen PR (2014) Synthetic promoter libraries for *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 98(6):2617–2623
203. Yim SS, An SJ, Kang M, Lee J, Jeong KJ (2013) Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. Biotechnol Bioeng 110 (11):2959–2969
204. Okibe N, Suzuki N, Inui M, Yukawa H (2010) Isolation, evaluation and use of two strong, carbon source-inducible promoters from *Corynebacterium glutamicum*. Lett Appl Microbiol 50(2):173–180
205. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136(4):615–628
206. Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. Curr Opin Microbiol 10(2):134–139
207. He W, Fu L, Li G, Andrew Jones J, Linhardt RJ, Koffas M (2015) Production of chondroitin in metabolically engineered *E. coli*. Metab Eng 27:92–100
208. Vecerek B, Moll I, Blasi U (2007) Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. Embo J 26(4):965–975
209. Chang H, Replogle JM, Vather N, Tsao-Wu M, Mistry R, Liu JM (2015) A cis-regulatory antisense RNA represses translation in *Vibrio cholerae* through extensive complementarity and proximity to the target locus. RNA Biol 12(2):136–148
210. Chae TU, Kim WJ, Choi S, Park SJ, Lee SY (2015) Metabolic engineering of *Escherichia coli* for the production of 1,3-diaminopropane, a three carbon diamine. Sci Rep 5:13040
211. Cho KH, Kim JH (2015) Cis-encoded non-coding antisense RNAs in streptococci and other low GC Gram (+) bacterial pathogens. Front Genet 6:110
212. Storz G, Opdyke JA, Zhang A (2004) Controlling mRNA stability and translation with small, noncoding RNAs. Curr Opin Microbiol 7(2):140–144
213. Brownlee G (1971) Sequence of 6S RNA of *E. coli*. Nature 229(5):147–149
214. Wassarman KM (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. Cell 109(2):141–144
215. Borujeni AE, Dong E, Salis HM (2011) Automated design of synthetic bacterial small RNAS. In: The 5th annual q-bio conference on cellular information processing, Santa Fe, USA, 10–14 August 2011
216. Sharma V, Yamamura A, Yokobayashi Y (2012) Engineering artificial small RNAs for conditional gene silencing in *Escherichia coli*. ACS Synth Biol 1(1):6–13
217. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotech 31(2):170–174
218. Meyer S, Chappell J, Sankar S, Chew R, Lucks JB (2015) Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. Biotechnol Bioeng 113:216–225
219. Pain A, Ott A, Amine H, Rochat T, Bouloc P, Gautheret D (2015) An assessment of bacterial small RNA target prediction programs. RNA Biol 12(5):509–513
220. Zemanová M, Kadeřábková P, Pátek M, Knoppová M, Šilar R, Nešvera J (2008) Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*. FEMS Microbiol Lett 279(2):195–201
221. Yang SM, Shim GY, Kim BG, Ahn JH (2015) Biological synthesis of coumarins in *Escherichia coli*. Microb Cell Fact 14:65
222. Breaker RR (2012) Riboswitches and the RNA world. Cold Spring Harbor perspectives in biology 4(2):a003566
223. Jiang M, Stephanopoulos G, Pfeifer BA (2012) Toward biosynthetic design and implementation of *Escherichia coli*-derived paclitaxel and other heterologous polyisoprene compounds. Appl Environ Microbiol 78(8):2497–2504
224. Barreteau H, Richard E, Drouillard S, Samain E, Priem B (2012) Production of intracellular heparosan and derived oligosaccharides by lyase expression in metabolically engineered *E. coli* K-12. Carbohydr Res 360:19–24

225. Zhou L-B, Zeng A-P (2015) Engineering lysine-ON riboswitch for metabolic control of lysine production in *Corynebacterium glutamicum*. ACS Synth Biol
226. Teramoto H, Watanabe K, Suzuki N, Inui M, Yukawa H (2011) High yield secretion of heterologous proteins in *Corynebacterium glutamicum* using its own Tat-type signal sequence. Appl Microbiol Biotechnol 91(3):677–687
227. Ravasi P, Peiru S, Gramajo H, Menzella HG (2012) Design and testing of a synthetic biology framework for genetic engineering of *Corynebacterium glutamicum*. Microb Cell Fact 11:147
228. Baumgärtner F, Jurzitza L, Conrad J, Beifuss U, Sprenger GA, Albermann C (2015) Synthesis of fucosylated lacto-N-tetraose using whole-cell biotransformation. Bioorg Med Chem 23 (21):6799–6806
229. Schneider J, Eberhardt D, Wendisch V (2012) Improving putrescine production by *Corynebacterium glutamicum* by fine-tuning ornithine transcarbamoylase activity using a plasmid addiction system. Appl Microbiol Biotechnol 95(1):169–178
230. Schwechheimer SK, Park EY, Revuelta JL, Becker J, Wittmann C (2016) Biotechnology of riboflavin. Appl Microbiol Biotechnol 100:2107–2019
231. Lee WH, Pathanibul P, Quarterman J, Jo JH, Han NS, Miller MJ, Jin YS, Seo JH (2012) Whole cell biosynthesis of a functional oligosaccharide, 2′-fucosyllactose, using engineered *Escherichia coli*. Microb Cell Fact 11:48
232. Zhang C, Zou R, Chen X, Stephanopoulos G, Too HP (2015) Experimental design-aided systematic pathway optimization of glucose uptake and deoxyxylulose phosphate pathway for improved amorphadiene production. Appl Microbiol Biotechnol 99(9):3825–3837
233. Restaino OF, Cimini D, De Rosa M, Catapano A, Schiraldi C (2011) High cell density cultivation of *Escherichia coli* K4 in a microfiltration bioreactor: a step towards improvement of chondroitin precursor production. Microb Cell Fact 10:10
234. Ikeda M, Mitsuhashi S, Tanaka K, Hayashi M (2009) Reengineering of a *Corynebacterium glutamicum* L-arginine and L-citrulline producer. Appl Environ Microbiol 75(6):1635–1641
235. Schneider J, Wendisch VF (2010) Putrescine production by engineered *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 88(4):859–868
236. Okai N, Miyoshi T, Takeshima Y, Kuwahara H, Ogino C, Kondo A (2015) Production of protocatechuic acid by *Corynebacterium glutamicum* expressing chorismate-pyruvate lyase from *Escherichia coli*. Appl Microbiol Biotechnol 100:135–145
237. Kawaguchi H, Sasaki K, Uematsu K, Tsuge Y, Teramura H, Okai N, Nakamura-Tsuruta S, Katsuyama Y, Sugai Y, Ohnishi Y, Hirano K, Sazuka T, Ogino C, Kondo A (2015) 3-Amino-4-hydroxybenzoic acid production from sweet sorghum juice by recombinant *Corynebacterium glutamicum*. Bioresour Technol 198:410–417
238. Cheng F, Gong Q, Yu H, Stephanopoulos G (2015) High-titer biosynthesis of hyaluronic acid by recombinant *Corynebacterium glutamicum*. Biotechnol J 11:574–584
239. Zhang S, Wang S, Zhan J (2016) Engineered biosynthesis of medicinally important plant natural products in microorganisms. Curr Top Med Chem 16(15):1740–1754
240. Feng L, Zhang Y, Fu J, Mao Y, Chen T, Zhao X, Wang Z (2015) Metabolic engineering of *Corynebacterium glutamicum* for efficient production of 5-aminolevulinic acid. Biotechnol Bioeng 113:1284–1293
241. Tanaka T, Kondo A (2015) Cell surface engineering of industrial microorganisms for biorefining applications. Biotechnol Adv 33(7):1403–1411
242. Cremer J, Eggeling L, Sahm H (1991) Control of the lysine biosynthesis sequence in *Corynebacterium glutamicum* as analyzed by overexpression of the individual corresponding genes. Appl Environ Microbiol 57(6):1746–1752
243. Ikeda M (2006) Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. Appl Microbiol Biotechnol 69(6):615–626
244. Kohlstedt M, Becker J, Wittmann C (2010) Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism. Appl Microbiol Biotechnol 88 (5):1065–1075

245. Dai Z, Nielsen J (2015) Advancing metabolic engineering through systems biology of industrial microorganisms. Curr Opin Biotechnol 36:8–15
246. Lee SY, Kim HU (2015) Systems strategies for developing industrial microbial strains. Nat Biotechnol 33(10):1061–1072
247. Lee JY, Seo J, Kim ES, Lee HS, Kim P (2013) Adaptive evolution of *Corynebacterium glutamicum* resistant to oxidative stress and its global gene expression profiling. Biotechnol Lett 35(5):709–717
248. Oide S, Gunji W, Moteki Y, Yamamoto S, Suda M, Jojima T, Yukawa H, Inui M (2015) Thermal and solvent stress cross-tolerance conferred to *Corynebacterium glutamicum* by adaptive laboratory evolution. Appl Environ Microbiol 81(7):2284–2298
249. Lessmeier L, Wendisch VF (2015) Identification of two mutations increasing the methanol tolerance of *Corynebacterium glutamicum*. BMC Microbiol 15(1):216
250. Mahr R, Gatgens C, Gatgens J, Polen T, Kalinowski O, Frunzke J (2015) Biosensor-driven adaptive laboratory evolution of l-valine production in *Corynebacterium glutamicum*. Metab Eng 32:184–194
251. Wendisch VF (2014) Microbial production of amino acids and derived chemicals: synthetic biology approaches to strain development. Curr Opin Biotechnol 30C:51–58
252. Woo HM, Park JB (2014) Recent progress in development of synthetic biology platforms and metabolic engineering of *Corynebacterium glutamicum*. J Biotechnol 180:43–51
253. Stäbler N, Oikawa T, Bott M, Eggeling L (2011) *Corynebacterium glutamicum* as a host for synthesis and export of D-amino acids. J Bacteriol 193(7):1702–1709
254. Shi F, Li Y (2011) Synthesis of gamma-aminobutyric acid by expressing *Lactobacillus brevis*-derived glutamate decarboxylase in the *Corynebacterium glutamicum* strain ATCC 13032. Biotechnol Lett 33(12):2469–2474
255. Takahashi C, Shirakawa J, Tsuchidate T, Okai N, Hatada K, Nakayama H, Tateno T, Ogino C, Kondo A (2012) Robust production of gamma-amino butyric acid using recombinant *Corynebacterium glutamicum* expressing glutamate decarboxylase from *Escherichia coli*. Enzyme Microb Technol 51(3):171–176
256. Hüser AT, Chassagnole C, Lindley ND, Merkamm M, Guyonvarch A, Elisakova V, Patek M, Kalinowski J, Brune I, Pühler A, Tauch A (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(6):3255–3268
257. Dickschat J, Wickel S, Bolten CJ, Nawrath T, Schulz S, Wittmann C (2010) Pyrazine biosynthesis in *Corynebacterium glutamicum*. Eur J Org Chem 2010(14):2687–2695
258. Heider SA, Peters-Wendisch P, Wendisch VF (2012) Carotenoid biosynthesis and overproduction in *Corynebacterium glutamicum*. BMC Microbiol 12:198
259. Heider SA, Peters-Wendisch P, Netzer R, Stafnes M, Brautaset T, Wendisch VF (2014) Production and glucosylation of C50 and C 40 carotenoids by metabolically engineered *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 98(3):1223–1235
260. Eggeling L, Bott M (2015) A giant market and a powerful metabolism: L-lysine provided by *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 99:3387–3394
261. Kalinowski J, Cremer J, Bachmann B, Eggeling L, Sahm H, Pühler A (1991) Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. Mol Microbiol 5(5):1197–1204
262. Vrljic M, Sahm H, Eggeling L (1996) A new type of transporter with a new type of cellular function: L-lysine export from *Corynebacterium glutamicum*. Mol Microbiol 22(5):815–826
263. Jiang LY, Chen SG, Zhang YY, Liu JZ (2013) Metabolic evolution of *Corynebacterium glutamicum* for increased production of L-ornithine. BMC Biotechnol 13:47
264. Kubota K, Onoda T, Kamijo H, Yoshinaga F, Okumura S (1973) Production of L-arginine by mutants of glutamic acid-producing bacteria. J Gen Appl Microbiol 19:339–352
265. Utagawa T (2004) Production of arginine by fermentation. J Nutr 134 (10 Suppl):2854S–2857S; discussion 2895S.

266. Udaka S, Kinoshita S (1958) Studies on L-ornithine fermentation I. – The biosynthetic pathway of L-ornithine in *Micrococcus glutamicus*. J Gen Appl Microbiol 4(4):272–275
267. Udaka S, Kinoshita S (1958) Studies on L-ornithine fermentation II. – The change of fermentation product by a feedback type mechanism. J Gen Appl Microbiol 4(4):283–288
268. Becker J, Wittmann C (2016) Diamines for bio-based materials. In: Wittmann C, Liao JC (eds) Industrial biotechnology. Advanced biotechnology. Wiley-VCH, Weinheim, 395–413
269. Petri K, Walter F, Persicke M, Rückert C, Kalinowski J (2013) A novel type of N-acetylglutamate synthase is involved in the first step of arginine biosynthesis in *Corynebacterium glutamicum*. BMC Genomics 14:713
270. Hwang GH, Cho JY (2014) Enhancement of L-ornithine production by disruption of three genes encoding putative oxidoreductases in *Corynebacterium glutamicum*. J Ind Microbiol Biotechnol 41(3):573–578
271. Jiang LY, Zhang YY, Li Z, Liu JZ (2013) Metabolic engineering of *Corynebacterium glutamicum* for increasing the production of L-ornithine by increasing NADPH availability. J Ind Microbiol Biotechnol 40(10):1143–1151
272. Pastor JM, Salvador M, Argandona M, Bernal V, Reina-Bueno M, Csonka LN, Iborra JL, Vargas C, Nieto JJ, Canovas M (2010) Ectoines in cell stress protection: uses and biotechnological production. Biotechnol Adv 28(6):782–801
273. Kunte HJ, Lentzen G, Galinski EA (2014) Industrial production of the cell protectant ectoine: protection mechanisms, processes, and products. Curr Biotechnol 3:10–25
274. Stöveken N, Pittelkow M, Sinner T, Jensen RA, Heider J, Bremer E (2011) A specialized aspartokinase enhances the biosynthesis of the osmoprotectants ectoine and hydroxyectoine in *Pseudomonas stutzeri* A1501. J Bacteriol 193(17):4456–4468
275. Kempf B, Bremer E (1998) Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol 170(5):319–330
276. Sauer T, Galinski EA (1998) Bacterial milking: a novel bioprocess for production of compatible solutes. Biotechnol Bioeng 57(3):306–313
277. Frohwitter J, Heider SA, Peters-Wendisch P, Beekwilder J, Wendisch VF (2014) Production of the sesquiterpene (+)-valencene by metabolically engineered *Corynebacterium glutamicum*. J Biotechnol 191:205–213
278. Kind S, Wittmann C (2011) Bio-based production of the platform chemical 1,5-diaminopentane. Appl Microbiol Biotechnol 91(5):1287–1296
279. Iles A, Martin AN (2013) Expanding bioplastics production: sustainable business innovation in the chemical industry. J Clean Prod 45:38–49
280. Becker J, Lange A, Fabarius J, Wittmann C (2015) Top value platform chemicals: bio-based production of organic acids. Curr Opin Biotechnol 36:168–175
281. Litsanov B, Brocker M, Oldiges M, Bott M (2014) Succinic acid. In: Bisaria VS, Kondo A (eds) Bioprocessing of renewable resources to commodity bioproducts. Wiley, Hoboken, pp 435–472
282. Mimitsuka T, Sawai H, Hatsu M, Yamada K (2007) Metabolic engineering of *Corynebacterium glutamicum* for cadaverine fermentation. Biosci Biotechnol Biochem 71(9):2130–2135
283. Okino S, Suda M, Fujikura K, Inui M, Yukawa H (2008) Production of D-lactic acid by *Corynebacterium glutamicum* under oxygen deprivation. Appl Microbiol Biotechnol 78 (3):449–454
284. Siebert D, Wendisch VF (2015) Metabolic pathway engineering for production of 1,2-propanediol and 1-propanol by *Corynebacterium glutamicum*. Biotechnol Biofuels 8:91
285. Niimi S, Suzuki N, Inui M, Yukawa H (2011) Metabolic engineering of 1,2-propanediol pathways in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90(5):1721–1729
286. Matsumoto K, Kitagawa K, Jo SJ, Song Y, Taguchi S (2011) Production of poly (3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Corynebacterium glutamicum* using propionate as a precursor. J Biotechnol 152(4):144–146
287. Matsumoto K, Yamada M, Leong CR, Jo SJ, Kuzuyama T, Taguchi S (2011) A new pathway for poly(3-hydroxybutyrate) production in *Escherichia coli* and *Corynebacterium glutamicum* by functional expression of a new acetoacetyl-coenzyme A synthase. Biosci Biotechnol Biochem 75(2):364–366

288. Matsumoto K, Tobitani K, Aoki S, Song Y, Ooi T, Taguchi S (2014) Improved production of poly(lactic acid)-like polyester based on metabolite analysis to address the rate-limiting step. AMB Express 4(1):83
289. Tsuge Y, Hasunuma T, Kondo A (2015) Recent advances in the metabolic engineering of *Corynebacterium glutamicum* for the production of lactate and succinate from renewable resources. J Ind Microbiol Biotechnol 42(3):375–389
290. Okino S, Noburyu R, Suda M, Jojima T, Inui M, Yukawa H (2008) An efficient succinic acid production process in a metabolically engineered *Corynebacterium glutamicum* strain. Appl Microbiol Biotechnol 81(3):459–464
291. Litsanov B, Brocker M, Bott M (2012) Toward homosuccinate fermentation: metabolic engineering of *Corynebacterium glutamicum* for anaerobic production of succinate from glucose and formate. Appl Environ Microbiol 78(9):3325–3337
292. Zhou Z, Wang C, Chen Y, Zhang K, Xu H, Cai H, Chen Z (2015) Increasing available NADH supply during succinic acid production by *Corynebacterium glutamicum*. Biotechnol Prog 31 (1):12–19
293. Zhu N, Xia H, Wang Z, Zhao X, Chen T (2013) Engineering of acetate recycling and citrate synthase to improve aerobic succinate production in *Corynebacterium glutamicum*. PLoS One 8(4), e60659
294. Litsanov B, Kabus A, Brocker M, Bott M (2012) Efficient aerobic succinate production from glucose in minimal medium with *Corynebacterium glutamicum*. Microb Biotechnol 5 (1):116–128
295. Yamauchi Y, Hirasawa T, Nishii M, Furusawa C, Shimizu H (2014) Enhanced acetic acid and succinic acid production under microaerobic conditions by *Corynebacterium glutamicum* harboring *Escherichia coli* transhydrogenase gene *pntAB*. J Gen Appl Microbiol 60 (3):112–118
296. Becker J, Reinefeld J, Stellmacher R, Schäfer R, Lange A, Meyer H, Lalk M, Zelder O, von Abendroth G, Schröder H, Haefner S, Wittmann C (2013) Systems-wide analysis and engineering of metabolic pathway fluxes in bio-succinate producing *Basfia succiniciproducens*. Biotechnol Bioeng 110(11):3013–3023
297. Lee SJ, Song H, Lee SY (2006) Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. Appl Environ Microbiol 72(3):1939–1948
298. Kind S, Jeong WK, Schröder H, Zelder O, Wittmann C (2010) Identification and elimination of the competing N-acetyldiaminopentane pathway for improved production of diaminopentane by *Corynebacterium glutamicum*. Appl Environ Microbiol 76 (15):5175–5180
299. Kind S, Kreye S, Wittmann C (2011) Metabolic engineering of cellular transport for overproduction of the platform chemical 1,5-diaminopentane in *Corynebacterium glutamicum*. Metab Eng 13(5):617–627
300. Blombach B, Riester T, Wieschalka S, Ziert C, Youn JW, Wendisch VF, Eikmanns BJ (2011) *Corynebacterium glutamicum* tailored for efficient isobutanol production. Appl Environ Microbiol 77(10):3300–3310
301. Heider SA, Wendisch VF (2015) Engineering microbial cell factories: metabolic engineering of *Corynebacterium glutamicum* with a focus on non-natural products. Biotechnol J 10 (8):1170–1184
302. Blombach B, Arndt A, Auchter M, Eikmanns BJ (2009) L-Valine production during growth of pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* in the presence of ethanol or by inactivation of the transcriptional regulator SugR. Appl Environ Microbiol 75(4):1197–1200
303. Blombach B, Schreiner ME, Bartek T, Oldiges M, Eikmanns BJ (2008) *Corynebacterium glutamicum* tailored for high-yield L-valine production. Appl Microbiol Biotechnol 79 (3):471–479

