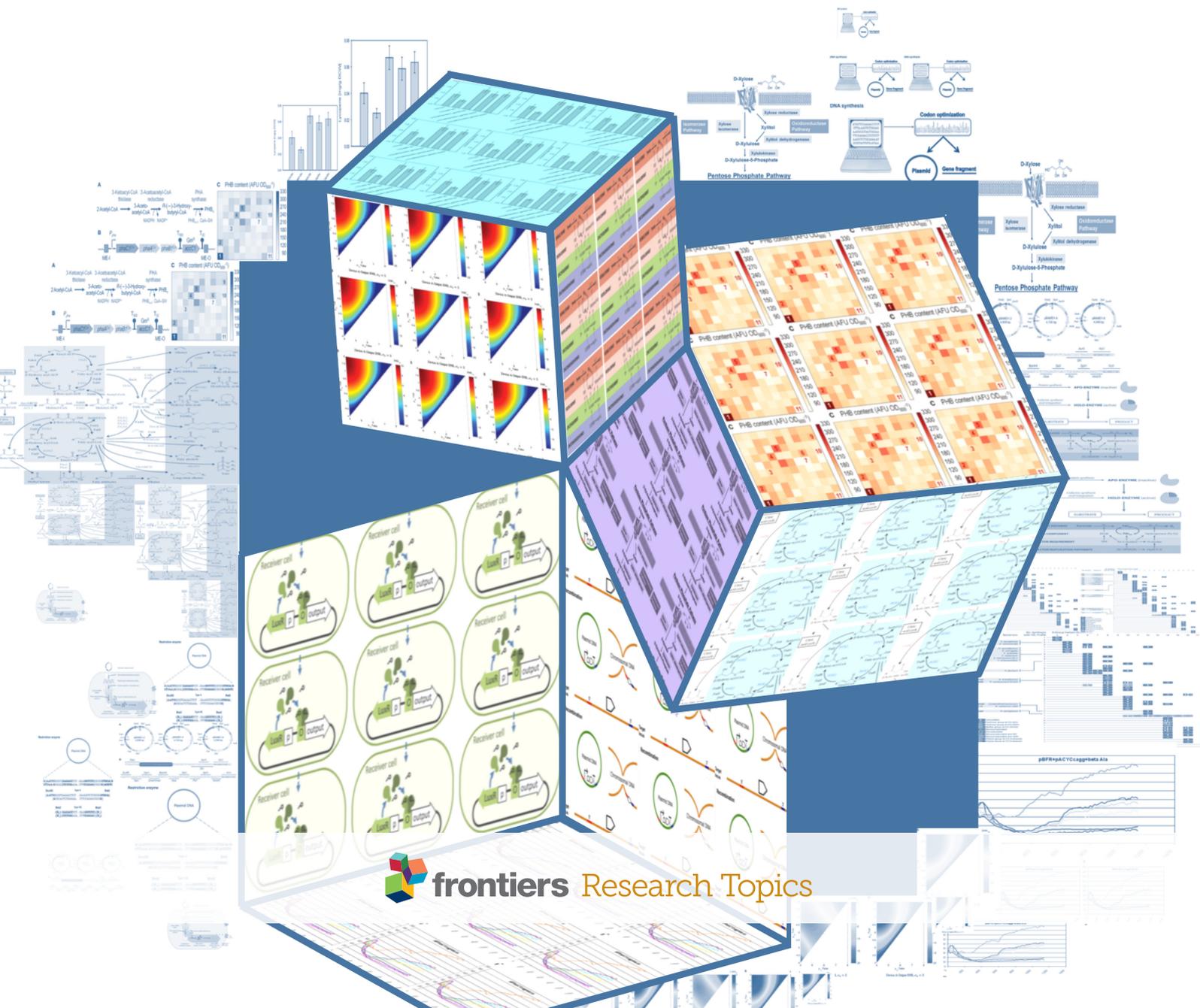
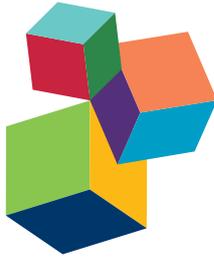


SYNTHETIC BIOLOGY: ENGINEERING COMPLEXITY AND REFACTORIZING CELL CAPABILITIES

EDITED BY: Francesca Ceroni, Karmella Ann Haynes, Pablo Carbonell
and Jean Marie François
PUBLISHED IN: Frontiers in Bioengineering and Biotechnology





frontiers

Frontiers Copyright Statement

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-685-2

DOI 10.3389/978-2-88919-685-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

SYNTHETIC BIOLOGY: ENGINEERING COMPLEXITY AND REFACTORIZING CELL CAPABILITIES

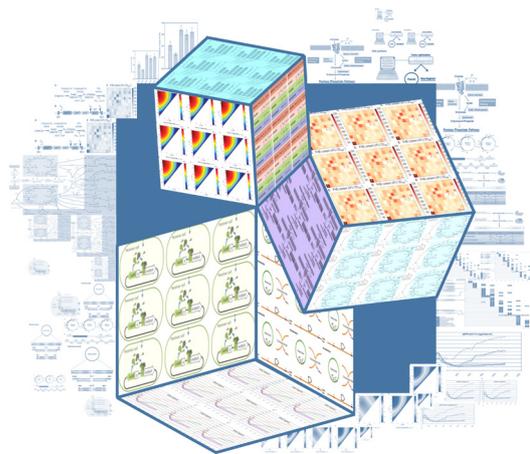
Topic Editors:

Francesca Ceroni, Imperial College London, UK

Karmella Ann Haynes, Arizona State University, USA

Pablo Carbonell, SYNBIOCHEM, University of Manchester, UK

Jean Marie François, Université de Toulouse, France



Tackling multifaceted complexity in synthetic biology.

This illustration by Karmella Haynes incorporates figures from the articles in this Special Topic into a graphical design that is inspired by the Frontiers logo.

One of the key features of biological systems is complexity, where the behavior of high level structures is more than the sum of the direct interactions between single components. Synthetic Biologists aim to use rational design to build new systems that do not already exist in nature and that exhibit useful biological functions with different levels of complexity. One such case is metabolic engineering, where, with the advent of genetic and protein engineering, by supplying cells with chemically synthesized non-natural amino acids and sugars as new building blocks, it is now becoming feasible to introduce novel physical and chemical functions and properties into biological entities.

The rules of how complex behaviors arise, however, are not yet well understood. For instance, instead of considering cells as inert chassis in which synthetic devices could be easily operated to impart new functions, the presence of these systems may impact cell physiology with reported effects on transcription, translation, metabolic fitness and optimal resource allocation. The result of these changes in the chassis may be failure of the synthetic device, unexpected or reduced device behavior, or perhaps a more permissive environment in which the synthetic device is allowed to function.

While new efforts have already been made to increase standardization and characterization of biological components in order to have well known parts as building blocks for the construction of more complex devices, also new strategies are emerging to better understand the biological dynamics underlying the phenomena we observe. For example, it has been shown that the features of single biological components [i.e. promoter strength, ribosome binding affinity, etc] change depending on the context where the sequences are allocated. Thus, new technical approaches have been adopted to preserve single components activity, as genomic insulation or the utilization of prediction algorithms able to take biological context into account.

There have been noteworthy advances for synthetic biology in clinical technologies, biofuel production, and pharmaceuticals production; also, metabolic engineering combined with microbial selection/adaptation and fermentation processes allowed to make remarkable progress towards bio-products formation such as bioethanol, succinate, malate and, more interestingly, heterologous products or even non-natural metabolites. However, despite the many progresses, it is still clear that ad hoc trial and error predominates over purely bottom-up, rational design approaches in the synthetic biology community. In this scenario, modelling approaches are often used as a descriptive tool rather than for the prediction of complex behaviors. The initial confidence on a pure reductionist approach to the biological world has left space to a new and deeper investigation of the complexity of biological processes to gain new insights and broaden the categories of synthetic biology.

In this Research Topic we host contributions that explore and address two areas of Synthetic Biology at the intersection between rational design and natural complexity: (1) the impact of synthetic devices on the host cell, or “chassis” and (2) the impact of context on the synthetic devices.

Particular attention will be given to the application of these principles to the rewiring of cell metabolism in a bottom-up fashion to produce non-natural metabolites or chemicals that should eventually serve as a substitute for petrol-derived chemicals, and, on a long-term view, to provide economical, ecological and ethical solutions to today’s energetic and societal challenges.

Citation: Ceroni, F., Haynes, K. A., Carbonell P. and François, J-M., eds. (2015). Synthetic biology: engineering complexity and refactoring cell capabilities. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-685-2

Table of Contents

- 05 Editorial – Synthetic biology: engineering complexity and refactoring cell capabilities**
Francesca Ceroni, Pablo Carbonell, Jean-Marie François and Karmella A. Haynes
- 07 Developments in the tools and methodologies of synthetic biology**
Richard Kelwick, James T. MacDonald, Alexander J. Webb and Paul Freemont
- 30 Production of fatty acid-derived valuable chemicals in synthetic microbes**
Ai-Qun Yu, Nina Kurniasih Pratomo Juwono, Susanna Su Jan Leong and Matthew Wook Chang
- 42 Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum***
Sabine A. E. Heider, Natalie Wolf, Arne Hofemeier, Petra Peters-Wendisch and Volker F. Wendisch
- 55 Engineering sugar utilization and microbial tolerance toward lignocellulose conversion**
Lizbeth M. Nieves, Larry A. Panyon and Xuan Wang
- 65 Cofactor engineering for enhancing the flux of metabolic pathways**
M. Kalim Akhtar and Patrik R. Jones
- 71 Can the natural diversity of quorum-sensing advance synthetic biology?**
René Michele Davis, Ryan Yue Muller and Karmella Ann Haynes
- 81 Signal-to-noise ratio measures efficacy of biological computing devices and circuits**
Jacob Beal
- 94 Obsolescence and intervention: on synthetic-biological entities**
Andrés Moya
- 97 A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli***
Tamás Fehér, Vincent Libis, Pablo Carbonell and Jean-Loup Faulon
- 111 New transposon tools tailored for metabolic engineering of Gram-negative microbial cell factories**
Esteban Martínez-García, Tomás Aparicio, Víctor de Lorenzo and Pablo I. Nikel

Editorial – Synthetic biology: engineering complexity and refactoring cell capabilities

Francesca Ceroni^{1,2*}, Pablo Carbonell³, Jean-Marie François^{4,5,6} and Karmella A. Haynes⁷

¹ Centre for Synthetic Biology and Innovation, Imperial College London, London, UK, ² Department of Bioengineering, Imperial College London, London, UK, ³ SYNBIOCHEM, Manchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, Manchester, UK, ⁴ LISBP, INSA, INP, UPS, Université de Toulouse, Toulouse, France, ⁵ MR792, Ingénierie des Systèmes Biologiques et des Bioprocédés, INRA, Toulouse, France, ⁶ UMR 5504, CNRS, Toulouse, France, ⁷ Ira A. Fulton School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA

Keywords: synthetic biology, complexity, metabolic engineering, emerging properties, crosstalk

Synthetic Biology is now in its second decade and many goals have been achieved toward the rational design of biological systems. This Research Topic features and reviews some of the latest progress in Synthetic Biology with a focus on research at the intersection between rational design and natural complexity with a potential outcome to concrete biotechnological applications. Kelwick et al. (2014) summarize the great expansion in the genetic toolkit and DNA assembly techniques that are currently available for synthetic biologists. These tools will advance the implementation of new functions and the production of useful metabolites in living cells in a controlled fashion. Using engineering formality, Synthetic Biology aims to identify biological design principles that can be used for practical applications. As one of the results, metabolic engineering is now becoming feasible to introduce novel functions and properties into an increasing number of microbial hosts. Examples come from Yu et al. (2014) and Heider et al. (2014) that describe the production of fatty-acid-derived chemicals and astaxanthin in microbes, respectively. Furthermore, bacteria can be engineered for the conversion of waste into renewable products, as Nieves et al. (2015) demonstrate with the bioconversion of lignocellulose.

Along with its great successes, Synthetic Biology is also encountering new challenges, represented by emerging behaviors in modified host cells (chassis) that are difficult to predict. Limitations in the robust prediction of gene networks arise from the lack of a proper understanding of the living systems used in synthetic biology. For instance, Akhtar and Jones (2014) appropriately present the evidence that the failure of a number of pathway engineering strategies are often due the lack of co-factors needed for the proper activity of the key enzymes. Co-factor production needs to be integrated in the system's design to achieve proper enzymatic activity. As synthetic network designs become more complex, emerging evidence shows that elements within these networks can exhibit crosstalk and lead to non-specific behavior. As presented in Davis et al. (2015), bacterial quorum sensing pathways, which are widely used in Synthetic Biology, exhibit crosstalk that can limit the number of nodes in a network, and therefore stifle efforts to build sophisticated systems. New efforts are needed to better understand the behavior of composable parts and to develop new orthogonal elements. Lastly, Beal (2015) addresses unresolved questions in the area of cell-based information processors and noise. He proposes a quantitative signal-to-noise ratio-based standard to assess circuit performance.

An important aspect in the engineering of living cells that only recently has been investigated in detail by the synthetic biology community is the interaction between the system and the chassis. The exploitation of the cell's resources for the operation of heterologous systems has proven to be deleterious, leading to non-robust gene expression and inefficient cellular

OPEN ACCESS

Edited by:

Pengcheng Fu,
Beijing University of Chemical
Technology, China

Reviewed by:

Qiang Wang,
Chinese Academy of Sciences, China

*Correspondence:

Francesca Ceroni
f.ceroni@imperial.ac.uk

Specialty section:

This article was submitted to
Synthetic Biology, a section of the
journal *Frontiers in Bioengineering and
Biotechnology*

Received: 06 July 2015

Accepted: 06 August 2015

Published: 21 August 2015

Citation:

Ceroni F, Carbonell P, François J-M
and Haynes KA (2015)

Editorial – Synthetic biology:
engineering complexity and
refactoring cell capabilities.

Front. Bioeng. Biotechnol. 3:120.

doi: 10.3389/fbioe.2015.00120

performance with decreased population growth. In that direction, Moya (2014) reflects on the need of controllable systems in synthetic entities preventing obsolescence, similarly as how living cells exhibit self-maintenance. In Fehér et al. (2015) the observation of the dynamic response of a malonyl-CoA biosensor in *Escherichia coli* was used to understand the toxicity of the overproduction of a synthetic compound, which interfered with the system's behavior. Martínez-García et al. (2014) present the development of new broad host range Tn5 vectors in order to relieve the burden of PHB production on the health of gram

negative bacteria (*E. coli*). These studies address the need to investigate and develop controlled production of the molecule of interest to avoid burden-related negative feedback from the chassis.

As illustrated by the works in this Research Topic, now is a critical moment for Synthetic Biology, where the initial enthusiasm for the major achievements attained gives way to a deeper and better understanding of the complexity of biological systems. Advancing in this direction will significantly improve the applicability of design principles for living organisms.

References

- Akhtar, M. K., and Jones, P. R. (2014). Cofactor engineering for enhancing the flux of metabolic pathways. *Front. Bioeng. Biotechnol.* 2:30. doi:10.3389/fbioe.2014.00030
- Beal, J. (2015). Signal-to-noise ratio measures efficacy of biological computing devices and circuits. *Front. Bioeng. Biotechnol.* 3:93. doi:10.3389/fbioe.2015.00093
- Davis, R. M., Muller, R. Y., and Haynes, K. A. (2015). Can the natural diversity of quorum-sensing advance synthetic biology? *Front. Bioeng. Biotechnol.* 3:30. doi:10.3389/fbioe.2015.00030
- Fehér, T., Libis, V., Carbonell, P., and Faulon, J.-L. (2015). A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*. *Front. Bioeng. Biotechnol.* 3:46. doi:10.3389/fbioe.2015.00046
- Heider, S. A. E., Wolf, N., Hofemeier, A., Peters-Wendisch, P., and Wendisch, V. F. (2014). Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum*. *Front. Bioeng. Biotechnol.* 2:28. doi:10.3389/fbioe.2014.00028
- Kelwick, R., MacDonald, J. T., Webb, A. J., and Freemont, P. (2014). Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* 2:60. doi:10.3389/fbioe.2014.00060
- Martínez-García, E., Aparicio, T., de Lorenzo, V., and Nikel, P. I. (2014). New transposon tools tailored for metabolic engineering of Gram-negative microbial cell factories. *Front. Bioeng. Biotechnol.* 2:46. doi:10.3389/fbioe.2014.00046
- Moya, A. (2014). Obsolescence and intervention: on synthetic-biological entities. *Front. Bioeng. Biotechnol.* 2:59. doi:10.3389/fbioe.2014.00059
- Nieves, L. M., Panyon, L. A., and Wang, X. (2015). Engineering sugar utilization and microbial tolerance toward lignocellulose conversion. *Front. Bioeng. Biotechnol.* 3:17. doi:10.3389/fbioe.2015.00017
- Yu, A.-Q., Pratomo Juwono, N. K., Leong, S. S. J., and Chang, M. W. (2014). Production of fatty acid-derived valuable chemicals in synthetic microbes. *Front. Bioeng. Biotechnol.* 2:78. doi:10.3389/fbioe.2014.00078

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Ceroni, Carbonell, François and Haynes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Developments in the tools and methodologies of synthetic biology

Richard Kelwick^{1,2*}, James T. MacDonald^{1,2}, Alexander J. Webb^{1,2} and Paul Freemont^{1,2*}

¹ Centre for Synthetic Biology and Innovation, Imperial College London, London, UK

² Department of Medicine, Imperial College London, London, UK

Edited by:

Karmella Ann Haynes, Arizona State University, USA

Reviewed by:

M. Kalim Akhtar, University College London, UK

Dong-Yup Lee, National University of Singapore, Singapore

*Correspondence:

Richard Kelwick and Paul Freemont, Department of Medicine, Centre for Synthetic Biology and Innovation, Sir Ernst Chain Building, South Kensington Campus, Exhibition Road, London SW7 2AZ, UK
e-mail: r.kelwick@imperial.ac.uk;
p.freemont@imperial.ac.uk

Synthetic biology is principally concerned with the rational design and engineering of biologically based parts, devices, or systems. However, biological systems are generally complex and unpredictable, and are therefore, intrinsically difficult to engineer. In order to address these fundamental challenges, synthetic biology is aiming to unify a “body of knowledge” from several foundational scientific fields, within the context of a set of engineering principles. This shift in perspective is enabling synthetic biologists to address complexity, such that robust biological systems can be designed, assembled, and tested as part of a biological design cycle. The design cycle takes a forward-design approach in which a biological system is specified, modeled, analyzed, assembled, and its functionality tested. At each stage of the design cycle, an expanding repertoire of tools is being developed. In this review, we highlight several of these tools in terms of their applications and benefits to the synthetic biology community.

Keywords: synthetic biology, engineering biology, design cycle, tools, standardization

INTRODUCTION

The synthetic biology toolkit has expanded greatly in recent years, which can be attributed to the efforts of a highly dynamic community of researchers, ambitious undergraduate students in the International Genetically Engineered Machine competition (iGEM), and the growing number of amateur scientists from the DIY BIO movement. Each of these groups has bold ambitions for the rapidly growing field of synthetic biology, which aims to rationally engineer biological systems for useful purposes (Purnick and Weiss, 2009; Anderson et al., 2012; Landrain et al., 2013; Jefferson et al., 2014). The merging of several foundational sciences, including molecular, cellular, and microbiology with a set of engineering principles, is a profound shift and is the key distinction between synthetic biology and genetic engineering (Andrianantoandro et al., 2006; Heinemann and Panke, 2006; Khalil and Collins, 2010; Kitney and Freemont, 2012). Indeed, many social scientists, who are themselves a part of the synthetic biology community, have extensively explored the ontological implications of this perspective (Scharck, 2012; Preston, 2013). Although many of the social aspects of synthetic biology are beyond the scope of this review, they will continue to shape the synthetic biology toolkit. In particular, society is an important stakeholder that has some influence over chassis (host cell) choice, the design of biosafety measures, biosecurity considerations, and long-term research applications (Marris and Rose, 2010; Anderson et al., 2012; Agapakis, 2013; Moe-Behrens et al., 2013; Wright et al., 2013; Douglas and Stemerding, 2014).

From a biological perspective, there have been important developments in the field across several areas, some of which have been reviewed elsewhere (Arpino et al., 2013; Lienert et al., 2014; Way et al., 2014). For instance, the number, quality, and availability

of biological parts (bioparts, e.g., promoters and ribosomal binding sites) have continued to increase. This is exemplified by the iGEM student registry of standard biological parts, which has increased its biopart collection to include over 12,000 parts, across 20 different categories (partsregistry.org). However, due to its open nature, the iGEM registry contains parts of variable quality that are mostly uncharacterized. There are also professional parts registries, such as those at BIOFAB, which include expansive libraries of characterized DNA-based regulatory elements (Mutalik et al., 2013a,b). Although libraries of bioparts are indeed useful, putting them together into predictable devices, pathways and systems are incredibly challenging as many biological design rules are not yet fully understood (Endy, 2005; Kitney and Freemont, 2012). Developing synthetic passive and active insulator sequences may help increase predictability and thus reduce context dependency (Davis et al., 2011; Lou et al., 2012; Qi et al., 2012; Mutalik et al., 2013a). Notwithstanding these challenges, the field is progressing across several areas. One such area is biopart characterization, which is critical to the field, primarily because it is fundamentally a realization of several of the core engineering principles adopted in synthetic biology, namely standardization, modularization, and abstraction. Discrete biological parts of known sequence and behavior can be abstracted based upon a descriptive function and thus, their true complexity can be masked behind a biological concept. For example, discrete DNA sequences (bioparts) that fit a standardized descriptive function, such as a promoter, can be functionally characterized and as a consequence bioparts become reusable (modularization) for use in other synthetic systems. Additionally, methods that provide standardized ways of assembling DNA parts such as the BioBrick standard can help establish platforms for the sharing and reuse of bioparts. At a higher

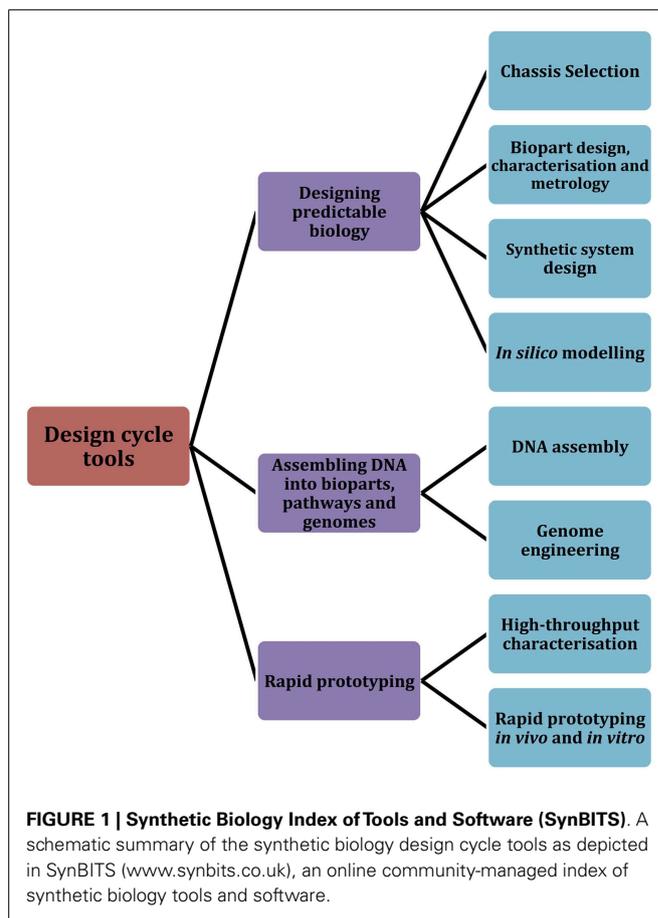
level, abstraction and standardization are important because they permit the separation of design from assembly (Endy, 2005).

A desirable consequence of this perspective is that these engineering principles enable the separation of labor, expertise, and complexity at each level of the design hierarchy (Endy, 2005). In practical terms, this separation of biological design from DNA assembly enables innovation within these hierarchies to occur at different rates. For instance, it is generally true that with more recent DNA assembly methods it is currently easier to assemble multi-part genetic circuits consisting of several bioparts, or even entire genomes, than it is to reliably predict how these bioparts will interact in the final system (Purnick and Weiss, 2009; Ellis et al., 2011; Arpino et al., 2013; Ellefson et al., 2014). However, it is envisioned that this will change, with the increasing adoption of high-throughput characterization platforms that can test entire biopart libraries in parallel. These platforms typically use automated liquid-handling robots, coupled with plate readers although microfluidics approaches are also gaining traction (Lin and Levchenko, 2012; Boehm et al., 2013; Benedetto et al., 2014). In either case, when coupled with automated data analysis, modeling, and sophisticated forward-design strategies (Marchisio and Stelling, 2009; Wang et al., 2009; Esvelt et al., 2011; Ellefson et al., 2014; Marchisio, 2014; Stanton et al., 2014), these high-throughput platforms provide the basis for the rapid prototyping workflows required to realize a synthetic biology design cycle (Kitney and Freemont, 2012).

In this review, we focus on several significant tools, both classical and emerging, that the field of synthetic biology employs as part of a typical design cycle workflow. Building upon a design cycle template, the review is organized to explore prominent tools and research methodologies across three core areas: designing predictable biology (design), assembling DNA into bioparts, pathways, and genomes (build), and rapid prototyping (test) (Figure 1). We first describe several of the core challenges that are associated with designing predictable biology, including the complexities associated with chassis selection, biopart design, engineering, and characterization. In parallel, we highlight relevant tools and methodologies that are particularly aligned with the engineering principles of synthetic biology. We then discuss established and newly developed DNA assembly methodologies, and group them according to four broad assembly strategies: restriction enzyme-based, overlap-directed, recombination-based, and DNA synthesis. Finally, we highlight several emerging rapid prototyping technologies that are set to significantly improve the field's capacity for testing synthetic parts, devices, and systems. We conclude with a summary of several of the core challenges that were described in each of the design, build, test sections of the review and discuss whether the synthetic biology toolbox is equipped to address them. In addition to this, we have also created an online community, the Synthetic Biology Index of Tools and Software (SynBITS) – synBITS.co.uk, which has also been structured according to the design cycle (Figure 1).

DESIGNING PREDICTABLE BIOLOGY

From an engineering perspective, living systems can be perceived as overly complex, inefficient, and unpredictable (Csete and Doyle, 2002). It is this perception that has driven the concept of the



biopart, in which a particular DNA sequence is defined by the function that it encodes (Endy, 2005). Thus, complex biological functions can be conceptually separated (abstracted) from the complexities of the sequence context from which they originated (Endy, 2005). As a consequence of this approach, biological pathways and circuits can potentially be redesigned into less complex and potentially more predictable designs. The defining examples of this perspective are the toggle switch (Gardner et al., 2000), a genetic circuit defined by two repressible promoters that were engineered to form a mutually inhibitory network, and the repressilator (Elowitz and Leibler, 2000), a type of oscillator (biological clock). What sets these examples apart from general genetic engineering is that modeling was used to predict and optimize the behavior of these genetic circuit designs prior to their construction.

While these forward-design approaches were hugely successful, the repressilator displayed noisy behavior as a result of stochastic fluctuations in components of the genetic circuit (Elowitz and Leibler, 2000). In other words, *in silico* modeling did not fully capture the true *in vivo* complexity of the synthetic circuit. Likewise, the toggle switch experienced natural fluctuations in gene expression that were sufficient to create variations in the level of inducer needed to switch the cells from one state to another. These variations were also not fully anticipated during *in silico* modeling (Gardner et al., 2000). While these genetic circuits have

been improved, with novel oscillator (Stricker et al., 2008; Olson et al., 2014) and toggle switch designs, including those designed for mammalian cells (Muller et al., 2014b) and plants (Muller et al., 2014a), it is clear that the modeling of biological systems still requires a concerted and long-term effort. Critical to this effort is the availability of new synthetically designed bioparts and experimental data that accurately captures the behavior of the components or bioparts that constitute a synthetic system (Arkin, 2013) as well as the characteristics or influence that the chassis/host cell enacts upon them.

CHASSIS SELECTION

As an engineering concept, the chassis refers to a physical internal framework or structure that supports the addition of other components that combine to form a finalized engineered structure. From a synthetic biology perspective, the concept invokes an understanding that a biological chassis is a tool to provide the structures that accommodate (host) the execution of a synthetic system, including the provision of a metabolic environment, energy sources, transcription, and translation machinery, as well as other minimal cellular functions (Acevedo-Rocha et al., 2013; Danchin and Sekowska, 2014). Chassis selection is therefore a critical design decision that synthetic biologists are required to take, particularly since the chassis will directly influence the behavior and function of a synthetic system. Essentially, the chassis determines which bioparts can be used since they must be compatible with the biological machinery that is present. This can result in a difficult choice for the synthetic biologist: either to use an established chassis and design the circuit to be orthogonal with that host, or design a synthetic system that fits a requirement and then choose a host chassis that is compatible with the resultant bioparts or system. These constraints can to some degree be designed around, either by engineering the chassis to knockout genes that optimize its orthogonality and reduce burden, through codon optimization (Chung and Lee, 2012) or through the use of insulator sequences that negate context dependency effects (Guye et al., 2013; Torella et al., 2014a). Ultimately, however, chassis selection will dictate the downstream design considerations for any given synthetic system, and therefore, chassis selection must be coordinated with biopart design efforts.

In order to rationalize which chassis selection strategy is most appropriate for an intended application, it is important to consider the consequences and advantages of each strategy. Where a chassis is selected as a priority above that of the design considerations of the synthetic system, it is important to consider whether the chassis has been extensively characterized in the literature and/or if the chassis has known intrinsic capabilities that complement the intended application (Table 1). Additionally, access to detailed biological knowledge of a chassis will aid modeling-guided design efforts and the implementation of chassis optimization strategies for dealing with burden or metabolic flux effects. Likewise, the wealth of knowledge acquired about model organisms across several biological disciplines may encourage synthetic biologists to consider them as a potential chassis in preference to established favorites (Table 1). Indeed, there are already several emerging chassis that are gaining traction and are set to be utilized more frequently in the field (Table 1).

Alternatively, a synthetic system could be specified and designed as a priority above that of chassis selection. As a consequence, there will be chassis, which are not compatible with the synthetic system and others that may require extensive engineering to accommodate its design. However, this approach is complementary to those chassis that are bespoke engineered. “Synthia,” the first organism to feature a fully synthetically manufactured genome, is indicative that the field of synthetic biology is shifting toward the development of rationally engineered chassis (Gibson et al., 2010a). Though it is important to recognize that the “Synthia” genome, while synthetic in origin, was not designed to significantly alter the characteristics of the chassis, and therefore, does not represent the first truly bespoke-engineered chassis. Yet, its successors, the synthetic yeast project (Annaluru et al., 2014), protocell developments (Xu et al., 2010), and even to some extent cell-free expression systems (Shin and Noireaux, 2012; Sun et al., 2013a) may all usher in an era in which the design of bespoke-engineered chassis is routine. Wholly rationally engineered chassis could conceivably be built around the specifications of a synthetic system, such that the chassis is both compatible with the synthetic system and the majority of its cellular resources are directed toward the execution of the synthetic system. In this sense, the function of the synthetic system would be free of chassis constraints; however, the full realization of this approach is still several decades away. Until then, chassis selection will remain a trade-off between which should be prioritized for each application, the chassis or the synthetic system? There are of course many other considerations to address, some of which we cover in the biopart design section of this review and others that have been previously discussed in the literature (Heinemann and Panke, 2006; Arpino et al., 2013; Danchin and Sekowska, 2014).

BIOPART DESIGN AND ENGINEERING

The field of synthetic biology continues to benefit from decades of biological research that has built a knowledge base of biological systems that can be deconstructed and re-engineered as bioparts and synthetic systems. Here, we highlight prominent bioparts that are particularly aligned with the engineering principles of synthetic biology. In most cases, existing natural biological parts can be reused in synthetic devices or systems. However, there are situations where new bioparts need to be designed and synthesized by modifying existing bioparts or by creating entirely new parts *de novo*. These novel parts could be enzymes that catalyze unnatural reactions (Jiang et al., 2008; Rothlisberger et al., 2008), molecular biosensors (Penchovsky and Breaker, 2005), protein scaffold (Koga et al., 2012; Heider et al., 2014), DNA or RNA scaffolds (Rothe-mund, 2006; Delebecque et al., 2011), ribosome-binding sites with specifically designed transcription rates (Salis et al., 2009), promoters with novel regulatory features and/or specific translation rates (Marples et al., 2000; Kelly et al., 2009).

Transcriptional circuits use RNA polymerase operations per second (PoPS) as the common signal carrier but, until recently only a small set of DNA-binding proteins and associated operator sequences were used to regulate the flux of RNA polymerase (RNAP) and construct synthetic circuits. The lack of a large set of orthogonal regulatory proteins has limited the complexity of synthetic systems (Purnick and Weiss, 2009), but a new wave of

Table 1 | Synthetic biology chassis.

Chassis	Advantages	Disadvantages
ESTABLISHED CHASSIS		
<i>Bacillus subtilis</i>	Model Gram-positive organism. Generally regarded as safe (GRAS) organism. Genetically tractable and genome sequences are available. Secretion of proteins. Extensive range of molecular biology tools are available, e.g., plasmids (Harwood et al., 2013; Radeck et al., 2013). Rapid growth, inexpensive to grow and maintain, can be induced to form heat and desiccation resistant spores (Harwood et al., 2013). Spores can be transported easily and cheaply. Suicide mechanisms are available (Wright et al., 2013).	Non-integrative plasmids are not always stably maintained between cell generations. Protease-deficient strains are required to minimize proteolytic degradation of expressed proteins.
Cell-free protein synthesis (CFPS)/transcription-translation coupled reactions (TX-TL)	Protein/metabolite production is decoupled from the need of the cell to survive and reproduce – ideal if product is toxic or inhibitory to living chassis. Amenable to high-throughput workflows (Sun et al., 2013a,b).	The biological system does not self-reproduce Reactions typically only last 4–6 h due to depletion of the reaction energy mix and/or the accumulation of inorganic phosphates. Reaction components can also be expensive Variability between cell extract batches
<i>Escherichia coli</i>	Genetically tractable and genome sequences are available. An extensive range of molecular biology tools are available, e.g., plasmids, phages, etc. Rapid growth, inexpensive to grow and maintain, extensive range of published data relating to this chassis, suicide mechanisms available (Wright et al., 2014). Whole-cell metabolic models have been developed and are being improved (Atlas et al., 2008).	Few post-translational modifications compared to eukaryotes, e.g., reduced protein glycosylation
<i>Saccharomyces cerevisiae</i>	Glycosylation of expressed proteins. Genetically tractable and genome sequenced. Molecular biology tools are available, e.g., plasmids	The core oligosaccharides that comprise the protein glycosylation events in <i>S. cerevisiae</i> are thought to be responsible for the hyper-antigenic nature of proteins expressed in this chassis rendering them potentially unsuitable for therapeutic uses (Hamilton and Gemgross, 2007; Cregg et al., 2009; Walsh, 2010).
EMERGING CHASSIS		
<i>Chlamydomonas reinhardtii</i>	An established model organism; eukaryotic photosynthetic organism	Slow cultivation time. Several strains have a cell wall, and are therefore difficult to transform. Low transformation frequency due to genome integration of plasmids
<i>Geobacillus sp.</i>	Several strains currently being developed as synthetic biology chassis including <i>G. thermoglucosidasius</i> (Bartosiak-Jentys et al., 2013). Enables the application of metabolic and enzymatic processes at higher temperatures (e.g., 55–65°C optimum for <i>G. thermoglucosidasius</i>) than is possible with alternative chassis	Few biological parts have been characterized. The majority of antimicrobial drugs are unstable at the high temperatures that these chassis can grow at, thus limited cloning strategies are available.
Induced pluripotent stem cells (iPSCs)	An ethical source of stem cells for therapeutic and other responsible innovation applications (Cachat and Davies, 2011; Ye et al., 2013). Potential platform for engineering complex synthetic systems across multi-tissue structures	Cellular differentiation programs are not yet fully understood and therefore rational engineering is difficult
<i>Marchantia polymorpha</i>	Compared with other plant model organisms, this chassis has a relatively simple, “streamlined” genomic architecture. Genome projects are underway and several molecular biology tools are in development. Can be cultured easily and grows rapidly (Sharma et al., 2014)	Molecular biology tools are still in development (Chiyoda et al., 2014).
<i>Physcomitrella patens</i>	An established model organism for research on plant evolution, development, and physiology (Schaefer and Zryd, 1997; Nishiyama, 2000). Genome sequence available. Does not have a codon usage bias	Slow growth; month timescale. Low transformation efficiency

(Continued)

Table 1 | Continued

Chassis	Advantages	Disadvantages
<i>Pichia pastoris</i>	Higher heterologous protein expression and reduced glycosylation compared to <i>S. cerevisiae</i> . Successful expression of more than 200 heterologous proteins has been published. Proteins expressed in this chassis are thought to be less antigenic in nature than those produced in <i>S. cerevisiae</i> making this organism more suitable for therapeutic protein generation. <i>P. pastoris</i> is a methylotroph, and can therefore, grow with methanol as a sole carbon source. Can grow to high cell densities with high growth rates on inexpensive media (Cereghino, 2000; Cregg et al., 2000; Vogl et al., 2013).	High level of clonal variation (Aw and Polizzi, 2013). Few plasmid vectors are available
<i>Protocells</i>	Enables a complete bottom-up approach in which the cellular machinery, metabolism, genome, etc., can all be bespoke engineered (Chen et al., 2004; Xu et al., 2010)	Still under development
<i>Synechocystis sp.</i>	<i>Cyanobacteria sp.</i> are model organisms for the study of photosynthesis, as well as carbon and nitrogen fixation (Heidorn et al., 2011). Synthetic biology approaches may enable coupling of photosynthesis with the generation of biofuels and natural products (Depaoli et al., 2014). Tools are available for use in this chassis (Heidorn et al., 2011; Berla et al., 2013).	Specialized growth conditions are required (Berla et al., 2013)
Synthetic yeast 2.0	First designer eukaryotic genome but based on an established chassis – <i>S. cerevisiae</i> . One of the first bespoke-engineered chassis, with a defined genetic context, that has been rationally engineered for the benefit of the synthetic biology community (Annaluru et al., 2014; Lin et al., 2014). As part of the project, LoxPsym recombination sites are being added to the 3' end of all non-essential genes to allow inducible genome shuffling using a system called SCRaMbLE (Dymond and Boeke, 2012)	The alteration of the natural genome structure may negatively affect genome stability Complex biosafety, biosecurity, and ethical challenges may arise as a consequence of alterations in the natural functions of <i>S. cerevisiae</i> Project not yet complete
POTENTIAL CHASSIS		
<i>Caenorhabditis elegans</i>	Genetically tractable (Redemann et al., 2011) and the genome have been sequenced. The number and position of every cell during development are known and therefore this organism has great potential for the engineering of whole-organism, developmentally organized synthetic systems. Used in synthetic screens (O'Reilly et al., 2014)	Genetic lines have to be maintained
<i>Danio rerio</i>	Regenerative abilities. The organism is largely transparent and therefore expression of fluorescent reporter systems can be used to characterize <i>in vivo</i> synthetic systems. Established systems biology model (Mushtaq et al., 2013)	There may be alternative chassis that are more appropriate for some applications due to the ethical and legal considerations associated with the use of vertebrates in research
<i>Drosophila melanogaster</i>	Genetically tractable, genome sequenced, and proven relevance to human disease models. <i>Drosophila</i> -derived cell lines can be engineered for constitutive and inducible expression of proteins (Yang and Reth, 2012)	Genetic lines have to be maintained
<i>Xenopus tropicalis</i>	Genome sequenced. Used in synthetic screens (White et al., 2011; Tomlinson et al., 2012).	There may be alternative chassis that are more appropriate for some applications due to the ethical and legal considerations associated with the use of vertebrates in research

engineered proteins has greatly increased the number of tools available to synthetic biology circuit designers. The clustered, regularly interspaced, short palindromic repeats (CRISPR)/Cas system consists of CRISPR and CRISPR associated genes (cas) coding for related proteins, which together constitute an adaptive prokaryotic immune system (Barrangou et al., 2007). The CRISPR

loci consist of repeats interspaced with spacer sequences, which are transcribed and processed into crRNAs containing individual spacer sequences that are complementary to foreign DNA. The crRNAs bind Cas9 nuclease and the resulting complex recognizes and cleaves sequences complementary to the spacer sequences. This natural system has been repurposed as a transcriptional

regulator by modifying the Cas proteins to deactivate nuclease activity and creating artificial guide RNA (gRNA) sequences to create the CRISPR interference (CRISPRi) system. Deactivated Cas9:gRNA complexes act as repressors by binding specific sites and inhibiting RNAP activity (Qi et al., 2013). Alternatively, Cas9 can be fused to domains that recruit RNAP in order to act as transcriptional activators (Bikard et al., 2013; Mali et al., 2013).

Transcription activator-like effectors (TALEs) are proteins secreted by *Xanthomonas* bacteria in order to activate expression of plant genes during the course of infection. They consist of tandem repeats of a small domain with two variable amino acid sites. The amino acid identities of the variable sites have a simple mapping to the DNA base recognized, enabling chains of domains to be strung together in order to bind specific sequences (Boch et al., 2009; Moscou and Bogdanove, 2009). The simple modular nature of TALEs has enabled the engineering of synthetic proteins such as TAL effector nucleases (TALENs) (Mahfouz et al., 2011) and artificial orthogonal activators and repressors (Morbitzer et al., 2010; Blount et al., 2012).

Translation initiation regulators are relatively easy to *de novo* design as they rely on the reasonably well-characterized thermodynamics of RNA structure (Liang et al., 2011). However, unlike transcriptional circuits, there is no common signal carrier and thus they cannot be as easily composed into complex regulatory designs. By repurposing a regulatory element from the *tnaCAB* operon of *E. coli*, Liu et al. (2012), have created an adapter to convert translational regulators into transcriptional regulators (Liu et al., 2012). The 5'-end of the operon codes for a short leader peptide, TnaC that stalls the ribosome in the presence of free tryptophan. The stalled ribosome then blocks a Rho factor-binding site located adjacent to the stop codon of *tnaC*, allowing the transcription of the downstream genes *tnaA* and *tnaB*. The ribosome-binding site of *tnaC* in the native operon is constitutive but replacing this with translational regulator sequences, such as the RNA-IN/OUT system (Ross et al., 2013), enables the control of transcription of downstream genes.

In recent years, there has been rapid progress in developing software algorithms to enable the design of synthetic proteins that can be controlled at the atomic level of resolution (Leaver-Fay et al., 2011). Computational protein design is generally split into two components. Initially, a backbone scaffold is either artificially generated or taken from an existing known structure. Secondly, the amino acid sequence is optimized such that it minimizes the free energy of folding. It appears that minimizing a potential energy function by trialing different amino acid identities and rotamers is sufficient to achieve this. There have been a number of dramatic successes using this approach including the *de novo* design of enzymes. The design of a novel enzyme requires knowledge about the transition state structure of the reaction to be catalyzed and a predicted spatial arrangement of chemical groups that are likely to stabilize the transition state. The transition state structure and the stabilizing constellation of chemical groups around it can be designed theoretically (theozyme) (Tantillo et al., 1998), and using this knowledge, known protein structures can be searched for sites capable of accommodating side chain functional groups in the desired geometry (Zanghellini et al., 2006). These methods

have resulted in a number of successful enzyme designs (Jiang et al., 2008; Rothlisberger et al., 2008).

In the cell, many biochemical processes are spatially organized in order to locally concentrate substrates or isolate toxic substances (e.g., the carboxysome or peroxisome) and reduce cross talk between components. Efforts to engineer high-level organization in synthetic biological systems is a major challenge with applications in encapsulating artificial organelles or protocells (Choi and Montemagno, 2005; Agapakis et al., 2012; Hammer and Kamat, 2012; Mali et al., 2013), the precise detection and delivery of payloads (Sukhorukov et al., 2005; Uchida et al., 2007). Methods based on the computational protein design methods described above have been applied to create new self-assembling biomaterials at the atomic level from protein subunits that do not naturally form into higher-order structures (King et al., 2012, 2014). Other work has focused on using hydrophobic patterning of peptides to produce higher-order structures based on coiled-coils (Rajagopal and Schneider, 2004; Woolfson and Mahmoud, 2010; Zaccai et al., 2011; Fletcher et al., 2013). However, these are not designed to atomic level accuracy, tend to be chemically synthesized and so have not yet been reported to assemble *in vivo*. An alternative approach is to reuse naturally occurring protein-protein interfaces and assemblies (Padilla et al., 2001; Howorka, 2011; Sinclair et al., 2011), although ultimately it may be more desirable to design completely artificial protein scaffolds that are more likely to be biologically neutral and avoid the Mullerian complexity of naturally evolved biological systems (Dutton and Moser, 2011).

Novel protein biomaterials have applications in metabolic engineering by co-locating enzymes in the same pathway on a structural scaffold. This has the advantage of increasing the local concentration of substrates improving reaction kinetics, helping to prevent the loss of intermediates to competing pathways and the accumulation of toxic intermediates (Dueber et al., 2009). Protein cages can be used to completely encapsulate metabolic pathways and create synthetic bacterial micro-compartments. In a recent example, genes from the propanediol utilization operon (*pdu*) encoding for an empty protein shell in *Salmonella enterica* were expressed in *E. coli*. Short peptide sequences known to bind to *pdu* shell proteins were used to target pyruvate decarboxylase and alcohol dehydrogenase to the micro compartment resulting in increased ethanol production (Lawrence et al., 2014). This approach promises to be particularly useful for biosynthetic pathways involving toxic metabolites.

Similarly, structural scaffolds can be constructed using nucleic acids. Base pairing in nucleic acids makes predicting and designing structures somewhat more tractable than for proteins. For example, 2D and 3D structures have been engineered *in vitro* using long single stranded DNA (ssDNA) and small ssDNA oligonucleotides called "staples," that direct the folding of the long ssDNA into a pre-designed structure (Rothemund, 2006; Douglas et al., 2009; Han et al., 2011). It has also been shown to be possible to express simpler nanostructures *in vivo* using RNA transcribed by the cell (Delebecque et al., 2011). These structures were used together with specific protein-binding aptamers to efficiently channel substrates from one enzyme to another and substantially increase hydrogen production. At short distances, substrate channeling has been

found to be more effective than expected by simple 3D Brownian diffusion models (Fu et al., 2012, 2014).

A number of tools for predicting and designing relative translation rates of ribosome-binding sites have been developed (Reeve et al., 2014) including the RBS Calculator (Salis et al., 2009; Salis, 2011), the RBS Designer (Na and Lee, 2010), and the UTR Designer (Seo et al., 2013), which can aid operon design (Arpino et al., 2013). These software tools are based on thermodynamic models of the pre-initiation complex of the 30S ribosomal subunit and the messenger RNA (mRNA) and include terms based on the free energy required to unfold the unbound mRNA, the free energy of hybridization of the mRNA and the 16S rRNA, and various other terms. If the pool of free 30S ribosomal subunits is assumed to be roughly constant then the translation initiation rate can be assumed to be proportional to $\exp(-\beta\Delta G)$. Mechanistic predictive models for promoters are somewhat more complicated as promoter strength is related to the binding of the sigma factor and RNAP, and also the efficiency of promoter escape. However, there has been some success in predicting the strength of promoters for the *E. coli* sigma factor σ^E using relatively simple position weight matrix models (Rhodius and Mutalik, 2010; Rhodius et al., 2012).

Most of the synthetic regulatory tools described above are used in the construction of transcriptional circuits. Nevertheless, post-transcriptional circuit design, particularly using RNA molecules, has attracted a great deal of interest in recent years (Liang et al., 2011; Wittmann and Suess, 2012). Unlike proteins, RNA molecules are somewhat easier to design due to their well-understood thermodynamics and the dominance of secondary structure formation on folding. One important application area is the use of RNA as switches (riboswitches) that respond to their environment. Riboswitches are RNA molecules that can regulate protein production in response to changes in the concentration of a small molecule and occur naturally as well as being synthetically designed. These molecules are composed of an RNA aptamer that binds a specific small-molecule ligand. On binding the small molecule, the RNA aptamer may then change conformation resulting in either the occlusion of the Shine–Dalgarno sequence or its increased accessibility. Expression of the downstream genes is then turned either on or off in response (Suess et al., 2004). Alternatively, aptamers may be coupled to a ribozyme that allosterically cleaves itself in response to ligand binding (Tang and Breaker, 1997; Penchovsky and Breaker, 2005). The *de novo* design of small-molecule binding RNA aptamers is a non-trivial task but novel aptamers can be evolved *in vitro* using methods such as SELEX (systematic evolution of ligands by exponential enrichment) (Ellington and Szostak, 1990; Tuerk and Gold, 1990; Jenison et al., 1994). Riboswitches may have uses in metabolic engineering such as down-regulating upstream genes if a metabolite reaches toxic levels (Zhang and Keasling, 2011). RNA aptamers have also found use as mimics of GFP by binding small-molecule fluorophores (Paige et al., 2011). As discussed in the metrology section, these aptamers can be used to monitor mRNA levels.

BIOPART CHARACTERIZATION

Biopart characterization describes the functional and experimental metadata that is required to sufficiently capture the biological behavior of a biopart and the context in which it is being tested.

The type and range of these characterization data have evolved over time as highlighted by refinements in biopart characterization data sheets (Arkin, 2008; Canton et al., 2008). Typically, these experimental metadata include information on the plasmid vector, the testing organism or strain, any relevant growth conditions, and the equipment or methodologies used to capture the bioparts functionality. The primary purpose of biopart characterization data is to provide the necessary experimental data for predictive *in silico* biological modeling. The determination of which biological data provide the greatest insight into the behavior of a given biological system is largely debatable; at least until more biological design rules are understood. The context dependency of bioparts *in vivo* provides significant challenges in predicting their function as modular components. Therefore, for biopart characterization, measurement standards should largely be defined by those biological data that can be measured (metrology), how relevant those data are for predicting the behavior of a biological process (modeling) and how widely these data can be adopted (standardization). This last point is particularly important since bioparts should ideally be reusable (modular) across multiple applications and contexts. To enable this, the formatting of these data should ideally be standardized to facilitate the measurement and use of biopart characterization data across different *in silico* design tools, forward-design strategies, and workflows.

The most concerted effort is the Synthetic Biology Open Language (SBOL) consortia, a group of life scientists, engineers, computer scientists and mathematicians that are actively building a set of standards that define a common data format for bioparts and their accompanying characterization data (Bower et al., 2010; Galdzicki et al., 2011; Quinn et al., 2013; Roehner and Myers, 2013). The concept is to create a file structure that can capture biopart sequence, characterization and experimental data in a format that is platform independent. Crucially, the format is designed to be extendable to include additional parameters as new characterization technologies and methodologies emerge. In combination with SBOL visual (SBOLv), which defines a standardized way to visually denote bioparts through symbols, the SBOL standard is set to enable the seamless sharing of genetic designs. Several bioinformatics and molecular cloning design tools have already adopted SBOL, and the intention for SBOL is to provide an interoperable standard between several *in silico* tools such that individuals can optimize their workflow as required, yet retain information between them. Several of these *in silico* tools have been extensively reviewed (MacDonald et al., 2011; Galdzicki et al., 2014); however, we include an updated list here, that combines *in silico* tools from these existing reviews, along with several new tools, in particular R2o designer and COOL (Table 2).

In contrast to SBOL, which is still under development, the iGEM registry of standard biological parts (<http://parts.igem.org/>) has provided a relatively large-scale and publically accessible repository of bioparts and some biopart characterization data for almost 10 years. Since its inception, the iGEM community has led, with mixed success, concerted efforts to improve the quality of its characterization data. The 2014 iGEM competition, for instance, has announced several specialist awards for teams that demonstrate advancements in metrology. This push for improvements in biopart characterization at the grassroots (undergraduate) level

Table 2 | Emerging tools for the forward-design of synthetic pathways and systems.

Software tool	Description
PATHWAY AND CIRCUIT DESIGN	
AutoBioCAD	Automated design of gene regulatory circuits (Rodrigo and Jaramillo, 2013).
Cell designer	Modeling of biochemical networks. http://www.celldesigner.org/
Genetic engineering of cells (GEC)	Biological programming language and visual simulator of biological systems. http://research.microsoft.com/en-us/projects/gec/
GenoCAD	GenoCAD is an open-source computer-assisted-design (CAD) application for synthetic biology. http://www.genocad.org/
Genome compiler – iGEM edition	Cloud based genetic design tool that is optimized for BioBrick assembly and the iGEM competition. http://igem.genomecompiler.com/join
MATLAB: Simbiology	SimBiology® provides an application and programmatic tools to model, simulate, and analyze dynamic biological systems. http://www.mathworks.co.uk/products/simbiology/
Operon calculator	Rational design of bacterial operons to control protein expression. https://salis.psu.edu/software/OperonCalculator_EvaluateMode
OptCom	A modeling framework for the flux balance analysis of microbial communities. http://maranas.che.psu.edu/software.htm
ProMoT	Process Modeling Tool, software for the construction and manipulation of complex technical and biological systems. http://www.mpi-magdeburg.mpg.de/projects/promot/
BIOPART DESIGN	
CaDNAano	Simplifies the process of designing three-dimensional DNA origami nanostructures. http://caDNAano.org/
COOL	Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design (Chin et al., 2014)
mfold/UNAFold	Prediction of nucleic acid secondary structure (Markham and Zuker, 2008). http://mfold.rna.albany.edu/
NUPAC	Prediction and design of nucleic acid secondary structure (Zadeh et al., 2011). http://www.nupack.org/
Promoter calculator	<i>E. coli</i> σ^E – In development (Rhodius and Mutalik, 2010; Rhodius et al., 2012).
RBS calculator	The Ribosome-Binding Site (RBS) Calculator is a design method for predicting and controlling translation initiation and protein expression in bacteria. https://salis.psu.edu/software
RBS designer	Computational design of synthetic ribosome-binding sites (RBS) to control gene expression levels. http://ssbio.cau.ac.kr/web/?page_id=195
RNA designer	Designs RNA secondary structure (Andronescu et al., 2004). http://www.rnasoft.ca/cgi-bin/RNASoft/RNAdesigner/rnadesign.pl
Rosetta	Tools for structure prediction, design, and remodeling of proteins and nucleic acids. http://maranas.che.psu.edu/software.htm
UTR designer	Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency (Seo et al., 2013). http://sbi.postech.ac.kr/utr_designer
MISCELLANEOUS	
R2oDNA designer	Designs orthogonal biologically neutral linker sequences for DNA assembly and other uses (Casini et al., 2013, 2014). http://r2oDNA.com/
SBOL	SBOL core provides an interoperable data format to transfer biopart characterization data between software programs and tools (Roehner and Myers, 2013). http://www.sbolstandard.org/
SBOLv	SBOL visual defines a standardized way to visually denote bioparts through symbols (Quinn et al., 2013). http://www.sbolstandard.org/visual

has permeated up to professional characterization efforts. For instance, early difficulties in the reproducibility of the behavior of DNA regulatory elements between iGEM teams and professional research groups provided the context for the emergence of the relative promoter unit (RPU) as a reference measurement standard (Kelly et al., 2009). The RPU standard compares the relative

activity of a promoter against a reference standard, tested under the same experimental conditions, with an RPU arbitrarily set to 1. The rationale underpinning this standard is that while the absolute activity of a promoter may differ between experimental repeats, the relative activity should be less prone to such variability. Essentially, a promoter that is twice the strength of the

standard should remain so, even between different experimental conditions and methodologies of different research groups. In agreement with this, Kelly et al. (2009) reported a 50% decrease in variability, when RPU were independently reported for a set of Anderson constitutive promoters. Inter-experimental variability and reproducibility of data are a significant problem facing all scientific endeavors (Collins and Tabak, 2014), and for synthetic biologists the RPU measurement standard has highlighted these issues within the context of biopart characterization.

There are, however, no universally agreed standards for advancing biopart characterization metrology, though in general the field is shifting away from relative measurements toward absolute measurements (Table 3). Many research groups are currently interested in measuring absolute numbers of cells, DNA molecules, proteins, or other components that constitute the synthetic system and its context. But this shift is largely incremental as certain types of biological data are very difficult to measure directly. These challenges are, however, worth addressing since it is assumed that such biological data are essential to improve the predictive capabilities of forward-design *in silico* models (Bower et al., 2010; Cooling et al., 2010). Yet, because of such data limitations, current-modeling approaches often depend upon inferred or assumed parameters that are derived from biological data that can be experimentally verified. One such modeling approach by Canton et al. (2008), proposed a set of standardized measurement units termed, PoPs and ribosomes per second (RIPS), even though the absolute biological data that underpin them has not been directly measured *in vivo* (Canton et al., 2008; Cooling et al., 2010; Marchisio, 2014). PoPs infers the flow of RNAP along a point of DNA per second and RIPS infers the flow of ribosomes across an mRNA molecule. As previously noted, PoPs and RIPS cannot be measured directly; instead they are calculated using fluorescence data from a reporter protein (e.g., GFP), growth data (OD), and largely assumed values for other parameters including protein or mRNA concentrations. These data are generally measured *in vivo* within a plate reader

setup, though flow cytometry-based characterization efforts are increasingly being adopted and are set to progress metrology at the single cell level (Díaz et al., 2010; Tracy et al., 2010; Choi et al., 2013; Zuleta et al., 2014). In either case, if experimental setups are sufficiently standardized, it is possible to convert measurements between several widely adopted standards: RPU, PoPs/RIPS, and absolute measurements such as GFP cell⁻¹ s⁻¹ (Kelly et al., 2009).

Notwithstanding the above limitations of PoPs and RIPS, these units were primarily designed to reflect the behavior of genetic circuits at the level of information flow (inputs/outputs) rather than at the truly mechanistic level (Gardner et al., 2000; Canton et al., 2008; Stricker et al., 2008; Marchisio, 2014). For biologists, however, these terms represent an abstract merger of several elements of the transcriptional and translational machinery, which does not accurately reflect the mechanistic underpinning biology. However, abstract and mechanistic modeling approaches are not necessarily mutually exclusive since both approaches can provide insightful information for the forward-design of predictable biological pathways and systems.