304. Yamamoto S, Suda M, Niimi S, Inui M, Yukawa H (2013) Strain optimization for efficient isobutanol production using *Corynebacterium glutamicum* under oxygen deprivation. Biotechnol Bioeng 110(11):2938–2948
305. Barrett E, Stanton C, Zelder O, Fitzgerald G, Ross RP (2004) Heterologous expression of lactose- and galactose-utilizing pathways from lactic acid bacteria in *Corynebacterium glutamicum* for production of lysine in whey. Appl Environ Microbiol 70(5):2861–2866
306. Brabetz W, Liebl W, Schleifer KH (1991) Studies on the utilization of lactose by *Corynebacterium glutamicum*, bearing the lactose operon of *Escherichia coli*. Arch Microbiol 155 (6):607–612
307. Matano C, Uhde A, Youn JW, Maeda T, Clermont L, Marin K, Krämer R, Wendisch VF, Seibold GM (2014) Engineering of *Corynebacterium glutamicum* for growth and L-lysine and lycopene production from N-acetyl-glucosamine. Appl Microbiol Biotechnol 98:5633–5643
308. Kim EM, Um Y, Bott M, Woo HM (2015) Engineering of *Corynebacterium glutamicum* for growth and succinate production from levoglucosan, a pyrolytic sugar substrate. FEMS Microbiol Lett 362(19)
309. Seibold G, Auchter M, Berens S, Kalinowski J, Eikmanns BJ (2006) Utilization of soluble starch by a recombinant *Corynebacterium glutamicum* strain: growth and lysine production. J Biotechnol 124(2):381–391
310. Tateno T, Fukuda H, Kondo A (2007) Direct production of L-lysine from raw corn starch by C*orynebacterium glutamicum* secreting *Streptococcus bovis* alpha-amylase using *cspB* promoter and signal sequence. Appl Microbiol Biotechnol 77(3):533–541
311. Tateno T, Fukuda H, Kondo A (2007) Production of L-lysine from starch by *Corynebacterium glutamicum* displaying alpha-amylase on its cell surface. Appl Microbiol Biotechnol 74 (6):1213–1220
312. Tsuge Y, Tateno T, Sasaki K, Hasunuma T, Tanaka T, Kondo A (2013) Direct production of organic acids from starch by cell surface-engineered *Corynebacterium glutamicum* in anaerobic conditions. AMB Express 3(1):72
313. Witthoff S, Schmitz K, Niedenfuhr S, Nöh K, Noack S, Bott M, Marienhagen J (2015) Metabolic engineering of *Corynebacterium glutamicum* for methanol metabolism. Appl Environ Microbiol 81(6):2215–2225
314. Meiswinkel TM, Rittmann D, Lindner SN, Wendisch VF (2013) Crude glycerol-based production of amino acids and putrescine by *Corynebacterium glutamicum*. Bioresour Technol 145:254–258
315. Rittmann D, Lindner SN, Wendisch VF (2008) Engineering of a glycerol utilization pathway for amino acid production by *Corynebacterium glutamicum*. Appl Environ Microbiol 74 (20):6216–6222
316. Rumbold K, van Buijsen HJ, Overkamp KM, van Groenestijn JW, Punt PJ, van der Werf MJ (2009) Microbial production host selection for converting second-generation feedstocks into bioproducts. Microb Cell Fact 8:64
317. Sakai S, Tsuchida Y, Nakamoto H, Okino S, Ichihashi O, Kawaguchi H, Watanabe T, Inui M, Yukawa H (2007) Effect of lignocellulose-derived inhibitors on growth of and ethanol production by growth-arrested *Corynebacterium glutamicum* R. Appl Environ Microbiol 73 (7):2349–2353
318. Tsuge Y, Hori Y, Kudou M, Ishii J, Hasunuma T, Kondo A (2014) Detoxification of furfural in *Corynebacterium glutamicum* under aerobic and anaerobic conditions. Appl Microbiol Biotechnol 98(20):8675–8683
319. Tsuge Y, Kudou M, Kawaguchi H, Ishii J, Hasunuma T, Kondo A (2015) FudC, a protein primarily responsible for furfural detoxification in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 100:2685–2692.
320. Liu YB, Chen C, Chaudhry MT, Si MR, Zhang L, Wang Y, Shen XH (2014) Enhancing *Corynebacterium glutamicum* robustness by over-expressing a gene, *mshA*, for mycothiol glycosyltransferase. Biotechnol Lett 36(7):1453–1459

321. den Haan R, van Rensburg E, Rose SH, Gorgens JF, van Zyl WH (2015) Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Curr Opin Biotechnol 33:32–38
322. Adachi N, Takahashi C, Ono-Murota N, Yamaguchi R, Tanaka T, Kondo A (2013) Direct L-lysine production from cellobiose by *Corynebacterium glutamicum* displaying beta-glucosidase on its cell surface. Appl Microbiol Biotechnol 97(16):7165–7172
323. Sasaki M, Jojima T, Inui M, Yukawa H (2008) Simultaneous utilization of D-cellobiose, D-glucose, and D-xylose by recombinant *Corynebacterium glutamicum* under oxygen-deprived conditions. Appl Microbiol Biotechnol 81(4):691–699
324. Kotrba P, Inui M, Yukawa H (2003) A single V317A or V317M substitution in Enzyme II of a newly identified beta-glucoside phosphotransferase and utilization system of *Corynebacterium glutamicum* R extends its specificity towards cellobiose. Microbiology 149 (Pt 6):1569–1580
325. Tsuchidate T, Tateno T, Okai N, Tanaka T, Ogino C, Kondo A (2011) Glutamate production from beta-glucan using endoglucanase-secreting *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90(3):895–901
326. Hyeon JE, Jeon WJ, Whang SY, Han SO (2011) Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant *Corynebacterium glutamicum*. Enzyme Microb Technol 48(4-5):371–377
327. Kim SJ, Hyeon JE, Jeon SD, Choi GW, Han SO (2014) Bi-functional cellulases complexes displayed on the cell surface of *Corynebacterium glutamicum* increase hydrolysis of lignocelluloses at elevated temperature. Enzyme Microb Technol 66:67–73
328. Song Y, Matsumoto K, Tanaka T, Kondo A, Taguchi S (2013) Single-step production of polyhydroxybutyrate from starch by using alpha-amylase cell-surface displaying system of *Corynebacterium glutamicum*. J Biosci Bioeng 115(1):12–14
329. Tateno T, Okada Y, Tsuchidate T, Tanaka T, Fukuda H, Kondo A (2009) Direct production of cadaverine from soluble starch using *Corynebacterium glutamicum* coexpressing alpha-amylase and lysine decarboxylase. Appl Microbiol Biotechnol 82(1):115–121
330. Parisutham V, Kim TH, Lee SK (2014) Feasibilities of consolidated bioprocessing microbes: from pretreatment to biofuel production. Bioresour Technol 161:431–440
331. Zheng P, Liu M, Liu XD, Du QY, Ni Y, Sun ZH (2012) Genome shuffling improves thermotolerance and glutamic acid production of *Corynebacteria glutamicum*. World J Microbiol Biotechnol 28(3):1035–1043
332. Yim SS, Choi JW, Lee SH, Jeong KJ (2016) Modular optimization of a hemicellulose-utilizing pathway in *Corynebacterium glutamicum* for consolidated bioprocessing of hemicellulosic biomass. ACS Synth Biol 5(4):334–343
333. Lee J, Saddler JN, Um Y, Woo HM (2016) Adaptive evolution and metabolic engineering of a cellobiose- and xylose-negative *Corynebacterium glutamicum* that co-utilizes cellobiose and xylose. Microb Cell Fact 15:20
334. Choi S, Song CW, Shin JH, Lee SY (2015) Biorefineries for the production of top building block chemicals and their derivatives. Metab Eng 28:223–239
335. Hasegawa S, Suda M, Uematsu K, Natsuma Y, Hiraga K, Jojima T, Inui M, Yukawa H (2013) Engineering of *Corynebacterium glutamicum* for high-yield L-valine production under oxygen deprivation conditions. Appl Environ Microbiol 79(4):1250–1257

Adv Biochem Eng Biotechnol (2018) 162: 265–300
DOI: 10.1007/10_2017_2

Published online: 20 April 2017

Synergizing ^{13}C Metabolic Flux Analysis and Metabolic Engineering for Biochemical Production

Weihua Guo, Jiayuan Sheng, and Xueyang Feng

Abstract Metabolic engineering of industrial microorganisms to produce chemicals, fuels, and drugs has attracted increasing interest as it provides an environment-friendly and renewable route that does not depend on depleting petroleum sources. However, the microbial metabolism is so complex that metabolic engineering efforts often have difficulty in achieving a satisfactory yield, titer, or productivity of the target chemical. To overcome this challenge, ^{13}C Metabolic Flux Analysis (^{13}C-MFA) has been developed to investigate rigorously the cell metabolism and quantify the carbon flux distribution in central metabolic pathways. In the past decade, ^{13}C-MFA has been widely used in academic labs and the biotechnology industry to pinpoint the key issues related to microbial-based chemical production and to guide the development of the appropriate metabolic engineering strategies for improving the biochemical production. In this chapter we introduce the basics of ^{13}C-MFA and illustrate how ^{13}C-MFA has been applied to synergize with metabolic engineering to identify and tackle the rate-limiting steps in biochemical production.

Keywords Biofuels, Bottleneck, Cell metabolism, Cofactor imbalance, Isotope, Synthetic biology

Contents

W. Guo, J. Sheng, and X. Feng (✉)
Department of Biological Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
e-mail: xueyang@vt.edu

1 Introduction

Producing chemicals from renewable resources would reduce strong dependence on petroleum and damage to the environment. Recently, with the development of metabolic engineering and synthetic biology, microbial production of a wide range of bulk chemicals [1–4], biofuels [5–9], and drugs [10–17] from renewable feedstock has been achieved successfully with many industrial microorganisms such as *Escherichia coli* [18–23] and *Saccharomyces cerevisiae* [24–27]. Among all the biosynthesized chemicals, however, only a few have achieved a satisfactory production level with a titer, yield, and productivity high enough for industrial commercialization [28, 29]. Therefore, it is crucial to develop novel strategies in metabolic engineering to improve microbial-based chemical production.

One of the main reasons for the low production level of engineered microorganisms is the complexity of cell metabolism [29]. Microbial production of chemicals is more than converting the precursors to the products. Rather, the microbial metabolism needs to coordinate the carbon flux [30, 31], cofactor supply [32–34], cell maintenance [10, 35, 36], and other factors [37–40] to achieve the production of target chemicals at a high level. The metabolic engineering strategies adopted to manipulate microbial metabolism often only focus on a few known challenges (e.g., poor gene expression) but also introduce new problems (e.g., metabolic burden) which prevent the microorganisms from achieving high-level chemical production. Such complex behavior of microbial physiology presents one of the biggest obstacles in current microbial-based chemical production.

To elucidate the metabolic rewiring of microorganisms and, more importantly, derive the appropriate strategy to engineer microorganisms for biochemical production, a technology named ^{13}C Metabolic Flux Analysis (^{13}C-MFA) was developed in the 1990s [41–46]. Basically, ^{13}C-MFA uses carbon isotopes to trace the cell metabolism and employs mathematical modeling to uncover the carbon flux distributions in metabolic networks of microorganisms [41, 42, 47–50]. By comparing the variations of metabolic fluxes among different engineered microorganisms, the key

issues, such as the bottleneck pathway, could often be discovered and hence guide the bioengineers to develop more appropriate metabolic engineering strategies [36, 51–56] to improve chemical production. In the past decade, we have witnessed many successful applications of ^{13}C-MFA to help metabolic engineers improve the microbial production of chemicals [31, 35, 52, 54, 57, 58], and ^{13}C-MFA has been widely recognized as one of the most important tools to diagnose microbial metabolism and develop novel metabolic engineering strategies [30, 31, 33, 35, 36, 54, 59–61].

In this chapter we summarize the synergistic tactics of ^{13}C-MFA and metabolic engineering from cases of improving microbial-based chemical production in the past decade. We first briefly introduce the principle of ^{13}C-MFA, and then categorize the ways in which ^{13}C-MFA synergizes with metabolic engineering into four groups: (1) uncovering the bottleneck steps in biochemical production, (2) identifying cofactor imbalance issues of host metabolism, (3) revealing cell maintenance requirement of industrial microorganisms, and (4) elucidating the mechanism of microbial resistance to fermentation inhibitors. We also point to emerging areas where breakthroughs of ^{13}C-MFA could potentially benefit rational metabolic engineering for improving microbial-based chemical production in the near future.

2 Technology Platform of ^{13}C Metabolic Flux Analysis

The technology platform of ^{13}C-MFA was first developed in the 1990s [41–46]. In the past two decades, mathematical algorithms and high-throughput mass spectrometry technology have been rapidly developed and have enabled more accurate quantitative analyses of metabolic fluxes for a broad scope of species. Because several protocols have been published to describe the procedures for both model and non-model organisms [47, 62], we focus on providing a concise introduction for ^{13}C-MFA, which includes cell culture and fermentation, isotopic analysis of metabolites, and ^{13}C-assisted pathway and flux analysis (Fig. 1).

2.1 Cell Culture and Fermentation

Cell culture on ^{13}C-labeled carbon substrates is the first step for ^{13}C-MFA and plays a vital role for the entire analysis. Three key factors have been recognized in this step, namely the composition of the medium, the cultivation mode, and the selection of ^{13}C-labeled substrates.

First, a strictly minimal medium with a single carbon source is often required for the ^{13}C-labeling experiments. This is because multiple carbon substrates and unlabeled nutrients could be assimilated by microorganisms, which "dilutes" the isotopic labeling of key metabolites, mystifies the carbon fate of metabolites of interests, and increases the difficulty of both carbon consumption measurements and accurate flux calculations [63]. However, it is worth mentioning that for certain

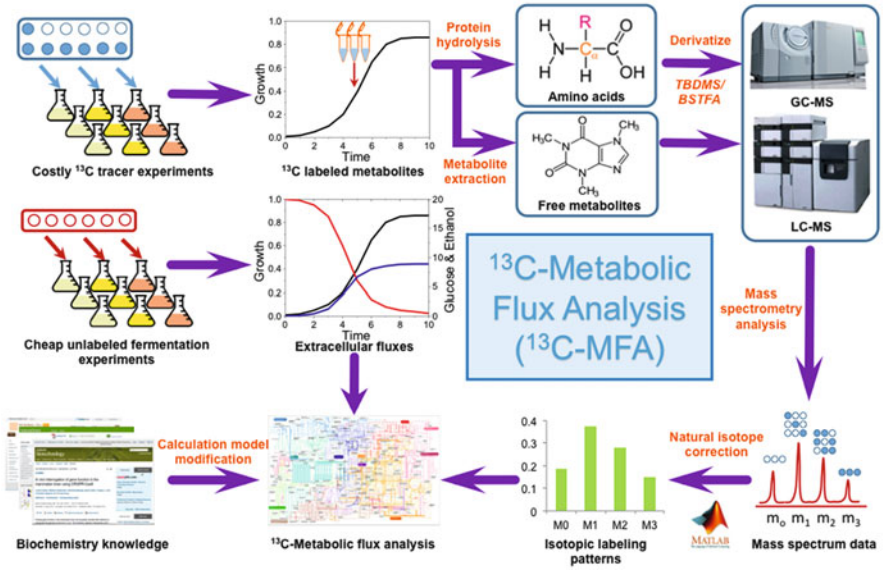

Fig. 1 Technology platform for ^{13}C-MFA

genetically engineered strains (e.g., *S. cerevisiae*) with auxotrophic markers, a trace amount of unlabeled exogenous amino acids could be supplemented into the minimal medium in order to support cell growth. Recently, several studies have reported alternative approaches for calculating the intracellular fluxes with complex medium composition and/or additional nutrients [64, 65].

Second, ^{13}C-MFA traditionally focuses on the metabolic flux distributions at the metabolic steady states, requiring both metabolic and isotopic steady states of microorganisms, that is, the concentration and isotopic labeling of intracellular metabolites do not change. Such requirements can be met by culturing microorganisms in either of the two modes: (1) batch mode, often using shaking flasks or culture tubes to culture microorganisms and harvesting biomass samples in log growth phase as a "pseudo" metabolic and isotopic steady state and (2) chemostat mode, often using bioreactors with continuous feeding to culture microorganisms and harvest biomass samples after two to three generations as the "real" metabolic and isotopic steady state. Although a chemostat setup can precisely control the desired metabolic status for metabolic flux analysis, the batch mode is simpler and more cost-effective. Till now, the majority of the ^{13}C-MFA in academic labs have been accomplished by sampling the ^{13}C-labeled biomass in late-log or early stationary growth phase when culturing microorganisms in batch mode [62]. Several advanced ^{13}C-assisted flux analysis approaches can also be implemented at either metabolic non-steady-state [30] or isotopic non-steady-state [66–70] to uncover the kinetic behaviors of intracellular metabolic rewiring by using novel computational tools (discussed in Sects. 4.2 and 4.3).

Third, the choice of ^{13}C-labeled substrate that should be used for ^{13}C-MFA is case-specific. In general, traditional ^{13}C-labeled glucose composition, that is, 80 wt % [1-^{13}C] and 20 wt% [U-^{13}C] glucose, can easily introduce sufficient ^{13}C carbons into the metabolites of interests for accurate mass spectrometry analysis and further flux analysis [47, 71–74]. On the other hand, pure and singly labeled carbon substrates are more sensitive for detection of novel pathways because it is easier to trace labeled carbons in intermediate metabolites. For example, the [3-^{13}C] lactate was used to investigate the biofilm metabolism of *Shewanella oneidensis* MR-1 and elucidated the heavy use of C1 metabolism in biofilm cells [75]. In brief, it was expected that ^{13}C would accumulate in most metabolites because it was difficult to remove the labeled carbon of the lactate from the *S. oneidensis* cells through well-known central pathways such as the TCA cycle. However, the high concentration of unlabeled metabolites in the biofilm cells indicated the high activity of C1 metabolism, which was the only known metabolic pathway to release the labeled carbon of lactate from *S. oneidensis* cells as $^{13}CO_2$. Additionally, multiple ^{13}C tracers are also used sometimes as they can also improve the flux resolution.

2.2 *Isotopic Analysis of Metabolites*

Experimental measurements of isotopic labeling of ^{13}C-labeled metabolites, for example, proteinogenic amino acids, are often achieved by mass spectrometry, which detects the fractions of the total population of any molecular fragment that is unlabeled, singly labeled, doubly labeled, etc. By correcting the effects of naturally labeled isotopes on the analysis of ^{13}C-labeled metabolites, we can obtain the isotopic distributions for the metabolites of interest with high sensitivity and use them as isotopic "fingerprints" to determine the metabolic fluxes [47, 76, 77]. Generally, three major procedures are commonly used to obtain such isotopic "fingerprints": metabolite extraction and separation, isotopic labeling detection, and correction of natural isotopomers.

Overall, many of the metabolite candidates for the flux analysis need to be extracted from cell biomass or culture medium. Sometimes the intracellular metabolites have low abundance and stability. Thus, a quick metabolite quenching method and sensitive mass spectrometry are often used to collect the isotopic labeling data [78]. The extracted metabolites could either be treated with a low heat derivatization group, *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) or *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA), followed by analysis in gas chromatography–mass spectrometry (GC-MS) [43], or directly injected without any treatment into liquid chromatography–mass spectrometry (LC-MS) [79], a machine that has much higher sensitivity than GC-MS.

For those metabolites with high abundance and stability, such as over-produced chemicals [79] and proteinogenic amino acids, the quenching step could be bypassed. Instead, the samples are often treated with *N*-*tert*-butyldimethylsilyl-*N*-

methyltrifluoroacetamide (MTBSTFA), a cheap and commonly used derivatization group, in a high heat process, followed by GC-MS analysis of the isotopic labeling. The derivatization process renders the molecules volatile enough to enter the GC column but also introduces considerable amounts of naturally labeled isotopes. Therefore, a systematic correction is required for the raw mass isotopomer spectrum prior to flux calculation. Several algorithms [80–82] have been well-established to curate the isotopic labeling and remove the effects of natural isotopes so that a mass distribution vector (MDV) for each metabolite can be generated and directly used for the pathway and flux analysis.