Advances in metrology and novel measurement standards that are accessible, and hence, more widely adopted will clearly benefit the whole field of synthetic biology. Yet, it is challenging to achieve consensus for developing measurement standards, since standards intrinsically empower those that promote them above those that have not adopted them (Calvert, 2012; Frow and Calvert, 2013). Conversely, it should be noted that consensus in measurement standards and metrology does not preclude innovation if such standards are flexible enough to accommodate developments in the tools and methodologies that enable researchers to easily share, reuse, and build upon existing genetic designs. Likewise, standardized biological information can still be combined with expert knowledge, or novel forward-design strategies for the construction of complex, robust, and efficient biological systems.

Metrology in biology has been enabled in part to continual advancements in microscopy and in synthetic biology,

Table 3 | Synthetic biology measurement standards.

Measurement standard	Advantages	Disadvantages
RELATIVE		
Relative promoter unit (RPU)	Reduces variability between promoter characterization data across different laboratory groups, equipment or slightly different experimental protocols. Concept may be applied in other contexts beyond promoter characterization.	The choice of reference standard promoter requires consensus
ABSTRACT		
Polymerase operations per second (PoPs) and ribosomes per second (RIPS)	Describes information flow (input/output) from transcriptional-based logic devices Abstract level modeling	Units cannot be directly measured May not capture biological processes at the mechanistic level Does not describe biological information that is sent through other mechanisms e.g., protein post-translational modifications
ABSOLUTE		
GFP cell ⁻¹ s ⁻¹	Direct measurement of the number of fluorescent reporter proteins produced Direct comparisons can be made between data sets Concept may be applied to other biological reporters	Requires careful consideration of the design and measurement of the calibration curve needed to compare fluorescence (arbitrary units) and known fluorescent protein concentrations

technologies such as microfluidics coupled with quantitative microscopy are continuing to gain traction (Lin and Levchenko, 2012; Song et al., 2013; Walter and Bustamante, 2014). Microfluidic technologies enable the precise manipulation of fluids at small-scales through engineered channels, chambers, and valves. Microfluidic chip designs are sufficiently advanced to enable a high-degree of spatial-temporal control of liquid-flows to and between individual cells or cell populations seeded within the chambers of prefabricated microfluidics chips. With this level of control, small molecules that induce gene expression or influence other biological processes can be precisely delivered to elicit acute, basal, or morphogenic responses. Within a synthetic biology context, such systems have been used to characterize DNA regulatory elements, intercellular communication, and synthetic pathways at high spatial-temporal resolution. One notable example shown by Hansen and O'Shea (2013), in which the microfluidic control of the delivery of a small molecule (1-NM-PP1) was used to control the nuclear localization of a Yeast stress-inducible transcription factor, Msn2. Deliberate alterations in the oscillatory or acute dynamics of Msn2 trans-nuclear localization revealed the extent to which promoters respond differently to transcriptional-activation dynamics. From this, promoters could be modeled *in silico*, according to the extent that they could elicit differential gene expression patterns, as a consequence of their ability to distinguish a genuine nuclear-influx of Msn2 from background "noise" (Hansen and O'Shea, 2013). Manipulation of these dynamics could be used to reduce promoter leakiness; or conversely to exploit different classes of promoter transcriptional-signal processing to coordinate multiple genetic programs, through the modulation of a single transcription factor.

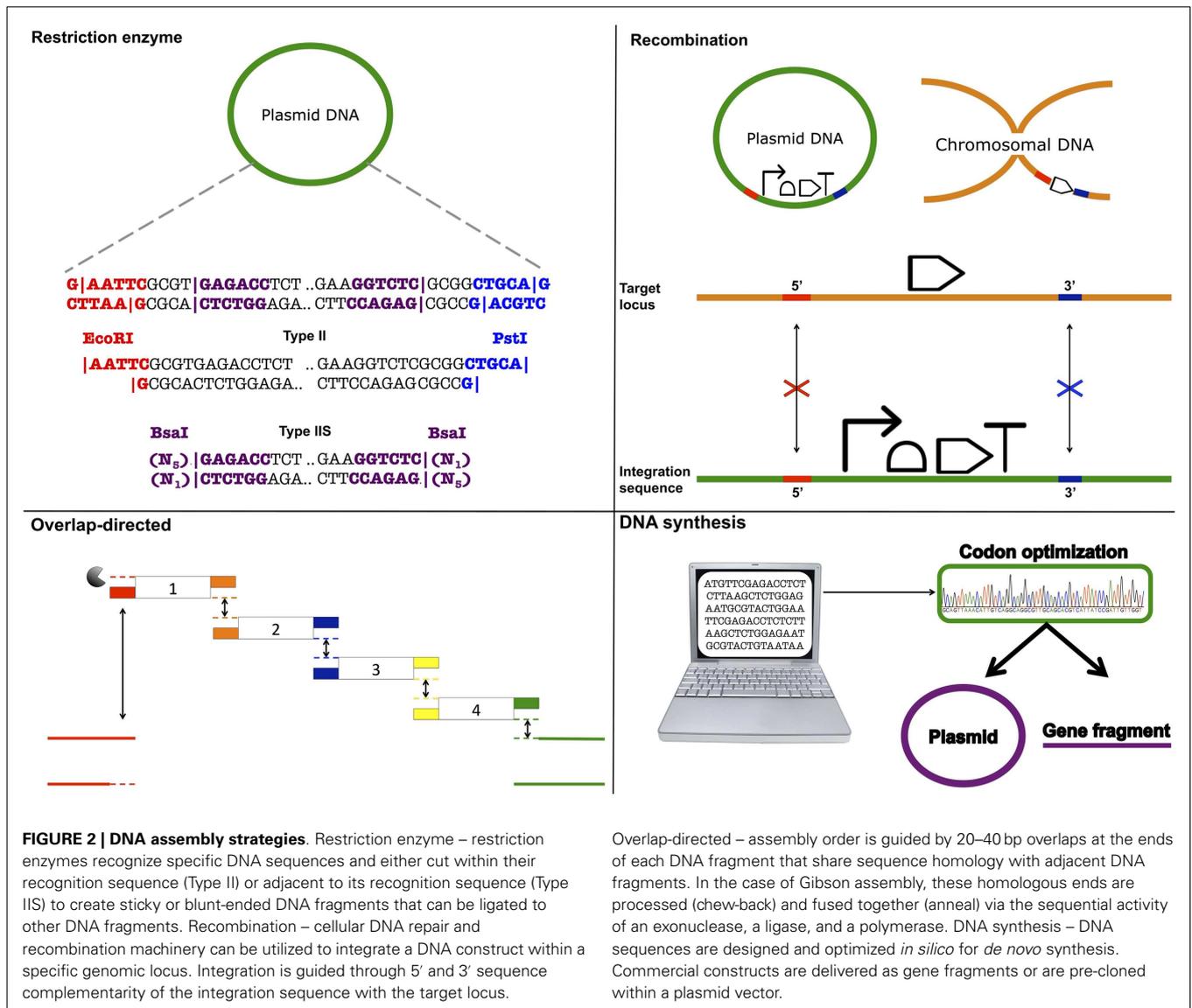
Another important technology for synthetic biology is flow cytometry, which relies upon hydrodynamic focusing to guide single cells through a fluidic channel where they are measured (Piyasena and Graves, 2014). Recent models of flow cytometers can simultaneously measure cell size, complexity, and up to 17 channels of fluorescence (Basiji et al., 2007; Piyasena and Graves, 2014), each of which could be used to capture data from different reporter outputs. Of the biological reporters available, RNA aptamers are particularly noteworthy, since they have the potential to increase the type and range of biological information that can be measured (Cho et al., 2013; Pothoulakis et al., 2014). For instance, several groups have reported the simultaneous measurement of both transcription (mRNA levels) and translation (protein levels) (Chizzolini et al., 2013; Pothoulakis et al., 2014). In both cases Spinach, an RNA aptamer that binds a fluorophore (Paige et al., 2011), was incorporated within the 3' untranslated region (UTR) of a fluorescent reporter protein, either GFP or RFP (Chizzolini et al., 2013; Pothoulakis et al., 2014). Providing there is no spectral-overlap between fluorophores, this strategy could conceivably be up-scaled to measure entire synthetic pathways, and thus inform operon design strategies (Hiroe et al., 2012; Chizzolini et al., 2013). Metabolic engineering efforts may also benefit from engineered RNA aptamer-hybrids that simultaneously bind cellular metabolites and a fluorophore, effectively enabling the real-time reporting of intracellular metabolic flux (Barrick and Breaker, 2007; Roth and Breaker, 2009; Sefah et al., 2013; Szeto et al., 2014). These exponential increases in biological data could significantly impact

whole-cell modeling (Atlas et al., 2008; Gama-Castro et al., 2011; Shuler et al., 2012; O'Brien et al., 2013) and pave the way for novel measurement standards or modeling approaches that are wholly based upon directly measured biological processes.

ASSEMBLING DNA INTO BIOPARTS, PATHWAYS, AND GENOMES

Recombinant DNA technology, in which DNA sequences are "cut and pasted" together via restriction enzymes and DNA ligases respectively, form the foundations of the 1970s biotechnological revolution and have greatly expanded the possibilities of genetic engineering (Zimmerman et al., 1967; Cohen et al., 1973; Lobban and Kaiser, 1973). Synthetic biology continues to benefit from these foundational advancements in recombinant DNA-based biotechnology. For example, the BioBrick DNA assembly standard, uses a set of standardized restriction sites, termed the prefix (*EcoRI XbaI*) and suffix (*SpeI PstI*), that flank each biopart (BioBrick) (Rokke et al., 2014). Digestion and ligation using these sites allow several parts to be assembled together in a standard fashion. The BioBrick standard was originally developed by Tom Knight in 2003 and is still used within the synthetic biology community, particularly during the iGEM competition (Rokke et al., 2014). The BioBrick assembly standard is beneficial to the synthetic biology community for several reasons. Firstly, the flanking restriction site sequences set a physical border that defines individual bioparts. As a result, the BioBrick assembly standard realizes the idea that DNA sequences encode discrete functions and that these individual blocks (BioBricks) can be assembled together like "legotm bricks." Additionally, the use of standardized restriction sites ensures that the cloning strategy for assembling BioBricks is standardized across the entire research community; thereby eliminating the requirement for some cloning-based tacit knowledge. Despite these advantages, a major limitation of the approach is that BioBrick sequences must not contain the prefix and suffix restriction sites, thus limiting the range of sequences that can be assembled. Additionally, when *XbaI* and *SpeI* sites are ligated together, the ligated sequence creates a "scar," which does not contain either an *XbaI* or *SpeI* restriction site (Speer and Richard, 2011; Rokke et al., 2014). Scar sequences may alter the behavior of the flanking bioparts or prevent the generation of fusion proteins, and therefore, can be undesirable (Anderson et al., 2010; Ellis et al., 2011).

BioBrick assembly is also an inefficient way to create large multi-part constructs since it is limited to the assembly of two bioparts per reaction, as defined by the three antibiotic (3A) assembly method (Speer and Richard, 2011). ePathBrick potentially overcomes this limitation through the use of an expansive set of BioBrick-compatible isocaudomer pairs of restriction sites (Xu et al., 2012). The combinatorial assembly of multiple inserts is possible through the restriction digestion and ligation of different isocaudomer pairs into an ePathBrick vector. Backwards compatibility with the BioBrick standard is certainly advantageous from the perspective of modularity (re-useable bioparts); however, ePathBrick is still subject to the BioBrick limitations of forbidden sequences and post-assembly scar sequences. With these limitations in mind, several DNA assembly methods have been developed to address them (Figure 2) (Chao et al., 2014).



RESTRICTION-DIRECTED ASSEMBLY: Bgl BRICKS, GOLDEN GATE, AND SEVA

Golden gate assembly (Engler et al., 2008; Engler and Marillonnet, 2011, 2013), Bgl Bricks (Anderson et al., 2010), and the Standard European Vector Architecture (SEVA) (Silva-Rocha et al., 2013) use a set of restriction sites to standardize DNA assembly. However, in contrast to the BioBrick standard, these assembly methods use rare restriction site sequences, and therefore, support a greater range of sequences. The Bgl Brick standard uses *BglII* and *BamHI* restriction sites. Annealed *BglII* and *BamHI* restriction sites generate an inert, glycine-serine encoding scar sequence, which in contrast to the BioBrick standard scar allows the assembly of protein fusions. Golden Gate assembly supports scar-less assembly through the use of Type IIS restriction enzymes that act by cleaving outside of their recognition sequence leaving a variable overhang, which directs the assembly order and ligation reaction. If cleavage sites are designed appropriately, these overhangs can be designed so that

the final assembled sequences are “scar-less.” More recently, combinatorial Golden Gate assembly methods have been described that allow multi-gene constructs, including synthetic pathways, to be assembled in parallel (Engler and Marillonnet, 2011, 2013). SEVA and to some extent ePathBrick, differ from the majority of assembly methods in that they are more correctly described as modular standards. SEVA describes a set of criteria for the physical assembly of plasmids according to a three-component architecture: an origin of replication segment, a selection marker segment, and a cargo segment (Silva-Rocha et al., 2013). These segments are flanked by insulator sequences and assembled together with a set of rare restriction sites. While the rationales for restriction site-based assembly methods support modularity, their limitations have led several research groups in the synthetic biology community to “trade-in” standardization and modularity, in favor of “bespoke” assembly methods that enable one-pot assembly of multiple DNA parts.

OVERLAP-DIRECTED ASSEMBLY: GIBSON, SLiC, CPEC, SLiCE, AND PAPERCLIP

Daniel Gibson developed a widely adopted DNA assembly method that allows multiple DNA fragments to be assembled in a one-pot *in vitro* reaction (Gibson et al., 2009; Gibson, 2011). The Gibson assembly uses a linearized destination vector and PCR generated inserts as its starting material. Inserts are generated with PCR primers that include 20–40 bp overlaps that share sequence homology to adjacent DNA fragments. As a result, the correct arrangement of several inserts entering the same destination vector can be defined. During the reaction, a T5 exonuclease acts to chew-back at the 5' ends of the linearized destination vector and inserts. The reaction occurs at 50°C and therefore the T5 exonuclease along with its activity is eventually inactivated. The destination vector and inserts anneal together, as defined by their exonuclease exposed homologous ends, and Phusion polymerase activity acts to fill in the gaps. Finally, Taq ligase seals nicks between the joined DNA fragments. Gibson assembly is simple, can assemble five or more parts in a single reaction, and the reaction itself only takes around 60 min, after which the final assembled product can be directly transformed into *E. coli*.

Sequence and Ligase-independent Cloning (SLiC) (Li and Elledge, 2007), Circular Polymerase Extension Cloning (CPEC) (Quan and Tian, 2009, 2011), Seamless Ligation Cloning Extract (SLiCE) (Zhang et al., 2012) are also overlap-directed DNA assembly methods that all result in the same final product. Therefore, inserts and destination vectors designed for Gibson assembly can also be used in SLiC, CPEC, and SLiCE assemblies. During SLiC reactions, the destination vector and inserts are independently treated *in vitro* with T4 DNA polymerase, which exhibits exonuclease activity in the absence of deoxynucleotide triphosphates (dNTPs). Exonuclease activity is subsequently inhibited with the addition of deoxycytidine triphosphate (dCTP) and the destination vector and inserts are then mixed together for annealing. However, because SLiC reactions do not include DNA ligase, gaps, or nicks in the DNA are repaired once the final product is transformed into *E. coli*. CPEC on the other hand, is a PCR-based approach in which the linearized destination vector and inserts are initially denatured to produce single DNA strands. These are then annealed together, as directed by the homologous DNA overlap regions. Once annealed, the destination vector and inserts act to prime each other for extension via the activity of Phusion DNA polymerase. A low number of PCR cycles act to prevent the propagation of PCR-based errors. SLiCE reactions markedly differ from the assembly methods just described in that they involve an *ex vivo* bacterial cell extract (PPY, *E. coli* DH10B λ -red) as the reaction mix. Since exogenous polymerases and DNA ligases are not required, this is a potentially cost-effective method and like Gibson, assembly reactions also typically take just 60 min, although at 37°C instead of 50°C as per Gibson assembly.

PaperClip DNA assembly is a relatively new overlap-directed assembly method that uses pairs of bridging oligonucleotides termed “Clips” to direct the assembly of multi-part constructs (Trubitsyna et al., 2014). Interestingly, PaperClip assembly protocols are derived from CPEC (PCR-based) and SLiCE (*ex vivo*-based) assembly methodologies. Yet, PaperClip assembly is advantageous over these assembly methods in that once the “Clips”

have been prepared, the required assembly order of parts can be determined in a single reaction. While, “Clips” introduce an alanine encoding scar sequence between each part, the bridging oligos used to assemble multi-part constructs in ligase cycling reaction (LCR) assembly are scar-less (Rouillard et al., 2004; de Kok et al., 2014). Though as we describe below, PaperClip assembly differentiates itself from Gibson, CPEC, SLiCE, and LCR assembly methods in that *de novo* assembly fragments do not need to be generated each time the order assembly is changed (Trubitsyna et al., 2014).

Overlap-directed assembly methods use sequence homology to guide assembly and are therefore largely sequence independent. This is a clear advantage over restriction site-based DNA assembly methods and their forbidden sequences. It should be noted that repeat and short DNA sequences, particularly those that give rise to DNA secondary structures, can reduce the efficiency of overlap-directed methods and are best avoided. On the other hand, CPEC denaturation PCR cycles mitigate the effect of DNA secondary structures to some degree. Overlap-directed methods are also efficient at assembling multiple parts in a predefined order within a single one-pot reaction. Gibson assembly, for example, has been used to assemble genome-scale DNA fragments, including the complete assembly of the *M. genitalium* genome (583 kb) and more recently the entire mouse mitochondrial genome (16.3 kb) (Gibson et al., 2008, 2010b). It is clear therefore that overlap-directed assembly methods can be scaled toward the assembly of large genetic constructs, including synthetic genomes (Gibson et al., 2010a). Yet, despite their proven utility, they are inherently “bespoke” and are thus in conflict with the ideals of embedding standardization and modularity concepts within DNA assembly strategies. For instance, custom primers are needed to generate inserts *de novo* each time the assembly order is changed and while it is now possible to automate overlap-directed assembly primer design (Hillson et al., 2012), these assembly methods still require tacit knowledge. To this end, additional methodologies are being developed with the aim of making overlap-directed DNA assembly modular.

OVERLAP-DIRECTED ASSEMBLY WITH BIOLOGICALLY NEUTRAL LINKER SEQUENCES

Modular overlap-directed assembly with linkers (MODAL) makes use of standardized flanking sequences and biologically neutral (orthogonal) linkers as part of a modular overlap-directed DNA assembly strategy (Casini et al., 2013). MODAL assembly requires bioparts to be standardized with the addition of a common prefix and suffix sequence. The prefix and suffix sequences do not contain restriction sites and are not directly required for the assembly process. Instead, these sequences serve as a consistent set of PCR primer “landing pads” that enable all MODAL bioparts to be generated using the same primer set. Additionally, these sites serve as priming sites for the PCR-directed addition of biologically neutral linker sequences that serve as homologous sequences for overlap-directed assembly. These sequences can be designed with R2oDNA Designer (Casini et al., 2013, 2014), an *in silico* tool that was developed to automatically design orthogonal linker sequences for use in MODAL and other applications. Similar strategies have also been developed in parallel, in which biologically inactive

unique nucleotide sequences (UNSeqs) were utilized to guide the Gibson assembly of insulated genetic circuits (Guye et al., 2013; Torella et al., 2014a,b). These neutral sequences are often standardized and may also incorporate BioBrick restriction sites, thus enabling modularity and standardization to be embedded within overlap-directed assembly strategies.

***IN VIVO* DNA ASSEMBLY AND GENOME ENGINEERING**

An array of chassis with a broad set of useful, extensively characterized genotypes and phenotypes are available to the synthetic biology community (Table 2). However, there are applications where it is appropriate to rationally engineer a chassis. For instance, an application may require a novel strain that is optimized, at the genome level, to fit a set of specific design requirements that may be difficult or otherwise impractical to bioprospect. Typically, genome-engineering efforts are geared toward maximizing compatibility between a chassis and a synthetic system, increasing the efficiency of the metabolic flux across a synthetic pathway or toward minimizing burden effects. The field is making progress in establishing rationally engineered genomes; of which the synthetic yeast 2.0 project (Dymond and Boeke, 2012; Annaluru et al., 2014; Lin et al., 2014) and minimal genome projects (Glass et al., 2006; Dewall and Cheng, 2011; Shuler et al., 2012), are currently the most prominent exemplars. These genome-engineering efforts are made possible due to the emergence and ongoing development of an expanding set of *in vivo* DNA assembly methods and genome-engineering tools.

Recombineering approaches, in which synthetic linear ds/ssDNA sequences are introduced into genomic regions through a process of homologous recombination, have proven utility as an efficient method to knockout or knock-in sequences of interest. Recombineering enables genomic engineering at all scales; from the introduction of single nucleotide polymorphisms, to the replacement of 40 kb+ DNA fragments or even toward the assembly of entire genomes (Narayanan and Chen, 2011; Zhao et al., 2011; Bonde et al., 2014; Song et al., 2014). *S. cerevisiae* transformation-associated recombination (TAR) cloning (Kouprina and Larionov, 2008), *Bacillus* Domino (Ohtani et al., 2012), and the *E. coli* Single-Selective-Marker Recombination Assembly System (SRAS) (Shi et al., 2013) uses the endogenous homologous recombination machinery of the indicated organisms to assemble DNA constructs *in vivo*. A variant of the yeast TAR method has successfully generated several genomes, including the first *in vivo* assembled synthetic genome of *M. genitalium* (Gibson et al., 2008). *Bacillus* domino has also shared similar successes in that this assembly method has also assembled DNA at the genomic scale, including the mouse mitochondrial genome and the rice chloroplast genome (Itaya et al., 2008; Ohtani et al., 2012; Iwata et al., 2013). While *E. coli* SRAS could potentially support the assembly of large DNA fragments, it is currently optimized for the assembly of multi-part constructs and their simultaneous integration into the *E. coli* genome (Shi et al., 2013).

The lambda-red (λ -red) recombinase system is another recombineering strategy, which is used for the integration of ssDNA or dsDNA constructs into the *E. coli* genome (Murphy, 1998; Murphy and Campellone, 2003). Optimized lambda-red recombination protocols can integrate linear DNA sequences into a

specific genomic target, guided by only 35–50 bases of flanking homologous sequence (Murphy and Campellone, 2003). Interestingly, lambda-red-mediated recombination events do not require endogenous recombination proteins (e.g., RecA) and instead linear ssDNA or dsDNA constructs are integrated into the *E. coli* genome via the action of three λ -red proteins; Gam, Exo, and Beta. Gam protects linear dsDNA from the exonuclease activity of the endogenous proteins RecBCD, thus increasing the efficiency at which the introduced dsDNA will be recombined into the genome. λ -red-mediated recombination itself is primarily mediated by Exo, a 5'–3'–dsDNA-specific exonuclease and Beta, a ssDNA annealing protein. It is interesting to note that Gam-associated protection of dsDNA is exploited in SLiCE *ex vivo* DNA assembly and as we discuss later, for *in vitro* transcription–translation (TX–TL) coupled reactions involving linear DNA as the input (Sitaraman et al., 2004).

The introduction of a large number of rationally engineered genomic changes is a potentially laborious process; however, multiplex automated genome engineering (MAGE) enables the automation of large-scale recombineering strategies. MAGE was originally characterized within EcNR2, a variant strain of *E. coli* MG1655. EcNR2 was modified to incorporate the λ -red recombination system and also to be deficient in DNA mismatch repair via the knockout of the *mutS* gene (Wang et al., 2009). MAGE relies upon the λ -red Beta protein-assisted incorporation of ssDNA oligonucleotides, typically 90mers, into the lagging strand during DNA replication (Wang et al., 2009). MAGE oligonucleotide pools can be designed to incorporate highly specific changes at a single genomic site, to introduce multiple changes across a single locus or to simultaneously target multiple genomic sites. These outcomes are largely defined through the diversity of the MAGE oligonucleotide pool, where mixtures of degenerate oligonucleotides can be designed to introduce divergent changes across a broad sequence and recombination efficiency space. Where a large number of simultaneous genomic changes are required, the process can be repeated through multiple MAGE cycles of cell growth, electroporation of oligonucleotides into the cell population, and phenotype/genotype characterization. MAGE cycles can be automated through a microfluidics-type setup and in combination with MODEST or optMAGE, which are *in silico* MAGE oligonucleotide design tools (Table 4), the directed evolution of a rationally designed chassis, can be accomplished within a timescale of several days. Indeed, MAGE has been used to optimize the DXP pathway in *E. coli*, such that isolated variants that are capable of a fivefold increase in lycopene production were engineered in just 3 days (Wang et al., 2009).

In parallel with MAGE, conjugative assembly genome engineering (CAGE) can be used to coordinate large-scale genomic engineering strategies across phases, such that subtle genetic combinations that are lethal can be screened out in a manner that does not impede overall progress toward the final strain. To achieve this, CAGE guides the conjugal transfer of MAGE genome modifications between hierarchical pairs of donor–recipient *E. coli*, such that a new strain emerges which incorporates all of the MAGE-optimized modifications from previous generations (Isaacs et al., 2011). Multiple MAGE–CAGE rounds enable a large set of genomic modifications to be generated and carefully

Table 4 | DNA assembly and genome-engineering tools.

Assembly method	Mechanism	Sequence independent*	Scar-less assembly	Software support tools
Bgl Brick	Type II restriction enzymes	No	No	Under development
BioBrick standard	Type II restriction enzymes	No	No	Registry of standard biological parts, an online and physical repository of BioBrick parts (http://parts.igem.org/)
ePathBrick	Type II restriction enzymes	No	No	–
SEVA	Type II restriction enzymes	No	Possible	SEVA-DB platform, an online repository of SEVA-compliant parts (Silva-Rocha et al., 2013)
Golden gate	Type IIS restriction enzymes	No	Possible	j5, an automated primer design tool can be adapted for Golden gate combinatorial assembly (Hillson et al., 2012)
Gibson	Overlap-directed	Yes – however, short or repeat sequences that give rise to secondary DNA structures are a problem	Yes	j5, an automated primer design tool (Hillson et al., 2012)
SLiC	Overlap-directed	Yes – however, short or repeat sequences that give rise to secondary DNA structures are a problem	Yes	j5, an automated primer design tool (Hillson et al., 2012)
CPEC	PCR-based overlap-directed	Yes – however, short or repeat sequences are a problem	Yes	j5, an automated primer design tool (Hillson et al., 2012)
SLiCE	<i>Ex vivo</i> overlap-directed	Yes – however, short or repeat sequences that give rise to secondary DNA structures are a problem	Yes	j5, an automated primer design tool (Hillson et al., 2012)
PaperClip	Overlap-directed with oligonucleotide pairs “Clips”	Yes – however, constructs cannot contain repetitive parts or more than 40 bases of identical regions	No	–
Ligase cycling reaction	Bridging oligo-directed assembly	Yes	Yes	Gene2Oligo: oligonucleotide design for <i>in vitro</i> gene synthesis (Rouillard et al., 2004). http://berry.engin.umich.edu/gene2oligo/
Gibson with UNSes	Overlap-directed with orthogonal linkers	Yes	No	R2oDNA designer: computational design of biologically neutral (orthogonal) synthetic DNA sequences (Casini et al., 2013, 2014). Computational design rules for UNSes (Guye et al., 2013; Torella et al., 2014a,b).
MODAL	Overlap-directed with orthogonal linkers	No	No	R2oDNA designer: computational design of biologically neutral (orthogonal) synthetic DNA sequences (Casini et al., 2013, 2014).
<i>Bacillus</i> domino	<i>In vivo</i> homologous recombination	Yes – however, cannot assemble <i>Bacillus</i> genomic sequences	Yes	–
<i>E. coli</i> (SRAS)	<i>In vivo</i> homologous recombination	Yes – however, homologous sequences are needed for recombination	Yes	–

(Continued)

Table 4 | Continued

Assembly method	Mechanism	Sequence independent*	Scar-less assembly	Software support tools
MAGE and CAGE	<i>In vivo</i> homologous recombination/conjugation	Yes – however, homologous sequences are needed for recombination	Yes	MAGE oligonucleotide design tools: MODEST (Colloms et al., 2014) http://modest.biosustain.dtu.dk/ ; optMAGE http://arep.med.harvard.edu/optMAGE/
Yeast TAR	<i>In vivo</i> homologous recombination	Yes	Yes	–
Engineered nucleases (zinc-finger nucleases, TALENs, and CRISPR/Cas9)	DNA cleavage and non-homologous end joining (NHEJ) or homology-directed repair (HDR)	Yes – cleavage can be directed toward sequence of interest	Yes*, However NHEJ can introduce random mutations.	E-CRISP: CRISPR target site identification (Heigwer et al., 2014). http://e-crisp-test.dkfz.de/E-CRISP/
SIRA	Serine integrase recombinational assembly	Yes – as long as ϕ C31 recombination sites are avoided	No	Software support tools are in development (Colloms et al., 2014)
DNA synthesis	Polymerase cycling assembly from pools of overlapping custom oligos	Yes – however, repeat sequences or high GC content can be problematic	Yes	Codon optimization, the removal of undesirable restriction sites and the specification of 5' and 3' sequences are possible during the order processes of several commercial companies. GeneDesigner (Villalobos et al., 2006) https://www.DNA20.com/resources/genedesigner

*Sequence-independent assembly strategies do not place restrictions upon which DNA sequences are permitted within assembly fragments.

integrated. As an example of such an approach, Isaacs et al. (2011) used a MAGE–CAGE strategy to replace 314 TAG stop codons with the synonymous TAA in *E. coli* across its entire genome (Isaacs et al., 2011).

Engineered nucleases, which cleave specific DNA sequences, creating double-stranded DNA breaks, can be used to introduce genomic changes. These strategies depend upon the random occurrence of perturbations in DNA repair mechanisms, where double-stranded breaks are inappropriately repaired, resulting in erroneous sequence insertions, deletions, or even significant chromosomal rearrangements. Screening strategies to identify cells that contain desirable genomic alterations can be subsequently isolated as an engineered population. Zinc-finger nucleases (Ellis et al., 2013), TALENs (Mahfouz et al., 2011), and the CRISPR/Cas system (Sander and Joung, 2014) have all been engineered for these types of genome editing applications. The CRISPR/Cas system is particularly interesting since as discussed above, a deactivated Cas9 nuclease:gRNA complex can also be fused with domains that act as transcriptional activators or repressors (Bikard et al., 2013; Mali et al., 2013; Qi et al., 2013). Nuclease-mediated genome editing strategies can also be combined with a recombineering-type approach, in which an engineered dsDNA can be introduced into the cell, which has sequence complementarity at the site of the nuclease breakage. Through the endogenous homologous recombination machinery (DNA repair mechanisms), it is possible

to rationally integrate the engineered dsDNA into the genome (Cong et al., 2013; Sander and Joung, 2014). Thus in combination, MAGE, CAGE, and engineered, targeted nucleases (Zinc, TALENs and Cas9) represent a set of molecular tools that enable genome editing and the transcriptional control of natural and synthetic genomes.

DNA SYNTHESIS

Synthetic biology has greatly benefited from the rapid decline in the cost of commercial gene synthesis, a phenomenon popularized by the Carlson curve (Carlson, 2009), which is analogous to Moore's law. Although the rate of decline has decreased in recent years, with DNA synthesis costs now relatively stable (Carlson, 2009, http://www.synthesis.cc/cgi-bin/mt/mt-search.cgi?blog_id=1&tag=CarlsonCurves&limit=20) it is likely that new disruptive technologies will decrease DNA synthesis costs in the near future. DNA synthesis costs are still sufficiently low that many research groups routinely order the synthesis of genes and gene fragments although still prohibitive for library generation or for the synthesis of large multi-part pathways. In these cases, gene synthesis can be combined with additional cloning techniques such as overlap-directed assembly or mutagenic PCR to generate large constructs or biopart libraries, respectively. It is likely that as DNA synthesis costs decline, there will be a continual shift away from DNA assembly toward *de novo* DNA synthesis,

which will have a transformative effect on synthetic biology and the design-build-test cycle.

RAPID PROTOTYPING

High-throughput platforms bring scalability to biopart characterization efforts, through the parallel characterization of function and context of entire biopart libraries (Arkin, 2013; Keren et al., 2013; Mutalik et al., 2013b). To ensure consistency at such scale, high-throughput workflows typically couple liquid-handling robots with plate readers (Keren et al., 2013), flow cytometry (Piyasena and Graves, 2014; Zuleta et al., 2014), or microfluidics (Lin and Levchenko, 2012; Benedetto et al., 2014) in order to automate the majority of the experimental workflow. Several high-throughput platforms have been described, the majority of which were used to characterize DNA regulatory elements (Keren et al., 2013; Mutalik et al., 2013a,b), however, this is expanding to include the characterization of enzymes (Choi et al., 2013), multi-gene operons (Chizzolini et al., 2013), and RNA aptamers (Cho et al., 2013; Szeto et al., 2014). When coupled with automated data analysis and modeling, these technologies and workflows could become rapid prototyping platforms, enabling a truly biological design cycle approach (Kitney and Freemont, 2012). At present, these high-throughput workflows are typically semi-rational design strategies in which thousands of biopart variants are tested and screened as part of a discovery workflow. Yet, at the same time, these approaches are simultaneously generating large data sets that provide useful insights into biological processes that may inform biological design rules. For example, characterization efforts have informed several systematic methodologies for the rational optimization of synthetic systems at the transcriptional, translational, and post-translational level (Table 2) (Arkin, 2008; Arpino et al., 2013; Reeve et al., 2014). In cases where synthetic systems could conceivably be rationally designed, it is still naïve to assume that the first iteration of a synthetic biological system will perfectly match the design specifications. Instead, multiple iterations of the design-build-test cycle will be needed until forward-design approaches are sufficiently advanced. Therefore, the requirements of interoperable standards in which researchers can apply the same protocols across different liquid-handling platforms are essential. To this end, Linshiz et al. (2013) have implemented a high-level robot programming language (PaR-PaR), which can translate biological protocols into instruction sets for an extendable range of liquid-handling robot platforms. As a consequence of this approach, the training requirements for end-users to implement the same biological protocol across different liquid-handlers are significantly reduced (Linshiz et al., 2013). If, as the authors propose, PaR-PaR is combined with SBOL, then the adoption of PaR-PaR scripts will enable researchers to share the same high-throughput DNA assembly or characterization protocols, but have them implemented across different experimental and equipment setups.

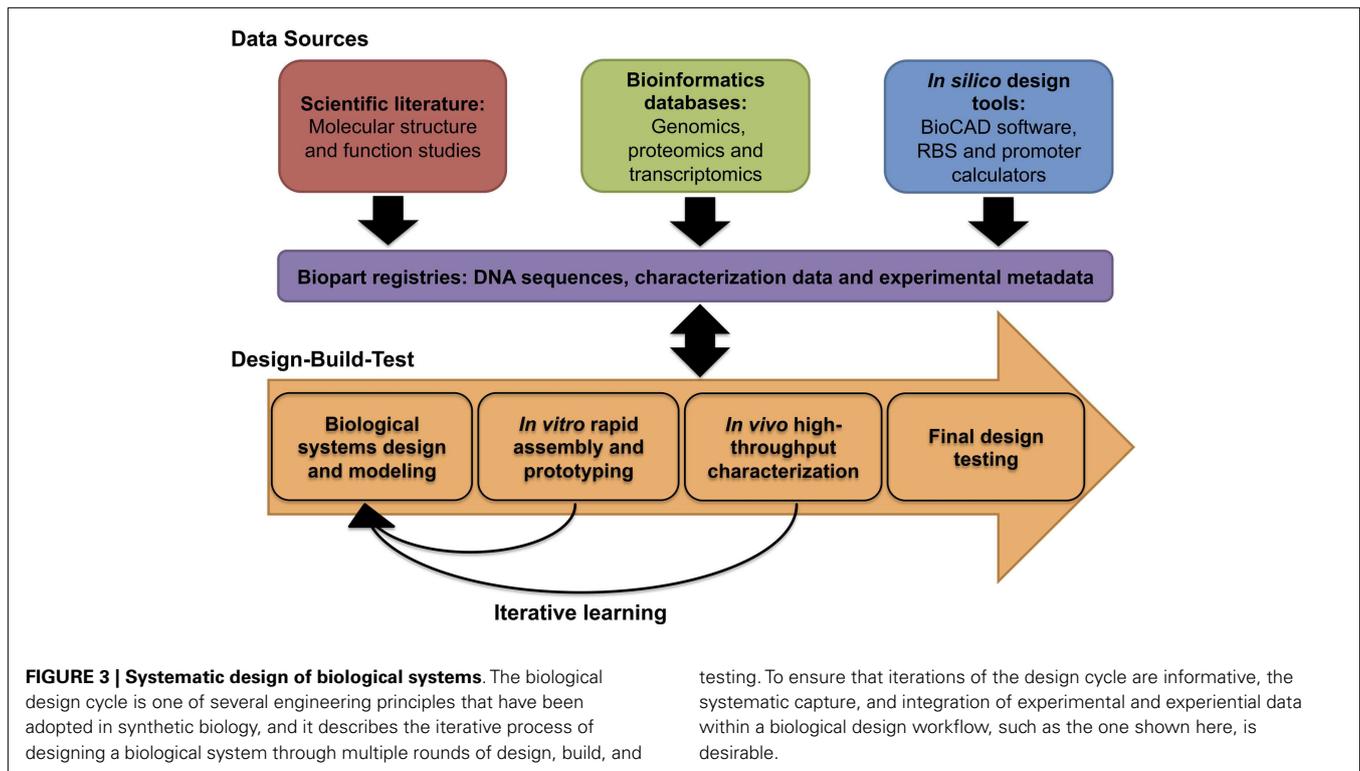
The majority of the rapid prototyping platforms that we have described so far have been optimized for testing biological parts, devices, and systems *in vivo*; however, *in vitro* systems are emerging as a useful testing platform. Cell-free protein synthesis (CFPS) systems based upon *E. coli* (Nirenberg and Matthaei, 1961; Sitaraman et al., 2004; Hong et al., 2014), *B. subtilis* (Zaghloul and

Doi, 1987), *S. cerevisiae* (Hodgman and Jewett, 2013; Gan and Jewett, 2014), or other cell extracts have been reported in the scientific literature for several decades. Several CFPS systems are commercially available and are principally marketed as protein expression systems. Optimized *E. coli* CFPS systems can synthesize up to 2.3 mg/ml of the target protein (Caschera and Noireaux, 2014), including those that are toxic *in vivo*. In recent years, the synthetic biology community has repurposed CFPS systems as *in vitro* transcription-translation (TX-TL) coupled characterization platforms. A typical TX-TL reaction combines a synthetic system encoded into plasmid, linear or closed circular DNA, with cell-free extract, and a reaction buffer, the contents of which can be optimized (Sun et al., 2013a). For instance, the addition of maltodextrin (Wang and Zhang, 2009) and to a lesser degree maltose (Caschera and Noireaux, 2014) as an additional energy source can increase protein production, essentially prolonging the duration of *in vitro* reactions for up to 10 h.

Transcription-translation characterization systems provide characterization data within a timescale of hours (Chappell et al., 2013), and are therefore, amenable to a rapid prototyping workflow (Chappell et al., 2013; Sun et al., 2013b). For instance, Chappell et al. (2013) characterized a panel of Anderson constitutive promoters, using a commercially available TX-TL system, within a 5-h workflow. Interestingly, the *in vitro* characterization data of a set of Anderson promoters correlated with their performance *in vivo* (Chappell et al., 2013). Likewise, in the same study, a panel of LasR responsive, AHL-inducible promoters, also behaved similarly *in vitro* and *in vivo*, although meaningful comparisons could only be made where constructs were encoded into plasmid or closed circular DNA (Chappell et al., 2013). PCR-generated linear DNA templates did not produce sufficient transcription and translation of the reporter protein (Chappell et al., 2013). Based upon several reports, it is likely that linear DNA templates are unstable *in vitro* due to the presence of exonuclease activity in the cell-free extract (Sitaraman et al., 2004; Sun et al., 2013b). Expression of the phage lambda protein Gam, an inhibitor of RecBCD (ExoV), along with other modifications, can minimize linear DNA degradation, thus restoring protein expression to levels that are comparable to plasmid DNA constructs (Sitaraman et al., 2004; Sun et al., 2013b). Yet, in disagreement with several other studies (Chappell et al., 2013; Iyer et al., 2013; Lu and Ellington, 2014), Sun et al. (2013b) reported that *in vitro* characterization data were not comparable to *in vivo* data, though they did describe a methodology to calibrate between them. While the comparability between *in vitro* and *in vivo* characterization requires further investigation, several reports have demonstrated that cell-free TX-TL systems have proven utility in the rapid prototyping of logic-based genetic circuits (Karig et al., 2012; Shin and Noireaux, 2012; Iyer et al., 2013) or synthetic operons (Lu and Ellington, 2014). Within a systematic design context, *in vitro* characterization approaches have the potential to complement *in vivo* prototyping efforts by rapidly providing the characterization data required to rationally select a smaller number of designs for final testing (Figure 3).

CONCLUSION

Synthetic biology is generally described as the “engineering of biology” yet since its inception, the field has faced the well-understood



reality that biological systems are complex, stochastic, and difficult to predict, and are therefore, intrinsically difficult to engineer. In order to address these fundamental challenges, synthetic biology must use and explore the existing large body of knowledge of biological systems at different scales from molecular to cellular to organismal. By establishing a systematic design framework in which existing biological knowledge can be adapted and utilized will ensure the rapid development of successful applications using synthetic biology. Furthermore, the accumulated measurements and acquired knowledge of many synthetic biology experiments will allow synthetic biologists to establish design rules that tackle biological complexity, such that robust biological systems can be designed, assembled, and prototyped as part of a biological design cycle. At each stage of the design cycle, an expanding repertoire of tools is being developed. In this review, we have highlighted several of these tools in terms of their applications and benefits to the synthetic biology community within the context of the synthetic biology design cycle namely, designing predictable biology (design), assembling DNA into bioparts, pathways, and genomes (build), and rapid prototyping (test).

Design encompasses the development of tools and methodologies that make it easier to forward-design predictable synthetic biological systems. While there are several areas that are critical to designing predictable biology including, chassis selection, biopart design, or engineering strategies, as well as, several accompanying *in silico* design tools, we would argue that measurement and characterization (metrology) of biological parts, devices, and systems is essential for the field of synthetic biology to fulfill its promise. It is only through improvements in our ability to measure and generate meaningful conclusions about the behavior of biological

processes that the field can progress in terms of unlocking additional biological design rules. The RBS calculator is the current exemplar of this perspective, though further work is required to equip the synthetic biology toolbox with the tools to make it easier to engineer radically complex synthetic biological parts, devices, and systems.

Build encompasses DNA assembly and genome-engineering methods that enable synthetic systems to be assembled. The field has benefited immensely from the BioBrick assembly standard. BioBrick assembly, effectively making bioparts reusable (modular) at the physical DNA level, creates a standard that enables multiple research groups to use and share an expanding library of bioparts, without the need for bespoke cloning strategies. While limitations in the BioBrick assembly standard led to the emergence of powerful overlap-directed assembly methods, including Gibson, these methods also shifted away from several of the core principals of synthetic biology since these methods rely on bespoke cloning strategies. However, emerging DNA assembly methods including MODAL or Gibson with UNSes, aim to unify the advantages of overlap-directed assembly with the engineering principle of modularity. However, advances in DNA synthesis and resultant reduction of costs could radically transform the field, such that more time could be diverted away from DNA assembly toward the designing or testing of synthetic systems.

Test encompasses elements of biopart characterization, since even the testing of non-functional designs may provide insights into our understanding of the biological design rules. Liquid-handling robot high-throughput characterization platforms, along with plate readers are equipped to test prototypes of synthetic bioparts, devices, and systems. However, these systems benefit

from the addition of flow cytometry and microfluidics, which bring single cell analysis to these platforms. Thus, individual cells could be analyzed and selected based upon preferred biological performance from a heterogeneous cell mix. Additionally, an array of emerging *in vitro* TX–TL cell-free characterization systems provide characterization data within a timescale of hours, and are therefore, amenable to a rapid prototyping workflow. Such systems are complementary to *in vivo* high-throughput approaches and may speed up iterations through the design cycle by reducing the number of final designs that need to be tested.

As the synthetic biology toolkit expands and more design rules are unlocked, the most successful forward-design strategies are likely to be those that encompass a diverse workflow that combines several interoperable tools at each stage of the design cycle.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the Engineering and Physical Science Research Council (EPSRC) and that of our colleagues in the Centre for Synthetic Biology and Innovation (CSynBI) at Imperial College. Funding: EPSRC [EP/K034359/1; EP/J02175X/1], Bill and Melinda Gates Foundation [OPP1046311], The Wellcome Trust [084369/Z/07/Z].

REFERENCES

- Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W., and Danchin, A. (2013). From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet.* 29, 273–279. doi:10.1016/j.tig.2012.11.001
- Agapakis, C. M. (2013). Designing synthetic biology. *ACS Synth. Biol.* 3, 121–128. doi:10.1021/sb4001068
- Agapakis, C. M., Boyle, P. M., and Silver, P. A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* 8, 527–535. doi:10.1038/nchembio.975
- Anderson, J., Strelkova, N., Stan, G. B., Douglas, T., Savulescu, J., Barahona, M., et al. (2012). Engineering and ethical perspectives in synthetic biology. Rigorous, robust and predictable designs, public engagement and a modern ethical framework are vital to the continued success of synthetic biology. *EMBO Rep.* 13, 584–590. doi:10.1038/embor.2012.81
- Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., et al. (2010). BglBricks: a flexible standard for biological part assembly. *J. Biol. Eng.* 4, 1. doi:10.1186/1754-1611-4-1
- Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* 2, 0028. doi:10.1038/msb4100073
- Andronescu, M., Fejes, A. P., Hutter, F., Hoos, H. H., and Condon, A. (2004). A new algorithm for RNA secondary structure design. *J. Mol. Biol.* 336, 607–624. doi:10.1016/j.jmb.2003.12.041
- Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science* 344, 55–58. doi:10.1126/science.1249252
- Arkin, A. (2008). Setting the standard in synthetic biology. *Nat. Biotechnol.* 26, 771–774. doi:10.1038/nbt0708-771
- Arkin, A. P. (2013). A wise consistency: engineering biology for conformity, reliability, predictability. *Curr. Opin. Chem. Biol.* 17, 893–901. doi:10.1016/j.cbpa.2013.09.012
- Arpino, J. A., Hancock, E. J., Anderson, J., Barahona, M., Stan, G. B., Papachristodoulou, A., et al. (2013). Tuning the dials of synthetic biology. *Microbiology* 159, 1236–1253. doi:10.1099/mic.0.067975-0
- Atlas, J. C., Nikolaev, E. V., Browning, S. T., and Shuler, M. L. (2008). Incorporating genome-wide DNA sequence information into a dynamic whole-cell model of *Escherichia coli*: application to DNA replication. *IET Syst. Biol.* 2, 369–382. doi:10.1049/iet-syb:20070079
- Aw, R., and Polizzi, K. M. (2013). Can too many copies spoil the broth? *Microb. Cell Fact.* 12, 128. doi:10.1186/1475-2859-12-128
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi:10.1126/science.1138140
- Barrick, J. E., and Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 8, R239. doi:10.1186/gb-2007-8-11-r239
- Bartosiak-Jentys, J., Hussein, A. H., Lewis, C. J., and Leak, D. J. (2013). Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology* 159, 1267–1275. doi:10.1099/mic.0.066332-0
- Basiji, D. A., Ortyu, W. E., Liang, L., Venkatachalam, V., and Morrissey, P. (2007). Cellular image analysis and imaging by flow cytometry. *Clin. Lab. Med.* 27, 653–670. doi:10.1016/j.cll.2007.05.008
- Benedetto, A., Accetta, G., Fujita, Y., and Charras, G. (2014). Spatiotemporal control of gene expression using microfluidics. *Lab. Chip* 14, 1336–1347. doi:10.1039/c3lc51281a
- Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S., and Pakrasi, H. B. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol.* 4:246. doi:10.3389/fmicb.2013.00246
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437. doi:10.1093/nar/gkt520
- Blount, B. A., Weenink, T., Vasylechko, S., and Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS ONE* 7:e33279. doi:10.1371/journal.pone.0033279
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi:10.1126/science.1178811
- Boehm, C. R., Freemont, P. S., and Ces, O. (2013). Design of a prototype flow microreactor for synthetic biology in vitro. *Lab. Chip* 13, 3426–3432. doi:10.1039/c3lc50231g
- Bonde, M. T., Klausen, M. S., Anderson, M. V., Wallin, A. I., Wang, H. H., and Sommer, M. O. (2014). MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Res.* 42, W408–W415. doi:10.1093/nar/gku428
- Bower, A. G., McClintock, M. K., and Fong, S. S. (2010). Synthetic biology: a foundation for multi-scale molecular biology. *Bioeng. Bugs* 1, 309–312. doi:10.4161/bbug.1.5.12391
- Cachat, E., and Davies, J. A. (2011). Application of synthetic biology to regenerative medicine. *J. Bioeng. Biomed. Sci.* S2:003. doi:10.4172/2155-9538.S2-003
- Calvert, J. (2012). Ownership and sharing in synthetic biology: a ‘diverse ecology’ of the open and the proprietary[quest]. *BioSocieties* 7, 169–187. doi:10.1057/biosoc.2012.3
- Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26, 787–793. doi:10.1038/nbt1413
- Carlson, R. (2009). The changing economics of DNA synthesis. *Nat. Biotechnol.* 27, 1091–1094. doi:10.1038/nbt1209-1091 Available at: http://www.synthesis.cc/cgi-bin/mt/mt-search.cgi?blog_id=1&tag=CarlsonCurves&limit=20
- Caschera, F., and Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie* 99, 162–168. doi:10.1016/j.biochi.2013.11.025
- Casini, A., Christodoulou, G., Freemont, P. S., Baldwin, G. S., Ellis, T., and MacDonald, J. T. (2014). R2oDNA designer: computational design of biologically neutral synthetic DNA Sequences. *ACS Synth. Biol.* 3, 525–528. doi:10.1021/sb4001323
- Casini, A., MacDonald, J. T., Jonghe, J. D., Christodoulou, G., Freemont, P. S., Baldwin, G. S., et al. (2013). One-pot DNA construction for synthetic biology: the modular overlap-directed assembly with linkers (MODAL) strategy. *Nucleic Acids Res.* 42, e7. doi:10.1093/nar/gkt915
- Cereghino, J. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 24, 45–66. doi:10.1016/s0168-6445(99)00029-7
- Chao, R., Yuan, Y., and Zhao, H. (2014). Recent advances in DNA assembly technologies. *FEMS Yeast Res.* doi:10.1111/1567-1364.12171
- Chappell, J., Jensen, K., and Freemont, P. S. (2013). Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* 41, 3471–3481. doi:10.1093/nar/gkt052
- Chen, I. A., Roberts, R. W., and Szostak, J. W. (2004). The emergence of competition between model protocells. *Science* 305, 1474–1476. doi:10.1126/science.1100757

- Chin, J. X., Chung, B. K., and Lee, D. Y. (2014). Codon optimization online (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* 30, 2210–2212. doi:10.1093/bioinformatics/btu192
- Chiyoda, S., Yamato, K. T., and Kohchi, T. (2014). Plastid transformation of sporelings and suspension-cultured cells from the liverwort *Marchantia polymorpha* L. *Methods Mol. Biol.* 1132, 439–447. doi:10.1007/978-1-62703-995-6_30
- Chizzolini, F., Forlin, M., Cecchi, D., and Mansy, S. S. (2013). Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.* 3, 363–371. doi:10.1021/sb4000977
- Cho, M., Soo Oh, S., Nie, J., Stewart, R., Eisenstein, M., Chambers, J., et al. (2013). Quantitative selection and parallel characterization of aptamers. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18460–18465. doi:10.1073/pnas.1315866110
- Choi, H. J., and Montemagno, C. D. (2005). Artificial organelle: ATP synthesis from cellular mimetic polymersomes. *Nano Lett.* 5, 2538–2542. doi:10.1021/nl051896e
- Choi, S. L., Rha, E., Lee, S. J., Kim, H., Kwon, K., Jeong, Y. S., et al. (2013). Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. *ACS Synth. Biol.* 3, 163–171. doi:10.1021/sb400112u
- Chung, B. K., and Lee, D. Y. (2012). Computational codon optimization of synthetic gene for protein expression. *BMC Syst. Biol.* 6:134. doi:10.1186/1752-0509-6-134
- Cohen, S. N., Chang, A. C., Boyer, H. W., and Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 70, 3240–3244. doi:10.1073/pnas.70.11.3240
- Collins, F. S., and Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature* 505, 612–613. doi:10.1038/505612a
- Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., et al. (2014). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* 42, e23. doi:10.1093/nar/gkt1101
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi:10.1126/science.1231143
- Cooling, M. T., Rouilly, V., Misirli, G., Lawson, J., Yu, T., Hallinan, J., et al. (2010). Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics* 26, 925–931. doi:10.1093/bioinformatics/btq063
- Cregg, J. M., Cereghino, J. L., Shi, J., and Higgins, D. R. (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16, 23–52. doi:10.1385/mb:16:1:23
- Cregg, J. M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., and Chappell, T. (2009). Chapter 13 expression in the yeast *Pichia pastoris*. *Meth. Enzymol.* 463, 169–189. doi:10.1016/s0076-6879(09)63013-5
- Csete, M. E., and Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science* 295, 1664–1669. doi:10.1126/science.1069981
- Danchin, A., and Sekowska, A. (2014). The logic of metabolism and its fuzzy consequences. *Environ. Microbiol.* 16, 19–28. doi:10.1111/1462-2920.12270
- Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* 39, 1131–1141. doi:10.1093/nar/gkq810
- de Kok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., et al. (2014). Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* 3, 97–106. doi:10.1021/sb4001992
- Delebecque, C. J., Lindner, A. B., Silver, P. A., and Aldaye, F. A. (2011). Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333, 470–474. doi:10.1126/science.1206938
- Depaoli, H. C., Borland, A. M., Tuskan, G. A., Cushman, J. C., and Yang, X. (2014). Synthetic biology as it relates to CAM photosynthesis: challenges and opportunities. *J. Exp. Bot.* 65, 3381–3393. doi:10.1093/jxb/eru038
- Dewall, M. T., and Cheng, D. W. (2011). The minimal genome: a metabolic and environmental comparison. *Brief. Funct. Genomics* 10, 312–315. doi:10.1093/bfpg/elt030
- Díaz, M., Herrero, M., García, L. A., and Quirós, C. (2010). Application of flow cytometry to industrial microbial bioprocesses. *Biochem. Eng. J.* 48, 385–407. doi:10.1016/j.bej.2009.07.013
- Douglas, C. M. W., and Stermerding, D. (2014). Challenges for the European governance of synthetic biology for human health. *Life Sci. Soc. Policy* 10, 6. doi:10.1186/s40504-014-0006-7
- Douglas, S. M., Dietz, H., Liedl, T., Hogberg, B., Graf, E., and Shih, W. M. (2009). Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 459, 414–418. doi:10.1038/nature08016
- Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759. doi:10.1038/nbt.1557
- Dutton, P. L., and Moser, C. C. (2011). Engineering enzymes. *Faraday Discuss.* 148, 443. doi:10.1039/c005523a
- Dymond, J., and Boeke, J. (2012). The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng. Bugs* 3, 168–171. doi:10.4161/bbug.19543
- Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. (2014). Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* 32, 97–101. doi:10.1038/nbt.2714
- Ellington, A. D., and Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature* 346, 818–822. doi:10.1038/346818a0
- Ellis, B. L., Hirsch, M. L., Porter, S. N., Samulski, R. J., and Porteus, M. H. (2013). Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by food and drug administration-approved drugs. *Gene Ther.* 20, 35–42. doi:10.1038/gt.2011.211
- Ellis, T., Adie, T., and Baldwin, G. S. (2011). DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. (Camb.)* 3, 109–118. doi:10.1039/c0ib00070a
- Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi:10.1038/35002125
- Endy, D. (2005). Foundations for engineering biology. *Nature* 438, 449–453. doi:10.1038/nature04342
- Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3:e3647. doi:10.1371/journal.pone.0003647
- Engler, C., and Marillonnet, S. (2011). Generation of families of construct variants using golden gate shuffling. *Methods Mol. Biol.* 729, 167–181. doi:10.1007/978-1-61779-065-2_11
- Engler, C., and Marillonnet, S. (2013). Combinatorial DNA assembly using golden gate cloning. *Methods Mol. Biol.* 1073, 141–156. doi:10.1007/978-1-62703-625-2_12
- Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature* 472, 499–503. doi:10.1038/nature09929
- Fletcher, J. M., Harniman, R. L., Barnes, F. R., Boyle, A. L., Collins, A., Mantell, J., et al. (2013). Self-assembling cages from coiled-coil peptide modules. *Science* 340, 595–599. doi:10.1126/science.1233936
- Frow, E., and Calvert, J. (2013). Opening up the future(s) of synthetic biology. *Futures* 48, 32–43. doi:10.1016/j.futures.2013.03.001
- Fu, J., Liu, M., Liu, Y., Woodbury, N. W., and Yan, H. (2012). Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *J. Am. Chem. Soc.* 134, 5516–5519. doi:10.1021/ja300897h
- Fu, J., Yang, Y. R., Johnson-Buck, A., Liu, M., Liu, Y., Walter, N. G., et al. (2014). Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* 9, 531–536. doi:10.1038/nnano.2014.100
- Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., et al. (2014). The synthetic biology open language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* 32, 545–550. doi:10.1038/nbt.2891
- Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M., and Gennari, J. H. (2011). Standard biological parts knowledgebase. *PLoS ONE* 6:e17005. doi:10.1371/journal.pone.0017005
- Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muniz-Rascado, L., Solano-Lira, H., et al. (2011). RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Sensor Units). *Nucleic Acids Res.* 39, D98–D105. doi:10.1093/nar/gkq1110
- Gan, R., and Jewett, M. C. (2014). A combined cell-free transcription-translation system from *Saccharomyces cerevisiae* for rapid and robust protein synthesis. *Biotechnol. J.* 9, 641–651. doi:10.1002/biot.201300545
- Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi:10.1038/35002131
- Gibson, D. G. (2011). Enzymatic assembly of overlapping DNA fragments. *Meth. Enzymol.* 498, 349–361. doi:10.1016/B978-0-12-385120-8.00015-2
- Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., et al. (2008). Complete chemical synthesis, assembly,