2.3 ^{13}C-Assisted Pathway and Flux Analysis

Based on the corrected MDV, the metabolic behaviors of microorganisms can be elucidated both qualitatively (i.e., pathway analysis) and quantitatively (i.e., flux analysis). On one hand, the ^{13}C-assisted pathway analysis often aims to answer whether a metabolic pathway is active in non-model microorganisms by measuring the ^{13}C-labeled patterns (i.e., MDVs) in key metabolites and determining the fate of biomolecule synthesis in the denoted biochemical pathways. One example of the ^{13}C-assisted pathway analysis is the discovery of C1 metabolism in biofilm *S. oneidensis* as mentioned above. On the other hand, the ^{13}C-assisted flux analysis aims to quantify the carbon fluxes in multiple metabolic pathways by simulating the ^{13}C-labeled patterns (i.e., MDVs) in key metabolites and searching for the "real" metabolic fluxes that could lead to the best fit of the measured ^{13}C-labeled patterns. Such quantitative analysis often reveals the network level rewiring of carbon fluxes in industrial workhorses (e.g., *E. coli*) when engineered for biochemical production. In short, although ^{13}C-assisted pathway analysis is suitable for pathway discovery in non-model microorganisms, ^{13}C-assisted flux analysis is more useful in identifying the metabolic rewiring in industrial microorganisms.

In the past decade, ^{13}C-assisted flux analysis has been widely applied to uncover the central metabolisms of various species. Accordingly to a curated database [82] recently developed to collect the central carbon metabolic flux distributions investigated by ^{13}C-MFA, over 500 metabolic flux analyses have been accomplished so far for 36 organisms (Fig. 2). Most ^{13}C-MFA studies focus on investigating metabolism of *E. coli* and *S. cerevisiae*. However, there is a trend that other industrial microorganisms, such as *Clostridium* and *Cyanobacteria*, can initiate more ^{13}C-MFA studies because of their importance in biochemical production and relatively less well-known cell metabolism. Also, with the wide application of ^{13}C-MFA, many ^{13}C-MFA software packages, such as OpenFLUX2 [83], 13CFLUX2 [84], Metran [85], INCA [86], FiatFLUX [87], and Biomet Toolbox 2.0 [88], have been developed by using highly efficient mathematical algorithms (e.g., elementary metabolite unit, EMU [85, 89]) to simulate ^{13}C-labeled patterns and calculating carbon fluxes in metabolic networks (Table 1). Thus, some of the difficulties, especially the computational load, of ^{13}C-MFA have been dramatically decreased.

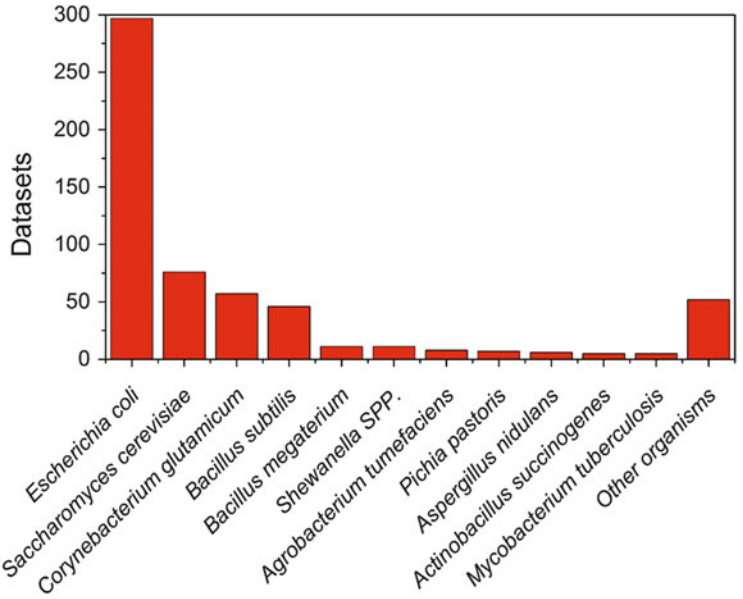

Fig. 2 Summary of current ^{13}C-MFA studies on different organisms

It is reasonable to believe that the numbers of ^{13}C-MFA studies could increase by orders of magnitudes in the next decade or two.

3 Synergy of ^{13}C Metabolic Flux Analysis and Metabolic Engineering

The ultimate goals of metabolic engineering are to design and build engineered biological systems that can produce chemicals, materials, food, and drugs at high yield [90]. However, the lack of fundamental understanding of cellular responses during industrial fermentation often prevents metabolic engineers achieving a satisfactory production of biochemical. In the past decade, ^{13}C-MFA has been widely used to provide insightful information about microbial metabolism and successfully helped metabolic engineers to improve biochemical production. Here, we have summarized recent successes on synergizing ^{13}C-MFA and metabolic engineering (Tables 2–4) and have organized them into four categories: (1) uncovering the bottleneck steps in biochemical production, (2) identifying cofactor imbalance issues of host metabolism, (3) revealing cell maintenance requirement of industrial microorganisms, and (4) elucidating the mechanism of microbial resistance to fermentation inhibitors.

Table 1 Summary of ^{13}C-MFA software

Software name	Capabilities			Labeled pattern				Key solver (algorithm)			Platform		Developer
	Steady-state	INST-^{13}C-MFA[a]	PLE integration[b]	EMU[c]	MDV[d]	SFL[e]	Cumomer	*fmincon*	IPOPT	Others	MATLAB	UNIX/Linux	
13CFLUX	●			●					●			●	Wiechert's group [174]
13CFLUX2	●			●					●			●	Wiechert's group [84]
Metran	●		●	●				●			●		Antoniewicz's group [85]
C13	●					●		●			●		Nielsen's group [175]
OpenFLUX	●				●			●			●		Krömer's group [139]
OpenFLUX2	●		●	●				●			●		Mashko's group [83]
FiatFLUX	●				●			●			●		Sauer's group [87]
FIA	●						●[f]			●[g]		●	Young's group [176]
INCA	●	●		●						●[h]	●		Young's group [86]
OpenMebius	●	●		●						●[i]	●		Shimizu's group [177]
influx_s	●						●			●[j]		●	Portais's group [178]

[a] *INST-^{13}C-MFA* Isotopic nonstationary ^{13}C metabolic flux analysis
[b] *PLE* Parallel labeling experiments
[c] *EMU* Elementary metabolite unit [85]
[d] *SFL* Summed fractional labeling [179]
[e] *MDV* Mass distribution vectors
[f] Fluxomer
[g] SNOPT [180]
[h] Customized differential equation solver [89]
[i] *Levenberg–Marquardt* method [181]
[j] NLSIC [178]

Table 2 Summary of synergistic tactics of ^{13}C-MFA and metabolic engineering in identifying and overcoming the bottleneck steps of biochemical production

Industrial microbial hosts	Target chemicals	Biological insights from ^{13}C-MFA						Strategies of genetic modifications		Outcomes of metabolic engineering
		Glycolysis	oxPP pathways[a]	Pyr-bypass pathways[b]	ED pathways[c]	PDC pathways[d]	NADPH supply[e]	Up-regulated	Down-regulated	
S. cerevisiae	*n*-Butanol	NC[f]	NC	UP [97]	NC	NC	NC	cyto-PDH [51], ACS	ADH, GPD [51]	300% increase [51]
S. cerevisiae	Amorphadiene	NC	NC	UP [97]	NC	NC	NC	ALD [95], ACS	None	70% increase [95]
S. cerevisiae	Acetyl-CoA	NC	NC	UP[g] [52]	NC	NC	NC	PHK pathways[g] [52]	None	~10% increase [52]
S. cerevisiae	Shikimic acid	UP [106]	NC	NC	NC	NC	NC	*aro1*, *aro4*, *tkl* [106]	None	600% increase [106]
E. coli	Fatty acid	NC	UP [57]	NC	UP [57]	DOWN [57]	UP	ACC, ACL, FASs [57]	FADs	200% increase

[a]*oxPP pathways* Oxidative pentose phosphate pathways
[b]*Pyr-bypass pathways* Pyruvate-bypass pathways, including pyruvate dehydrogenase pathways
[c]*ED pathway* Entner–Doudoroff pathways
[d]NADPH supply represents the overall NADPH net production, including oxidative PP pathways and transdehydrogenase pathways
[e]*PDC pathways* Pyruvate decarboxylase pathways
[f]*NC* Not changed
[g]*PHK pathways* Heterologous phosphoketolase (PHK) pathways. Certain metabolic engineering works were related to the ^{13}C-MFA studies but did not sequentially follow the ^{13}C-MFA studies

Table 3 Summary of synergistic tactics of ^{13}C-MFA and metabolic engineering in uncovering and solving the cofactor imbalance issues[a]

Industrial microbial hosts	Target chemicals	Biological insights from ^{13}C-MFA			Strategies of metabolic engineering			Outcomes of metabolic engineering
		TCA cycle[b]	oxPP pathways	NADPH supply	Up-regulated	Down-regulated	Others	
E. coli	Lycopene	NC	UP [53]	UP [53]	G6PDH, transdehydrogenase [134, 135]	*pgi* [133]	None	~100% increase [53]
C. glutamicum	L-Valine	NC	UP [54]	UP [54]	Transdehydrogenase [54]	None	None	>200% increase [54]
S. cerevisiae	Ethanol (xylose utilization)	UP [36]	UP [36]	NC	None	None	Alternate the cofactor specificity of XR to NADH [113–115, 117]	~40% increase [117]

[a]All abbreviations and table headers in Table 3 have the same definitions as in Table 2. Certain metabolic engineering works were related to the ^{13}C-MFA studies but did not sequentially follow the ^{13}C-MFA studies

[b]*TCA cycle* Tricarboxylic acid cycle

Table 4 Summary of synergistic tactics of ^{13}C-MFA and metabolic engineering in revealing and compensating cell maintenance requirement of industrial microorganisms[a]

Industrial microbial hosts	Target chemicals	Biological insights from ^{13}C-MFA				Strategies of metabolic engineering	Outcomes of metabolic engineering
		TCA cycle	oxPP pathways	PDC pathways	NADPH supply		
P. pastoris	*R. oryzae* lipase	UP [182]	NC	NC	NC	• Co-substrate culture [182]	• Supply additional ~30% ATP to compensate the requirement of maintenance energy [182].
S. cerevisiae	SAM[b]	UP [35]	NC	NC	NC	• Nutrient medium optimization [183]	• 250% increase of SAM production [183]
S. cerevisiae	(Xylose utilization)	UP [36]	UP [36]	NC	UP [36]	• Nutrient medium optimization [36]	• ~50% decrease the requirement of maintenance energy [36]
S. cerevisiae	Cell growth	UP [39]	UP [39]	NC	UP [39]	• Overexpress NADPH-dependent oxireductase in parent strain	• More than 20% decrease of biomass production [39]
E. coli	Cell growth	DOWN [38]	NC	UP [38]	UP [38]	• Decrease the cofactor (NADH/NAD^+) sensitivity of relevant enzymes • Overexpress the proteins of electron transport chain • Add other electron acceptors	• ~50% increase of growth rate [38]

[a]All abbreviations and table headers in Table 3 have the same definitions in Tables 2 and 3. Certain metabolic engineering works were related to the ^{13}C-MFA studies but did not sequentially follow the ^{13}C-MFA studies

[b]*SAM* S-Adenosyl-L-Methionine

3.1 Uncovering the Bottleneck Steps in Biochemical Production

3.1.1 Bottlenecks in Acetyl-CoA Synthesis

As a central metabolite, acetyl-CoA plays an important role in a series of cellular functions. In metabolic engineering, acetyl-CoA is a key precursor in the biosynthesis of sterols, amino acids, fatty acid-derived chemicals, polyketides, and isoprenoid-derived drugs [51]. To accommodate the cellular requirement, the organisms use a variety of routes for acetyl-CoA synthesis (Fig. 3), such as the oxidative decarboxylation of pyruvate, the oxidation of long-chain fatty acids, and the oxidative degradation of certain amino acids. The most common way to produce the acetyl-CoA is the direct conversion from pyruvate by pyruvate dehydrogenase

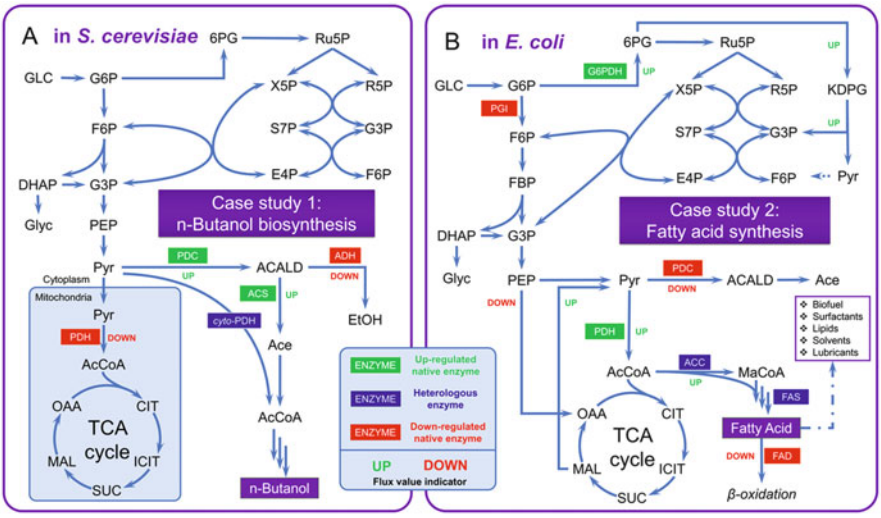

Fig. 3 Case studies that identify key bottleneck steps of biochemical production via ^{13}C-MFA and the corresponding metabolic engineering strategies. (**a**) *n*-Butanol biosynthesis in *S. cerevisiae*. (**b**) Fatty acid synthesis in *E. coli*. Please note that the pathways shown in Fig. 2 are schematic and there could be missing pathways. Abbreviations: *G6P* Glucose 6-phosphate, *F6P* Fructose 6-phosphate, *6PG* 6-Phosphogluconate, *Ru5P* Ribulose 5-phosphate, *X5P* Xylulose 5-phosphate, *R5P* Ribose 5-phosphate, *E4P* Erythrose 4-phosphate, *PEP* Phosphoenolpyruvate, *SA* Shikimic acid, *MAA* Mycosporine-like amino acids, *AA* Amino acids, *PGI* Phosphoglucose isomerase, *G6PDH* G6P dehydrogenase, *TKL* Transketolase, *ARO* Pentafunctional protein ARO1p, *DHAP* Dihydroxyacetone phosphate, *G3P* Glyceraldehyde 3-phosphate, *Glyc* Glycerol, *AceP* Acetyl-P, *EtOH* Ethanol, *Pyr* Pyruvate, *ACAL* Acetaldehyde, *Ace* Acetate, *AcCoA* Acetyl-CoA, *OAA* Oxaloacetate, *CIT* Citrate, *ICIT* Isocitrate, *SUC* Succinate, *MAL* Malate, *Glox* Glyoxylate, *GPD* Glycerol-3-phosphate dehydrogenase, *XpkA* Phosphoketolase, *ACK* acetate kinase, *PDC* Pyruvate decarboxylase, *ADH* Alcohol dehydrogenase, *PDH* Pyruvate dehydrogenase, *cyto-PDH* Cytosolic pyruvate dehydrogenase, *ACS* Acetyl-CoA synthetase, *ACL* ATP citrate lyase, *ICL* Isocitrate lyase, *MLS* Malate synthetase, *MaCoA* Malonyl-CoA, *ACC* Acetyl-CoA carboxylase, *FAS* Fatty acid synthesis enzymes, *FAD* Fatty acids degradation enzymes

(PDH) [91, 92], pyruvate ferredoxin oxidoreductase (PFO), pyruvate NADPH oxidoreductase (PNO) [93], or pyruvate formate lyase (PFL) [94] under anaerobic conditions. Other acetyl-CoA synthesis pathways, for example, acetyl-CoA synthetase (ACS) [95] and citrate lyase (ACL) [96], also play important roles in acetyl-CoA supplements for different organisms, especially for supplying the cytosolic acetyl-CoA as the precursor for various biochemical products.

To investigate acetyl-CoA biosynthesis in living cells, ^{13}C-MFA was used to compare *S. cerevisiae* strains growing under purely oxidative, respiro-fermentative and predominantly fermentative conditions [97]. Based on the flux distributions, the activated pyruvate bypass pathway, that is, converting pyruvate to acetaldehyde and then to acetate for synthesis of cytosolic acetyl-CoA, was found to be the main pathway used by *S. cerevisiae* to supply cytosolic acetyl-CoA. However, the flux in the pyruvate bypass pathway was not strong enough to supply sufficient cytosolic acetyl-CoA when engineering *S. cerevisiae* to produce acetyl-CoA-derived chemicals, such as *n*-butanol. To increase the capability of producing cytosolic acetyl-CoA in *S. cerevisiae*, various metabolic engineering strategies have been adopted to enhance further cytosolic acetyl-CoA availability, including the disruption of competing pathways [51] and the introduction of heterologous biosynthetic pathways with higher catalytic efficiency and lower energy input requirement, such as cytosolic localized PDHs (cytoPDHs) [51] and ATP-citrate lyase (ACLs) [96]. In one of the studies that evaluated the effects of various acetyl-CoA synthesis pathways on *n*-butanol production, the cytoPDHs was found to work best and led to threefold increased *n*-butanol production in the engineered *S. cerevisiae* (Fig. 3a and Table 2) [51].

In addition to uncovering the bottleneck of cytosolic acetyl-CoA biosynthesis for wild-type yeast, ^{13}C-MFA was also used to elucidate the effect of a heterogonous acetyl-CoA enhanced pathway, that is, phosphoketolase pathway (PHK), in a genetically modified yeast strain in which the genes *xpkA* and *ack* from *Aspergillus nidulans* were introduced [52]. The PHK pathway was originally utilized by several bacterial species [98] and filamentous fungi for glucose dissimilation as an alternative to the Embden–Meyerhof–Parnas pathway (EMP). For example, in *A. nidulans*, the utilization of this metabolic pathway led to increased carbon flow toward acetate and acetyl-CoA through the action of a phosphotransacetylase [99]. Flux distribution in the central metabolic pathways showed the positive role of the PHK pathway on improving the supply of cytosolic acetyl-CoA in the *S. cerevisiae* strain, which also accounted for the improved acetate yield. Encouraged by this discovery, the same PHK pathway was co-expressed together with a wax ester synthase (*ws2*) and successfully improved the titer of fatty acid ethyl esters by 1.7-fold [100]. Such proof-of-concept studies indicated that the PHK pathway could be established as a stand-alone route to divert flux from glycolysis to cytosolic acetyl-CoA supply, and holds great potential for future improvement of the production of acetyl-CoA-derived chemicals.

3.1.2 Bottlenecks in Fatty Acid Synthesis

Fatty acids are the precursors to produce transportation fuels and industrial chemicals including surfactants, solvents, and lubricants [57]. The microbial production of fatty acid-derived chemicals has recently been achieved in many industrial applications. *Escherichia coli* can serve as an excellent host for fatty acids production because of its fast growth, simple nutrient requirements, well-understood metabolism, and well-established genetic tools. However, only a small amount of free fatty acids is detectable under normal conditions in the wild-type *E. coli*. The synthesis of saturated fatty acid starts with the conversion of acetyl-CoA into malonyl-CoA catalyzed by ATP-dependent acetyl-CoA carboxylase and the transesterification of malonyl-CoA into an acyl carrier protein (ACP) catalyzed by malonyl-CoA ACP transacylase (*fabD*), followed by cyclic chain elongation (Table 2) [101].

In spite of various fatty acid over-producing strains that have been created, most studies focus on engineering terminal enzymes in fatty acid biosynthesis pathways and little is known about how central metabolism responds to fatty acid production. To reveal the metabolic bottlenecks in fatty acid production, ^{13}C-MFA has been performed by using an engineered fatty acid over-producing *E. coli* DH1 strain with over-expression of *tesA*, and *fadR* genes and knock-out of *fadE* gene (Fig. 3b) [57]. This ^{13}C-MFA study clearly showed that the *E. coli* metabolic flux was redistributed in response to over-production of fatty acid. Basically, compared to the wild-type *E. coli* strain, the flux in the engineered strain was significantly diverted from acetate synthesis to fatty acid synthesis, indicating that an increase in the supply of key precursors in fatty acid synthesis is crucial to increasing subsequent fatty acid synthesis. The fluxes of the pentose phosphate pathway (PPP) also dramatically increased to supply large amounts of reduction power, mostly NADPH, to support the fatty acid production in the engineered strain. Finally, the flux of the anaplerotic pathway into the TCA cycle decreased 1.7-fold in the engineered strain and, consequently, more carbon fluxes were diverted to supply cytosolic acetyl-CoA, the starting point of fatty acid biosynthesis. Overall, as indicated by ^{13}C-MFA, the supply of fatty acid precursor and NADPH was recognized as the key bottleneck in microbial engineering for fatty acid production.

To improve fatty acid production, numerous engineering strategies have been suggested and explored. For example, to overcome the challenge of the limited supply of fatty acid precursors, the acetyl-CoA carboxylase was over-expressed to provide more malonyl-CoA, a key precursor for fatty acid synthesis, which successfully enhanced the production of fatty acids [102, 103]. In another study, the fatty acid degradation pathway was removed by knocking out *fadE* in *E. coli*. Together with the over-expression of *tesA* and *fabF*, the yield of fatty acids was enhanced by nearly threefold [104]. Similarly, another study showed that by co-expressing *fabZ* and a thioesterase from *Ricinus communis* in a *fadD* (a key gene in fatty acid degradation) deletion mutant, the fatty acid titer was enhanced by nearly threefold (Table 2) [105]. In total, the precursor issue identified by ^{13}C-MFA

has now been well-addressed in microbial engineering for fatty acid production. The inadequate supply of NADPH, another issue revealed by ^{13}C-MFA in fatty acid production, could be the next matter to which metabolic engineers need to pay attention.

3.1.3 Bottlenecks in Pentose Phosphate Pathway

Pentose phosphate pathway, well-known as the limitation step in providing sufficient NADPH for biochemical synthesis (discussed in Sect. 3.2), is also the essential pathway to provide the precursors for the synthesis of nucleotides and nucleic acids from ribose 5-phosphate and aromatic amino acids (e.g., phenylalanine and tyrosine). Several aromatic compounds, such as shikimic acid, a valuable drug precursor, could be produced from metabolites in the pentose phosphate pathway. To investigate the bottleneck in the pentose phosphate pathway for the biosynthesis process of shikimic acid, ^{13}C-MFA was recently applied to four different engineered *S. cerevisiae* strains which were engineered to produce shikimic acids in different amounts [106]. By comparing flux distributions of the four strains with different shikimic acid productions, a higher flux through the pentose phosphate pathway was positively correlated with higher production of shikimic acid. This analysis indicated that the low flux into the PP pathway could be the bottleneck for the shikimic acid production. Indeed, it was found that when removing the original phenylalanine and tyrosine synthesis pathway, and overexpressing *aro1*, *aro4*, and *tkl* genes to improve the metabolic fluxes in the PP pathway, the shikimic acid was increased by nearly twofold in *S. cerevisiae* (Table 2).

Similarly, riboflavin is an important industrial bio-product from the PP pathway, which has been commercially produced by engineering *Bacillus subtilis* strains [107]. ^{13}C-MFA has also been implemented for both wild-type and engineered *B. subtilis* strains in the past two decades [108, 109] to unravel the metabolic rewiring in the riboflavin producing strain, which could further improve the riboflavin production. The intracellular flux distributions of a riboflavin-producing *B. subtilis* strain has been rigorously investigated via ^{13}C-MFA at three different dilution rates in chemostats [108]. It was found that the PP pathway was activated, which not only supplied sufficient precursors but also produced sufficient NADPH. More interestingly, cofactor NADPH was always excessively produced in *B. subtilis* strains under all the three dilution rates based on the flux analysis, especially in the low dilution rate without riboflavin production. In other words, the estimated amount of NADPH required was found to be less than the NADPH formations for both biomass and riboflavin production. Thus, the high production of riboflavin and purine nucleotides is attributed to the sufficient precursor supply from the PP pathway in *B. subtilis*. It is worth noting that the transhydrogenase, which catalyzed the reversible conversion of NADPH to NADH, played an important role to re-oxidize the excessive NADPH that was generated because of the highly activated PP pathways.

3.2 *Identifying Cofactor Imbalance Issues of Host Metabolism*

Cofactors, for example, NADH/NAD^+ and NADPH/$NADP^+$, play a major role as the redox carriers for catabolic and anabolic reactions as well as the important agents in transfer of energy for the cell. NADH/NAD^+ functions as a cofactor pair in over 300 redox reactions and regulates various enzymes and genetic processes [56]. Under aerobic growth, NADH acts as an electron carrier for the transportation of electrons from the carbon source to the final electron acceptor, oxygen. Under anaerobic growth with the absence of an alternative oxidizing agent, the regeneration of NAD^+ is of great significance for the redox balance, which is achieved through fermentation by using NADH to reduce metabolic intermediates [110, 111]. Thus, the balance issue of NADH/NAD^+ is crucial for both aerobic and anaerobic conditions. NADPH/$NADP^+$, as the phosphorylation products of NADH/NAD^+, drives the anabolic reactions. The enzymatic synthesis of some important compounds, for example, fatty acids and amino acids, depends heavily on cofactor NADPH as the reducing equivalents. The major pathways supplying NADPH during heterotrophic growth on glucose are the oxidative pentose phosphate pathway, the Entner–Doudoroff pathway, and $NADP^+$-dependent isocitrate dehydrogenase in the TCA cycle [53]. Additionally, the balance between NADH and NADPH also plays an important role to provide sufficient NADPH for the anabolic reactions. NAD(P) transhydrogenase can catalyze the reversible conversion between NADH and NADPH to balance the cofactors [112]. With the expense of 1 mol ATP, NAD kinase also can catalyze the conversion from NADH to NADPH [55].

It is conceivable that in cofactor-dependent production systems, cofactor availability and balance issue play a vital role in dictating the overall process yield. Hence, the identification of key pathways to balance the cofactor levels would be helpful for the rational design of metabolic engineering strategies to increase biochemical production further. ^{13}C-MFA is one of few analytical tools that can rigorously determine the cofactor usage in cell metabolism and has been used with many industrial microorganisms to reveal the cofactor imbalance issues related to biochemical production. In this section, we have summarized several important discoveries of the cofactor balance issues via ^{13}C-MFA and the corresponding metabolic engineering strategies to solve such issues (Fig. 4 and Table 3).

3.2.1 Cofactor Imbalance in Xylose Fermentation of *S. cerevisiae*

S. cerevisiae with the ability to ferment sugars anaerobically to ethanol at high rates has been domesticated for millennia and continuously selected as a workhorse for bioethanol production. However, *S. cerevisiae* cannot utilize xylose anaerobically, which is the second most abundant composition in lignocellulose, a renewable and non-food-competitive resource. One of the commonly used strategies to engineer

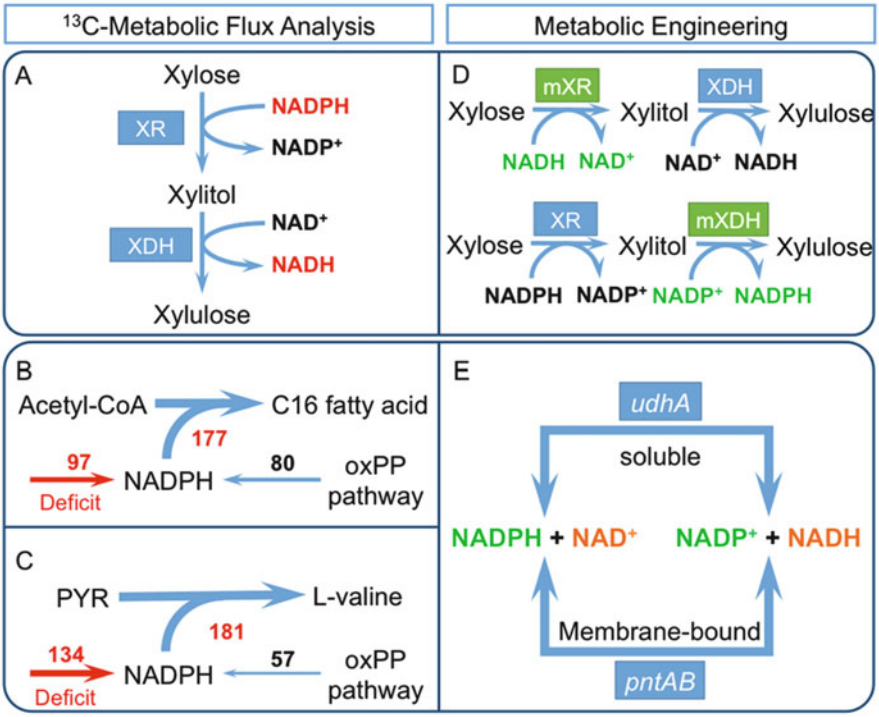

Fig. 4 Cofactor imbalance issues identified by ^{13}C-MFA and the corresponding metabolic engineering strategies. *Left column*: cofactor imbalance issues in (**a**) xylose utilization of *S. cerevisiae* strains, (**b**) fatty acid and fatty acid-derived chemical production, and (**c**) L-valine production. *Right column*: the corresponding metabolic engineering strategies to tackle cofactor imbalance issues: (**d**) altering the co-factor specificities of xylose reductase (XR) or xylitol dehydrogenase (XDH); (**e**) overexpressing transhydrogenase to balance the NADH and NADPH. Abbreviations: *mXR* Mutated xylose reductase, *mXDH* Mutated xylitol dehydrogenase, *oxPP pathway* Oxidative pentose phosphate pathway

S. cerevisiae to utilize xylose is the introduction of a fungal xylose pathway from xylose-utilizing yeasts such as *Pichia stipites*. Through this fungal xylose pathway, xylose could be converted to fermentable xylulose through the consecutive redox reactions catalyzed by NADPH-dependent xylose reductase (XR) and NAD^+-dependent xylitol dehydrogenases (XDH), with xylitol produced as the intermediate. However, the use of different cofactors in the fungal xylose pathway brings about a notorious cofactor imbalance issue and severely limits the xylose utilization in *S. cerevisiae*. More importantly, the cofactor imbalance issue is not standalone. Rather, as shown in a few ^{13}C-MFA studies, it is intertwined with the central metabolism to induce network-level rewiring of carbon fluxes. Basically, a systematic investigation of xylose utilization of recombinant *S. cerevisiae* strains in oxygen-limited conditions for ethanol production was accomplished by ^{13}C-MFA [36]. By implementing the ^{13}C tracer experiments and running metabolic flux

analysis for six recombinant strains with different origins of XR and XDH in the xylose pathway [36], a universally high activity of the oxidative pentose phosphate pathway was found to supply the NADPH for the XR. The strong activities in the TCA cycle was also found and indicated that huge amounts of NADH needed to be consumed by oxidative phosphorylation. Concurrent with the global metabolic rewiring, only a small amount of the carbon fluxes was diverted to ethanol production.

To solve the cofactor imbalance issues in the xylose utilization of *S. cerevisiae*, numerous efforts in metabolic engineering have been devoted. One of the strategies is the partial alteration of the cofactor preference for these two enzymes, that is, altering the preference of XR to use NADH or altering XDH to use $NADP^+$ as the cofactors, which would generate a cofactor balance cycle for the initial two steps of xylose utilization to balance the cofactor utilization. The cofactor engineering strategy has proved to be functional in several studies. For example, by replacing the native *Pichia stipitis* XR with a mutated XR with increased preference of NADH, the ethanol yield was improved by ~40% with the decreased xylitol production [113–119]. Similar successes have also been achieved in several other attempts to increase the $NADP^+$ preference for the XDH, which have successfully improved the ethanol production by 28–41% [120–123]. The other strategy used to tackle the cofactor imbalance issue is to engineer the cofactor-dependent metabolic pathways which could decrease the xylitol production and enhance ethanol yield. For instance, lowering the flux through the NADPH-producing pentose phosphate pathway could lead to increasing ethanol yield and decreasing xylitol production. This is attributed to an inadequate supplement of NADPH which could improve the NADH preference of XR, and thus partially balance the cofactor usage of XR and XDH [124]. In addition, replacing the NADPH-producing PP pathway (glucose-6-phosphate dehydrogenase) with a fungal $NADP^+$-dependent D-glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH) could produce NADPH for the XR without losing any carbon, and could provide more carbon for ethanol production [125]. Improving the NAD^+ regeneration directly by introducing heterogeneous genes could also decrease xylitol production and increase ethanol production [126]. Beside the two-step xylose utilization pathway, using xylose isomerase is another efficient approach for the xylose fermentation as it does not require cofactors when converting xylose to xylulose, and hence bypasses the cofactor imbalance issue (Table 3). However, as indicated by a recent ^{13}C-MFA study, the lower glycolysis activity that led to inefficient re-oxidation of NADH could potentially be a new bottleneck step when using xylose isomerase in *S. cerevisiae* [127].

3.2.2 Cofactor Imbalance in Chemical Biosynthesis

The production of many chemicals, such as fatty acids and amino acids, requires a large amount of cofactors. Thus, cofactor imbalance issues are tightly related not only to the sugar utilization but also to the chemical production. For example, by using ^{13}C-MFA to analyze the cell metabolism in wild-type and fatty acid over-

producing *E. coli* strains, it was found that the engineered strain requires excessive NADPH compared to the wild-type strain, that is, 255 units compared to 179 units NADPH with the flux of glucose uptake normalized at 100 units. However, the sum of NADPH supplied from central metabolism could only reach 100 units [57], which clearly indicated that more NADPH production would be needed to increase fatty acid production in *E. coli*. To balance the NADPH usage, an alternative transdehydrogenase pathway that converted NADH to NADPH was activated in the engineered *E. coli*. The flux of the transdehydrogenase pathway was increased by 70% compared to that in wild-type strains (i.e., from 90 to 153 units) to support fatty acid biosynthesis.

Realizing the importance of cofactor balance, particularly the NADPH supply, in biochemical production, metabolic engineers have adopted various strategies to overcome this challenge. One strategy is to switch the specificities of glycolytic enzymes, for example, GAPDH, from NAD^+-dependence to $NADP^+$-dependence, which could build an NADPH-producing glycolysis pathway to increase the bioavailability of NADPH and further improve the NADPH-dependent lycopene production by ~100% (Table 3) [53]. It is also used to redirect the metabolic flux from the glycolysis pathway into the pentose phosphate pathway to enhance NADPH supply by the overexpression of *zwf* that encodes glucose-6-phosphate dehydrogenase (G6PDH) [128–130], deletion of *pfkA* and *pfkB* that encode the phosphofructokinase (PFK) [131], or deletion of *pgi* that encodes phosphoglucose isomerase [132, 133]. In addition, transhydrogenase [134, 135] or NAD kinase [53] was overexpressed to boost further the NADPH/$NADP^+$ availability in *E. coli* and other microorganisms.

In addition to the production of fatty acids, biosynthesis of amino acids, such as L-lysine and L-valine, requires NADPH as cofactor of the enzymatic reactions. *Corynebacterium glutamicum*, one of the industrial workhorses for producing amino acids, has been considered an important microorganism with extensive ^{13}C-MFA studies [46, 136–139]. Many metabolic engineering strategies provided by ^{13}C-MFA have been developed to improve the amino acids production. For example, L-lysine is one of the major products of *C. glutamicum*, which is synthesized from the pyruvate and oxaloacetate consuming 4 mol NADPH for 1 mol L-lysine. To study the intracellular metabolic rewiring of the L-lysine producing *C. glutamicum* strains, several ^{13}C-MFA studies have been implemented and uncovered the fact that the PP pathway was increased to supply NADPH. The anaplerotic carboxylation pathway was also enhanced to provide enough precursors for L-lysine synthesis [137, 138, 140]. Similarly, a significant increase in the PP pathway flux was also found to be associated with L-valine production in a pyruvate decarboxylase-deficient *C. glutamicum* strain via ^{13}C-MFA, which again indicated that the NADPH supply was the key issue in L-valine production [54].

Based on the insightful information from ^{13}C-MFA studies, various metabolic strategies have been developed to improve L-lysine and L-valine production. First, to overcome the inadequate NADPH supply, the enzymes in PP pathway, such as glucose-6-phosphate dehydrogenase [141], transketolase, and transaldolase [142], as well as 1,6-bisphosphatase [143], were overexpressed to redirect fluxes toward

the PP pathway to improve L-lysine production. In addition to the strategy of overexpressing the native enzymes in the PP pathway, the alteration of cofactor specificity of GAPDH from NAD^+-dependence to $NADP^+$-dependence has successfully improved the L-lysine production by ~50% without decreasing cell growth rate [144]. To investigate further the metabolic responses of such alteration, ^{13}C-MFA was performed again to compare the intracellular flux distributions between wild-type strains and mutants with alternated GAPDH. It was found that the mutated GAPDH pathway was the major source of the NADPH in the mutated strain with the similar PP pathway flux, but with higher L-lysine production. Last but not least, by cloning a transhydrogenase from *E. coli* to enhance NADPH supply in *C. glutamicum*, L-valine yield in *C. glutamicum* strain was dramatically improved by ~200% (Fig. 4c and Table 3) [54]. These discoveries were consistent with the expectations of the metabolic engineering strategies and, more importantly, demonstrated that ^{13}C-MFA could indeed rationally guide the metabolic engineering and improve microbial performance.

3.3 *Revealing Cell Maintenance Requirement of Industrial Microorganisms*

Metabolic engineering is frequently equated with the heterologous production of a series of recombinant proteins. Nowadays, with the development of synthetic biology approaches, more and more heterologous pathways have been introduced into a host cell to produce non-natural products with multiple genes inserted, deleted, replaced, or overexpressed. On one hand, the genetic manipulation could modify cell metabolism and divert more carbon flux into the desired chemicals. On the other hand, the metabolic engineering, particularly heterologous protein overexpression, could interfere with the host metabolism and generate severe metabolic burdens because the protein expression could be energetically expensive during transcription and translation. Such issues, however, have not yet been studied very much in the field of metabolic engineering.

The metabolic burden of industrial microorganisms is often reflected as elevated cell maintenance energy of industrial microorganisms, as revealed in several pioneering ^{13}C-MFA studies. In one of the ^{13}C-MFA studies, *Pichia pastoris*, a methylotrophic yeast with an attractive ability to produce various heterologous proteins [64–67], was investigated by introducing a mock plasmid, a low-copy plasmid to express *Rhizopus oryzae* lipase, and a high-copy plasmid to express *R. oryzae* lipase, respectively. It was found that the TCA cycle fluxes of both protein-expressing *P. pastoris* strains were much higher than the control strain in producing more ATP to sustain cell growth, confirming that the protein folding and conformational stress indeed imposed a metabolic burden on the microbial host. The similar metabolic rewiring, that is, elevated TCA cycle fluxes to provide more ATP for cell maintenance, was also found in an *S*-adenosyl-L-methionine (SAM)

producing an *S. cerevisiae* strain [35] and a xylose-utilizing *S. cerevisiae* strain (Table 4) [36].

To avoid the introduction of metabolic burdens, metabolic engineers have explored three strategies: medium optimization, use of low-copy plasmids, and promoter engineering. Optimization of cultural medium and fermentation condition could potentially remove stresses such as nutrient limitation, and hence reduce the requirement for cell maintenance energy. For example, by using several novel feeding strategies with cultural SAM-producing *P. pastoris*, the production of SAM was found to be improved by ~35% [145, 146]. In addition, it has been found that the utilization of a high-copy plasmid may increase risk of plasmid instability and metabolic burden [147], as the protein over-expression requires tremendous amounts of building blocks and energy, which could jeopardize the normal cell growth and increase the metabolic burden. It was found that using a low-copy plasmid sometimes could be a better choice for chemical production. For example, in the study aiming to engineer *E. coli* to produce lycopene [148], the cell density of the engineered *E. coli* with high-copy plasmid at stationary phase was approximately 24% lower than the one with low-copy plasmid and 30% lower than the control culture. Similarly, the titer of lycopene in the *E. coli* with high-copy plasmid was 20% lower than that with low-copy plasmid. Another method to decrease metabolic burden is to tune the promoter strengths of various genes to balance the pathways and to avoid the accumulation of certain toxic intermediates as growth inhibitors. One example uses this method to engineer a more efficient production of taxadiene in *E. coli* [10]. In general, by tuning the expression levels of two modules in the taxadiene pathway, a native upstream methylerythritol phosphate (MEP) pathway forming isopentenyl pyrophosphate and a heterologous downstream pathway forming terpenoid, an inhibitory intermediate compound for cell growth, indole, was achieved at the minimal accumulation by expressing the upstream pathway at a very low level. Correspondingly, the taxadiene production was improved by ~15-fold.

3.4 Elucidating the Mechanism of Microbial Resistance to Fermentation Inhibitors

Environmental stresses, such as physical heat shock [149] and chemical acidity [150], could affect the physiology and viability of microbial cells and decrease or even stop the bioprocess productivity. For example, the lignocellulosic biofuels hold promises for a sustainable fuel economy. However, the chemical stresses from the toxic compounds in processed lignocellulosic hydrolysates, for example, weak acids, furans, and phenolic compounds [151], have hampered the economic feasibility of biofuels. Thus, it is important to identify the intracellular metabolic responses of industrial microorganisms to various stresses in order to improve their resistance to inhibitors rationally [40]. Compared to other commonly used approaches, such as transcriptomics and proteomics analysis, ^{13}C-MFA is more

intuitive and provides direct and quantitative readouts of the metabolic rewiring under stress conditions. In this section, we introduce recent advances in the study of stress response using ^{13}C-MFA and the corresponding metabolic engineering strategies to improve microbial resistance to different inhibitors (Fig. 5).