- and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220. doi:10.1126/science.1151721
- Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010a). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi:10.1126/science.1190719
- Gibson, D. G., Smith, H. O., Hutchison, C. A. III, Venter, J. C., and Merryman, C. (2010b). Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* 7, 901–903. doi:10.1038/nmeth.1515
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318
- Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., et al. (2006). Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 103, 425–430. doi:10.1073/pnas.0510013103
- Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013). Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* 41, e156. doi:10.1093/nar/gkt605
- Hamilton, S. R., and Gerngross, T. U. (2007). Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr. Opin. Biotechnol.* 18, 387–392. doi:10.1016/j.copbio.2007.09.001
- Hammer, D. A., and Kamat, N. P. (2012). Towards an artificial cell. *FEBS Lett.* 586, 2882–2890. doi:10.1016/j.febslet.2012.07.044
- Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y., and Yan, H. (2011). DNA origami with complex curvatures in three-dimensional space. *Science* 332, 342–346. doi:10.1126/science.1202998
- Hansen, A. S., and O’Shea, E. K. (2013). Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.* 9, 704. doi:10.1038/msb.2013.56
- Harwood, C. R., Pohl, S., Smith, W., and Wipat, A. (2013). *Bacillus subtilis*: model Gram-positive synthetic biology chassis. *Comput. Sci.* 40, 87–117. doi:10.1016/b978-0-12-417029-2.00004-2
- Heider, S. A. E., Wolf, N., Hofemeier, A., Peters-Wendisch, P., and Wendisch, V. F. (2014). Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum*. *Front. Bioeng. Biotechnol.* 2:28. doi:10.3389/fbioe.2014.00028
- Heidorn, T., Camsund, D., Huang, H. H., Lindberg, P., Oliveira, P., Stensjo, K., et al. (2011). Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Meth. Enzymol.* 497, 539–579. doi:10.1016/B978-0-12-385075-1.00024-X
- Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat. Methods* 11, 122–123. doi:10.1038/nmeth.2812
- Heinemann, M., and Panke, S. (2006). Synthetic biology – putting engineering into biology. *Bioinformatics* 22, 2790–2799. doi:10.1093/bioinformatics/btl469
- Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012). j5 DNA assembly design automation software. *ACS Synth. Biol.* 1, 14–21. doi:10.1021/sb2000116
- Hiroe, A., Tsuge, K., Nomura, C. T., Itaya, M., and Tsuge, T. (2012). Rearrangement of gene order in the phaCAB operon leads to effective production of ultrahigh-molecular-weight poly[(R)-3-hydroxybutyrate] in genetically engineered *Escherichia coli*. *Appl. Environ. Microbiol.* 78, 3177–3184. doi:10.1128/AEM.07715-11
- Hodgman, C. E., and Jewett, M. C. (2013). Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. *Biotechnol. Bioeng.* 110, 2643–2654. doi:10.1002/bit.24942
- Hong, S. H., Kwon, Y.-C., and Jewett, M. C. (2014). Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. *Front. Chem.* 2:34. doi:10.3389/fchem.2014.00034
- Howorka, S. (2011). Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* 22, 485–491. doi:10.1016/j.copbio.2011.05.003
- Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., et al. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333, 348–353. doi:10.1126/science.1205822
- Itaya, M., Fujita, K., Kuroki, A., and Tsuge, K. (2008). Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods* 5, 41–43. doi:10.1038/nmeth1143
- Iwata, T., Kaneko, S., Shiwa, Y., Enomoto, T., Yoshikawa, H., and Hirota, J. (2013). *Bacillus subtilis* genome vector-based complete manipulation and reconstruction of genomic DNA for mouse transgenesis. *BMC Genomics* 14:300. doi:10.1186/1471-2164-14-300
- Iyer, S., Karig, D. K., Norred, S. E., Simpson, M. L., and Doktycz, M. J. (2013). Multi-input regulation and logic with T7 promoters in cells and cell-free systems. *PLoS ONE* 8:e78442. doi:10.1371/journal.pone.0078442
- Jefferson, C., Lentzos, F., and Marris, C. (2014). Synthetic biology and biosecurity: challenging the “myths”. *Front. Public Health* 2:115. doi:10.3389/fpubh.2014.00115
- Jenison, R. D., Gill, S. C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science* 263, 1425–1429. doi:10.1126/science.7510417
- Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., et al. (2008). De novo computational design of retro-aldol enzymes. *Science* 319, 1387–1391. doi:10.1126/science.1152692
- Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. (2012). Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res.* 40, 3763–3774. doi:10.1093/nar/gkr1191
- Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., et al. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* 3, 4. doi:10.1186/1754-1611-3-4
- Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* 9, 701. doi:10.1038/msb.2013.59
- Khalil, A. S., and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat. Rev. Genet.* 11, 367–379. doi:10.1038/nrg2775
- King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., et al. (2014). Accurate design of co-assembling multi-component protein nanomaterials. *Nature* 510, 103–108. doi:10.1038/nature13404
- King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., Andre, L., et al. (2012). Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 336, 1171–1174. doi:10.1126/science.1219364
- Kitney, R., and Freemont, P. (2012). Synthetic biology – the state of play. *FEBS Lett.* 586, 2029–2036. doi:10.1016/j.febslet.2012.06.002
- Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., et al. (2012). Principles for designing ideal protein structures. *Nature* 491, 222–227. doi:10.1038/nature11600
- Kouprina, N., and Larionov, V. (2008). Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* 3, 371–377. doi:10.1038/nprot.2008.5
- Landrain, T., Meyer, M., Perez, A. M., and Sussan, R. (2013). Do-it-yourself biology: challenges and promises for an open science and technology movement. *Syst. Synth. Biol.* 7, 115–126. doi:10.1007/s11693-013-9116-4
- Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W. F., et al. (2014). Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* 3, 454–465. doi:10.1021/sb4001118
- Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth. Enzymol.* 487, 545–574. doi:10.1016/B978-0-12-381270-4.00019-6
- Li, M. Z., and Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* 4, 251–256. doi:10.1038/nmeth1010
- Liang, J. C., Bloom, R. J., and Smolke, C. D. (2011). Engineering biological systems with synthetic RNA molecules. *Mol. Cell* 43, 915–926. doi:10.1016/j.molcel.2011.08.023
- Lienert, F., Lohmueller, J. J., Garg, A., and Silver, P. A. (2014). Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat. Rev. Mol. Cell Biol.* 15, 95–107. doi:10.1038/nrm3738
- Lin, B., and Levchenko, A. (2012). Microfluidic technologies for studying synthetic circuits. *Curr. Opin. Chem. Biol.* 16, 307–317. doi:10.1016/j.cbpa.2012.04.012
- Lin, Q., Jia, B., Mitchell, L. A., Luo, J., Yang, K., Zeller, K. I., et al. (2014). RADOM, an efficient in vivo method for assembling designed DNA fragments up to 10 kb long in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* doi:10.1021/sb500241e
- Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J. D., and Hillson, N. J. (2013). PaR-PaR laboratory automation platform. *ACS Synth. Biol.* 2, 216–222. doi:10.1021/sb300075t
- Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., Mutalik, V. K., et al. (2012). An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods* 9, 1088–1094. doi:10.1038/nmeth.2184

- Lobban, P. E., and Kaiser, A. D. (1973). Enzymatic end-to end joining of DNA molecules. *J. Mol. Biol.* 78, 453–471. doi:10.1016/0022-2836(73)90468-3
- Lou, C., Stanton, B., Chen, Y. J., Munsky, B., and Voigt, C. A. (2012). Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* 30, 1137–1142. doi:10.1038/nbt.2401
- Lu, W.-C., and Ellington, A. D. (2014). Design and Selection of a Synthetic Operon. *ACS Synth. Biol.* 3, 410–415. doi:10.1021/sb400160m
- MacDonald, J. T., Barnes, C., Kitney, R. I., Freemont, P. S., and Stan, G. B. (2011). Computational design approaches and tools for synthetic biology. *Integr. Biol. (Camb.)* 3, 97–108. doi:10.1039/c0ib00077a
- Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J. K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2623–2628. doi:10.1073/pnas.1019533108
- Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838. doi:10.1038/nbt.2675
- Marchisio, M. A. (2014). In silico design and in vivo implementation of yeast gene Boolean gates. *J. Biol. Eng.* 8, 6. doi:10.1186/1754-1611-8-6
- Marchisio, M. A., and Stelling, J. (2009). Computational design tools for synthetic biology. *Curr. Opin. Biotechnol.* 20, 479–485. doi:10.1016/j.copbio.2009.08.007
- Markham, N. R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.* 453, 3–31. doi:10.1007/978-1-60327-429-6_1
- Marples, B., Scott, S. D., Hendry, J. H., Embleton, M. J., Lashford, L. S., and Margison, G. P. (2000). Development of synthetic promoters for radiation-mediated gene therapy. *Gene Ther.* 7, 511–517. doi:10.1038/sj.gt.3301116
- Marris, C., and Rose, N. (2010). Open engagement: exploring public participation in the biosciences. *PLoS Biol.* 8:e1000549. doi:10.1371/journal.pbio.1000549
- Moe-Behrens, G. H., Davis, R., and Haynes, K. A. (2013). Preparing synthetic biology for the world. *Front. Microbiol.* 4:5. doi:10.3389/fmicb.2013.00005
- Morbitzer, R., Romer, P., Boch, J., and Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21617–21622. doi:10.1073/pnas.1013133107
- Moscou, M. J., and Bogdanov, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501. doi:10.1126/science.1178817
- Muller, K., Siegel, D., Rodriguez Jahnke, F., Gerrer, K., Wend, S., Decker, E. L., et al. (2014a). A red light-controlled synthetic gene expression switch for plant systems. *Mol. Biosyst.* 10, 1679–1688. doi:10.1039/c3mb70579j
- Muller, K., Zurbriggen, M. D., and Weber, W. (2014b). Control of gene expression using a red- and far-red light-responsive bi-stable toggle switch. *Nat. Protoc.* 9, 622–632. doi:10.1038/nprot.2014.038
- Murphy, K. C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.* 180, 2063–2071.
- Murphy, K. C., and Campellone, K. G. (2003). Lambda red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.* 4:11. doi:10.1186/1471-2199-4-11
- Mushtaq, M. Y., Verpoorte, R., and Kim, H. K. (2013). Zebrafish as a model for systems biology. *Biotechnol. Genet. Eng. Rev.* 29, 187–205. doi:10.1080/02648725.2013.801238
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., et al. (2013a). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360. doi:10.1038/nmeth.2404
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., et al. (2013b). Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* 10, 347–353. doi:10.1038/nmeth.2403
- Na, D., and Lee, D. (2010). RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* 26, 2633–2634. doi:10.1093/bioinformatics/btq458
- Narayanan, K., and Chen, Q. (2011). Bacterial artificial chromosome mutagenesis using recombineering. *J. Biomed. Biotechnol.* 2011, 971296. doi:10.1155/2011/971296
- Nirenberg, M. W., and Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 47, 1588–1602. doi:10.1073/pnas.47.10.1588
- Nishiyama, T. (2000). Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Res.* 7, 9–17. doi:10.1093/dnares/7.1.9
- O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. O. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* 9, 693. doi:10.1038/msb.2013.52
- Ohtani, N., Hasegawa, M., Sato, M., Tomita, M., Kaneko, S., and Itaya, M. (2012). Serial assembly of Thermus megaplasmid DNA in the genome of *Bacillus subtilis* 168: a BAC-based domino method applied to DNA with a high GC content. *Biotechnol. J.* 7, 867–876. doi:10.1002/biot.201100396
- Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R., and Tabor, J. J. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* 11, 449–455. doi:10.1038/nmeth.2884
- O'Reilly, L. P., Luke, C. J., Perlmutter, D. H., Silverman, G. A., and Pak, S. C. (2014). *C. elegans* in high-throughput drug discovery. *Adv. Drug Deliv. Rev.* 6, 247–253. doi:10.1016/j.addr.2013.12.001
- Padilla, J. E., Colovos, C., and Yeates, T. O. (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2217–2221. doi:10.1073/pnas.041614998
- Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011). RNA mimics of green fluorescent protein. *Science* 333, 642–646. doi:10.1126/science.1207339
- Penchovsky, R., and Breaker, R. R. (2005). Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nat. Biotechnol.* 23, 1424–1433. doi:10.1038/nbt1155
- Piyasena, M. E., and Graves, S. W. (2014). The intersection of flow cytometry with microfluidics and microfabrication. *Lab. Chip* 14, 1044–1059. doi:10.1039/c3lc51152a
- Pothoulakis, G., Ceroni, F., Reeve, B., and Ellis, T. (2014). The spinach RNA aptamer as a characterization tool for synthetic biology. *ACS Synth. Biol.* 3, 182–187. doi:10.1021/sb400089c
- Preston, B. (2013). Synthetic biology as red herring. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 649–659. doi:10.1016/j.shpsc.2013.05.012
- Purnick, P. E., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422. doi:10.1038/nrm2698
- Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A., and Arkin, A. P. (2012). RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* 30, 1002–1006. doi:10.1038/nbt.2355
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi:10.1016/j.cell.2013.02.022
- Quan, J., and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* 4:e6441. doi:10.1371/journal.pone.0006441
- Quan, J., and Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* 6, 242–251. doi:10.1038/nprot.2010.181
- Quinn, J., Beal, J., Bhatia, S., Cai, P., Chen, J., Clancy, K., et al. (2013). Synthetic biology open language visual (SBOL visual), version 1.0.0, BBF RFC #93. doi:1721.1/78249
- Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Durr, F., Emenegger, J., et al. (2013). The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J. Biol. Eng.* 7, 29. doi:10.1186/1754-1611-7-29
- Rajagopal, K., and Schneider, J. P. (2004). Self-assembling peptides and proteins for nanotechnological applications. *Curr. Opin. Struct. Biol.* 14, 480–486. doi:10.1016/j.sbi.2004.06.006
- Redemann, S., Schloissnig, S., Ernst, S., Pozniakowsky, A., Ayloo, S., Hyman, A. A., et al. (2011). Codon adaptation-based control of protein expression in *C. elegans*. *Nat. Methods* 8, 250–252. doi:10.1038/nmeth.1565
- Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014). Predicting translation initiation rates for designing synthetic biology. *Front. Bioeng. Biotechnol.* 2:1. doi:10.3389/fbioe.2014.00001
- Rhodius, V. A., and Mutalik, V. K. (2010). Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2854–2859. doi:10.1073/pnas.0915066107
- Rhodius, V. A., Mutalik, V. K., and Gross, C. A. (2012). Predicting the strength of UP-elements and full-length *E. coli* sigmaE promoters. *Nucleic Acids Res.* 40, 2907–2924. doi:10.1093/nar/gkr1190
- Rodrigo, G., and Jaramillo, A. (2013). AutoBioCAD: full biodesign automation of genetic circuits. *ACS Synth. Biol.* 2, 230–236. doi:10.1021/sb300084h

- Roehner, N., and Myers, C. J. (2013). A methodology to annotate systems biology markup language models with the synthetic biology open language. *ACS Synth. Biol.* 3, 57–66. doi:10.1021/sb400066m
- Rokke, G., Korvald, E., Pahr, J., Oyas, O., and Lale, R. (2014). BioBrick assembly standards and techniques and associated software tools. *Methods Mol. Biol.* 1116, 1–24. doi:10.1007/978-1-62703-764-8_1
- Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013). Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA* 19, 670–684. doi:10.1261/rna.037747.112
- Roth, A., and Breaker, R. R. (2009). The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* 78, 305–334. doi:10.1146/annurev.biochem.78.070507.135656
- Rothmund, P. W. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature* 440, 297–302. doi:10.1038/nature04586
- Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., Dechance, J., Betker, J., et al. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* 453, 190–195. doi:10.1038/nature06879
- Rouillard, J. M., Lee, W., Truan, G., Gao, X., Zhou, X., and Gulari, E. (2004). Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res.* 32, W176–W180. doi:10.1093/nar/gkh401
- Salis, H. M. (2011). The ribosome binding site calculator. *Meth. Enzymol.* 498, 19–42. doi:10.1016/B978-0-12-385120-8.00002-4
- Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950. doi:10.1038/nbt.1568
- Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi:10.1038/nbt.2842
- Schaefer, D. G., and Zryd, J.-P. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* 11, 1195–1206. doi:10.1046/j.1365-313X.1997.11061195.x
- Schark, M. (2012). Synthetic biology and the distinction between organisms and machines. *Environ. Values* 21, 19–41. doi:10.3197/096327112x13225063227943
- Sefah, K., Yang, Z., Bradley, K. M., Hoshika, S., Jimenez, E., Zhang, L., et al. (2013). In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1449–1454. doi:10.1073/pnas.1311778111
- Seo, S. W., Yang, J. S., Kim, I., Yang, J., Min, B. E., Kim, S., et al. (2013). Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* 15, 67–74. doi:10.1016/j.ymben.2012.10.006
- Sharma, N., Jung, C. H., Bhalla, P. L., and Singh, M. B. (2014). RNA sequencing analysis of the gametophyte transcriptome from the liverwort, *Marchantia polymorpha*. *PLoS One* 9:e97497. doi:10.1371/journal.pone.0097497
- Shi, Z., Wedd, A. G., and Gras, S. L. (2013). Parallel in vivo DNA assembly by recombination: experimental demonstration and theoretical approaches. *PLoS ONE* 8:e56854. doi:10.1371/journal.pone.0056854
- Shin, J., and Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* 1, 29–41. doi:10.1021/sb200016s
- Shuler, M. L., Foley, P., and Atlas, J. (2012). Modeling a minimal cell. *Methods Mol. Biol.* 881, 573–610. doi:10.1007/978-1-61779-827-6_20
- Silva-Rocha, R., Martinez-Garcia, E., Calles, B., Chavarría, M., Arce-Rodriguez, A., De Las Heras, 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. doi:10.1093/nar/gks1119
- Sinclair, J. C., Davies, K. M., Venien-Bryan, C., and Noble, M. E. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* 6, 558–562. doi:10.1038/nnano.2011.122
- Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S. F., Hartley, J. L., and Chatterjee, D. K. (2004). A novel cell-free protein synthesis system. *J. Biotechnol.* 110, 257–263. doi:10.1016/j.jbiotec.2004.02.014
- Song, C. W., Lee, J., and Lee, S. Y. (2014). Genome engineering and gene expression control for bacterial strain development. *Biotechnol. J.* doi:10.1002/biot.201400057
- Song, Y., Sauret, A., and Cheung Shum, H. (2013). All-aqueous multiphase microfluidics. *Biomicrofluidics* 7, 61301. doi:10.1063/1.4827916
- Speer, M. A., and Richard, T. L. (2011). Amplified insert assembly: an optimized approach to standard assembly of BioBrick™ genetic circuits. *J. Biol. Eng.* 5, 17. doi:10.1186/1754-1611-5-17
- Stanton, B. C., Nielsen, A. A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* 10, 99–105. doi:10.1038/nchembio.1411
- Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–519. doi:10.1038/nature07389
- Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* 32, 1610–1614. doi:10.1093/nar/gkh321
- Sukhorukov, G. B., Rogach, A. L., Zebli, B., Liedl, T., Skirtach, A. G., Kohler, K., et al. (2005). Nanoengineered polymer capsules: tools for detection, controlled delivery, and site-specific manipulation. *Small* 1, 194–200. doi:10.1002/smll.200400075
- Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013a). Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 79:e50762. doi:10.3791/50762
- Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2013b). Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397. doi:10.1021/sb400131a
- Szeto, K., Reinholt, S. J., Duarte, F. M., Pagano, J. M., Ozer, A., Yao, L., et al. (2014). High-throughput binding characterization of RNA aptamer selections using a microplate-based multiplex microcolumn device. *Anal. Bioanal. Chem.* 406, 2727–2732. doi:10.1007/s00216-014-7661-7
- Tang, J., and Breaker, R. R. (1997). Rational design of allosteric ribozymes. *Chem. Biol.* 4, 453–459. doi:10.1016/S1074-5521(97)90197-6
- Tantillo, D. J., Chen, J., and Houk, K. N. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* 2, 743–750. doi:10.1016/S1367-5931(98)80112-9
- Tomlinson, M. L., Hendry, A. E., and Wheeler, G. N. (2012). Chemical genetics and drug discovery in *Xenopus*. *Methods Mol. Biol.* 917, 155–166. doi:10.1007/978-1-61779-992-1_9
- Torella, J. P., Boehm, C. R., Lienert, F., Chen, J. H., Way, J. C., and Silver, P. A. (2014a). Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res.* 42, 681–689. doi:10.1093/nar/gkt860
- Torella, J. P., Lienert, F., Boehm, C. R., Chen, J. H., Way, J. C., and Silver, P. A. (2014b). Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat. Protoc.* 9, 2075–2089. doi:10.1038/nprot.2014.145
- Tracy, B. P., Gaida, S. M., and Papoutsakis, E. T. (2010). Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Curr. Opin. Biotechnol.* 21, 85–99. doi:10.1016/j.copbio.2010.02.006
- Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., and French, C. E. (2014). Paper-Clip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res.* 42, e154. doi:10.1093/nar/gku829
- Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510. doi:10.1126/science.2200121
- Uchida, M., Klem, M. T., Allen, M., Suci, P., Flenniken, M., Gilitzer, E., et al. (2007). Biological containers: protein cages as multifunctional nanoplatforms. *Adv. Mater. Weinheim* 19, 1025–1042. doi:10.1002/adma.200601168
- Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7:285. doi:10.1186/1471-2105-7-285
- Vogl, T., Hartner, F. S., and Glieder, A. (2013). New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. *Curr. Opin. Biotechnol.* 24, 1094–1101. doi:10.1016/j.copbio.2013.02.024
- Walsh, G. (2010). Post-translational modifications of protein biopharmaceuticals. *Drug Discov. Today* 15, 773–780. doi:10.1016/j.drudis.2010.06.009
- Walter, N. G., and Bustamante, C. (2014). Introduction to single molecule imaging and mechanics: seeing and touching molecules one at a time. *Chem. Rev.* 114, 3069–3071. doi:10.1021/cr500059w
- Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi:10.1038/nature08187
- Wang, Y., and Zhang, Y. H. (2009). Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnol.* 9:58. doi:10.1186/1472-6750-9-58

- Way, J. C., Collins, J. J., Keasling, J. D., and Silver, P. A. (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
- White, R. M., Cech, J., Ratanasirinrawoot, S., Lin, C. Y., Rahl, P. B., Burke, C. J., et al. (2011). DHODH modulates transcriptional elongation in the neural crest and melanoma. *Nature* 471, 518–522. doi:10.1038/nature09882
- Wittmann, A., and Suess, B. (2012). Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* 586, 2076–2083. doi:10.1016/j.febslet.2012.02.038
- Woolfson, D. N., and Mahmoud, Z. N. (2010). More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem. Soc. Rev.* 39, 3464–3479. doi:10.1039/c0cs00032a
- Wright, O., Delmans, M., Stan, G. B., and Ellis, T. (2014). GeneGuard: a modular plasmid system designed for biosafety. *ACS Synth. Biol.* doi:10.1021/sb500234s
- Wright, O., Stan, G. B., and Ellis, T. (2013). Building-in biosafety for synthetic biology. *Microbiology* 159, 1221–1235. doi:10.1099/mic.0.066308-0
- Xu, J., Sigworth, F. J., and Lavan, D. A. (2010). Synthetic protocells to mimic and test cell function. *Adv. Mater.* 22, 120–127. doi:10.1002/adma.200901945
- Xu, P., Vansiri, A., Bhan, N., and Koffas, M. A. (2012). ePathBrick: a synthetic biology platform for engineering metabolic pathways in *E. coli*. *ACS Synth. Biol.* 1, 256–266. doi:10.1021/sb300016b
- Yang, J., and Reth, M. (2012). Drosophila S2 Schneider cells: a useful tool for rebuilding and redesigning approaches in synthetic biology. *Methods Mol. Biol.* 813, 331–341. doi:10.1007/978-1-61779-412-4_20
- Ye, H., Aubel, D., and Fussenegger, M. (2013). Synthetic mammalian gene circuits for biomedical applications. *Curr. Opin. Chem. Biol.* 17, 910–917. doi:10.1016/j.cbpa.2013.10.006
- Zaccai, N. R., Chi, B., Thomson, A. R., Boyle, A. L., Bartlett, G. J., Bruning, M., et al. (2011). A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* 7, 935–941. doi:10.1038/nchembio.692
- Zadeh, J. N., Wolfe, B. R., and Pierce, N. A. (2011). Nucleic acid sequence design via efficient ensemble defect optimization. *J. Comput. Chem.* 32, 439–452. doi:10.1002/jcc.21633
- Zaghoul, T., and Doi, R. (1987). In vitro expression of a Tn9-derived chloramphenicol acetyltransferase gene fusion by using a *Bacillus subtilis* system. *J. Bacteriol.* 169, 1212–1216.
- Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., et al. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* 15, 2785–2794. doi:10.1110/ps.062353106
- Zhang, F., and Keasling, J. (2011). Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol.* 19, 323–329. doi:10.1016/j.tim.2011.05.003
- Zhang, Y., Werling, U., and Edelman, W. (2012). SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* 40, e55. doi:10.1093/nar/gkr1288
- Zhao, Y., Wang, S., and Zhu, J. (2011). A multi-step strategy for BAC recombineering of large DNA fragments. *Int J Biochem Mol Biol* 2, 199–206. Available from <http://www.ijbmb.org/files/IJBMB1104005.pdf>
- Zimmerman, S. B., Little, J. W., Oshinsky, C. K., and Gellert, M. (1967). Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proc. Natl. Acad. Sci. U.S.A.* 57, 1841–1848. doi:10.1073/pnas.57.6.1841
- Zuleta, I. A., Aranda-Diaz, A., Li, H., and El-Samad, H. (2014). Dynamic characterization of growth and gene expression using high-throughput automated flow cytometry. *Nat. Methods* 11, 443–448. doi:10.1038/nmeth.2879

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 September 2014; accepted: 12 November 2014; published online: 10 December 2014.

Citation: Kelwick R, MacDonald JT, Webb AJ and Freemont P (2014) Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* 2:60. doi: 10.3389/fbioe.2014.00060

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Kelwick, MacDonald, Webb and Freemont. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Production of fatty acid-derived valuable chemicals in synthetic microbes

Ai-Qun Yu^{1,2}, Nina Kurniasih Pratomo Juwono^{1,2}, Susanna Su Jan Leong^{1,2,3} and Matthew Wook Chang^{1,2*}

¹ Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

² Synthetic Biology Research Program, National University of Singapore, Singapore, Singapore

³ Singapore Institute of Technology, Singapore, Singapore

Edited by:

Jean Marie François, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés UMR-CNRS 5504, France

Reviewed by:

Zongbao K. Zhao, Chinese Academy of Sciences, China
Taek Soon Lee, Lawrence Berkeley National Laboratory, USA

*Correspondence:

Matthew Wook Chang, Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 28 Medical Drive, 117456 Singapore
e-mail: bchcmw@nus.edu.sg

Fatty acid derivatives, such as hydroxy fatty acids, fatty alcohols, fatty acid methyl/ethyl esters, and fatty alka(e)nes, have a wide range of industrial applications including plastics, lubricants, and fuels. Currently, these chemicals are obtained mainly through chemical synthesis, which is complex and costly, and their availability from natural biological sources is extremely limited. Metabolic engineering of microorganisms has provided a platform for effective production of these valuable biochemicals. Notably, synthetic biology-based metabolic engineering strategies have been extensively applied to refactor microorganisms for improved biochemical production. Here, we reviewed: (i) the current status of metabolic engineering of microbes that produce fatty acid-derived valuable chemicals, and (ii) the recent progress of synthetic biology approaches that assist metabolic engineering, such as mRNA secondary structure engineering, sensor-regulator system, regulatable expression system, ultrasensitive input/output control system, and computer science-based design of complex gene circuits. Furthermore, key challenges and strategies were discussed. Finally, we concluded that synthetic biology provides useful metabolic engineering strategies for economically viable production of fatty acid-derived valuable chemicals in engineered microbes.

Keywords: synthetic biology, metabolic engineering, fatty acid biosynthesis pathway, biochemical production, *Escherichia coli*, yeast

INTRODUCTION

Fatty acids are one of the major components found in all organisms, usually in the intracellular forms of fatty acyl–acyl carrier protein (acyl-ACP), fatty acyl-coenzyme A ester (acyl-CoA), storage lipids, eicosanoids, and unesterified free fatty acids. In industry, applications of free fatty acids are generally limited due to the ionic nature of their carboxyl group (Peralta-Yahya et al., 2012). Comparatively, fatty acid derivatives have wider applications such as biofuels, biomaterials, and other biochemicals (Lennen and Pfleger, 2013; Rungtaphan and Keasling, 2014).

The low abundance or yield of fatty acid-derived chemicals in organisms renders their isolation from natural sources non-economically viable. The synthesis of fatty acid derivatives by chemical means also suffers from low efficiency and often requires harsh reaction conditions, prolonged times, and expensive equipment footprint (Song et al., 2013). The production of fatty acid-derived chemicals by engineering microbial cells into microbial factories is becoming an attractive alternative approach that can overcome the aforementioned bottlenecks associated with the other synthesis routes (Keasling and Chou, 2008; Schirmer et al., 2010; Lee et al., 2012).