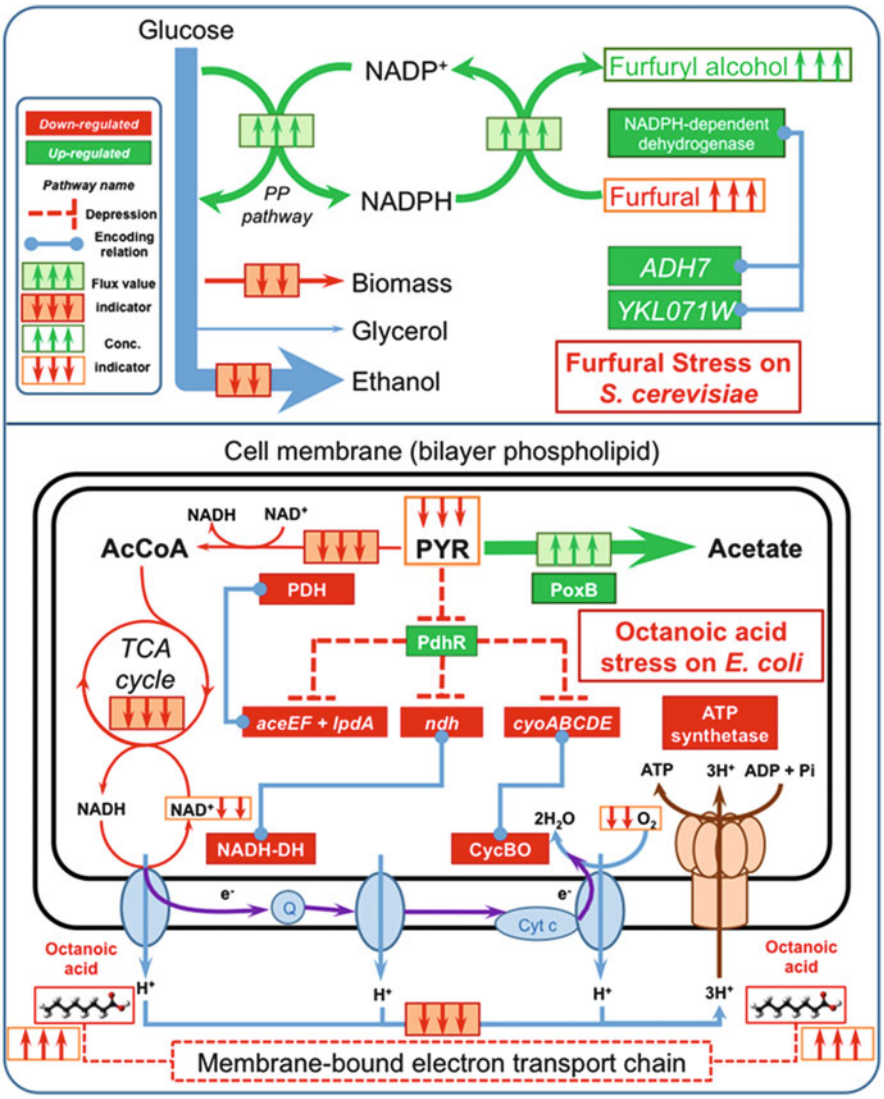

Fig. 5 Mechanisms of microbial stress responses identified by ^{13}C-MFA. *Top*: stress responses of *S. cerevisiae* to the furfural. *Bottom*: stress responses of *E. coli* to octanoic acid. Abbreviations: *ADH* Alcohol dehydrogenase, *PDH* Pyruvate dehydrogenase, *PoxB* Pyruvate oxidase, *PdhR* Pyruvate dehydrogenase regulator, *NADH-DH* NADH dehydrogenase, *CycBO* Cytochrome *bo* oxidase, *Cyc c* Cytochrome C

Among the various toxic compounds from the lignocellulose pretreatment and hydrolysis, furfural is an important contributor to the toxicity for *S. cerevisiae*. Although it has been found that *S. cerevisiae* has a weak intrinsic ability to reduce furfural to the less toxic furfuryl alcohol, the holistic view of metabolic responses to furfural is still missing. To investigate further the flux distribution of *S. cerevisiae* under the increasing strengths (concentrations) of furfural stress, ^{13}C-MFA has been applied for both wild-type and several evolved furfural-resistant strains in micro-aerobic and glucose-limited chemostats [39]. As revealed by ^{13}C-MFA, NADH-dependent oxireductases, which catalyzed the reduction of furfural, were the main defense mechanisms at lower concentration of furfural (<15 mM), whereas NADPH-dependent oxireductases became the major resistance mechanism at high concentration of furfural (>15 mM). Thus, the carbon flux of pentose phosphate increased as the main physiological response to high concentrations of furfural, which indicated that the NADPH supply was the key to help *S. cerevisiae* better resist furfural stress. Inspired by this discovery, metabolic engineers overexpressed several NADPH-dependent oxireductases, particularly ADH7 and YKL071W, and successfully increased furfural resistance in the parent *S. cerevisiae* strain by 200% [39].

In another study, ^{13}C-MFA was used to examine the metabolic responses of *E. coli* to octanoic acid stress [38]. When comparing the flux distributions of stressed and unstressed *E. coli* strains, a decreased flux in the TCA cycle and an increased flux in the pyruvate oxidative pathway for producing acetate were observed. It was hypothesized that octanoic acid triggered the membrane disruption and led to NAD^+ deficiency because of the destabilization of membrane-bound proteins, such as NADH dehydrogenase, which would down-regulate several key NAD^+-dependent pathways, such as the malate dehydrogenase pathways in the TCA cycle, and the pyruvate dehydrogenase multi-enzyme complex pathway. The pyruvate pool also shrank under octanoic acid stress condition, which could be attributed to the repression of the *pdhR* regulator, a regulator with high sensitivity to pyruvate in controlling the expressions of the PDH complex, NADH dehydrogenase II, and cytochrome *bo*-type oxidase encoded by *aceEF* and *lpdA*, *ndh*, and *cyoABCDE*, respectively [152]. Based on the discussion of ^{13}C-MFA results, several possible strategies to enhance further the C8 acid tolerance were proposed, including the supplementing of pyruvate in the medium and the replacement of $NADH/NAD^+$-sensitive enzymes.

4 Perspectives of Synergizing ^{13}C Metabolic Flux Analysis with Metabolic Engineering

The conventional ^{13}C-MFA has been widely applied to determine microbial metabolism and to guide metabolic engineers in the development of numerous strategies to improve biochemical production. However, there are still several technique

limitations that restrict the accuracy and flexibility of ^{13}C-MFA. For example, the ^{13}C-MFA can only be applied at metabolic and isotopic steady states [62], which could be difficult to use when the target chemicals are produced in a non-steady state (e.g., drug synthesis in the stationary growth phase). In addition, most of the conventional ^{13}C-MFA studies are limited in central metabolism [153], which has very limited use when analyzing the secondary metabolism of microorganisms. To overcome these challenges, novel experimental and computational methods have recently been developed to empower ^{13}C-MFA studies. In this section we summarize recent breakthroughs in ^{13}C-MFA and provide a perspective for novel routes to achieve synergy of ^{13}C-MFA and metabolic engineering.

4.1 Expand ^{13}C-MFA into Genome Scale

The conventional ^{13}C-MFA can only be applied to determine flux distribution in the central metabolic network, mainly because of the difficulties in (1) measuring the isotopic labeling of the numerous low-abundant metabolites and (2) the huge computational burden of simulating isotopic labeling of all metabolites in genome-scale metabolic networks. However, with the rapid development of high-resolution mass spectrometry, the accurate measurement of isotopic labeling of low-abundant metabolites becomes possible, as reported by several groups [154–156]. For computational simulation, an *E. coli* genome-scale model (imPR90068) has recently been constructed for ^{13}C-MFA [153], which spans 1,039 metabolites and 2,077 reactions. To calculate the genome-scale metabolic flux distribution, a total of 1.37×10^{157} isotopomers need to be simulated [153]. Thanks to the implementation of the EMU method [85], the computational burden was decreased by one to two orders of magnitude and, for the first time, the fluxes in all the metabolic pathways of *E. coli* were elucidated. Compared to the conventional ^{13}C-MFA, the genome-scale ^{13}C-MFA could rigorously determine the metabolic rewiring in secondary metabolism, from which many high-value chemicals, such as drugs, could be produced. The genome-scale ^{13}C-MFA could provide valuable information about the metabolic rewiring in response to the production of these secondary metabolites and guide the development of rational metabolic engineering strategies in a similar way to that used for improving bulk chemical production.

4.2 Isotopic Non-stationary ^{13}C-MFA (^{13}C-INST-MFA)

^{13}C-INST-MFA is a cutting-edge technology recently developed [66–70] to enable the application of ^{13}C-MFA for various autotrophic systems including cyanobacteria and plants. In brief, instead of collecting ^{13}C-labeling patterns at the isotopic steady state, ^{13}C-INST-MFA tracks the dynamics of ^{13}C-labeling in intracellular metabolites and applies computational algorithms to calculate the

steady-state metabolic fluxes that can best fit the ^{13}C-labeling kinetics. ^{13}C-INST-MFA has been applied to determine the photosynthetic metabolism of *Synechocystis* sp. PCC6803 [157] and *Arabidopsis thaliana* [158]. Such autotrophic metabolism is unable to be examined by conventional ^{13}C-MFA because all of the metabolites are universally labeled at the isotopic steady state when feeding with $^{13}CO_2$ and the information about pathway usage is completely lost. The merit of ^{13}C-INST-MFA for metabolic engineering lies in the fact that the metabolisms of numerous autotrophic systems, which used to be mysterious, can now be rigorously determined. Because many autotrophic systems are promising cell factories [159–161] that convert CO_2 into valuable chemicals, we can envision that ^{13}C-INST-MFA could guide metabolic engineers to understand better and to modify more rationally such systems for improving the production of autotrophic products.

4.3 ^{13}C-Based Dynamic Metabolic Flux Analysis (^{13}C-DMFA)

^{13}C-DMFA has recently been developed as an approach to investigate microbial metabolism at a metabolic non-steady state [30]. Compared to conventional kinetic models to describe microbial dynamics [162–165], ^{13}C-DMFA could reveal the dynamic reprogramming of intracellular fluxes and thus provides in-depth understanding of microbial metabolism in the pathway level. In one of the proof-of-concept studies, the *E. coli* metabolism in a fed-batch fermentation process for overproduction of 1,3-propanediol was investigated. By introducing several additional parameters to describe the fed-batch fermentation process, a time-resolved flux map was generated and showed that the intracellular flux associated with the PDO pathway increased by 10% and the split ratio between glycolysis and the pentose phosphate pathway decreased from 70/30 to 50/50. ^{13}C-DMFA has provided a way for metabolic engineer to investigate the dynamic metabolism during industrial fermentation, especially fed-batch fermentation. It is also expected that ^{13}C-DMFA could be further extended to study microbial metabolism at stationary growth phase, during which numerous high-value secondary metabolites are often produced. With the insightful information about the metabolic rewiring at non-steady state, metabolic engineers could develop more appropriate strategies to improve the biochemical production, particularly microbial-based drug production.

4.4 Improve Flux Resolution of ^{13}C-MFA via the Integration of Isotopic Patterns from Parallel Labeling Experiments

Parallel labeling experiment design has been widely applied in ^{13}C-MFA to improve the observability of global metabolic network by conducting multiple

labeling experiments with different isotopic tracers simultaneously [166–169]. Another recent advance in ^{13}C-MFA is the integration of the data (i.e., isotopic labeling patterns) from parallel labeling experiments to improve flux resolution [170–172]. The integration of the data from parallel labeling experiments has been combined with rapid development of the high-throughput measure techniques and computational algorithms [170–172] which would offer unique advantages compared to conventional ^{13}C-MFA [173], particularly by improving the precision of flux estimation [170, 171] and reducing the time of labeling experiments. With the more precise measurement of intracellular carbon fluxes, it is reasonable to conclude that higher resolution of microbial metabolism should be provided for metabolic engineers in the near future and fine-tuned engineering strategies should be developed for general applications in improving biochemical production.

Acknowledgements We thank Herbert Huttanus in Virginia Tech for improving the language of the chapter. This study was supported by start-up fund (#175323) from Virginia Tech and Junior Faculty Award from Institute for Critical Technology and Applied Science.

References

1. Sheng J, Feng X Metabolic engineering of yeast to produce fatty acid-derived biofuels: bottlenecks and solutions. Front Microbiol 6:554
2. Shin JH, Kim HU, Kim DI, Lee SY Production of bulk chemicals via novel metabolic pathways in microorganisms. Biotechnol Adv 31(6):925–935
3. Weusthuis RA, Lamot I, van der Oost J, Sanders JPM Microbial production of bulk chemicals: development of anaerobic processes. Trends Biotechnol 29(4):153–158
4. Hermann BG, Blok K, Patel MK (2007) Producing bio-based bulk chemicals using industrial biotechnology saves energy and combats climate change. Environ Sci Technol 41 (22):7915–7921
5. Stephanopoulos G (2007) Challenges in engineering microbes for biofuels production. Science 315(5813):801–804
6. Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. Biotechnol J 5(2):147–162
7. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD (2012) Microbial engineering for the production of advanced biofuels. Nature 488(7411):320–328
8. Lee SK, Chou H, Ham TS, Lee TS, Keasling JD (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Curr Opin Biotechnol 19(6):556–563
9. Stephanopoulos G (2008) Metabolic engineering: enabling technology for biofuels production. Metab Eng 10(6):293–294
10. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science 330(6000):70–74
11. Lee SY, Kim HU, Park JH, Park JM, Kim TY (2009) Metabolic engineering of microorganisms: general strategies and drug production. Drug Discov Today 14(1–2):78–88
12. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21 (7):796–802

13. Chang MCY, Keasling JD (2006) Production of isoprenoid pharmaceuticals by engineered microbes. Nat Chem Biol 2(12):674–681
14. Demain AL, Sanchez S (2009) Microbial drug discovery: 80 years of progress. J Antibiot 62 (1):5–16
15. Ferrer-Miralles N, Domingo-Espín J, Corchero JL, Vázquez E, Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Factories 8(1):1–8
16. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330 (6009):1355–1358
17. Huttanus H, Sheng J, Feng X (2016) Metabolic engineering for production of small molecule drugs: challenges and solutions. Fermentation 2(1):4
18. Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, Chou KJY, Hanai T, Liao JC (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab Eng 10(6):305–311
19. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277(5331):1453–1462
20. Huang Jr C, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. J Ind Microbiol Biotechnol 39(3):383–399
21. 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(5):612–616
22. Farmer WR, Liao JC (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat Biotechnol 18(5):533–537
23. Clomburg J, Gonzalez R (2010) Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. Appl Microbiol Biotechnol 86(2):419–434
24. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7(2): e1001287
25. Ostergaard S, Olsson L, Nielsen J (2000) Metabolic engineering of *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev 64(1):34–50
26. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418(6896):387–391
27. Nielsen J, Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res 8:122
28. Chen G-Q (2012) New challenges and opportunities for industrial biotechnology. Microb Cell Factories 11:111
29. Kwok R (2010) Five hard truths for synthetic biology. Nature 463(7279):288
30. Antoniewicz M (2015) Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol 42(3):317–325
31. Young JD (2014) ^{13}C metabolic flux analysis of recombinant expression hosts. Curr Opin Biotechnol 30:238–245
32. Wang Y, San K-Y, Bennett GN (2013) Cofactor engineering for advancing chemical biotechnology. Curr Opin Biotechnol 24(6):994–999
33. Wasylenko TM, Ahn WS, Stephanopoulos G (2015) The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. Metab Eng 30:27–39
34. Hollinshead WD, Henson WR, Abernathy M, Moon TS, Tang YJ (2015) Rapid metabolic analysis of *Rhodococcus opacus* PD630 via parallel ^{13}C-metabolite fingerprinting. Biotechnol Bioeng 113:91–100
35. Hayakawa K, Kajihata S, Matsuda F, Shimizu H (2015) ^{13}C-metabolic flux analysis in S-adenosyl-l-methionine production by *Saccharomyces cerevisiae*. J Biosci Bioeng 120:532–538

36. Feng X, Zhao H (2013) Investigating xylose metabolism in recombinant Saccharomyces cerevisiae via 13C metabolic flux analysis. Microb Cell Factories 12(1):114
37. Lam FH, Ghaderi A, Fink GR, Stephanopoulos G (2014) Engineering alcohol tolerance in yeast. Science 346(6205):71–75
38. Fu Y, Yoon J, Jarboe L, Shanks J (2015) Metabolic flux analysis of *Escherichia coli* MG1655 under octanoic acid (C8) stress. Appl Microbiol Biotechnol 99(10):4397–4408
39. Heer D, Heine D, Sauer U (2009) Resistance of *Saccharomyces cerevisiae* to high concentrations of furfural is based on NADPH-dependent reduction by at least two oxireductases. Appl Environ Microbiol 75(24):7631–7638
40. Çakar ZP, Seker UOS, Tamerler C, Sonderegger M, Sauer U (2005) Evolutionary engineering of multiple-stress resistant *Saccharomyces cerevisiae*. FEMS Yeast Res 5:569–578
41. Wittmann C, Heinzle E (1999) Mass spectrometry for metabolic flux analysis. Biotechnol Bioeng 62(6):739–750
42. Wiechert W (2001) ^{13}C metabolic flux analysis. Metab Eng 3(3):195–206
43. Dauner M, Sauer U (2000) GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. Biotechnol Prog 16:642–649
44. de Graaf AA (2000) Use of ^{13}C labelling and NMR spectroscopy in metabolic flux analysis. In: Barbotin JN, Portais JC (eds) NMR in biotechnology: theory and applications, vol 4. Horizon Scientific Press, Norwich
45. Christensen B, Nielsen J (1999) Isotopomer analysis using GC-MS. Metab Eng 1(4):282–290
46. Szyperski T (1998) ^{13}C-NMR, MS and metabolic flux balancing in biotechnology research. Q Rev Biophys 31(01):41–106
47. Feng X, Zhuang WQ, Colletti P, Tang YJ (2012) Metabolic pathway determination and flux analysis in nonmodel microorganisms through ^{13}C-isotope labeling. Methods Mol Biol (Clifton) 881:309–330
48. You L, Page L, Feng X, Berla B, Pakrasi HB, Tang YJ (2012) Metabolic pathway confirmation and discovery through ^{13}C-labeling of proteinogenic amino acids. J Vis Exp 59:e3583
49. Sauer U (2006) Metabolic networks in motion: ^{13}C-based flux analysis. Mol Syst Biol 2:62
50. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EEK, Keasling JD (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. Mass Spectrom Rev 28 (2):362–375
51. Lian J, Si T, Nair NU, Zhao H (2014) Design and construction of acetyl-CoA overproducing *Saccharomyces cerevisiae* strains. Metab Eng 24:139–149
52. Papini M, Nookaew I, Siewers V, Nielsen J (2012) Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: validation of activity through ^{13}C-based metabolic flux analysis. Appl Microbiol Biotechnol 95(4):1001–1010
53. Wang Y, San K-Y, Bennett GN (2013) Improvement of NADPH bioavailability in *Escherichia coli* by replacing NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase *GapA* with $NADP^+$-dependent *GapB* from *Bacillus subtilis* and addition of NAD kinase. J Ind Microbiol Biotechnol 40(12):1449–1460
54. Bartek T, Blombach B, Lang S, Eikmanns BJ, Wiechert W, Oldiges M, Nöh K, Noack S (2011) Comparative ^{13}C metabolic flux analysis of pyruvate dehydrogenase complex-deficient, L-valine-producing *Corynebacterium glutamicum*. Appl Environ Microbiol 77 (18):6644–6652
55. Hou J, Vemuri G, Bao X, Olsson L (2009) Impact of overexpressing NADH kinase on glucose and xylose metabolism in recombinant xylose-utilizing *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 82(5):909–919
56. Berrios-Rivera SJ, Bennett GN, San K-Y (2002) Metabolic engineering of *Escherichia coli*: increase of NADH availability by overexpressing an NAD^+-dependent formate dehydrogenase. Metab Eng 4(3):217–229

57. He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, Peng L (2014) Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on ^{13}C-metabolic flux analysis. Biotechnol Bioeng 111(3):575–585
58. Ranganathan S, Tee TW, Chowdhury A, Zomorrodi AR, Yoon JM, Fu Y, Shanks JV, Maranas CD (2012) An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. Metab Eng 14(6):687–704
59. Guo W, Chen Y, Wei N, Feng X (2016) Investigate the metabolic reprogramming of *Saccharomyces cerevisiae* for enhanced resistance to mixed fermentation inhibitors via ^{13}C metabolic flux analysis. PLoS One 11(8):e0161448
60. Quarterman J, Skerker JM, Feng X, Liu IY, Zhao H, Arkin AP, Jin YS (2016) Rapid and efficient galactose fermentation by engineered *Saccharomyces cerevisiae*. J Biotechnol 229:13–21
61. Guo W, Sheng J, Feng X (2016) ^{13}C-metabolic flux analysis: an accurate approach to demystify microbial metabolism for biochemical production. Bioengineering 3(1):3
62. Zamboni N, Fendt S-M, Ruhl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protoc 4(6):878–892
63. Tang YJ, Martin HG, Deutschbauer A, Feng X, Huang R, Llora X, Arkin A, Keasling JD (2009) Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations. Biotechnol Prog 25(5):1254–1259
64. Adler P, Bolten CJ, Dohnt K, Hansen CE, Wittmann C (2013) Core fluxome and metafluxome of lactic acid bacteria under simulated cocoa pulp fermentation conditions. Appl Environ Microbiol 79(18):5670–5681
65. Klein T, Lange S, Wilhelm N, Bureik M, Yang T-H, Heinzle E, Schneider K (2014) Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe*–a quantitative approach using ^{13}C-based metabolic flux analysis. Metab Eng 21:34–45
66. Jazmin L, O'Grady J, Ma F, Allen D, Morgan J, Young J (2014) Isotopically nonstationary MFA (INST-MFA) of autotrophic metabolism. In: Dieuaide-Noubhani M, Alonso AP (eds) Plant metabolic flux analysis, vol 1090. Humana Press, Totowa, pp. 181–210
67. Murphy TA, Dang CV, Young JD (2013) Isotopically nonstationary ^{13}C flux analysis of Myc-induced metabolic reprogramming in B-cells. Metab Eng 15:206–217
68. Jazmin L, Young J (2013) Isotopically nonstationary ^{13}C metabolic flux analysis. In: Alper HS Systems metabolic engineering, vol 985. Humana Press, Totowa, pp 367–390
69. Wiechert W, Nöh K (2013) Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr Opin Biotechnol 24(6):979–986
70. Wiechert W, Nöh K (2005) From stationary to instationary metabolic flux analysis. In: Kragl U (ed) Technology transfer in biotechnology, vol 92. Springer, Berlin, pp 145–172
71. Pingitore F, Tang Y, Kruppa GH, Keasling JD (2007) Analysis of amino acid isotopomers using FT-ICR MS. Anal Chem 79(6):2483–2490
72. Wiechert W, de Graaf AA (1997) Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng 55 (1):101–117
73. Wiechert W, Siefke C, de Graaf AA, Marx A (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. Biotechnol Bioeng 55(1):118–135
74. Wittmann C, Heinzle E (2001) MALDI-TOF MS for quantification of substrates and products in cultivations of *Corynebacterium glutamicum*. Biotechnol Bioeng 72(6):642–647
75. Guo W, Luo S, He Z, Feng X (2015) 13C pathway analysis of biofilm metabolism of shewanella oneidensis MR-1. RSC Adv 5(50):39840–39843
76. Christensen B, Karoly Gombert A, Nielsen J (2002) Analysis of flux estimates based on ^{13}C-labelling experiments. Eur J Biochem 269(11):2795–2800
77. Christensen B, Nielsen J (2000) Metabolic network analysis of *Penicillium chrysogenum* using ^{13}C-labeled glucose. Biotechnol Bioeng 68(6):652–659
78. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc 3 (8):1299–1311

79. Iwatani S, Van Dien S, Shimbo K, Kubota K, Kageyama N, Iwahata D, Miyano H, Hirayama K, Usuda Y, Shimizu K, et al. (2007) Determination of metabolic flux changes during fed-batch cultivation from measurements of intracellular amino acids by LC-MS/MS. J Biotechnol 128(1):93–111
80. Millard P, Letisse F, Sokol S, Portais J-C (2012) IsoCor: correcting MS data in isotope labeling experiments. Bioinformatics 28(9):1294–1296
81. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in 13C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng 85(3):259–268
82. Zhang Z, Shen T, Rui B, Zhou W, Zhou X, Shang C, Xin C, Liu X, Li G, Jiang J, et al. (2014) CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by 13C-fluxomics. Nucleic Acids Res 43:D549–D557
83. Shupletsov MS, Golubeva LI, Rubina SS, Podvyaznikov DA, Iwatani S, Mashko SV (2014) OpenFLUX2: ^{13}C-MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. Microb Cell Factories 13:152
84. Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W (2013) 13CFLUX2—high-performance software suite for 13C-metabolic flux analysis. Bioinformatics 29(1):143–145
85. Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab Eng 9(1):68–86
86. Young JD (2014) INCA: a computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics 30(9):1333–1335
87. Zamboni N, Fischer E, Sauer U (2005) FiatFlux–a software for metabolic flux analysis from ^{13}C-glucose experiments. BMC Bioinf 6:209–209
88. Garcia-Albornoz M, Thankaswamy-Kosalai S, Nilsson A, Väremo L, Nookaew I, Nielsen J (2014) BioMet toolbox 2.0: genome-wide analysis of metabolism and omics data. Nucleic Acids Res 42(Web server issue):W175–W181
89. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G (2008) An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. Biotechnol Bioeng 99(3):686–699
90. Chopra P, Kamma A (2006) Engineering life through synthetic biology. In Silico Biol 6 (5):401–410
91. Chen Y, Daviet L, Schalk M, Siewers V, Nielsen J (2013) Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism. Metab Eng 15:48–54
92. Jing F, Cantu DC, Tvaruzkova J, Chipman JP, Nikolau BJ, Yandeau-Nelson MD, Reilly PJ (2011) Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochem 12:44
93. Inui HOK, Miyatake K, Nakano Y, Kitaoka S (1987) Purification and characterization of pyruvate: NADP$^+$ oxidoreductase in *Euglena gracilis*. J Biol Chem 262(19):6
94. Kozak BU, van Rossum HM, Benjamin KR, Wu L, Daran JM, Pronk JT, van Maris AJ (2014) Replacement of the *Saccharomyces cerevisiae* acetyl-CoA synthetases by alternative pathways for cytosolic acetyl-CoA synthesis. Metab Eng 21:46–59
95. Shiba Y, Paradise EM, Kirby J, Ro D-K, Keasling JD (2007) Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. Metab Eng 9(2):160–168
96. Zaidi N, Swinnen JV, Smans K (2012) ATP-citrate lyase: a key player in cancer metabolism. Cancer Res 72(15):3709–3714
97. Frick O, Wittmann C (2005) Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative ^{13}C flux analysis. Microb Cell Factories 4:30
98. Meile L, Rohr LM, Geissmann TA, Herensperger M, Teuber M (2001) Characterization of the D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase gene (*xfp*) from *Bifidobacterium lactis*. J Bacteriol 183(9):2929–2936

99. Panagiotou G, Andersen MR, Grotkjaer T, Regueira TB, Hofmann G, Nielsen J, Olsson L (2008) Systems analysis unfolds the relationship between the phosphoketolase pathway and growth in *Aspergillus nidulans*. PLoS One 3(12):e3847
100. de Jong BW, Shi S, Siewers V, Nielsen J (2014) Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. Microb Cell Factories 13:39–39
101. Feng Y, Cronan JE (2009) *Escherichia coli* unsaturated fatty acid synthesis: complex transcription of the fabA gene and in vivo identification of the essential reaction catalyzed by FabB. J Biol Chem 284(43):29526–29535
102. Davis MS, Solbiati J, Cronan Jr JE (2000) Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. J Biol Chem 275 (37):28593–28598
103. Lu X, Vora H, Khosla C (2008) Overproduction of free fatty acids in *E. coli*: implications for biodiesel production. Metab Eng 10(6):333–339
104. Zhang F, Ouellet M, Batth TS, Adams PD, Petzold CJ, Mukhopadhyay A, Keasling JD (2012) Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. Metab Eng 14(6):653–660
105. Subrahmanyam S, Cronan Jr JE (1998) Overproduction of a functional fatty acid biosynthetic enzyme blocks fatty acid synthesis in *Escherichia coli*. J Bacteriol 180(17):4596–4602
106. Suástegui M, Guo W, Feng X, Shao Z (2016) Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*. Biotechnol Bioeng 113 (12):2676–2685
107. Perkins JB, Sloma A, Hermann T, Theriault K, Zachgo E, Erdenberger T, Hannett N, Chatterjee NP, Williams VII, Rufo GAJ, et al. (1999) Genetic engineering of *Bacillus subtilis* for the commercial production of riboflavin. J Ind Microbiol Biotechnol 22(1):8–18
108. Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wuthrich K (1997) Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. Nat Biotechnol 15(5):448–452
109. Dauner M, Bailey JE, Sauer U (2001) Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. Biotechnol Bioeng 76(2):144–156
110. Berrios-Rivera SJ, Bennett GN, San K-Y (2002) The effect of increasing NADH availability on the redistribution of metabolic fluxes in *Escherichia coli* chemostat cultures. Metab Eng 4 (3):230–237
111. San K-Y, Bennett GN, Berrios-Rivera SJ, Vadali RV, Yang Y-T, Horton E, Rudolph FB, Sariyar B, Blackwood K (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. Metab Eng 4(2):182–192
112. Sonderegger M, Jeppsson M, Hahn-Hägerdal B, Sauer U (2004) Molecular basis for anaerobic growth of *Saccharomyces cerevisiae* on xylose, investigated by global gene expression and metabolic flux analysis. Appl Environ Microbiol 70(4):2307–2317
113. Watanabe S, Abu Saleh A, Pack SP, Annaluru N, Kodaki T, Makino K (2007) Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from *Pichia stipitis*. Microbiology 153 (9):3044–3054
114. Jeppsson M, Bengtsson O, Franke K, Lee H, Hahn-Hägerdal B, Gorwa-Grauslund MF (2006) The expression of a *Pichia stipitis* xylose reductase mutant with higher KM for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. Biotechnol Bioeng 93(4):665–673
115. Watanabe S, Pack SP, Saleh AA, Annaluru N, Kodaki T, Makino K (2007) The positive effect of the decreased NADPH-preferring activity of xylose reductase from *Pichia stipitis* on ethanol production using xylose-fermenting recombinant *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 71(5):1365–1369
116. Runquist D, Hahn-Hägerdal B, Bettiga M (2009) Increased expression of the oxidative pentose phosphate pathway and gluconeogenesis in anaerobically growing xylose-utilizing *Saccharomyces cerevisiae*. Microb Cell Factories 8:49–49

117. Petschacher B, Nidetzky B (2008) Altering the coenzyme preference of xylose reductase to favor utilization of NADH enhances ethanol yield from xylose in a metabolically engineered strain of *Saccharomyces cerevisiae*. Microb Cell Factories 7:9–9
118. Bengtsson O, Hahn-Hägerdal B, Gorwa-Grauslund MF (2009) Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*. Biotechnol Biofuels 2:9–9
119. Runquist D, Hahn-Hägerdal B, Bettiga M (2010) Increased ethanol productivity in xylose-utilizing *Saccharomyces cerevisiae* via a randomly mutagenized xylose reductase. Appl Environ Microbiol 76(23):7796–7802
120. Watanabe S, Saleh AA, Pack SP, Annaluru N, Kodaki T, Makino K (2007) Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein engineered $NADP^+$-dependent xylitol dehydrogenase. J Biotechnol 130(3):316–319
121. Matsushika A, Watanabe S, Kodaki T, Makino K, Inoue H, Murakami K, Takimura O, Sawayama S (2008) Expression of protein engineered $NADP^+$-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 81(2):243–255
122. Krahulec S, Klimacek M, Nidetzky B (2009) Engineering of a matched pair of xylose reductase and xylitol dehydrogenase for xylose fermentation by *Saccharomyces cerevisiae*. Biotechnol J 4(5):684–694
123. Matsushika A, Inoue H, Watanabe S, Kodaki T, Makino K, Sawayama S (2009) Efficient bioethanol production by a recombinant flocculent *Saccharomyces cerevisiae* strain with a genome-integrated $NADP^+$-dependent xylitol dehydrogenase gene. Appl Environ Microbiol 75(11):3818–3822
124. Jeppsson M, Johansson B, Hahn-Hägerdal B, Gorwa-Grauslund MF (2002) Reduced oxidative pentose phosphate pathway flux in recombinant xylose-utilizing *Saccharomyces cerevisiae* strains improves the ethanol yield from xylose. Appl Environ Microbiol 68 (4):1604–1609
125. Verho R, Londesborough J, Penttilä M, Richard P (2003) Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*. Appl Environ Microbiol 69(10):5892–5897
126. Zhang G-C, Liu J-J, Ding W-T (2012) Decreased xylitol formation during xylose fermentation in *Saccharomyces cerevisiae* due to overexpression of water-forming NADH oxidase. Appl Environ Microbiol 78(4):1081–1086
127. Wasylenko TM, Stephanopoulos G (2015) Metabolomic and ^{13}C-metabolic flux analysis of a xylose-consuming *Saccharomyces cerevisiae* strain expressing xylose isomerase. Biotechnol Bioeng 112(3):470–483
128. Kim YM, Cho H-S, Jung GY, Park JM (2011) Engineering the pentose phosphate pathway to improve hydrogen yield in recombinant *Escherichia coli*. Biotechnol Bioeng 108 (12):2941–2946
129. Lee W-H, Park J-B, Park K, Kim M-D, Seo J-H (2007) Enhanced production of ε-caprolactone by overexpression of NADPH-regenerating glucose 6-phosphate dehydrogenase in recombinant *Escherichia coli* harboring cyclohexanone monooxygenase gene. Appl Microbiol Biotechnol 76(2):329–338
130. Chin JW, Cirino PC (2011) Improved NADPH supply for xylitol production by engineered *Escherichia coli* with glycolytic mutations. Biotechnol Prog 27(2):333–341
131. Wang Y, San K-Y, Bennett G (2013) Improvement of NADPH bioavailability in *Escherichia coli* through the use of phosphofructokinase deficient strains. Appl Microbiol Biotechnol 97 (15):6883–6893
132. Chemler JA, Fowler ZL, McHugh KP, Koffas MAG (2010) Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. Metab Eng 12 (2):96–104
133. Kim S, Lee CH, Nam SW, Kim P (2011) Alteration of reducing powers in an isogenic phosphoglucose isomerase (*pgi*)-disrupted *Escherichia coli* expressing NAD(P)-dependent

malic enzymes and NADP-dependent glyceraldehyde 3-phosphate dehydrogenase. Lett Appl Microbiol 52(5):433–440
134. Sánchez AM, Andrews J, Hussein I, Bennett GN, San K-Y (2006) Effect of overexpression of a soluble pyridine nucleotide transhydrogenase (*udhA*) on the production of poly (3-hydroxybutyrate) in *Escherichia coli*. Biotechnol Prog 22(2):420–425
135. Chou H-H, Marx CJ, Sauer U (2015) Transhydrogenase promotes the robustness and evolvability of *E. coli* deficient in NADPH production. PLoS Genet 11(2):e1005007
136. Wiechert W, de Graaf AA (1996) In vivo stationary flux analysis by ^{13}C labeling experiments. In: Sahm H, Wandrey C (eds) Metabolic engineering, vol 54. Springer, Berlin, pp 109–154
137. Wittmann C, Heinzle E (2001) Application of MALDI-TOF MS to lysine-producing *Corynebacterium glutamicum*. Eur J Biochem 268(8):2441–2455
138. Klapa MI, Aon J-C, Stephanopoulos G (2003) Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. Eur J Biochem 270 (17):3525–3542
139. Quek L-E, Wittmann C, Nielsen L, Kromer J (2009) OpenFLUX: efficient modelling software for ^{13}C-based metabolic flux analysis. Microb Cell Factories 8(1):25
140. Krömer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C (2004) In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J Bacteriol 186(6):1769–1784
141. Becker J, Klopprogge C, Herold A, Zelder O, Bolten CJ, Wittmann C (2007) Metabolic flux engineering of l-lysine production in *Corynebacterium glutamicum*—over expression and modification of G6P dehydrogenase. J Biotechnol 132(2):99–109
142. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C (2011) From zero to hero—design-based systems metabolic engineering of *Corynebacterium glutamicum* for l-lysine production. Metab Eng 13(2):159–168
143. Becker J, Klopprogge C, Zelder O, Heinzle E, Wittmann C (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(12):8587–8596
144. Bommareddy RR, Chen Z, Rappert S, Zeng A-P (2014) A *de novo* NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. Metab Eng 25:30–37
145. Hu X-Q, Chu J, Zhang S-L, Zhuang Y-P, Wang Y-H, Zhu S, Zhu Z-G, Yuan Z-Y (2007) A novel feeding strategy during the production phase for enhancing the enzymatic synthesis of S-adenosyl-l-methionine by methylotrophic *Pichia pastoris*. Enzym Microb Technol 40 (4):669–674
146. Hu X-Q, Chu J, Zhang Z, Zhang S-L, Zhuang Y-P, Wang Y-H, Guo M-J, Chen H-X, Yuan Z-Y (2008) Effects of different glycerol feeding strategies on S-adenosyl-l-methionine biosynthesis by PGAP-driven *Pichia pastoris* overexpressing methionine adenosyltransferase. J Biotechnol 137(1–4):44–49
147. Birnbaum S, Bailey JE (1991) Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. Biotechnol Bioeng 37 (8):736–745
148. Jones KL, Kim SW, Keasling JD (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab Eng 2(4):328–338
149. Guyot S, Gervais P, Young M, Winckler P, Dumont J, Davey HM (2015) Surviving the heat: heterogeneity of response in *Saccharomyces cerevisiae* provides insight into thermal damage to the membrane. Environ Microbiol 17:2982–2992
150. Nugroho RH, Yoshikawa K, Shimizu H (2015) Metabolomic analysis of acid stress response in *Saccharomyces cerevisiae*. J Biosci Bioeng 120:396–404

151. Parawira W, Tekere M (2011) Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review. Crit Rev Biotechnol 31(1):20–31
152. King T, Lucchini S, Hinton JCD, Gobius K (2010) Transcriptomic analysis of *Escherichia coli* O157:H7 and K-12 cultures exposed to inorganic and organic acids in stationary phase reveals acidulant- and strain-specific acid tolerance responses. Appl Environ Microbiol 76 (19):6514–6528
153. Ravikirthi P, Suthers PF, Maranas CD (2011) Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. Biotechnol Bioeng 108(6):1372–1382
154. Zamboni N, Sauer U (2009) Novel biological insights through metabolomics and 13C-flux analysis. Curr Opin Microbiol 12(5):553–558
155. Büscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N (2009) Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. Anal Chem 81 (6):2135–2143
156. Christen S, Sauer U (2011) Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. FEMS Yeast Res 11(3):263–272
157. Young JD, Shastri AA, Stephanopoulos G, Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary 13C flux analysis. Metab Eng 13(6):656–665
158. Ma F, Jazmin LJ, Young JD, Allen DK (2014) Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proc Natl Acad Sci 111(47):16967–16972
159. Brennan L, Owende P (2010) Biofuels from microalgae—a review of technologies for production, processing, and extractions of biofuels and co-products. Renew Sust Energ Rev 14(2):557–577
160. Varman A, Yu Y, You L, Tang Y (2013) Photoautotrophic production of D-lactic acid in an engineered cyanobacterium. Microb Cell Factories 12(1):117
161. Zhou J, Li Y (2010) Engineering cyanobacteria for fuels and chemicals production. Protein Cell 1(3):207–210
162. Feng X, Tang Y, Dolan KD (2014) Construction of a parsimonious kinetic model to capture microbial dynamics via parameter estimation. Inverse Probl Sci Eng 22(2):309–324
163. Xiao Y, Feng X, Varman AM, He L, Yu H, Tang YJ (2012) Kinetic modeling and isotopic investigation of isobutanol fermentation by two engineered *Escherichia coli* strains. Ind Eng Chem Res 51(49):15855–15863
164. Hoefnagel MHN, Starrenburg MJC, Martens DE, Hugenholtz J, Kleerebezem M, Van Swam II, Bongers R, Westerhoff HV, Snoep JL (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology 148(4):1003–1013
165. Gombert AK, Nielsen J (2000) Mathematical modelling of metabolism. Curr Opin Biotechnol 11(2):180–186
166. Wittmann C, Kiefer P, Zelder O (2004) Metabolic fluxes in *Corynebacterium glutamicum* during lysine production with sucrose as carbon source. Appl Environ Microbiol 70 (12):7277–7287
167. Becker J, Klopprogge C, Wittmann C (2008) Metabolic responses to pyruvate kinase deletion in lysine producing *Corynebacterium glutamicum*. Microb Cell Factories 7(1):8
168. Kiefer P, Heinzle E, Zelder O, Wittmann C (2004) Comparative metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* cultured on glucose or fructose. Appl Environ Microbiol 70(1):229–239
169. Wittmann C, Heinzle E (2002) Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing *Corynebacteria*. Appl Environ Microbiol 68(12):5843–5859
170. Leighty RW, Antoniewicz MR (2013) COMPLETE-MFA: complementary parallel labeling experiments technique for metabolic flux analysis. Metab Eng 20:49–55
171. Crown SB, Long CP, Antoniewicz MR (2015) Integrated ^{13}C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. Metab Eng 28:151–158