To date, synthetic enzymatic pathways that lead to the production of fatty acid-derived valuable chemicals including fatty alkanes, fatty acid methyl/ethyl esters, fatty alcohols, hydroxy fatty acids, and lactones have been constructed in microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. However,

it remains a challenge to achieve high yield, titer, and productivity of these fatty acid-derived chemicals. The major challenges faced in maximizing product titer are associated with: (1) improving the low enzyme activity of an entire metabolic pathway, (2) increasing the inadequate tolerance of the used microorganisms toward toxic target compounds, (3) recycling or replacing insufficient cofactors for enzymatic reactions, (4) enriching precursors and eliminating byproducts, and (5) optimizing and balancing the fluxes of whole metabolic networks to reduce burden on the host, and remove negative feedback regulation. Recently, advanced synthetic biology approaches have provided potential to address these challenging problems in re-engineering microbial systems for fatty acid-derived chemicals production (Clomburg and Gonzalez, 2010; Siddiqui et al., 2012; Zhang et al., 2012a), which narrows the gap toward realizing full-scale commercialization and industrialization of this manufacturing route.

In this review, we focus on the recent progress in metabolic engineering efforts to convert fatty acids to valuable chemicals using microbes as hosts, and advancement in synthetic biology approaches for further optimizing biochemical production in microbial biofactories.

METABOLISMS OF FATTY ACIDS IN ORGANISMS

Fatty acids are an integral part of all living organisms, and are generally composed of a hydrophobic hydrocarbon chain ending in one hydrophilic carboxylic acid functional group. The metabolic

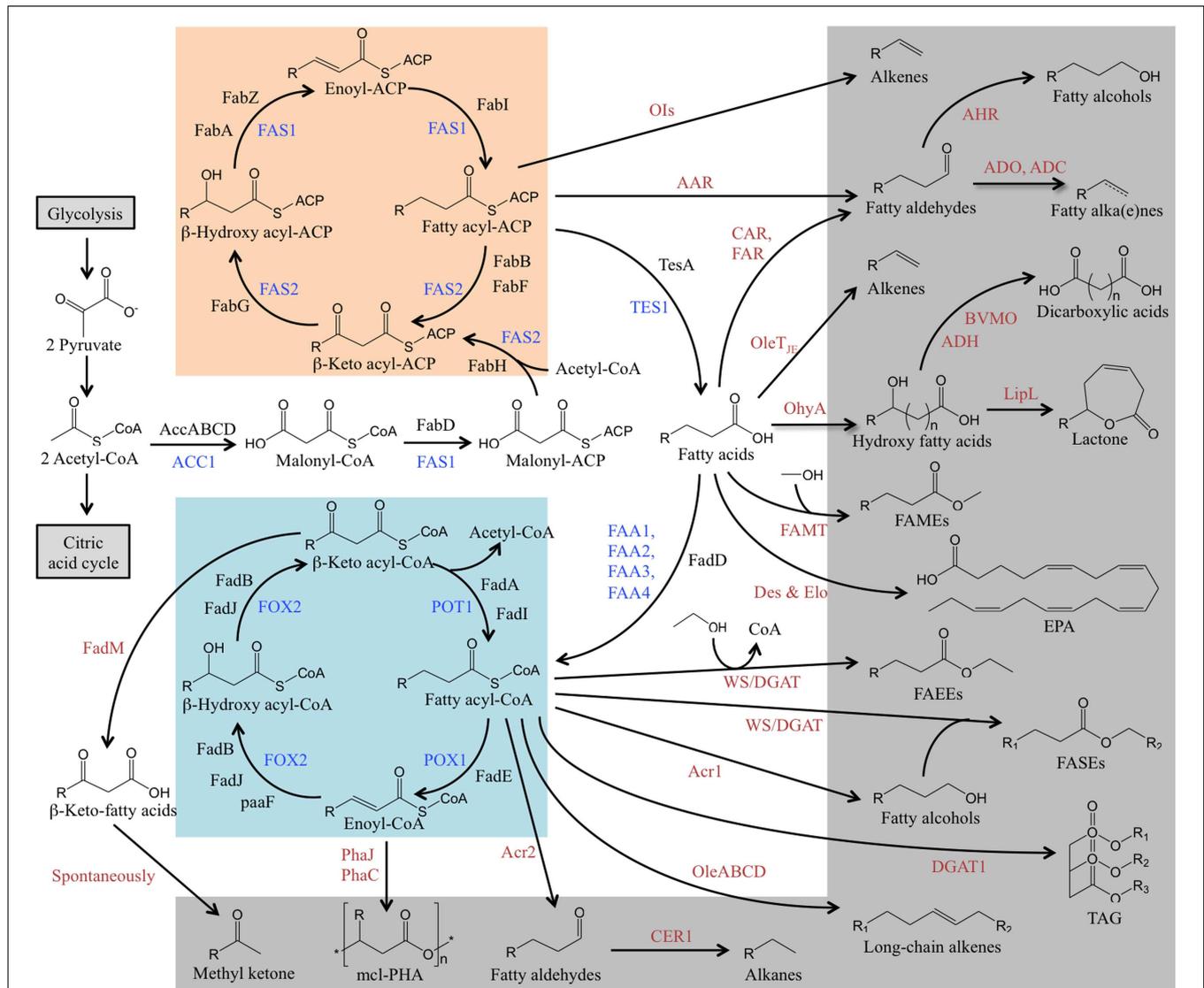


FIGURE 1 | Overview of metabolic pathways that lead to the production of fatty acids and fatty acid-derived chemicals.

The fatty acid biosynthesis (orange), β -oxidation cycle (blue), and the biosynthesis pathway of fatty acid-derived chemicals (gray) are presented. The enzymes of fatty acid metabolism in *S. cerevisiae* is in blue, in *E. coli* is in black, and the enzymes for conversion of fatty acids to their derivatives from other organisms is in red. AAR, acyl-ACP reductase; ACC1, acetyl-CoA carboxylase; AccABCD, a four subunits, biotin carboxyl carrier protein (AccB), biotin carboxylase (AccC), and acetyl-CoA carboxyltransferase (AccA, AccD); Acr1 & Acr2, acyl-CoA reductase; ADO, aldehyde decarbonylase; ADH, alcohol dehydrogenase; AHR, aldehyde-deformylating oxygenase; AOR, aldehyde reductase; BVMO, Baeyer-Villiger mono-oxygenase; CAR, carboxylic acid reductase; CER1, fatty aldehyde decarbonylase; Des, fatty acid desaturase; DGAT1, acyl-CoA:diacylglycerol acyltransferase; Elo, fatty acid elongase; FAA1 & FAA4, long-chain fatty acyl-CoA synthetase; FAA2 & FAA3, fatty acyl-CoA synthetase; FabA & FabZ, β -hydroxy acyl-ACP dehydratase; FabB, β -keto acyl-ACP synthase I; FabD, malonyl-CoA:ACP transacylase; FabF, β -keto

acyl-ACP synthase II; FabG, β -keto acyl-ACP reductase; FabH, β -keto acyl-ACP synthase III; FabI, enoyl acyl-ACP reductase; FadA & FadI, β -keto acyl-CoA thiolase; FadB & FadJ, enoyl-CoA hydratase/ β -hydroxy acyl-CoA dehydrogenase; FadD, fatty acyl-CoA synthase; FadE, acyl-CoA dehydrogenase; FadM, long-chain acyl-CoA thioesterase III; FADMT, fatty acid methyltransferase; FAR, fatty acid reductase; FAS1, acyl-CoA:ACP transferase/ β -hydroxyl acyl-ACP dehydratase/acyl-ACP reductase; FAS2, acyl-ACP synthase/ β -keto acyl-ACP synthase; FOX2, enoyl-CoA hydratase/ β -hydroxyl acyl-CoA dehydrogenase; LipL, lactonizing lipase; OhyA, oleate hydratase; OleABCD, a four protein families for long-chain olefin biosynthesis; OleT_{EF}, *Jeotgalicoccus* sp terminal olefin-forming fatty acid decarboxylase; PaaF, 2,3-dehydroadipyl-CoA hydratase; PhaJ & PhaC, polyhydroxyalkanoate (PHA) synthases to yield medium-chain length polyester (mcl-PHA); POX1, fatty acyl-CoA oxidase; POT1, β -keto acyl-CoA thiolase; TE, acyl-ACP thioesterase; WS/DGAT, wax ester synthase/acyl-CoA:diacylglycerol acyltransferase.

pathway of fatty acid metabolism in organisms is well-studied (Figure 1). Fatty acids are commonly built via *de novo* synthesis and elongation. Figure 1 shows that the *de novo* fatty acid synthesis

starts from the primer acetyl-CoA and the extender malonyl-CoA through a cyclic series of reactions catalyzed by fatty acid synthases. The synthesized fatty acids are almost entirely composed of

even-length and straight carbon chains that have various numbers of carbon atoms (<6, short chain; 6–12, medium chain; >14, long chain) and different degrees of unsaturation (saturated, monounsaturated, and polyunsaturated). Fatty acid breakdown takes place mainly via the β -oxidation pathway, which is like the *de novo* synthesis pathway running in a reverse direction (Figure 1).

The fatty acid metabolic pathway generates both fatty acids and their derivatives. The fatty acids and their derivatives from the synthesis and breakdown pathways can ultimately be converted to desirable value-added chemicals through metabolic engineering.

METABOLIC ENGINEERING

Metabolic engineering is undoubtedly an essential tool in biocatalytic systems because it can develop new cell factories or improve existing cell factories to produce non-native compounds. The primary objective of metabolic engineering is to improve the cellular properties by intentional modification of organisms through redirecting metabolic fluxes. Traditionally, metabolic engineering is performed by introducing completely new pathways for production of novel proteins, drugs, chemicals, or modifying native pathways to achieve desired metabolic goals such as high productivity of metabolites and high robustness of host strains. Here, metabolic engineering relies on directed genetic perturbations, usually in terms of modifying the promoter activity of a given gene, performing over-expression or deletion of endogenous genes/enzymes/pathways, and utilizing heterologous expression of genes/enzymes/pathways (Ostergaard et al., 2000).

However, traditional metabolic engineering approaches frequently fail to lead to the desired phenotypes because of unclear or complex gene structures, functions, and regulations in cellular metabolic networks. Hence, more efforts are required to achieve an integrative and holistic view of the overall network of pathways in organisms rather than individual pathways, which can then guide rational design strategies.

It is challenging to reconstruct certain biochemical pathways in a dynamic metabolic network without having the entire information on intracellular gene regulatory, metabolic, and signaling networks. Thus, fundamental knowledge on cellular genetics, biochemistry, and physiology is critical. Recently, multiple analytical and modeling tools, such as genomics, transcriptomics, proteomics, metabolomics, fluxomics, high-throughput screening, and *in silico* studies, have been utilized to elucidate metabolic engineering workflows, which provide useful information to predict the altered behaviors of metabolic networks, guide strain design, and maximize the efficacy of metabolic engineering.

MICROBIAL HOSTS FOR THE PRODUCTION OF FATTY ACID-DERIVED CHEMICALS

Metabolic engineering of microbial systems provides a renewable route to produce desired organic molecules such as fuels, materials, and chemicals. Many different types of microbes can naturally produce and accumulate varying levels of fatty acids efficiently. Some of them exhibit properties advantageous to the production of fatty acid-derived compounds through metabolic engineering.

Escherichia coli and *S. cerevisiae* are the most intensively studied and widely used model microorganisms in the development of metabolic engineering strategies aimed at providing heterologous

bioproduction of value-added metabolites. They have several key advantages such as lower safety risks, faster growth rates, good tractability, more well-studied, and more industrially relevant. So far, a number of fatty acid-derived chemicals have been successfully produced in metabolically engineered *E. coli* and *S. cerevisiae* (for references, see Table 1 below). Compared to *E. coli*, *S. cerevisiae* can be cultured at higher cell density and has a better fermentation performance at low temperature and pH (Aronsson and Ronner, 2001; Ageitos et al., 2011). *S. cerevisiae* is also more suited for the functional expression of eukaryotic enzymes (many enzymes involved in fatty acid production are from the plant kingdom) due to its endomembrane systems and post-translational modifications (Ageitos et al., 2011). However, in many cases, the production yields of fatty acid-derived chemicals from the engineered *S. cerevisiae* are much lower than those of *E. coli* when overexpressing identical heterologous genes. The reasons for this are not clearly understood.

Oleaginous microorganisms, which include bacteria, yeast, cyanobacteria, microalgae, and filamentous fungi, can accumulate intracellular lipids to at least 20% of their cellular dry mass. Thus they are considered attractive next-generation host candidates for production of fatty acid-derived chemicals because these oleaginous species have the ability to provide fatty acids or lipids as precursors (Ratledge, 1994). Oleaginous bacteria have been less studied to date because the lipid content in oleaginous bacteria is relatively lower than that in yeast, cyanobacteria, microalgae, and filamentous fungi, and they are also limited by lower growth rates. Oleaginous cyanobacteria and microalgae are attractive hosts for fatty acid-derived chemical production mainly because of their unique photosynthesis capability that directly converts solar energy and recycles CO₂ into fuels (Parmar et al., 2011). For instance, cyanobacteria *Synechococcus elongatus* sp. strain PCC 7942 have already been successfully engineered to produce a number of different biofuel related compounds, including 1-butanol (Lan and Liao, 2012), isobutanol (Atsumi et al., 2009), isobutyraldehyde (Atsumi et al., 2009), and 2-methyl-1-butanol (Shen and Liao, 2012). However, they are both technically difficult to manipulate genetically, and their cultivation and growth processes are more complicated and expensive than bacteria, yeast, and fungi. These hurdles have hampered their use in the production of fatty acid-derived chemicals through metabolic engineering. Similarly, the exploitation of oleaginous filamentous fungi as production hosts is also impeded by the lack of efficient genetic transformation techniques.

In comparison, oleaginous yeast has many advantages over other oleaginous microbial sources that makes this class of microbes the most promising cell factories for the production of fatty acid-derived chemicals. They can grow to high cell densities in simple and inexpensive culture, reaching extremely high levels of lipid accumulation of more than 70% of their dry weight (Beopoulos et al., 2008; Santamauro et al., 2014). They are also able to use different kinds of residues in waste resources as nutrients (Papanikolaou et al., 2003; Fickers et al., 2005). They are more genetically tractable than oleaginous cyanobacteria, microalgae, and filamentous fungi with relatively well-developed genetic tools (Madzak et al., 2004). Oleaginous yeast candidates, which show great potential as hosts for fatty acid-derived chemical production,

Table 1 | Examples of valuable fatty acid-derived chemicals produced by metabolically engineered microorganisms.

Chemical	Organism	Titer	Reference
FAMEs (C12–18)	<i>E. coli</i>	16 mg/L	Nawabi et al. (2011)
FAEEs (C12–20)	<i>E. coli</i>	1.5 g/L	Zhang et al. (2012a)
	<i>S. cerevisiae</i>	47.6 mg/L	Shi et al. (2014)
FASEs (C12–18)	<i>E. coli</i>	1.05 g/L	Guo et al. (2014)
Butanol	<i>E. coli</i>	0.81 mg/L	Mattam and Yazdani (2013)
	<i>E. coli</i>	30 g/L	Shen et al. (2011)
	<i>S. cerevisiae</i>	2.5 mg/L	Steen et al. (2008)
	<i>S. cerevisiae</i>	242.8 mg/L	Si et al. (2014)
	<i>C. saccharoperbutylacetonicum</i>	32.8 g/L	Richter et al. (2012)
3-Methyl-1-pentanol	<i>E. coli</i>	384.3 mg/L	Zhang et al. (2008)
Fatty alcohols (C4–5)	<i>E. coli</i>	4035 mg/L	Huo et al. (2011)
Fatty alcohols (C6–10)	<i>E. coli</i>	0.33 g/L	Dellomonaco et al. (2011)
Fatty alcohols (C12, C14)	<i>E. coli</i>	0.45 g/L	Zheng et al. (2012)
Fatty alcohols (C12–18)	<i>E. coli</i>	1.725 g/L	Liu et al. (2013)
Fatty alcohols (C16–18)	<i>E. coli</i>	0.1 g/L	Zheng et al. (2012)
	<i>S. cerevisiae</i>	0.1 g/L	Runguphan and Keasling (2014)
Pentane	<i>Y. lipolytica</i>	4.98 mg/L	Blazeck et al. (2013)
Alkanes	<i>E. coli</i>	580.8 mg/L	Choi and Lee (2013)
Iso-alkanes	<i>E. coli</i>	5 mg/L	Howard et al. (2013)
Alkanes/Alkenes	<i>E. coli</i>	300 mg/L	Schirmer et al. (2010)
Alkenes	<i>E. coli</i>	97.6 mg/L	Liu et al. (2014)
Long-chain alkenes	<i>E. coli</i>	40 μ g/L	Beller et al. (2010)
Hydroxy fatty acid (C18)	<i>S. pombe</i>	137 μ g/L	Holic et al. (2012)
Hydroxy fatty acid (C18)	<i>Y. lipolytica</i>	60 mg/g DCW	Beopoulos et al. (2014)
Hydroxy fatty acid (C14)	<i>C. tropicalis</i>	174 g/L	Lu et al. (2010)
ω -1-Hydroxy fatty acid	<i>B. pumilus</i>	570 mg/L	Dellomonaco et al. (2011)
Dicarboxylic acid (C14)	<i>C. tropicalis</i>	210 g/L	Picataggio et al. (1992)
Methyl ketone	<i>E. coli</i>	500 mg/L	Park et al. (2012)
Lactone	<i>Y. lipolytica</i>	11 g/L	Wache et al. (2003)
ω -3-eicosapentaenoic acid (EPA)	<i>Y. lipolytica</i>	0.15 g/g DCW	Xue et al. (2013)
Triacylglyceride (TAG)	<i>E. coli</i>	1.1 mg/L	Rucker et al. (2013)
Poly-3-hydroxylalkanoates (mcl-PHA)	<i>E. coli</i>	0.4 g/L	Wang et al. (2012)
Medium-chain fatty acids (MCFAs)	<i>E. coli</i>	263 mg/L	Torella et al. (2013)

DCW, dry cell weight.

include *Yarrowia lipolytica* (Blazeck et al., 2014), *Lipomyces starkeyi* (Tapia et al., 2012), *Lipomyces tetrasporus* (Lomascolo et al., 1994), *Rhodotorula glutinis* (Saenge et al., 2011), *Rhodospiridium toruloides* (Li et al., 2007), *Cryptococcus albidus* (Fei et al., 2011), *Cryptococcus curvatus* (Gong et al., 2014), *Metschnikowia pulcherrima* (Santamauro et al., 2014), *Trichosporon pullulans* (Huang et al., 2011), and *Waltomyces lipofer* (Raschke and Knorr, 2009). In particular, the model oleaginous yeast *Y. lipolytica* provides a promising platform as an oleaginous cell factory to convert fatty acids to more valuable metabolites. This oleaginous platform has the ability to utilize wide-scale renewable materials as substrates (Papanikolaou et al., 2003; Fickers et al., 2005) and multiple cheap carbon sources for growth (Papanikolaou et al., 2002; Athanastadt et al., 2006). Furthermore, it is more competitive than the non-oleaginous yeast *S. cerevisiae* in terms of lipid yield and heterologous protein yield (Gellissen et al., 2005; Papanikolaou and Aggelis, 2009). All of these features make *Y. lipolytica* very attractive for use in the production of fatty acid-derived products. Recently,

production of various fatty acid-derived biofuel and bioproducts using engineered *Y. lipolytica* has been investigated, including compounds such as triglycerides (Tai and Stephanopoulos, 2013), alkanes (Blazeck et al., 2013), lactones (Wache et al., 2003), hydroxy fatty acids (Beopoulos et al., 2014), dicarboxylic acids (Wache, 2013), and polyunsaturated fatty acids (Xue et al., 2013). However, transport mechanisms, transcriptional regulatory, and signal transduction pathways involved in lipid accumulation and degradation in *Y. lipolytica* need further exploration. This will pave the way to better utilization of this platform.

METABOLIC ENGINEERING OF MICROBES FOR PRODUCING FATTY ACID-DERIVED CHEMICALS

As discussed above, most fatty acid-derived chemicals are hard to obtain efficiently from natural sources or through native metabolic pathways. Recent efforts of metabolic engineering have been made in developing microbial chemical factories for the production of target chemicals. **Figure 1** shows that the chemicals derived

from fatty acids are generated by introducing the corresponding conversion steps associated with native fatty acid metabolic pathways. In this section, we describe pathway engineering for biochemical synthesis and review applications of metabolic engineering in the production of various fatty acid-derived chemicals, including: (1) fatty acid esters; (2) fatty alkanes and alkenes; and (3) fatty alcohols and other chemicals such as fatty ketones and lactones.

METABOLIC ENGINEERING TOWARD FATTY ACID ESTER PRODUCTION

Fatty acid methyl esters (FAMEs) and fatty acid ethyl esters (FAEEs) can be used as “biodiesel” fuel. The key enzyme to synthesize FAEEs in engineered microbes is wax ester synthase, which is responsible for catalyzing the esterification reaction of acyl-CoAs and alcohols.

In *S. cerevisiae*, by expressing heterologous wax ester synthase from *Marinobacter hydrocarbonoclasticus* DSM 8798 and up-regulating endogenous acetyl-CoA carboxylase, FAEEs were produced at a final titer of 8.2 mg/L (Shi et al., 2012). By further eliminating pathways for triacylglycerols (TAG) formation, steryl esters (SE) formation, and β -oxidation that compete with FAEE forming pathway, the production of FAEEs was at 17.2 mg/L in the strain lacking these non-essential fatty acid utilization pathways (Valle-Rodriguez et al., 2014). The corresponding FAEE production increased up to 34 mg/L after integrating the wax ester synthase gene cassette into the yeast genome. To further improve FAEE production, endogenous acyl-CoA binding protein, and NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans* were overexpressed in the final integration strain. The highest FAEE titer of 47.6 mg/L was achieved (Shi et al., 2014). In *E. coli*, FAEEs at 674 mg/L were produced by using combinatorial approaches: (1) over-expression of wax ester synthases from *Acinetobacter baylyi* for conversion of fatty acids to FAEEs, native acyl-ACP thioesterases and acyl-CoA ligases for acyl-CoA production, pyruvate decarboxylase and alcohol dehydrogenase from *Zymomonas mobilis* for non-native ethanol-forming, and (2) deletion of the competing fatty acid β -oxidation pathway (knockouts are *fadE*) (Steen et al., 2010). It was reported that over-expression of acetyl-CoA carboxylase and optimization of cultivation conditions further improved the yield of FAEEs to 922 mg/L (Duan et al., 2011). A recent work demonstrated that a dynamic sensor-regulator system increased the FAEEs titer to 1.5 g/L in genetically engineered *E. coli* strain (Zhang et al., 2012a). Fed-batch pilot scale cultivation of the engineered *E. coli* p(Microdiesel) strain could yield 15 g/L FAEEs, by first using glycerol as sole carbon source for biomass production before glucose and oleic acid were added as carbon sources (Elbahloul and Steinbüchel, 2010).

In *E. coli*, FAMEs were formed from free fatty acids and S-adenosylmethionine through expressing fatty acid methyltransferases from *Mycobacterium marinum* and *Mycobacterium smegmatis*. Over-expression of heterologous thioesterases can increase free fatty acids, and further result in increased FAME synthesis. It was reported that over-expression of thioesterases such as thioesterase II from *E. coli*, acyl-ACP thioesterases from *Clostridium phytofermentans*, *Clostridium sporogenes*, *Clostridium tetani* and *M. marinum*, 3-hydroxyacyl ACP:CoA transacylases from

Pseudomonas putida, and methionine adenosyltransferases from rat, combined with deletion of a global methionine regulator *metJ*, led to the production of FAMEs at up to 16 mg/L (Nawabi et al., 2011).

METABOLIC ENGINEERING TOWARD FATTY ALKA(E)NE PRODUCTION

Fatty alka(e)nes can exist as straight or branched chains. Both straight- and branched-chain alka(e)nes have the potential to serve as advanced biofuels. There are two primary pathways for alka(e)ne biosynthesis: (1) a pathway that starts from acyl-ACP, followed by reducing acyl-ACPs to form fatty aldehydes catalyzed by reductases, and then converting fatty aldehydes to alka(e)nes by aldehyde decarbonylases; and (2) a pathway that starts from free fatty acids, followed by reduction and decarboxylation to generate alka(e)nes.

Over-expression of acyl-ACP reductases and aldehyde decarbonylases from cyanobacteria in *E. coli* and *Synechocystis* sp. PCC 7002 achieved alka(e)ne concentration at 300 mg/L (Schirmer et al., 2010) and 5% of cell dry weight (Reppas et al., 2010), respectively. Recently, in *E. coli*, free fatty acids were catalyzed to form fatty aldehydes by expressing fatty acid reductase complex from *Photobacterium luminescens*. Coupled with aldehyde decarbonylases from *Nostoc punctiforme*, fatty aldehydes were converted further to alka(e)nes. In this study, production of branched-chain alka(e)nes from branched-chain fatty acids at a titer of 2–5 mg/L was also reported by over-expression of branched-chain α -keto acid dehydrogenase complex and β -ketoacyl-ACP synthase III from *B. subtilis* (Howard et al., 2013).

Terminal alkenes can also be produced in microorganisms via two pathways: (1) conversion of free fatty acids to terminal alkenes by cytochrome P450 peroxygenase (Rude et al., 2011); and (2) conversion of acyl-ACP to terminal alkenes by a large multi-domain type I polyketide synthases (Mendez-Perez et al., 2011). However, the pathways involving free fatty acids and acyl-ACP need to be further optimized to improve the efficiency and yield. Very long-chain alkenes can be generated by a head-to-head condensation of two acyl-CoAs catalyzed by the OleABCD protein families. In a previous study, heterologous expression of the Ole cluster from *Micrococcus luteus* ATCC 4698 in *E. coli* led to the production of very long-chain alkenes at a total concentration of 40 μ g/L (Beller et al., 2010).

METABOLIC ENGINEERING TOWARD PRODUCTION OF FATTY ALCOHOLS AND OTHER CHEMICALS

Fatty alcohols (or long-chain alcohols) can be formed by reduction from fatty aldehyde intermediates using aldehyde reductases, for example, from cyanobacterium *Synechocystis* sp. PCC 680 (Steen et al., 2010). Fatty alcohols can also be directly produced by acyl-CoA reductases from *M. aquaeolei*, mouse, jojoba, and *Arabidopsis thaliana*. Another fatty aldehyde reductase from *M. aquaeolei* was found to possess the ability to catalyze not only fatty aldehydes but also acyl-CoA or acyl-ACP to corresponding fatty alcohols (Hofvander et al., 2011; Liu et al., 2013). In these pathways, fatty aldehyde intermediates can be bypassed (Tan et al., 2011). In addition, another synthetic pathway leading to 1-butanol (short-chain fatty alcohol) production from *Clostridium* species was functionally constructed in *E. coli* (Shen et al., 2011),

S. cerevisiae (Steen et al., 2008), and *Thermoanaerobacterium saccharolyticum* (Bhandiwad et al., 2014). This pathway begins with a CoA-dependent Claisen condensation reaction of two acetyl-CoA followed by reduction, dehydration, and hydrogenation. Thus, this sequence of chemical reactions is the reverse direction of that in β -oxidation pathway. Recently, this CoA-dependent 1-butanol synthesis pathway has been extended to produce other linear short-chain fatty alcohols (C6–C8) in *E. coli* (Zhang et al., 2008; Tseng and Prather, 2012).

In addition, chemicals derived from fatty acids also include methyl ketones, hydroxy fatty acids, lactones, and dicarboxylic acids. Methyl ketones can be synthesized through conversion of fatty acids to β -keto acyl-CoAs in β -oxidation, and hydrolysis of β -keto acyl-CoAs by thioesterases to form β -keto fatty acids, followed by decarboxylation of β -keto fatty acids to methyl ketones (Goh et al., 2012). Hydroxy fatty acids can be synthesized by diverse kinds of fatty acid-hydroxylation enzymes, including P450, lipoxygenase, hydratase, 12-hydroxylase, and diol synthase (Kim and Oh, 2013). Lactones can be generally obtained by one-step biotransformation of the precursors hydroxy fatty acids (Wache et al., 2003). To generate dicarboxylic acids, hydroxy fatty acids can be oxidized to fatty ketones by alcohol dehydrogenases, followed by further oxidation of the fatty ketones to esters by Baeyer–Villiger monooxygenases. The esters are subsequently hydrolyzed by esterases to yield dicarboxylic acids (Song et al., 2013). The representatives of valuable chemicals derived from fatty acids in engineered microbes are listed in **Table 1**.

Taken together, metabolic engineering of microorganisms serves as a good platform for effective production of desired fatty acid-derived valuable chemicals. However, more research efforts are required to achieve industrially relevant titers of these chemicals.