172. Leighty RW, Antoniewicz MR (2012) Parallel labeling experiments with [U-^{13}C]glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis. Metab Eng 14 (5):533–541
173. Crown SB, Antoniewicz MR (2013) Parallel labeling experiments and metabolic flux analysis: past, present and future methodologies. Metab Eng 16:21–32
174. Wiechert W, Möllney M, Petersen S, de Graaf AA (2001) A universal framework for ^{13}C metabolic flux analysis. Metab Eng 3(3):265–283
175. Cvijovic M, Olivares-Hernández R, Agren R, Dahr N, Vongsangnak W, Nookaew I, Patil KR, Nielsen J (2010) BioMet toolbox: genome-wide analysis of metabolism. Nucleic Acids Res 38(suppl 2):W144–W149
176. Srour O, Young JD, Eldar YC (2011) Fluxomers: a new approach for ^{13}C metabolic flux analysis. BMC Syst Biol 5:129–129
177. Kajihata S, Furusawa C, Matsuda F, Shimizu H (2014) OpenMebius: an open source software for isotopically nonstationary ^{13}C-based metabolic flux analysis. Biomed Res Int 2014:10
178. Sokol S, Millard P, Portais J-C (2012) influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. Bioinformatics 28(5):687–693
179. Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J (2001) Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. J Bacteriol 183(4):1441–1451
180. Gill PE, Murray W, Saunders MA (2005) SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev 47(1):99–131
181. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical recipes in C: the art of scientific computing. Cambridge University Press, New York
182. Jordà J, Jouhten P, Cámara E, Maaheimo H, Albiol J, Ferrer P (2012) Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures. Microb Cell Factories 11:57–57
183. Hayakawa K, Matsuda F, Shimizu H (2016) Metabolome analysis of *Saccharomyces cerevisiae* and optimization of culture medium for S-adenosyl-l-methionine production. AMB Express 6(1):38

Adv Biochem Eng Biotechnol (2018) 162: 301–316
DOI: 10.1007/10_2016_14

Published online: 1 June 2017

Xenobiology: State-of-the-Art, Ethics, and Philosophy of New-to-Nature Organisms

Markus Schmidt, Lei Pei, and Nediljko Budisa

Abstract The basic chemical constitution of all living organisms in the context of carbon-based chemistry consists of a limited number of small molecules and polymers. Until the twenty-first century, biology was mainly an analytical science and has now reached a point where it merges with engineering science, paving the way for synthetic biology. One of the objectives of synthetic biology is to try to change the chemical compositions of living cells, that is, to create an artificial biological diversity, which in turn fosters a new sub-field of synthetic biology, xenobiology. In particular, the genetic code in living systems is based on highly standardized chemistry composed of the same "letters" or nucleotides as informational polymers (DNA, RNA) and the 20 amino acids which serve as basic building blocks for proteins. The universality of the genetic code enables not only vertical gene transfer within the same species but also horizontal gene transfer across biological taxa, which require a high degree of standardization and interconnectivity. Although some minor alterations of the standard genetic code are found in nature (e.g., proteins containing non-conical amino acids exist in nature, and some organisms use alternated coding systems), all structurally deep chemistry changes within living systems are generally lethal, making the creation of artificial biological system an extremely difficult challenge.

In this context, one of the great challenges for bioscience is the development of a strategy for expanding the standard basic chemical repertoire of living cells. Attempts to alter the meaning of the genetic information stored in DNA as an informational polymer by changing the chemistry of the polymer (i.e., xeno-nucleic acids) or by changes in the genetic code have already yielded successful results. In

M. Schmidt (✉) and L. Pei
Biofaction KG, Kundmanngasse 39/12, Vienna 1030, Austria
e-mail: schmidt@biofaction.com

N. Budisa
AK Biokatalyse, Institut für Chemie, Technische Universität Berlin, Müller-Breslau-Straße 10, 10623 Berlin, Germany

the future this should enable the partial or full redirection of the biological information flow to generate "new" version(s) of the genetic code derived from the "old" biological world.

In addition to the scientific challenges, the attempt to increase biochemical diversity also raises important ethical and philosophical issues. Although promotors of this branch of synthetic biology highlight the many potential applications to come (e.g., novel tools for diagnostics and fighting infection diseases), such developments could also bring risks affecting social, political, and other structures of nearly all societies.

Keywords Ethics, New-to-nature, Non-canonical amino acids, Philosophy, Synthetic biology, Xenobiology

Contents

1 The Future of Biology

The genetic program of living cells is considered as the "software of life" [1, 2]. However, are we able to read and interpret this "software" correctly? If so, we would be able to understand how life works and we could try to change and even improve the "software" with man-made versions. From the viewpoint of synthetic biology, living cells are small, programmable production units (e.g., "similar to" a robot or a chemical machine). Researchers on the frontlines of this field are seeking ways to understand and create new types of cells for useful purposes, such as engineering cells to produce nearly every imaginable chemical compound for utilization (not only natural compounds but also synthetic compounds) in the advancing fields of medicine and technology.

From the beginning of agricultural domestication (e.g., wheat cultivation which can be dated back a thousand years) [3], and especially from the onset of genetic engineering (since the 1970s to 1980s and the development of molecular cloning technologies) [4], the gap between natural and modified organisms is steadily increasing, such that the modified organisms are not only those harboring heterogeneous genes from other natural organisms but also those with totally artificial genetic makeup. At the end there awaits artificial life, genetically as well as metabolically distant from its natural origin. The new life forms will probably be genetically isolated, that is, they possess a kind of a genetic firewall serving as a

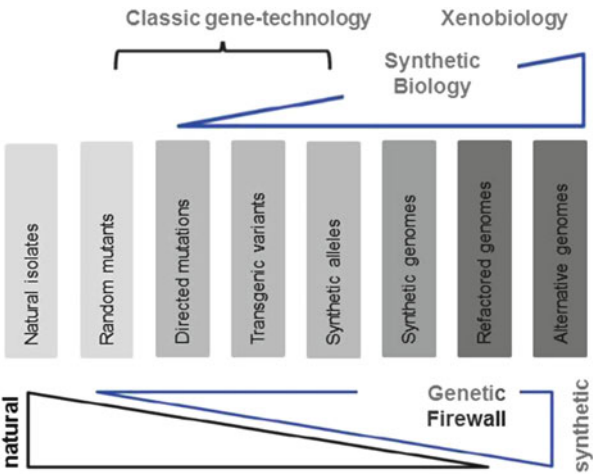

Fig. 1 De Lorenzo [6] model of the eight stages explaining the transition from naturally occurring organisms to completely artificial microbes

biological containment strategy to prevent horizontal gene transfer [5]. Horizontal gene transfer is a common cause of a gene spreads from one species to another, facilitated mainly by phage mediated transduction, sequence independent uptake of free DNA (transformation), or pili mediated conjugation. Accordingly, a stepwise model with eight developmental phases toward the creation of alternative forms of life has been developed [6], as shown in Fig. 1. This model was supplemented with the categories of xenobiology and genetic firewall, and placed in the current scientific development of xenobiology somewhere between steps 6 (synthetic genomes), 7 refactored genomes, and eventually 8 (alternative genomes).

One of the keystones of Darwinism is the fact that geographically (and hence genetically) isolated species tend to evolve unique and heritable changes over time. The classical example is Darwin's finches, which illustrates the way that gene pools of the finch have adapted to take advantage of feeding conditions in different ecologic settings for long-term survival. What is true for Darwin's finches also applies to cells in general. Through man-made, directed evolution of life forms we can attempt to achieve the implementation of new and sophisticated chemistries (elements, reactions, metabolic pathways) into the protoplasm of desired life forms [7]. It is the ambition of a number of synthetic biologists to find out experimentally how far we could go toward this objective [8]. As for the idea of a genetic firewall, it is crucial to learn whether the chemical standard composition of terrestrial life forms (invariant for almost four billion years!) could be changed in principle and whether we could open the door to a parallel biological world. Meanwhile, some caveats need to be considered for those experiments to create new species, as issues brought up by Buckling et al. for experiments on evolution, e.g., the simplicity of "testing tube" conditions, the homogeneity of the testing population, and other unpredictable factors. All these caveats for experimental evolution are also applicable to xenobiological approaches to create the parallel biological system [9].

In nature, the energy flux on Earth mandates a cyclic material flow with a simultaneously continuous maintenance of order, resulting in the formation of living systems. Morowitz stated that such a process was essential and deterministic

[10]. All the necessary information about the energy and material flows is encoded in the genomes of organisms. Following this paradigm, the chemical compositions and the choice for the fundamental units both affect the processes of life. The molecules participating in life on Earth cover primarily amino acid polymers (proteins), nucleic acids (mainly DNA and RNA), lipids, and other small molecules which act as coenzymes and cofactors. Although the number of monomeric building blocks of life is rather small, their transient combinations lead to very diverse molecules [11]. The genetic program of all living cells (and viruses) is mainly based on information encoded in nucleic acid structures and most biological activities are determined by protein structures. Besides, huge repertoires of other macromolecules such as fatty acids, carbohydrates, and small molecules, and metabolic pathways and modes of information processing, are common to the cells [12].

2 Biology Can Be Synthesized Biologically and/or Chemically

In the first years of today's ubiquitous synthetic chemistry, the synthesis of complex substances, originally produced from plants and animals, was assumed to be an impossible task. Additionally, a lot of physiological conditions were experimentally inaccessible in those days. This left space for the appearance of metaphysical concepts such as the idea that organic compounds were just formed in the presence of a special, vital power ("vis vitalis") acting exclusively in creatures. Accordingly, metaphysical concepts were used as the main criteria to decide between animate and inanimate matter [2, 13]. Yet at the beginning of the nineteenth century this metaphysical viewpoint was proven wrong by chemical synthesis of organic molecules (e.g., urea in Woehler's *Harnstoffsynthese* in 1828) [14]. Although this was not the first milestone for the synthesis of naturally occurring organic compounds, starting from then the awareness of the accessibility of natural, organic molecules increased. Complex compounds could be manufactured starting from simple structures in a stepwise and controlled manner. Less than 50 years later, organic synthetic chemistry has turned into an engineering discipline with the ambition to synthesize all naturally occurring organic substances [15]. Nowadays, synthetic biology has a similar goal: to define biological parts of living systems as modules, standardize them, and combine these standardized parts into a novel organism. Xenobiology goes a step further, aiming at the compositional (chemical) redesign of these particular modules [16–18], which goes beyond the concept of building novel system based on naturally existing or modified modules for synthetic biology. In general, xenobiology aims to design biological systems endowed with unusual biochemistries.

The concept of modularity, a prerequisite for synthetic biology, arose out of the observed successes in other engineering fields (e.g., software or electronic engineering) [19, 20]. A modular approach should facilitate the simplification of

biological systems to make it possible to define first principles for a biological hierarchy until the creation of biological systems from scratch is achievable (bottom-up principle). As soon as these modular units, for example, synthetic networks on the levels of transcription, translation, and signal transduction, as well as the metabolism, are orthogonalized (uncoupled) from their biological context, it should be feasible to add new parts to a system without facing unwanted side effects or cross reactions. Of course, this is an idealization of biological engineering, because in reality unforeseen effects emerge as soon as the complexity of the desired organism increases.

Biomolecules, and especially genes, have plenty of undetermined degrees of freedom on the molecular level and on the level of interactions and functions, which explains the difficulty when orthogonalizing these molecules within the biological context [21]. This is best illustrated by recently reported construct of a new synthetic "minimal" *Mycoplasma* containing 149 genes of unknown function which are somehow essential for growth on a defined growth medium [22]. It is indeed surprising that almost one-third of the "minimal" genome corresponds to unknown functions. This was however not obvious in the initial report used for creating artificial life. The construction of the "first self-replicating synthetic bacterial cell" was accomplished via copying a natural genome, in which additional synthetic but essentially inactive DNA sequences were inserted [23]. Nonetheless, the successful construction of synthetic cells proves that chemically synthesized modules could turn into living cells, which paves way for turning synthetic genomes containing unnatural genetic letters into living organism as well.

3 Motives for the Development of New Biological Systems

Synthetic biology offers a perspective for the development of a multitude of novel, chemically diverse biocatalysts for the production of fuel, additives, or medicines, amongst others [24]. Although synthetic biology mainly works with naturally existing building blocks and a canonical chemistry, xenobiological applications use non-natural building blocks and non-canonical chemistry. Thus the aim of xenobiology covers the implementation of these man-made chemical syntheses in living cells, e.g., engineered organisms that could conduct metathesis pathways [25] or similar chemical transformations which are still the exclusive domain of the synthetic organic chemist.

The simplest biological models used in xenobiology are microorganisms, which can be seen as ready-made production systems primarily ruled by their genetic programs [26–28]. By introducing small changes in the genetic program of an organism, a bioengineer can reach significant changes in terms of the production's results. This is done experimentally with refunctionalization, reprogramming, or recoding of natural processes [29, 30].

The fundamental difference between synthetic biology and xenobiology is that in synthetic biology living systems are restructured via exchange and combination

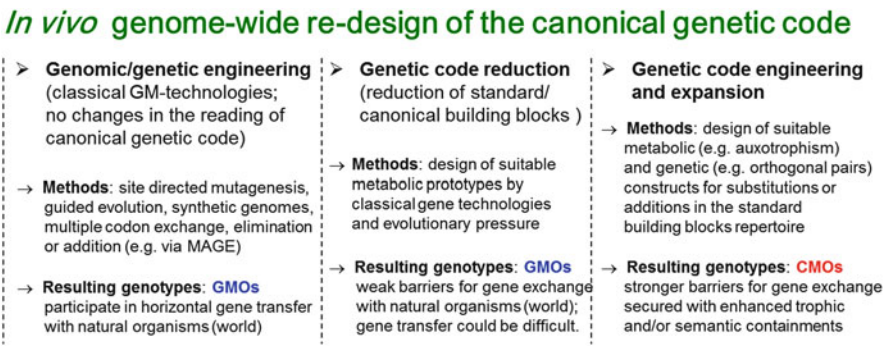

Fig. 2 Differences between GMOs and CMOs in a hypothetical experiment of genome-wide redesign of standardized genetic code in a given organism. (*GM* genetic modification, *MAGE* multiplex automated genome engineering [31])

of standardized parts (modules, biobricks). In contrast, xenobiology uses non-natural (or so-called non-canonical) molecules to create CMOs (chemically modified organisms) [16]. These CMOs manage to use up-to-date unused chemical elements (e.g., fluorine or boron), novel "letters," building blocks, or scaffolds (the differences between GMOs and CMOs are shown in Fig. 2). To achieve this, researchers plan to come up with an alternative genetic code, necessitating a conversion of the whole flow of genetic information [1, 5, 7, 11, 12, 32].

Other than serving as novel building blocks for genomes, non-canonical DNA bases can develop into diagnostic tools for infectious diseases [33]. The unnatural base pair system consists of an expanded genetic alphabet built into oligo nucleotide fragments on specific sites, or via enzymatic incorporation of extra, functional components into nucleic acids. These fragments containing unnatural base pairs can be obtained via PCR amplifications. Diagnostic molecular beacons with fluorescent dye linked to the unnatural bases can serve as molecular diagnostic tools, for example, to target infectious diseases of interest [34]. Furthermore, aptamers containing unnatural bases are considered valuable for pharmaceutical applications because of their unique features in affinity, thermo stability, and resistance to nucleases [35].

If we manage to change the way the genetic code is read in a living organism and to add new "letters" or building blocks, the corresponding cell constitutes a genetic enclave because the genetic exchange with natural cells is impaired. This is an important aspect for biological safety, because the risk of horizontal gene transfer to natural cells is supposed to be strongly reduced [16, 21, 24, 36–38]. Therefore, xenobiology seeks conditions in which the cells can be cultivated in the laboratory but stay genetically isolated from naturally occurring species [39].

4 Present State of Xenobiology

Currently, xenobiology and even synthetic biology is not widely harnessed to tackle modern technological questions because of the overwhelming diversity of existing structures and information transmission pathways (e.g., horizontal gene transfer, mutation, recombination) already present in nature. Reprogrammed cells or proteins equipped with synthetic structures are just considered as useful tools for academic research or small applications, if any. Interestingly, however, xenobiology is not a more recent development. Yet in the 1950s and 1960s (albeit under a different name), the incorporation of, for example, ncAAs into the proteome of organisms was demonstrated to be approachable [40]. Back then, auxotrophic microbial strains were used, which lost their ability to synthesize a particular, essential nutrient and forced them – by feeding a structurally similar artificial compound – to adjust to this certain substitute, leaving them just the choice to "take it or leave it" [41–43].

The current synthesis of alternative biological systems within the framework of genome engineering is in particular focused on the three universal biomolecules DNA, RNA, and amino acids, and on the genetic code redesign via directed evolution of microbial strains. All basic constituent parts of DNA, that is, the nucleobases, the deoxyribose, and the phosphate backbone, can be exchanged with alternative chemical structures such as xeno (noncanonical) nucleic (XNAs) [32, 44–49]. According to this rationale, we certainly face progressive advancement in the construction of novel biological systems running with XNA in the near future. For example, experimental evolution has been successfully used to engineer bacterial genomes with XNA [50] or proteomes with ncAAs [51, 52] (see below). Recently, Issacs and Church [45, 53] also showed that the incorporation of various ncAAs into some *E. coli* essential genes can serve as a promising biosafety tactic: As long as the ncAAs is absent from the medium, no bacterial growth could be detected. Obviously, substitution of canonical amino acids with ncAAs and the expansion of the genetic code in essential genes with ncAAs can be promising strategies to isolate further synthetic organisms from natural ones [54, 55]. Such strains can even have practical importance when applications such as bioremediation (in open systems) or industrial biocatalysis (closed systems) are considered [56].

In this context, it is attractive to reassign some of the (degenerated and rare) codons of the genetic code as recently reported by Budisa and Bohlke [57]. They succeeded "to deprive" *Escherichia coli* of the capacity to read one of its own triplets, the AUA codon, and thus "emancipated" these bacteria to translate all 5,797 AUA triplets with synthetic or alien amino acids. This presents the first step toward a so-called codon reassignment during which new amino acids, which do not occur in nature, are inserted into the genetic code. These cells feature a different genetic code relative to all other living organisms and present, therefore, a preliminary stage to completely synthetic cells.

Most recently, the group of Budisa [51] reported a long-term evolution which led to 20,899 reassignments in the genetic code of the bacterium *Escherichia coli*. In particular, a long-term cultivation experiment in defined synthetic media resulted in the evolution of cells capable of surviving full tryptophan to thienopyrrole-alanine substitutions in their proteomes in response to all TGG codons in the genome. These evolved bacteria with their new-to-nature amino acid composition are capable of robust growth in the complete absence of the canonical (natural) amino acid tryptophan. Doubtless such experimental results not only reveal that translational ambiguity is essential for the evolution of alternative genetic codes; they also pinpoint a strategy for the evolution of synthetic cells with alternative biochemistries [58]. It should be noted, however, that 20,899 UGG codons in Budisa's evolution experiments could be defined as trophically reassigned (i.e., the meaning of a codon is redefined throughout the whole translationary machinery for the evolved cells only in the defined synthetic medium). However, supplementation of cells in such a media with canonical substrate tryptophan reverses them to "natural" ones as they still favor the incorporation of the canonical building block. To achieve a nutrient-independent reassignment (i.e., "real" codon reassignment) for all the genome UGG codons in *E. coli*, an experimental strategy for biocontainment needs to be developed and executed.

5 Trophic and Semantic Containment, Astrobiochemistry and the Origin of Life

The genetic code is almost universal on Earth and its way of being read is a crucial step for the transfer of genetic information. With the creation of a modified microbial strain showing reprogrammed codons throughout its genome, the experimental change of this unity would be achieved. The natural limitation of the genetic code exemplified by a limited repertoire of amino acids as building units could be transcended via trophic and/or semantic containment [59]. The trophic containment is to make microorganisms to be dependent on unnatural nutrients (xeno-nutrients). In more detail, trophic containment means the implementation of xeno-nutrients and the prevention by cross-feeding of natural alternative nutrients or analogues (59), whereas semantic containment is based upon the prevention of genetic information exchange, for example, by using different interpretation systems such as those using XNA or those with an alternative genetic code [21].

An alternative genetic code may decipher for a smaller or larger number of amino acids. In addition, a selection of amino acids could be replaced by ncAAs (genetic code engineering) or a selection of ncAAs could be added to the genetic code's repertoire (genetic code expansion). To equip the genetic code with novel chemical functionalities, some yet occupied codons have to be released from their original function, hence uncoupling cells from the canonical reading of individual codon triplets (codon emancipation) [19, 57, 60].

Meanwhile, xenobiology is an emergent area at the interface of synthetic biology and synthetic organic chemistry which aims to construct biological systems endowed with novel biochemistries such as XNAs and/or "xeno" amino acids (usually called noncanonical amino acids, ncAAs). Xenobiology aims to answer the fundamental questions in the chemistry of life: can biological systems also function with an alternative genetic code composed of XNA or ncAAs as building blocks or both? Xenobiological research is therefore closely linked to studies on the origin of life, including the development of the genetic code [5, 17, 61, 62].

Consequently, a conversation with astrobiochemistry or astrobiology yields important insights. For instance, the intensive chemical analysis of carbon-containing meteorites such as the Murchison meteorite demonstrated the presence of more than 70 extraterrestrial amino acids of which the L-enantiomers were more dominant [46]. By testing the suitability of these amino acids as building blocks for the production of proteins in terrestrial life forms (viruses, archaea, eubacteria, fungi, plants, and eukaryotes), we would gain new knowledge regarding the experimental rules and determinants by which the universal genetic code on Earth is limited to 20 AAs.

Xenobiology reached a highly significant milestone by demonstrating the synthesis of an organism that in at least some aspect and under some environmental conditions has a higher evolutionary fitness than a natural organism [63, 64]. In this context, the above-mentioned experiment of the Budisa group [51] represents "the most structurally disturbing deviation introduced into Life so far" (P. Marliere, personal communication). It shows how far the experimental evolution of the bacterium *Escherichia coli* can be pushed by demonstrating that complete replacement of one the endogenous building blocks, tryptophan (20,899 TGG codons) by an exogenous/synthetic one (thienylpyrrole) is possible. On the other hand, the general importance of such an engineering experiment is enormous: it suggests that ncAAs may indeed be potentially advantageous in some artificial media (environment) which is also important to keep in mind when we search for life elsewhere in the universe. Namely, most likely it "would have a biochemistry different from life on our planet" [58].

6 Ethical and Philosophic Considerations on Xenobiology

The abilities of xenobiology to construct microorganisms with new-to-nature biological systems and functions by cellular tinkering and experimental evolution require all stakeholders to use this technology safely and responsibly [36]. The issues include both ethical and philosophical considerations. Responsible research calls for engagement from all involved stakeholders ranging from researchers from both academy and industry to the regulatory authorities to the public at large [65, 66]. Regarding ethical considerations, we should bear in mind that xenobiology is not a fundamental challenge to bioethics. As a subfield of synthetic biology which gains a lot of attention regarding bioethics, both are all still part of biology.

The first ethical consideration is safety. This concern comes in multiple facets, ranging from biosafety, risk assessment to impact on health and environment [67, 68]. The biosafety challenges of xenobiology have been reviewed as part of synthetic biology risk assessment [69–71]. Possible impacts on health needing careful attention include novel toxicity, allergenicity, and pathogenicity of the new-to-nature molecules (and eventually life forms) generated by xenobiology, and the potential impacts on the environment include ecological competitiveness and the degree of horizontal gene transfer (e.g., lack of metric to measure the escape frequency of these types of containment) [72]. One could not rule out the challenges raised from xenobiology for biosecurity, directly or indirectly. Examples are the potential to develop novel pathogens with no available treatment option because they might have a different makeup, making them resistant to available drugs or native defense mechanisms of the recipient hosts as a direct challenge, or techniques developed by xenobiological research that would be abused to produce toxins or restricted chemicals as an indirect challenge.

The second consideration is the viewpoint of novel organisms as machines [73–76]. Although the morality of microorganisms is in question, some scholars insist that microbes should be granted the moral status that they should have the right to exist as they have been [77]. In a reverse scenario, xenobiology, instead of posing a threat to microbes, speculates about adding novel species to the world (although contained) that are different from existing ones on a biochemical level (see, e.g., [56]). For the time being, the speed of bringing up new "species" is foreseeably much lower than the rate by which existing biodiversity is eliminated as a consequence of human behavior in the so-called anthropocene. The argument of adding new species should not be understood as an excuse to curb down the efforts of stopping the ongoing global biodiversity decline because it is hardly possible to compare the (ecological) importance of an existing species with a newly created one.

The third consideration is related to intellectual property (IP). The IP issue gravitates around how the knowledge and technologies accumulating in research can be shared and translated to the market [65, 76, 78–80]. This type of technology, however, might also be developed into a stringent control of productions based on microbes equipped with controlled synthetic auxotrophy, which could lead to the establishment of terminator-like technological solutions (where a certain chemical has to be purchased to guarantee the survival of the cells). Xenobiology could, however, also be explored and developed by open source biologists into a realm of biotechnology that is free of IP restrictions.

Another scenario is the development of different IP regimes in distinct xenobiological fields (e.g., restrictive in HNA and permissive in CeNA variants of nucleic acids, or vice versa). It is unclear to what extent current access and benefit sharing agreements, such as the UN Nagoya protocol [81], apply to xenobiology at all, opening the way for alternative forms of sharing [82].

For philosophic considerations, other than the "playing God" hubris concern which has been discussed extensively, the considerations for xenobiological research deal with scientific attempts to alter or redirect evolution [83, 84]. It is

known that microbes equipped with novel genetic makeup (e.g., using XNA as genetic information carrier) have been engineered, which are usually equipped with an error-prone mechanism in their genetic replication [85]. However, natural evolution has taken millions of years to reach the less error-prone biological systems we know today. The rationale to alter or redirect this natural process to fit the research purposes may be in question. Meanwhile, one would probably need to ask the question as to whether it makes sense to redirect evolution within an extremely short time frame to achieve something comparable to natural evolution?

Xenobiological research should learn from the lessons of other research fields (such as agriculture research, stem cell research, and nanotechnology) to take into account the ethical and philosophic considerations relating to the research and development of the field. In addition, the research should also embed the Responsible Research and Innovation framework to serve as an optimal model for an emerging technology, to avoid developing a technology that does not benefit society, and to build trust in research and innovation. More open (and open ended) debates on xenobiology are needed to produce the broad stakeholder involvement as foreseen in Responsible Research and Innovation.

Acknowledgments The thoughts and ideas presented here are largely results of our interaction with Philippe Marlière, Sven Panke, Piet Herdewijn, Carlos-Acevedo Rocha, and Dirk Schulze-Makuch. Another very fortunate circumstance was that we worked together in EU-FP7 founded project METACODE (289572) whereby we could start to implement some of our conceptual ideas in the field of xenobiology. MS and NB also acknowledge support form EC FP7 project SYNPEPTIDE (613981) and MS acknowledges EC FP7 project SYNENERGENE (321488).

References

1. Budisa N (2014) Life at the speed of light. From the double helix to the dawn of digital life. By J. Craig Venter. Angew Chem Int Ed 53(36):9421–9422
2. Venter JC (2013) Life at the speed of light – from the from the double helix to the dawn of digital life. Viking Penguin, New York
3. Balter M (2015) Farming was so nice, it was invented at least twice. Science news, from http://www.sciencemag.org/news/2013/07/farming-was-so-nice-it-was-invented-least-twice
4. Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. Nat Rev Microbiol 12(5):381–390
5. Schmidt M (2010) Xenobiology: a new form of life as the ultimate biosafety tool. BioEssays 32(4):322–331
6. de Lorenzo V (2010) Environmental biosafety in the age of synthetic biology: do we really need a radical new approach? Environmental fates of microorganisms bearing synthetic genomes could be predicted from previous data on traditionally engineered bacteria for in situ bioremediation. BioEssays 32(11):926–931
7. Wiltschi B, Budisa N (2007) Natural history and experimental evolution of the genetic code. Appl Microbiol Biotechnol 74(4):739–753
8. Heinemann M, Panke S (2006) Synthetic biology – putting engineering into biology. Bioinformatics 22(22):2790–2799
9. Buckling A, Craig Maclean R, Brockhurst MA, Colegrave N (2009) The Beagle in a bottle. Nature 457(7231):824–829

10. Morowitz HJ, Heinz B, Deamer DW (1988) The chemical logic of a minimum protocell. Orig Life Evol Biosph 18(3):281–287
11. Budisa N (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. Angew Chem Int Ed Engl 43(47):6426–6463
12. Budisa N (2005) Expanding the amino acid repertoire for the design of novel proteins. Willey-VHC, Weinheim/New York/Brisbane/Singapore/Toronto
13. Church G, Regis E (2012) Regenesis: how synthetic biology will reinvent nature and ourselves. Basic Books, New York
14. Multhauf RP (1966) The origins of chemistry. Oldbourne, London
15. Fisher E (1907) Synthetic chemistry in its relation to biology (Faraday Lecture). J Chem Soc Chem Commun 91:1749–1765
16. Acevedo-Rocha CG, Budisa N (2011) On the road towards chemically modified organisms endowed with a genetic firewall. Angew Chem Int Ed Engl 50(31):6960–6962
17. Mampel J, Buescher JM, Meurer G, Eck J (2013) Coping with complexity in metabolic engineering. Trends Biotechnol 31(1):52–60
18. Schmidt M, de Lorenzo V (2012) Synthetic constructs in/for the environment: managing the interplay between natural and engineered biology. FEBS Lett 586(15):2199–2206
19 Acevedo-Rocha CG (2016) The synthetic nature of biology. In: Hagen K, Engelhard M, Toepfer G (eds) Ambivalences of creating life: societal and philosophical dimensions of synthetic biology. Springer, Switzerland, pp 9–53
20. Agapakis CM, Silver PA (2009) Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. Mol Biosyst 5(7):704–713
21. Budisa N (2014) Xenobiology, new-to-nature synthetic cells and genetic firewall. Curr Org Chem 18(8):936–943
22. Hutchison CA III, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC (2016) Design and synthesis of a minimal bacterial genome. Science 351(6280):aad6253
23. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56
24. SCHER, SCENIHR, SCCS (2014) Opinion on synthetic biology I definition. Available at http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_044.pdf
25. Mayer C, Gillingham DG, Ward TR, Hilvert D (2011) An artificial metalloenzyme for olefin metathesis. Chem Commun (Camb) 47(44):12068–12070
26. Alterovitz G, Muso T, Ramoni MF (2010) The challenges of informatics in synthetic biology: from biomolecular networks to artificial organisms. Brief Bioinform 11:80–95
27. Danchin A (2009) Information of the chassis and information of the program in synthetic cells. Syst Synth Biol 3(1-4):125–134
28. Landrain TE, Carrera J, Kirov B, Rodrigo G, Jaramillo A (2009) Modular model-based design for heterologous bioproduction in bacteria. Curr Opin Biotechnol 20(3):272–279
29. Keasling JD (2008) Synthetic biology for synthetic chemistry. ACS Chem Biol 3(1):64–76
30. Noirel J, Ow SY, Sanguinetti G, Wright PC (2009) Systems biology meets synthetic biology: a case study of the metabolic effects of synthetic rewiring. Mol Biosyst 5(10):1214–1223
31. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342(6156):357–360
32. Herdewijn P, Marliere P (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. Chem Biodivers 6(6):791–808
33. Benner SA, Sismour AM (2005) Synthetic biology. Nat Rev Genet 6(7):533–543

34. Kimoto M, Cox RS 3rd, Hirao I (2011) Unnatural base pair systems for sensing and diagnostic applications. Expert Rev Mol Diagn 11(3):321–331
35. Matsunaga K, Kimoto M, Hanson C, Sanford M, Young HA, Hirao I (2015) Architecture of high-affinity unnatural-base DNA aptamers toward pharmaceutical applications. Sci Rep 5:18478
36. SCHER, SCENIHR, SCCS (2015) Opinion on synthetic biology II - risk assessment methodologies and safety aspects. Available at http://ec.europa.eu/health/scientific_committees/consultations/public_consultations/scenihr_consultation_26_en.htm
37. Wright O, Delmans M, Stan GB, Ellis T (2015) GeneGuard: a modular plasmid system designed for biosafety. ACS Synth Biol 4(3):307–316
38. Wright O, Stan GB, Ellis T (2013) Building-in biosafety for synthetic biology. Microbiology 159(Pt 7):1221–1235
39. Schmidt M (2013) Safeguarding the genetic firewall with xenobiology. 21st century borders/synthetic biology: focus on responsibility and governance, Institute on Science for Global Policy, Tucson, Arizona
40. Brookes P (1959) Studies on the incorporation of an unnatural amino acid, p-di-(2-hydroxy [14C2]ethyl)amino-L-phenylalanine, into proteins. Br J Cancer 13:313–317
41. Beiboer SH, van den Berg B, Dekker N, Cox RC, Verheij HM (1996) Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A2. Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH. Protein Eng 9(4):345–352
42. Budisa N, Minks C, Alefelder S, Wenger W, Dong F, Moroder L, Huber R (1999) Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. FASEB J 13(1):41–51
43. Lemeignan B, Sonigo P, Marlière P (1993) Phenotypic suppression by incorporation of an alien amino acid. J Mol Biol 231(2):161–166
44. Jang MY, Song XP, Froeyen M, Marliere P, Lescrinier E, Rozenski J, Herdewijn P (2013) A synthetic substrate of DNA polymerase deviating from the bases, sugar, and leaving group of canonical deoxynucleoside triphosphates. Chem Biol 20(3):416–423
45. Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM (2015) Biocontainment of genetically modified organisms by synthetic protein design. Nature 518(7537):55–60
46. Pinheiro VB, Holliger P (2012) The XNA world: progress towards replication and evolution of synthetic genetic polymers. Curr Opin Chem Biol 16(3-4):245–252
47. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew SY, McLaughlin SH, Herdewijn P, Holliger P (2012) Synthetic genetic polymers capable of heredity and evolution. Science 336(6079):341–344
48. Steele FR, Gold L (2012) The sweet allure of XNA. Nat Biotechnol 30(7):624–625
49. Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, Weeks KM, Herdewijn P, Holliger P (2015) Catalysts from synthetic genetic polymers. Nature 518:427–430
50. Marliere P, Patrouix J, Doring V, Herdewijn P, Tricot S, Cruveiller S, Bouzon M, Mutzel R (2011) Chemical evolution of a bacterium's genome. Angew Chem Int Ed Engl 50 (31):7109–7114
51. Hoesl MG, Oehm S, Durkin P, Darmon E, Peil L, Aerni HR, Rappsilber J, Rinehart J, Leach D, Soll D, Budisa N (2015) Chemical evolution of a bacterial proteome. Angew Chem Int Ed Engl 54(34):10030–10034
52. Ma Y, Biava H, Contestabile R, Budisa N, di Salvo ML (2014) Coupling bioorthogonal chemistries with artificial metabolism: intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. Molecules 19(1):1004–1022
53. Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, Isaacs FJ (2015) Recoded organisms engineered to depend on synthetic amino acids. Nature 518:89–93

54. Dolgin E (2015) Safety boost for GM organisms. Nature 517:423
55. Nunes-Alves C (2015) GMOs in lockdown. Nat Rev Microbiol 13:3443
56. Schmidt M, de Lorenzo V (2016) Synthetic bugs on the loose: containment options for deeply engineered (micro)organisms. Curr Opin Biotechnol 38:90–96
57. Bohlke N, Budisa N (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. FEMS Microbiol Lett 351(2):133–144
58. Acevedo-Rocha CG, Schulze-Makuch D (2015) How many biochemistries are available to build a cell? Chembiochem 16(15):2137–2139
59. Marliere P (2009) The farther, the safer: a manifesto for securely navigating synthetic species away from the old living world. Syst Synth Biol 3(1-4):77–84
60. Hoesl MG, Budisa N (2012) Recent advances in genetic code engineering in Escherichia coli. Curr Opin Biotechnol 23(5):751–757
61. Acevedo-Rocha CG, Fang G, Schmidt M, Ussery DW, Danchin A (2013) From essential to persistent genes: a functional approach to constructing synthetic life. Trends Genet 29:273–279
62. Popa R (2010) Necessity, futility and the possibility of defining life are all embedded in its origin as a punctuated-gradualism. Orig Life Evol Biosph 40(2):183–190
63. Pezo V, Metzgar D, Hendrickson TL, Waas WF, Hazebrouck S, Doring V, Marliere P, Schimmel P, De Crecy-Lagard V (2004) Artificially ambiguous genetic code confers growth yield advantage. Proc Natl Acad Sci U S A 101(23):8593–8597
64. Xiao H, Nasertorabi F, Choi SH, Han GW, Reed SA, Stevens RC, Schultz PG (2015) Exploring the potential impact of an expanded genetic code on protein function. Proc Natl Acad Sci U S A 112(22):6961–6966
65. Konig H, Dorado-Morales P, Porcar M (2015) Responsibility and intellectual property in synthetic biology: a proposal for using Responsible Research and Innovation as a basic framework for intellectual property decisions in synthetic biology. EMBO Rep 16:1055–1059
66. Owen R, Stilgoe J, Macnaghten P, Gorman M, Fisher E, Guston D (2013) A framework for responsible innovation. In: Owen R, Bessant J, Heintz M (eds) Responsible innovation, vol 1. John Wiley & Sons, London, pp 27–50
67. EGE (2009) Ethically speaking. Available at http://ec.europa.eu/archives/bepa/european-group-ethics/docs/ethic_speak_n13_en.pdf
68. EGE (2009) Ethics of synthetic biology. Available at http://ec.europa.eu/bepa/european-group-ethics/docs/opinion25_en.pdf
69. De Vriend H (2006) Constructing life. Early social reflections on the emerging field of synthetic biology. Available at http://www.rathenau.nl/uploads/tx_tferathenau/WED97_Constructing_Life_2006.pdf
70. European Commission (2010) Synthetic biology from science to governance. Workshop organised by the European Commission's Directorate-General for Health & Consumers, Brussels
71. Schmidt M (2008) Diffusion of synthetic biology: a challenge to biosafety. Syst Synth Biol 2 (1–2):1–6
72. SCHER, SCENIHR, SCCS (2015) Opinion on synthetic biology III: research priorities. Available at http://ec.europa.eu/health/scientific_committees/consultations/public_consultations/scenihr_consultation_28_en.htm
73. Baertschi B (2013) Defeating the argument from hubris. Bioethics 27(8):435–441
74. Boldt J, Müller O, Maio G (2009) Synthetische Biologie Eine ethisch-philosophische Analyse. Bundesamt für Bauten und Logistik, Bern
75. Deplazes A (2009) Piecing together a puzzle. An exposition of synthetic biology. EMBO Rep 10(5):428–432
76. Douglas T, Savulescu J (2010) Synthetic biology and the ethics of knowledge. J Med Ethics 36 (11):687–693
77. Cockell CS (2011) Microbial rights? EMBO J 12:181

78. Frank D, Heil R, Coenen C, Konig H (2015) Synthetic biology's self-fulfilling prophecy - dangers of confinement from within and outside. Biotechnol J 10(2):231–235
79. Heavey P (2013) Synthetic biology ethics: a deontological assessment. Bioethics 27 (8):442–452
80. Minssen T, Rutz B, van Zimmeren E (2015) Synthetic biology and intellectual property rights: six recommendations. Biotechnol J 10(2):236–241
81. CBD (2012) Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. Available at https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf
82 Trojok RD (2014) Bio-commons whitepaper. Available at http://bioartsociety.fi/Bio-Commons_Whitepaper.pdf
83. Dabrock P (2009) Playing God? Synthetic biology as a theological and ethical challenge. Syst Synth Biol 3(1–4):47–54
84. van den Belt H (2009) Playing God in Frankenstein's footsteps: synthetic biology and the meaning of life. Nanoethics 3(3):257–268
85. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR Jr, Romesberg FE (2014) A semi-synthetic organism with an expanded genetic alphabet. Nature 509 (7500):385–388

Index