FACILITATION OF FATTY ACID-DERIVED CHEMICAL BIOPRODUCTION WITH ADVANCED SYNTHETIC BIOLOGY TOOLS

Successful production of fatty acid-derived chemicals by metabolic engineering of microbial systems has already been achieved. However, the productivity and titers of each of these processes remain to be improved. Further improvement in production efficiency is critical because high productivity and product yield for cost-effective production are the most important pre-requisites for large-scale industrial production of fatty acid-derived chemicals that is also financially viable.

Recent years have witnessed the emergence and marked progress in synthetic biology. Many advanced synthetic biology tools have offered a variety of applications to improve the ability to re-engineer microbial cells for achieving high yields of valuable chemicals, e.g., modular control over metabolic flux in mevalonate biosynthesis pathway using synthetic protein scaffolds in *E. coli* (Dueber et al., 2009), enhancement in production of fatty acid-derived biofuels by using dynamic sensor-regulator system in *E. coli* (Zhang et al., 2012a) and improvement of tolerance against alkane biofuels by transporter engineering in *S. cerevisiae* (Chen et al., 2013a). Although these tools are not widely used in metabolic engineering of microorganisms aiming to produce fatty

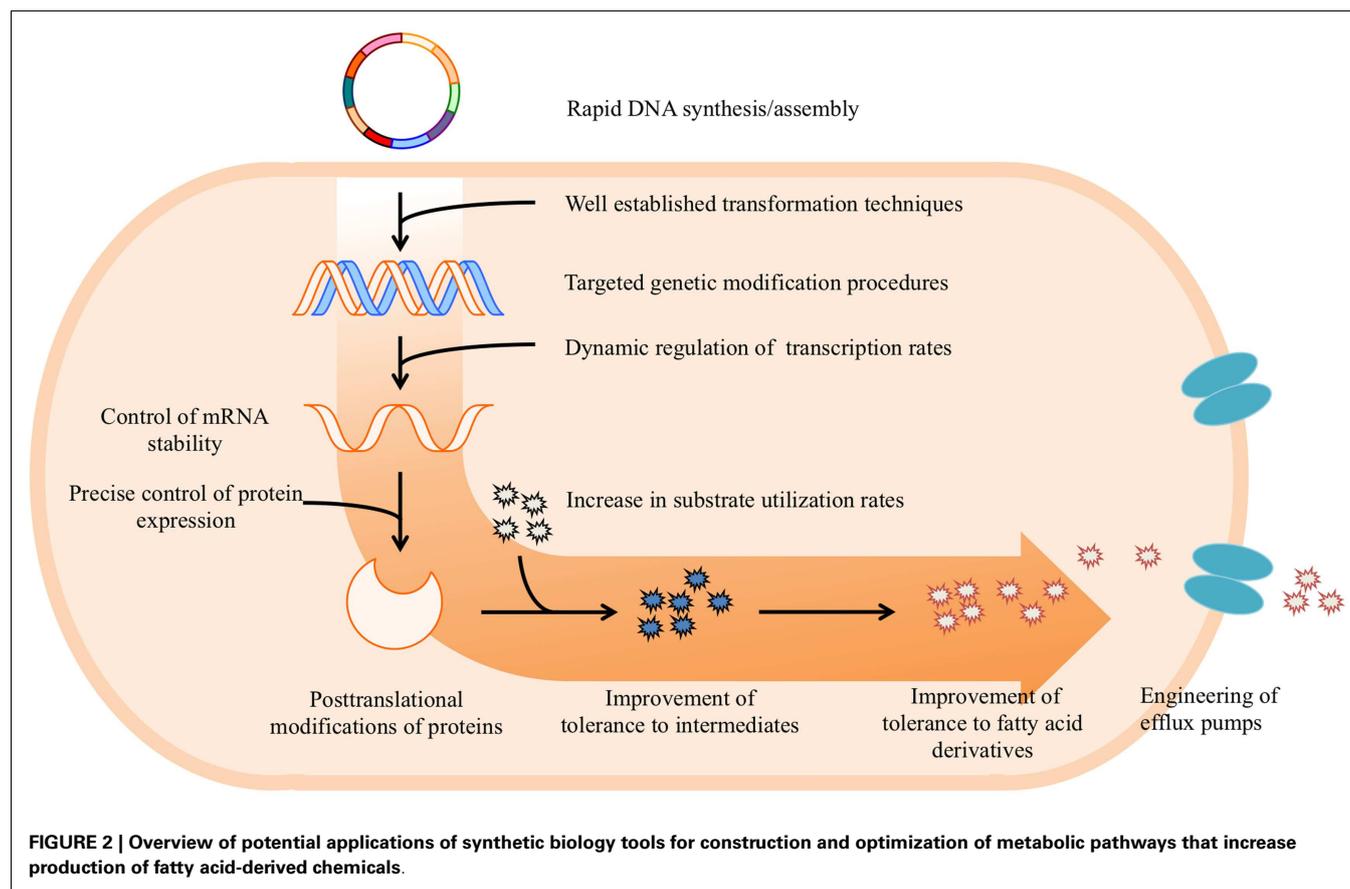
acid-derived chemicals, there is no doubt that these innovations would facilitate tremendous potential for improved metabolic engineering of microbial systems in the production of various fatty acid-derived products.

In summary, advanced synthetic biology approaches for pathway optimization show great promise in enhancing the speed and efficiency of creating improved microbial strains in combination with common metabolic engineering efforts. The production of fatty acid-derived chemicals could benefit from the integration of synthetic biology tools with the work already accomplished through metabolic engineering. Thus implementation of advanced synthetic biology tools in redesigning fatty acid biosynthesis pathway and heterologous metabolic pathways for the production of fatty acid-derived targets will guide rational manipulation for production of our target at high yields and titers. In this section, we will briefly review the recent development of synthetic biology methodologies and possible applications for construction and optimization of metabolic pathways in microbes at DNA, transcription, translation, and post-translation levels (**Figure 2**).

DNA ENGINEERING

The first step of most metabolic engineering and synthetic biology studies is to reconstruct a completely or partially synthetic pathway. Therefore, rapid assembly of heterologous pathways with many enzymatic steps is a major challenge in metabolic engineering. Traditional DNA molecular cloning approaches, which are tedious, time-consuming and mainly limited by template-based synthesis, restriction digestion, and ligation-based cloning, are increasingly being replaced with *de novo* DNA synthesis and more sophisticated assembly capabilities. Many simple, rapid, high-throughput, high-fidelity and low-cost DNA synthesis, and assembly methods in synthetic biology have been developed, including programmable microfluidic chips (Tian et al., 2004), BioBricks assembly (Sleight et al., 2010), BglBricks assembly (Anderson et al., 2010), In-Fusion assembly (Zhu et al., 2007), Gibson DNA assembly (Gibson et al., 2009), TAR-based assembly (Benders et al., 2010), Circular polymerase extension cloning (CPEC) (Quan and Tian, 2009), Sequence and ligase independent cloning (SLIC) (Li and Elledge, 2007), Seamless Ligation Cloning Extract (SLICE) (Zhang et al., 2012c), DNA assembler (Shao et al., 2009), Uracil-specific excision reagent cloning (USER) (Gulig et al., 2009), Methylation-assisted tailorable ends rational ligation (MASTER) (Chen et al., 2013b), Site-specific recombination-based tandem assembly (SSRTA) (Zhang et al., 2011), PCR-based two-step DNA synthesis (PTDS) (Xiong et al., 2004), Golden Gate assembly (Cermak et al., 2011), and Polymerase incomplete primer extension cloning (PIPE) (Liu and Naismith, 2008). These approaches together enable the efficient synthesis of synthetic DNA fragments with no apparent limits on either sequence or length. Therefore, these powerful and efficient toolboxes allow efficient manufacture of genes, regulatory elements, circuits, gene clusters, and metabolic pathways for the production of novel chemicals.

The laborious and site-specific gene targeting by homologous recombination techniques, which have limited applicability for genome wide modification are now being increasingly displaced with such genome-scale engineering techniques as multiplex



automated genome engineering (MAGE), conjugative assembly of genome engineering (CAGE), and transcription activator-like effector nucleases (TALENs). MAGE simultaneously targets multiple locations on chromosomes to introduce small modifications in a single cell or across a population of cells, facilitating rapid generation of a diverse set of genetic changes. 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway in *E. coli* was optimized by this technique. Twenty-four genetic components in the DXP pathway were modified simultaneously using a complex pool of synthetic oligonucleotides, creating over 4.3 billion combinatorial genomic variants per day and achieving a more than fivefold increase in lycopene production within 3 days (Wang et al., 2009). CAGE enabled large-scale assembly of many modified genomes on the basis of MAGE (Isaacs et al., 2011). TALENs is another powerful tool created to target double-strand breaks at specific locations in the genome (Christian et al., 2010).

TRANSCRIPTIONAL ENGINEERING

Transcription is the first dedicated phase of gene expression and therefore, different toolsets have been developed in synthetic biology for controlling gene expression and modulating RNA levels in the engineered cells. The primary goal of transcriptional engineering is synthetic control of RNA transcription and transcript levels by controlling gene copy number, transcription initiation rate, transcription termination efficiency, and transcript decay rate. Modifications of gene copy number can be achieved by

changing the origin of replication of recombinant expression plasmids or the number of chromosomally integrated gene copies (in particular, strategies for chromosomal integrations at multiple loci). In addition, promoter engineering can be applied to regulate the rate of transcription initiation by using different types of promoters such as constitutive promoters, inducible promoters, specific promoters, hybrid promoters, synthetic promoters, and synthetic promoter libraries (De Mey et al., 2007). Transcription termination efficiency can be regulated as well by changing terminator sequence contexts (Cambray et al., 2013). Studies on mRNA folding and degradation rates determined by mRNA message itself (primary sequences and/or secondary-structures) and on the genomic region of 5'- and 3'-UTR allowed for further control of transcript abundances of genes of interest (Dori-Bachash et al., 2011; Zaborske et al., 2013).

Based on the principles above, increasing attempts have been recently made to further improve the sensitivity and precision of transcription regulation. First, RNA control system by engineered RNA hairpins enables conditional activation of an endogenous pathway capable of operating in autonomous mode within a complex cellular regulatory network (Venkataraman et al., 2010). Second, dynamic sensor-regulator system uses a transcription factor to specifically sense key intermediates and dynamically regulates the expression of genes. In biodiesel biosynthetic pathways in *E. coli*, this system substantially improved the stability of biodiesel-producing strains and increased the yield by threefold.

This strategy can also be extended to other biosynthetic pathways to balance metabolism, thereby increasing product titers and conversion yields and stabilizing production hosts (Zhang et al., 2012a). Third, regulatable expression system has been developed for modulating gene expression in *Corynebacterium glutamicum*. Furthermore, this work provided a synthetic promoter library that enabled the selection of strong promoters. This technology should have many future applications for optimizing bioproduction in *C. glutamicum* and other organisms (Rytter et al., 2014). In addition, transcription factor engineering (Lee et al., 2011) and global transcription machinery engineering study (Zhang et al., 2012b) also serve as a good example for using synthetic biology tools to re-engineer transcriptional regulation in organisms. All these efforts have already shown promise and could lead to highly optimized expression of synthetic pathways at the transcriptional level.

TRANSLATIONAL ENGINEERING

After gene transcription is complete, translational engineering tools can be used to speed translation rates, lower degradation rates, and tune protein yields. Synthetic ribosome binding sites (Salis, 2011), antisense RNA (Chang et al., 2012), ribozymes (Meaux and Van Hoof, 2006), translation machinery (rRNA, tRNA, and amino acid) (Harris and Jewett, 2012), peptide tags, and codon optimization method have been proved effective in control of cellular protein levels at the translational level. mRNA secondary structure engineering is a newly developed method for translational regulation of gene expression. The engineered mRNA molecules that exhibit diverse activities including sensing, regulatory, information processing, and scaffolding activities has been implemented as key control elements in synthetic genetic networks to program biological function (Liang et al., 2011). Compared with DNA engineering and transcriptional engineering, translational engineering tools have not yet been extensively developed. Although translational regulation in cellular systems is not as well-studied, these advances have shown to be effective in removing translation-level limitations.

POST-TRANSLATIONAL ENGINEERING

Post-translational modification of proteins also takes place after translation and include phosphorylation, glycosylation, ubiquitination, methylation, acetylation, and proteolysis. Regulation of this process in the field of synthetic biology is especially important to either prolong or shorten the half-life of desirable proteins. To this end, addition of a synthetic ligand that binds to the destabilizing domains of specific proteins shields them from degradation, allowing fused proteins to perform their cellular functions in mammalian cells (Banaszynski et al., 2006). In addition, a synthetic gene network for tunable degradation of a tagged protein has been constructed in *S. cerevisiae* using components of the *E. coli* degradation machinery (Grilly et al., 2007), opening the door foreengineering, and optimization of protein degradation for a variety of future applications in microbial cell factories.

PATHWAY ENGINEERING

Once the enzymes are expressed from specific genes, the last major challenge lies in optimizing gene expression, protein abundance,

enzyme activities, synthetic pathways, and metabolic products as a system, especially in a dynamic manner. To address this problem, researchers have recently developed an array of tools, including global regulator engineering (Hong et al., 2010), computational protein design (Samish et al., 2011), protein engineering (Bommarius et al., 2011), protein trafficking (Hou et al., 2012), protein scaffolds (Dueber et al., 2009), transporter engineering (Chen et al., 2013a), cellular efflux pump engineering (Dunlop et al., 2011), ultrasensitive input/output control system (Dueber et al., 2007), and computer-based complex gene circuits (Daniel et al., 2013). For example, transporter engineering through expression of heterologous ABC transporters from *Y. lipolytica* has been utilized successfully to significantly improve tolerance of *S. cerevisiae* against alkanes. In particular, the tolerance limit of *S. cerevisiae* against decane was increased about 80-fold (Chen et al., 2013a). Ultrasensitive switches with a non-linear input/output function can be effectively harnessed to control many complex biological behaviors in higher-order regulatory systems. These switches approximate digital behavior, providing an input detection threshold at which small changes in input concentration lead to large changes in output behavior. Another successful example of pathway engineering is computer-based complex gene circuits. Synthetic analog gene circuits were engineered to execute sophisticated computational functions in living cells using three transcription factors. Such circuits could lead to new applications for synthetic biology and biotechnology that require complex computations with limited parts (Daniel et al., 2013). These methods and technologies can be combined to optimize the metabolic pathway and significantly boost the production of target compounds in a controllable, scalable, and effective way within host cells.

CONCLUSION AND FUTURE PERSPECTIVES

Fatty acid-derived diverse valuable chemicals are in great demand. This class of chemicals has recently been successfully produced by introducing different biosynthesis genes, enzymes, and pathways into various microbial hosts. Although much progress has been made in the use of metabolic engineering of microbes for the production of fatty acid-derived chemicals, the sub-optimal product yields, and productivities render these platforms far from reaching large-scale commercial exploitation.

Conventional metabolic engineering efforts on the microbial production of fatty acid-derived chemicals predominantly rely on identifying the activity of related enzymes isolated from different sources. In this regard, future efforts should be invested in finding and adopting novel sources of enzymes either in existing pathways or from completely novel producing pathways with such desired features as higher enzyme activity, stability, and specificity. High-throughput enzyme screening methods and bioinformatics tools could be used to screen these enzymes from vastly different organisms.

However, many attempts have demonstrated that the simple import of heterologous pathways into microbial hosts without a good understanding of complex regulatory networks underlying their biosynthesis pathways, will unlikely yield high-level production of target fatty acid-derived chemicals. Hence, the exploration of such metabolic and regulatory information is crucial for the

heterologous production of these chemicals. Due to the complexity of regulatory networks, difficulties can be formidable. Synthetic biology-based tools can help to elucidate complex regulatory networks, enhance gene expression, increase enzyme activities and substrate specificity, improve metabolic flux, and boost product titer in heterologous microbial hosts. Taken together, combinatorial approaches encompassing metabolic engineering and synthetic biology together with more detailed knowledge of metabolic and genetic regulatory mechanisms, will be effective in overcoming bottlenecks inherent in the production of fatty acid-derived valuable chemicals in microbes. Ultimately, successful engineering strategies will be key to push efficient microbial-based production of the fatty acid-derived valuable chemicals forward toward industrialization.

ACKNOWLEDGMENTS

We gratefully acknowledge funding support from the Competitive Research Program of the National Research Foundation of Singapore (NRF-CRP5-2009-03), the Agency for Science, Technology and Research of Singapore (1324004108), the National Environment Agency of Singapore (ETRP 1201102), and Global R&D Project Program, the Ministry of Knowledge Economy, the Republic of Korea (N0000677).

REFERENCES

- Ageitos, J. M., Vallejo, J. A., Veiga-Crespo, P., and Villa, T. G. (2011). Oily yeasts as oleaginous cell factories. *Appl. Microbiol. Biotechnol.* 90, 1219–1227. doi:10.1007/s00253-011-3200-z
- Anderson, J., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., et al. (2010). BglBricks: a flexible standard for biological part assembly. *J. Biol. Eng.* 4, 1–12. doi:10.1186/1754-1611-4-1
- Aronsson, K., and Ronner, U. (2001). Influence of pH, water activity and temperature on the inactivation of *Escherichia coli* and *Saccharomyces cerevisiae* by pulsed electric fields. *Innov. Food Sci. Emerg. Technol.* 2, 105–112. doi:10.1016/S1466-8564(01)00030-3
- Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M., et al. (2006). Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics* 6, 1450–1459. doi:10.1002/pmic.200500339
- Atsumi, S., Higashide, W., and Liao, J. C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* 27, 1177–1180. doi:10.1038/nbt.1586
- Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A., and Wandless, T. J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126, 995–1004. doi:10.1016/j.cell.2006.07.025
- Beller, H. R., Goh, E. B., and Keasling, J. D. (2010). Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl. Environ. Microbiol.* 76, 1212–1223. doi:10.1128/AEM.02312-09
- Benders, G. A., Noskov, V. N., Denisova, E. A., Lartigue, C., Gibson, D. G., Assad-Garcia, N., et al. (2010). Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* 38, 2558–2569. doi:10.1093/nar/gkq119
- Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., Hapala, I., Papanikolaou, S., et al. (2008). Control of lipid accumulation in the yeast *Yarrowia lipolytica*. *Appl. Environ. Microbiol.* 74, 7779–7789. doi:10.1128/aem.01412-08
- Beopoulos, A., Verbeke, J., Bordes, F., Guicherd, M., Bressy, M., Marty, A., et al. (2014). Metabolic engineering for ricinoleic acid production in the oleaginous yeast *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* 98, 251–262. doi:10.1007/s00253-013-5295-x
- Bhandiwad, A., Shaw, A. J., Guss, A., Guseva, A., Bahl, H., and Lynd, L. R. (2014). Metabolic engineering of *Thermoanaerobacterium saccharolyticum* for n-butanol production. *Metab. Eng.* 21, 17–25. doi:10.1016/j.ymben.2013.10.012
- Blazeck, J., Hill, A., Liu, L., Knight, R., Miller, J., Pan, A., et al. (2014). Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat. Commun.* 5, 3131. doi:10.1038/ncomms4131
- Blazeck, J., Liu, L., Knight, R., and Alper, H. S. (2013). Heterologous production of pentane in the oleaginous yeast *Yarrowia lipolytica*. *J. Biotechnol.* 165, 184–194. doi:10.1016/j.jbiotec.2013.04.003
- Bommarius, A. S., Blum, J. K., and Abrahamson, M. J. (2011). Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* 15, 194–200. doi:10.1016/j.copbio.2010.11.011
- Cambay, G., Guimaraes, J. C., Mutalik, V. K., Lam, C., Mai, Q. A., Thimmaiah, T., et al. (2013). Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* 41, 5139–5148. doi:10.1093/nar/gkt163
- Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82. doi:10.1093/nar/gkr218
- Chang, A. L., Wolf, J. J., and Smolke, C. D. (2012). Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Curr. Opin. Biotechnol.* 23, 679–688. doi:10.1016/j.copbio.2012.01.005
- Chen, B., Ling, H., and Chang, M. W. (2013a). Transporter engineering for improved tolerance against alkane biofuels in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 6, 21. doi:10.1186/1754-6834-6-21
- Chen, W. H., Qin, Z. J., Wang, J., and Zhao, G. P. (2013b). The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Res.* 41, e93. doi:10.1093/nar/gkt122
- Choi, Y. J., and Lee, S. Y. (2013). Microbial production of short-chain alkanes. *Nature* 502, 571–574. doi:10.1038/nature12536
- Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. doi:10.1534/genetics.110.120717
- Clomburg, J. M., and Gonzalez, R. (2010). Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* 86, 419–434. doi:10.1007/s00253-010-2446-1
- Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature* 497, 619–623. doi:10.1038/nature12148
- De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., and Vandamme, E. J. (2007). Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol.* 7:34. doi:10.1186/1472-6750-7-34
- Dellomonaco, C., Clomburg, J. M., Miller, E. N., and Gonzalez, R. (2011). Engineered reversal of the β -oxidation cycle for the synthesis of fuels and chemicals. *Nature* 476, 355–359. doi:10.1038/nature10333
- Dori-Bachash, M., Shema, E., and Tirosh, I. (2011). Coupled evolution of transcription and mRNA degradation. *PLoS Biol.* 9:e1001106. doi:10.1371/journal.pbio.1001106
- Duan, Y., Zhu, Z., Cai, K., Tan, X., and Lu, X. (2011). De novo biosynthesis of biodiesel by *Escherichia coli* in optimized fed-batch cultivation. *PLoS ONE* 6:e20265. doi:10.1371/journal.pone.0020265
- Dueber, J. E., Mirsky, E. A., and Lim, W. A. (2007). Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* 25, 660–662. doi:10.1038/nbt1308
- Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759. doi:10.1038/nbt.1557
- Dunlop, M. J., Dossani, Z. Y., Szmidi, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., et al. (2011). Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* 7, 487. doi:10.1038/msb.2011.21
- Elbahloul, Y., and Steinbüchel, A. (2010). Pilot-scale production of fatty acid ethyl esters by an engineered *Escherichia coli* strain harboring the p (Microdiesters) plasmid. *Appl. Environ. Microbiol.* 76, 4560–4565. doi:10.1128/AEM.00515-10
- Fei, Q., Chang, H. N., Shang, L., Kim, N., and Kang, J. (2011). The effect of volatile fatty acids as a sole carbon source on lipid accumulation by *Cryptococcus albidus* for biodiesel production. *Bioresour. Technol.* 102, 2695–2701. doi:10.1016/j.biortech.2010.10.141
- Fickers, P., Benetti, P. H., Wache, Y., Marty, A., Mauersberger, S., Smit, M., et al. (2005). Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its

- potential applications. *FEMS Yeast Res.* 5, 527–543. doi:10.1016/j.femsyr.2004.09.004
- Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J. M., Berardi, E., Veenhuis, M., et al. (2005). New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adenivorans* and *Yarrowia lipolytica*-a comparison. *FEMS Yeast Res.* 5, 1079–1096. doi:10.1016/j.femsyr.2005.06.004
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318
- Goh, E. B., Baidoo, E. E., Keasling, J. D., and Beller, H. R. (2012). Engineering of bacterial methyl ketone synthesis for biofuels. *Appl. Environ. Microbiol.* 78, 70–80. doi:10.1128/AEM.06785-11
- Gong, Z. W., Shen, H. W., Yang, X. B., Wang, Q., Xie, H. B., and Zhao, Z. (2014). Lipid production from corn stover by the oleaginous yeast *Cryptococcus curvatus*. *Biotechnol. Biofuels* 7, 158. doi:10.1186/s13068-014-0158-y
- Grilly, C., Stricker, J., Pang, W. L., Bennett, M. R., and Hasty, J. (2007). A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.* 3, 127. doi:10.1038/msb4100168
- Gulig, P. A., Tucker, M. S., Thiaville, P. C., Joseph, J. L., and Brown, R. N. (2009). USER friendly cloning coupled with chitin-based natural transformation enables rapid mutagenesis of *Vibrio vulnificus*. *Appl. Environ. Microbiol.* 75, 4936–4949. doi:10.1128/AEM.02564-08
- Guo, D., Zhu, J., Deng, Z., and Liu, T. (2014). Metabolic engineering of *Escherichia coli* for production of fatty acid short-chain esters through combination of the fatty acid and 2-keto acid pathways. *Metab. Eng.* 22, 69–75. doi:10.1016/j.ymben.2014.01.003
- Harris, D. C., and Jewett, M. C. (2012). Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Curr. Opin. Biotechnol.* 23, 672–678. doi:10.1016/j.copbio.2012.02.002
- Hofvander, P., Doan, T. T., and Hamberg, M. (2011). A prokaryotic acyl-CoA reductase performing reduction of fatty acyl-CoA to fatty alcohol. *FEBS Lett.* 585, 3538–3543. doi:10.1016/j.febslet.2011.10.016
- Holic, R., Yazawa, H., Kumagai, H., and Uemura, H. (2012). Engineered high content of ricinoleic acid in fission yeast *Schizosaccharomyces pombe*. *Appl. Microbiol. Biotechnol.* 95, 179–187. doi:10.1007/s00253-012-3959-6
- Hong, S. H., Lee, J., and Wood, T. K. (2010). Engineering global regulator Hha of *Escherichia coli* to control biofilm dispersal. *Microb. Biotechnol.* 3, 717–728. doi:10.1111/j.1751-7915.2010.00220.x
- Hou, J., Tyo, K., Liu, Z., Petranovic, D., and Nielsen, J. (2012). Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Metab. Eng.* 14, 120–127. doi:10.1016/j.ymben.2012.01.002
- Howard, T. P., Middelhaufe, S., Moore, C., Edner, C., Kolak, D. M., Taylor, G. N., et al. (2013). Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7636–7641. doi:10.1073/pnas.1215966110
- Huang, X., Wang, Y., Liu, W., and Bao, J. (2011). Biological removal of inhibitors leads to the improved lipid production in the lipid fermentation of corn stover hydrolysate by *Trichosporon cutaneum*. *Bioresour. Technol.* 102, 9705–9709. doi:10.1016/j.biortech.2011.08.024
- Huo, Y. X., Cho, K. M., Rivera, J. G. L., Monte, E., Shen, C. R., Yan, Y., et al. (2011). Conversion of proteins into biofuels by engineering nitrogen flux. *Nat. Biotechnol.* 29, 346–351. doi:10.1038/nbt
- Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., et al. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333, 348–353. doi:10.1126/science.1205822
- Keasling, J. D., and Chou, H. (2008). Metabolic engineering delivers next-generation biofuels. *Nat. Biotechnol.* 26, 298–299. doi:10.1038/nbt0308-298
- Kim, K. R., and Oh, D. K. (2013). Production of hydroxy fatty acids by microbial fatty acid-hydroxylation enzymes. *Biotechnol. Adv.* 31, 1473–1485. doi:10.1016/j.biortechadv.2013.07.004
- Lan, E. I., and Liao, J. C. (2012). ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6018–6023. doi:10.1073/pnas.1200074109
- Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8, 536–546. doi:10.1038/nchembio.970
- Lee, J. Y., Yang, K. S., Jang, S. A., Sung, B. H., and Kim, S. C. (2011). Engineering butanol-tolerance in *Escherichia coli* with artificial transcription factor libraries. *Biotechnol. Bioeng.* 108, 742–749. doi:10.1002/bit.22989
- Lennen, R. M., and Pfleger, B. F. (2013). Microbial production of fatty acid-derived fuels and chemicals. *Curr. Opin. Biotechnol.* 24, 1044–1053. doi:10.1016/j.copbio.2013.02.028
- Li, M. Z., and Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* 4, 251–256. doi:10.1038/nmeth1010
- Li, Y., Zhao, Z. K., and Bai, F. (2007). High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme Microb. Technol.* 41, 312–317. doi:10.1016/j.enzmictec.2007.02.008
- Liang, J. C., Bloom, R. J., and Smolke, C. D. (2011). Engineering biological systems with synthetic RNA molecules. *Mol. Cell* 43, 915–926. doi:10.1016/j.molcel.2011.08.023
- Liu, A., Tan, X., Yao, L., and Lu, X. (2013). Fatty alcohol production in engineered *E. coli* expressing *Marinobacter* fatty acyl-CoA reductases. *Appl. Microbiol. Biotechnol.* 97, 7061–7071. doi:10.1007/s00253-013-5027-2
- Liu, H., and Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 8:91. doi:10.1186/1472-6750-8-91
- Liu, Y., Wang, C., Yan, J., Zhang, W., Guan, W., Lu, X., et al. (2014). Hydrogen peroxide-independent production of α -alkenes by OleTJE P450 fatty acid decarboxylase. *Biotechnol. Biofuels* 7, 28. doi:10.1186/1754-6834-7-28
- Lomascolo, A., Dubreucq, E., Perrier, V., and Galzy, P. (1994). Study of lipids in *Lipomyces* and *Waltomyces*. *Can. J. Microbiol.* 40, 724–729. doi:10.1139/m94-115
- Lu, W., Ness, J. E., Xie, W., Zhang, X., Minshull, J., and Gross, R. A. (2010). Biosynthesis of monomers for plastics from renewable oils. *J. Am. Chem. Soc.* 132, 15451–15455. doi:10.1021/ja107707v
- Madzak, C., Gaillardin, C., and Beckerich, J. M. (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J. Biotechnol.* 109, 63–81. doi:10.1016/j.jbiotec.2003.10.027
- Mattam, A. J., and Yazdani, S. S. (2013). Engineering *E. coli* strain for conversion of short chain fatty acids to bioalcohols. *Biotechnol. Biofuels* 6, 128. doi:10.1186/1754-6834-6-128
- Meaux, S., and Van Hoof, A. (2006). Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly (A) tail in translation and mRNA decay. *RNA* 12, 1323–1337. doi:10.1261/rna.46306
- Mendez-Perez, D., Begemann, M. B., and Pfleger, B. F. (2011). Modular synthase-encoding gene involved in α -olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. *Appl. Environ. Microbiol.* 77, 4264–4267. doi:10.1128/AEM.00467-11
- Nawabi, P., Bauer, S., Kyripides, N., and Lykidis, A. (2011). Engineering *Escherichia coli* for biodiesel production utilizing a bacterial fatty acid methyltransferase. *Appl. Environ. Microbiol.* 77, 8052–8061. doi:10.1128/AEM.05046-11
- Ostergaard, S., Olsson, L., and Nielsen, J. (2000). Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* 64, 34–50. doi:10.1128/MMBR.64.1.34-50.2000
- Papanikolaou, S., and Aggelis, G. (2009). Biotechnological valorization of biodiesel derived glycerol waste through production of single cell oil and citric acid by *Yarrowia lipolytica*. *Lipid Technol.* 21, 83–87. doi:10.1002/lite.200900017
- Papanikolaou, S., Chevlat, I., Komaitis, M., Marc, I., and Aggelis, G. (2002). Single cell oil production by *Yarrowia lipolytica* growing on an industrial derivative of animal fat in batch cultures. *Appl. Microbiol. Biotechnol.* 58, 308–312. doi:10.1007/s00253-001-0897-0
- Papanikolaou, S., Muniglia, L., Chevlat, I., Aggelis, G., and Marc, I. (2003). Accumulation of a cocoa-butter-like lipid by *Yarrowia lipolytica* cultivated on agro-industrial residues. *Curr. Microbiol.* 46, 124–130. doi:10.1007/s00284-002-3833-3
- Park, J., Rodriguez-Moya, M., Li, M., Pichersky, E., San, K. Y., and Gonzalez, R. (2012). Synthesis of methyl ketones by metabolically engineered *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 39, 1703–1712. doi:10.1007/s10295-012-1178-x
- Parmar, A., Singh, N. K., Pandey, A., Gnansounou, E., and Madamwar, D. (2011). Cyanobacteria and microalgae: a positive prospect for biofuels. *Bioresour. Technol.* 102, 10163–10172. doi:10.1016/j.biortech.2011.08.030

- Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328. doi:10.1038/nature11478
- Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J., et al. (1992). Metabolic engineering of *Candida tropicalis* for the production of long-chain dicarboxylic acids. *Nat. Biotechnol.* 10, 894–898. doi:10.1038/nbt0892-894
- Quan, J., and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* 4:e6441. doi:10.1371/journal.pone.0006441
- Raschke, D., and Knorr, D. (2009). Rapid monitoring of cell size, vitality and lipid droplet development in the oleaginous yeast *Waltomyces lipofer*. *J. Microbiol. Methods* 79, 178–183. doi:10.1016/j.mimet.2009.08.011
- Ratledge, C. (1994). “Yeasts, moulds, algae and bacteria as sources of lipids,” in *Technological Advances in Improved and Alternative Sources of Lipids*, ed. B. S. Kamel (New York: Springer), 235–291.
- Reppas, N., Ridley, C., Reppas, N., Ridley, C., and Rodley, C. (2010). *Producing Hydrocarbons Comprises Culturing Engineered Cyanobacterium in Culture Medium and Exposing Engineered Cyanobacterium to Light and Carbon Dioxide*. Boston: Joule Unlimited Inc, 7794969–B7794961.
- Richter, H., Qureshi, N., Heger, S., Dien, B., Cotta, M. A., and Angenent, L. T. (2012). Prolonged conversion of n-butyrate to n-butanol with *Clostridium saccharoperbutylacetonicum* in a two-stage continuous culture with in-situ product removal. *Biotechnol. Bioeng.* 109, 913–921. doi:10.1002/bit.24380
- Rucker, J., Paul, J., Pfeifer, B. A., and Lee, K. (2013). Engineering *E. coli* for triglyceride accumulation through native and heterologous metabolic reactions. *Appl. Microbiol. Biotechnol.* 97, 2753–2759. doi:10.1007/s00253-013-4714-3
- Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., and Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl. Environ. Microbiol.* 77, 1718–1727. doi:10.1128/AEM.02580-10
- Runguphan, W., and Keasling, J. D. (2014). Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* 21, 103–113. doi:10.1016/j.ymben.2013.07.003
- Rytter, J. V., Helmark, S., Chen, J., Lezyk, M. J., Solem, C., and Jensen, P. R. (2014). Synthetic promoter libraries for *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 98, 2617–2623. doi:10.1007/s00253-013-5481-x
- Saenge, C., Cheirsilp, B., Suksaroge, T. T., and Bourtoom, T. (2011). Potential use of oleaginous red yeast *Rhodotorula glutinis* for the bioconversion of crude glycerol from biodiesel plant to lipids and carotenoids. *Process Biochem.* 46, 210–218. doi:10.1016/j.procbio.2010.08.009
- Salis, H. M. (2011). The ribosome binding site calculator. *Meth. Enzymol.* 498, 19–42. doi:10.1016/B978-0-12-385120-8.00002-4
- Samish, I., Macdermaid, C. M., Perez-Aguilar, J. M., and Saven, J. G. (2011). Theoretical and computational protein design. *Annu. Rev. Phys. Chem.* 62, 129–149. doi:10.1146/annurev-physchem-032210-103509
- Santamauro, F., Whiffin, F. M., Scott, R. J., and Chuck, C. J. (2014). Low-cost lipid production by an oleaginous yeast cultured in non-sterile conditions using model waste resources. *Biotechnol. Biofuels* 7, 34–43. doi:10.1186/1754-6834-7-42
- Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science* 329, 559–562. doi:10.1126/science.1187936
- Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37, e16. doi:10.1093/nar/gkn991
- Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011). Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 77, 2905–2915. doi:10.1128/AEM.03034-10
- Shen, C. R., and Liao, J. C. (2012). Photosynthetic production of 2-methyl-1-butanol from CO₂ in cyanobacterium *Synechococcus elongatus* PCC7942 and characterization of the native acetohydroxyacid synthase. *Energy Environ. Sci.* 5, 9574–9583. doi:10.1039/C2EE23148D
- Shi, S., Valle-Rodriguez, J. O., Khoomrung, S., Siewers, V., and Nielsen, J. (2012). Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. *Biotechnol. Biofuels* 5, 7. doi:10.1186/1754-6834-5-7
- Shi, S., Valle-Rodriguez, J. O., Siewers, V., and Nielsen, J. (2014). Engineering of chromosomal wax ester synthase integrated *Saccharomyces cerevisiae* mutants for improved biosynthesis of fatty acid ethyl esters. *Biotechnol. Bioeng.* 111, 1740–1747. doi:10.1002/bit.25234
- Si, T., Luo, Y., Xiao, H., and Zhao, H. (2014). Utilizing an endogenous pathway for 1-butanol production in *Saccharomyces cerevisiae*. *Metab. Eng.* 22, 60–68. doi:10.1016/j.ymben.2014.01.002
- Siddiqui, M. S., Thodey, K., Trenchard, I., and Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* 12, 144–170. doi:10.1111/j.1567-1364.2011.00774.x
- Sleight, S. C., Bartley, B. A., Lieviant, J. A., and Sauro, H. M. (2010). In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Res.* 38, 2624–2636. doi:10.1093/nar/gkq179
- Song, J. W., Jeon, E. Y., Song, D. H., Jang, H. Y., Bornscheuer, U. T., Oh, D. K., et al. (2013). Multistep enzymatic synthesis of long-chain α,ω -dicarboxylic and ω -hydroxycarboxylic acids from renewable fatty acids and plant oils. *Angew. Chem. Int. Ed. Engl.* 52, 2534–2537. doi:10.1002/anie.201209187
- Steen, E. J., Chan, R., Prasad, N., Myers, S., Petzold, C. J., Redding, A., et al. (2008). Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* 7, 36. doi:10.1186/1475-2859-7-36
- Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., et al. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559–562. doi:10.1038/nature08721
- Tai, M., and Stephanopoulos, G. (2013). Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab. Eng.* 15, 1–9. doi:10.1016/j.ymben.2012.08.007
- Tan, X., Yao, L., Gao, Q., Wang, W., Qi, F., and Lu, X. (2011). Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab. Eng.* 13, 169–176. doi:10.1016/j.ymben.2011.01.001
- Tapia, E., Anschau, A., Coradini, A. L., Franco, T. T., and Deckmann, A. C. (2012). Optimization of lipid production by the oleaginous yeast *Lipomyces starkeyi* by random mutagenesis coupled to cerulenin screening. *AMB Express* 2, 1–8. doi:10.1186/2191-0855-2-64
- Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., et al. (2004). Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* 432, 1050–1054. doi:10.1038/nature03151
- Torella, J. P., Ford, T. J., Kim, S. N., Chen, A. M., Way, J. C., and Silver, P. A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11290–11295. doi:10.1073/pnas.1307129110
- Tseng, H. C., and Prather, K. L. (2012). Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17925–17930. doi:10.1073/pnas.1209002109
- Valle-Rodriguez, J. O., Shi, S., Siewers, V., and 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. doi:10.1016/j.apenergy.2013.10.003
- Venkataraman, S., Dirks, R. M., Ueda, C. T., and Pierce, N. A. (2010). Selective cell death mediated by small conditional RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16777–16782. doi:10.1073/pnas.1006377107
- Wache, Y. (2013). “Production of dicarboxylic acids and fragrances by *Yarrowia lipolytica*,” in *Yarrowia lipolytica*, ed. G. Barth (Berlin: Springer), 151–170.
- Wache, Y., Aguedo, M., Nicaud, J. M., and Belin, J. M. (2003). Catabolism of hydroxyacids and biotechnological production of lactones by *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* 61, 393–404. doi:10.1007/s00253-002-1207-1
- Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi:10.1038/nature08187
- Wang, Q., Tappel, R. C., Zhu, C., and Nomura, C. T. (2012). Development of a new strategy for production of medium-chain-length polyhydroxyalkanoates by recombinant *Escherichia coli* via inexpensive non-fatty acid feedstocks. *Appl. Environ. Microbiol.* 78, 519–527. doi:10.1128/AEM.07020-11
- Xiong, A. S., Yao, Q. H., Peng, R. H., Li, X., Fan, H. Q., Cheng, Z. M., et al. (2004). A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res.* 32, e98. doi:10.1093/nar/gnh094
- Xue, Z., Sharpe, P. L., Hong, S. P., Yadav, N. S., Xie, D., Short, D. R., et al. (2013). Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nat. Biotechnol.* 31, 734–740. doi:10.1038/nbt.2622

- Zaborske, J. M., Zeitler, B., and Culbertson, M. R. (2013). Multiple transcripts from a 3'-UTR reporter vary in sensitivity to nonsense-mediated mRNA decay in *Saccharomyces cerevisiae*. *PLoS ONE* 8:e80981. doi:10.1371/journal.pone.0080981
- Zhang, F., Carothers, J. M., and Keasling, J. D. (2012a). 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
- Zhang, H., Chong, H., Ching, C. B., Song, H., and Jiang, R. (2012b). Engineering global transcription factor cyclic AMP receptor protein of *Escherichia coli* for improved 1-butanol tolerance. *Appl. Microbiol. Biotechnol.* 94, 1107–1117. doi:10.1007/s00253-012-4012-5
- Zhang, Y., Werling, U., and Edlmann, W. (2012c). SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* 40, e55. doi:10.1093/nar/gkr1288
- Zhang, K., Sawaya, M. R., Eisenberg, D. S., and Liao, J. C. (2008). Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20653–20658. doi:10.1073/pnas.0807157106
- Zhang, L., Zhao, G., and Ding, X. (2011). Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci. Rep.* 1, 141. doi:10.1038/srep00141
- Zheng, Y. N., Li, L. L., Liu, Q., Yang, J. M., Wang, X. W., Liu, W., et al. (2012). Optimization of fatty alcohol biosynthesis pathway for selectively enhanced production of C12/14 and C16/18 fatty alcohols in engineered *Escherichia coli*. *Microb. Cell Fact* 11, 65. doi:10.1186/1475-2859-11-65
- Zhu, B., Cai, G., Hall, E. O., and Freeman, G. J. (2007). In-Fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* 43, 354–359. doi:10.2144/000112536

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 August 2014; accepted: 10 December 2014; published online: 23 December 2014.

Citation: Yu A-Q, Pratomo Juwono NK, Leong SSJ and Chang MW (2014) Production of fatty acid-derived valuable chemicals in synthetic microbes. *Front. Bioeng. Biotechnol.* 2:78. doi: 10.3389/fbioe.2014.00078

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Yu, Pratomo Juwono, Leong and Chang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum*

Sabine A. E. Heider, Natalie Wolf, Arne Hofemeier, Petra Peters-Wendisch and Volker F. Wendisch*

Faculty of Biology and Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

Edited by:

Jean Marie François, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés UMR-CNRS 5504, France

Reviewed by:

Klaas J. Jan Hellingwerf, University of Amsterdam, Netherlands
Tiangang Liu, Wuhan University, China

*Correspondence:

Volker F. Wendisch, Faculty of Biology and CeBiTec, Bielefeld University, Universitätsstr. 25, Bielefeld 33615, Germany
e-mail: volker.wendisch@uni-bielefeld.de

The biotechnologically relevant bacterium *Corynebacterium glutamicum*, currently used for the million ton-scale production of amino acids for the food and feed industries, is pigmented due to synthesis of the rare cyclic C50 carotenoid decaprenoxanthin and its glucosides. The precursors of carotenoid biosynthesis, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate, are synthesized in this organism via the methylerythritol phosphate (MEP) or non-mevalonate pathway. Terminal pathway engineering in recombinant *C. glutamicum* permitted the production of various non-native C50 and C40 carotenoids. Here, the role of engineering isoprenoid precursor supply for lycopene production by *C. glutamicum* was characterized. Overexpression of *dxs* encoding the enzyme that catalyzes the first committed step of the MEP-pathway by chromosomal promoter exchange in a prophage-cured, genome-reduced *C. glutamicum* strain improved lycopene formation. Similarly, an increased IPP supply was achieved by chromosomal integration of two artificial operons comprising MEP pathway genes under the control of a constitutive promoter. Combined overexpression of *dxs* and the other six MEP pathway genes in *C. glutamicum* strain LYC3-MEP was not synergistic with respect to improving lycopene accumulation. Based on *C. glutamicum* strain LYC3-MEP, astaxanthin could be produced in the milligrams per gram cell dry weight range when the endogenous genes *crtE*, *crtB*, and *crtI* for conversion of geranylgeranyl pyrophosphate to lycopene were coexpressed with the genes for lycopene cyclase and β -carotene hydroxylase from *Pantoea ananatis* and carotene C(4) oxygenase from *Brevundimonas aurantiaca*.

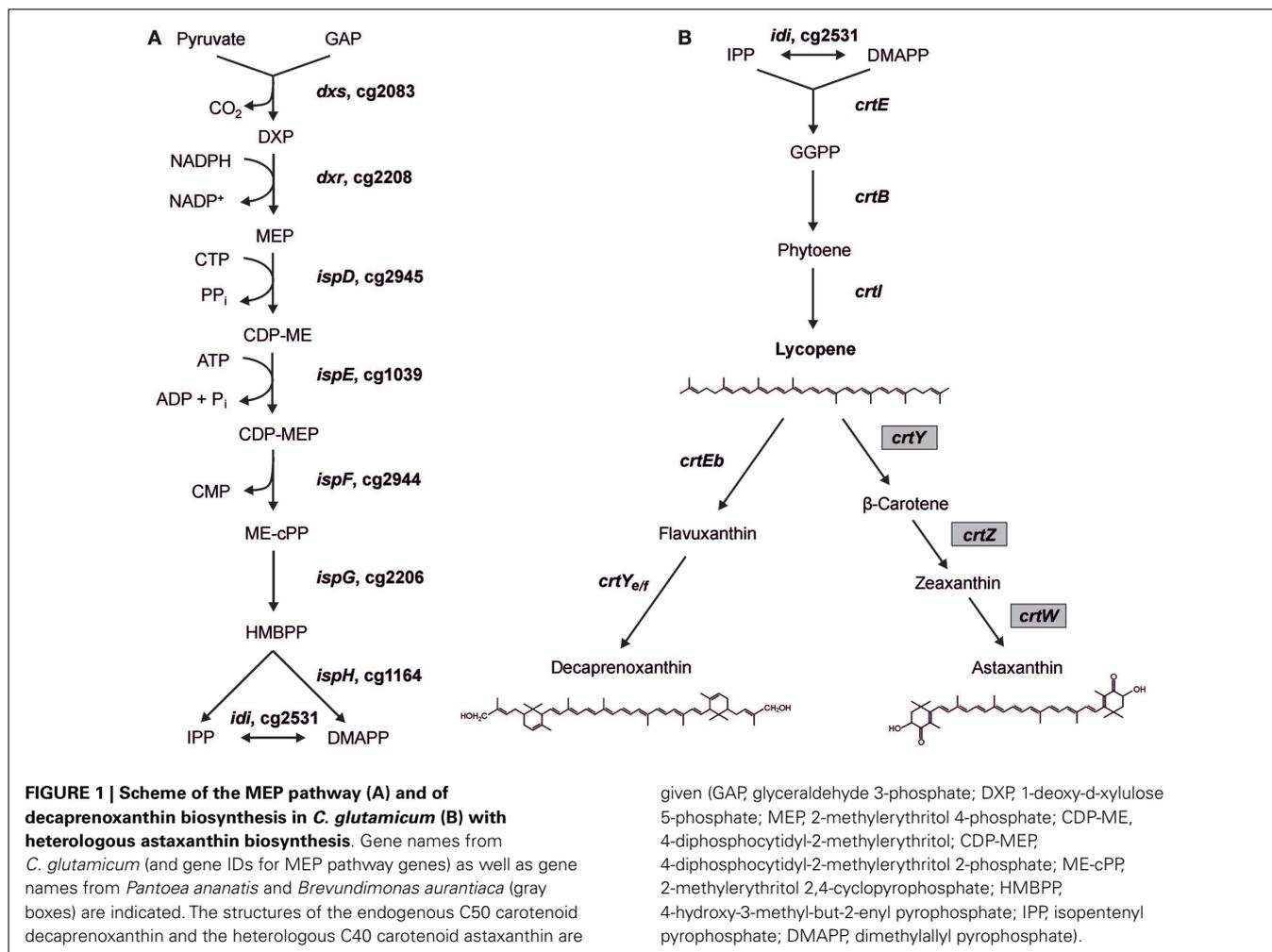
Keywords: carotenoid production, genome-reduced *Corynebacterium glutamicum*, MEP pathway, synthetic operons, astaxanthin

INTRODUCTION

Carotenoids are ubiquitous natural pigments with colors ranging from yellow to red. They are composed of isoprene units and belong to the family of terpenoids. These pigments do not only play important and versatile roles in their biological hosts, but are also suggested to have a beneficial effect on human health. Furthermore, they are intensively applied for food and beverage coloration (Downham and Collins, 2000; Gassel et al., 2013). Hence, carotenoids have received extensive considerable attention and especially the interest for an efficient and environmental-friendly production by microbial hosts is increasing (Lee and Schmidt-Dannert, 2002; Das et al., 2007; Harada and Misawa, 2009; Cutzu et al., 2013). In order to compete with already existing production processes, such as chemical synthesis or extraction from organic material, the large-scale production in microbial hosts requires process as well as strain optimization. One of the most common strategies for enhanced production is the efficient supply of precursor molecules as all carotenoids derive from the universal C5 precursor molecule IPP and its isomer DMAPP. IPP and DMAPP can be synthesized via two independent pathways, the mevalonate (MVA) and the 2-methylerythritol 4-phosphate (MEP) pathway (Rodriguez-Concepcion and Boronat, 2002). The MVA pathway

starts from acetyl-CoA and operates mainly in eukaryotes (mammals, fungi, in the cytoplasm of plant cells), archaea, and a limited number of bacteria. The MEP pathway that starts from pyruvate and glyceraldehyde 3-phosphate and proceeds via the eponymous intermediate MEP was identified much later (Rohmer et al., 1993) and is found in most bacteria as well as in plant plastids (Rohmer, 1999; Lange et al., 2000; Lee and Schmidt-Dannert, 2002). Both pathways also differ regarding redox and energy requirements (Steinbüchel, 2003). As the MEP pathway is present in several pathogens such as *Plasmodium falciparum* and *Mycobacterium tuberculosis*, but not in mammals, it is considered a drug target (Jomaa et al., 1999; Testa and Brown, 2003).

The MEP pathway consists of nine reactions catalyzed by eight enzymes (Figure 1) starting with the transfer of an acetaldehyde group derived from pyruvate to GAP, forming 1-deoxy-D-xylulose 5-phosphate (DXP), in the reaction of DXP synthase Dxs (EC 2.2.1.7). The intermediate DXP is also the precursor for thiamine (vitamine B₁) (Begley et al., 1999) and pyridoxol (vitamine B₆) (Hill et al., 1996) biosynthesis. Subsequently, DXP reductoisomerase Dxr (EC 1.1.1.267) converts DXP to MEP using NADPH as cofactor. MEP is then converted to the cyclic diphosphate 2C-methyl-D-erythritol-2,4-cyclodiphosphate (ME-cPP) by the three



enzymes IspD, IspE, and IspF (Gräwert et al., 2011). ME-cPP is then converted to IPP and DMAPP by a reduction and elimination reaction catalyzed by the two iron-sulfur proteins IspG and IspH (Rohdich et al., 2004). It is proposed that flavodoxin is an essential redox partner for one of the enzymes (Adam et al., 2002; Gräwert et al., 2004; Puan et al., 2005). IPP and DMAPP can be synthesized independently by IspH (Gräwert et al., 2004). IPP and DMAPP often do not occur in the same ratio as for example in *Escherichia coli* IPP is synthesized in a 5:1 proportion to DMAPP (Rohdich et al., 2002; Gräwert et al., 2004; Xiao et al., 2008). The IPP:DMAPP isomerase Idi (EC 5.3.3.2) facilitates the isomerization between IPP and DMAPP. In the case of microorganisms using the MVA pathway produce/synthesize IPP exclusively, isomerases are essential enzymes, whereas in bacteria possessing the MEP pathway *idi* is not essential for the survival of the cells (Hahn et al., 1999; Jülsing et al., 2007).

Corynebacterium glutamicum is a pigmented Gram-positive bacterium with a long and safe history in the food and feed sector as it is used for the fermentative production of amino acids. Annually, about 2.6 million tons of L-glutamate and about 1.95 million tons of L-lysine are produced biotechnologically worldwide (Ajinomoto, Food Products Business. Available from

<http://www.ajinomoto.com/en/ir/pdf/Food-Oct2012.pdf> and [/Feed-useAA-Oct2013.pdf](http://www.ajinomoto.com/en/ir/pdf/Food-Oct2012.pdf), Cited 18 March 2014). Besides amino acids, the diamines cadaverine and putrescine (Mimitsuka et al., 2007; Schneider and Wendisch, 2010) and the alcohols ethanol and isobutanol (Sakai et al., 2007; Blombach and Eikmanns, 2011), among others, can be produced from sugars by recombinant *C. glutamicum* strains. Furthermore, access of *C. glutamicum* to alternative feed stocks like glycerol from the biodiesel process (Meiswinkel et al., 2013), pentoses from lignocellulosics (Gopinath et al., 2011), amino sugars (Uhde et al., 2013; Matano et al., 2014), starch (Seibold et al., 2006), and β -glucans (Tsuchidate et al., 2011) has been engineered.

Recently, the potential of *C. glutamicum* for production of carotenoids has been explored. *C. glutamicum* synthesizes the cyclic C50 carotenoid decaprenoxanthin and its glucosides (Figure 1). Its carotenogenic pathway and the respective genes have been elucidated (Krubasik et al., 2001; Heider et al., 2012, 2014a) and overproduction of the C50 carotenoids decaprenoxanthin, sarcinaxanthin, and C.p. 450 in the milligrams per gram cell dry weight (DCW) range by *C. glutamicum* was achieved by metabolic engineering of the terminal carotenoid pathway (Heider et al., 2014a). Moreover, the heterologous production of the C40

carotenoids β -carotene and zeaxanthin could be established (Heider et al., 2014a) and hydroxylated carotenoids could be produced either as aglycons or as di-glucosides (Heider et al., 2014a). Engineering of *C. glutamicum* for the production of a sesquiterpene, (+)-valencene, was possible as well (Frohwitter et al., 2014).

Based on its genome sequence, all genes of the MEP pathway of *C. glutamicum* have been putatively assigned. However, neither have the respective genes or enzymes of the MEP pathway been functionally analyzed nor has engineering for an increased IPP supply been reported. The MEP pathway genes are distributed over the genome of *C. glutamicum*. The MEP pathway genes *dxs* (cg2083), *ispH* (cg1164), and *idi* (cg2531) are monocistronic, while *dxr* (cg2208), *ispD* (cg2945), *ispE* (cg1039), *ispF* (cg2944), and *ispG* (cg2206) belong to operons. *IspE* is the third gene of the operon cg1037-*ksgA-ispE-cg1040-pdxK* with genes for a putative resuscitation-promoting factor (cg1037), putative dimethyladenosine transferase *KsgA*, and putative pyridoxamine kinase *PdxK*. *IspD* and *ispF* are encoded in the cg2946-*ispDF* operon with cg2946, which codes for a CarD-like transcriptional regulator. *Dxr* and *ispG* are organized in a transcriptional unit separated by an uncharacterized gene (cg2207) putatively encoding a membrane-embedded Zn-dependent protease. In bacteria, two bottlenecks in the MEP pathway were proposed. On the one hand, DXP synthase, which catalyzes the first reaction is claimed to be rate-limiting (Sprenger et al., 1997; Xiang et al., 2007) and is essential in *E. coli* (Sauret-Gueto et al., 2003) and *Bacillus subtilis* (Julsing et al., 2007) and possibly further bacteria. On the other hand, overproduction of *Idi*, which is not essential in bacteria possessing the MEP pathway (Hahn et al., 1999; Julsing et al., 2007), improved carotenoid production (Harker and Bramley, 1999; Kim and Keasling, 2001).

In this study, two synthetic operons (*ispDFE* and *dxr-ispGH*) under control of the strong promoter P_{tuf} of the *C. glutamicum* translation elongation factor EF-Tu gene were integrated into the prophage-cured, genome-reduced *C. glutamicum* strain MB001 (Baumgart et al., 2013). Furthermore, *dxs* was overexpressed from the chromosome by exchanging the endogenous promoter with the P_{tuf} promoter. Finally, *idi* was overexpressed from an IPTG-inducible plasmid. The genome-reduced strain overexpressing all of the eight MEP pathway genes was then shown to be suitable for production of lycopene and endogenous decaprenoxanthin as well as for production of the non-native astaxanthin.

MATERIALS AND METHODS

BACTERIAL STRAINS, MEDIA AND GROWTH CONDITIONS

The strains and plasmids used in this work are listed in Table 1. *C. glutamicum* ATCC13032 was used as wild type (WT), for metabolic engineering the prophage-cured *C. glutamicum* MB001 (Baumgart et al., 2013) was used as platform strain. Precultivation of *C. glutamicum* strains was performed in LB medium or LB with glucose. For cultivation in CGXII medium (Eggeling and Reyes, 2005), precultivated cells were washed once with CGXII medium without carbon source and inoculated to an initial OD₆₀₀ of 1. Glucose was added as carbon and energy source to a concentration of 100 mM. Standard cultivations of *C. glutamicum* were performed at 30°C in a volume of 50 ml in 500 ml flasks with two baffles shaking at 120 rpm. The OD₆₀₀ was measured in dilutions using a Shimadzu UV-1202 spectrophotometer (Duisburg,

Germany). Alternatively, cultivations were performed in 1 ml volume in microtiterplates at 1100 rpm at 30°C using Biolector® micro fermentation system (m2p-labs GmbH, Baesweiler, Germany). For cloning, *E. coli* DH5 α was used as host and cultivated in LB medium at 37°C. When appropriate, kanamycin or spectinomycin was added to concentrations of 25 and 100 $\mu\text{g ml}^{-1}$, respectively. Gene expression was induced by adding 50 μM and 1 mM IPTG, respectively, at inoculation of the main culture.

RECOMBINANT DNA WORK

Plasmids were constructed in *E. coli* DH5 α from PCR-generated fragments (KOD, Novagen, Darmstadt, Germany) and isolated with the QIAprep spin miniprep kit (QIAGEN, Hilden, Germany). Oligonucleotides used in this study were obtained from Eurofins MWG Operon (Ebersberg, Germany) and are listed in Table 2. Standard reactions like restriction, ligation, and PCR were performed as described previously (Sambrook and Russell, 2001). Besides the common ligation reaction, the Gibson assembly has been applied for the construction of plasmids (Gibson et al., 2009). If applicable, PCR products were purified using the PCR purification kit or MinElute PCR purification kit (QIAGEN, Hilden, Germany). For transformation of *E. coli*, the RbCl method was used (Hanahan, 1983) and *C. glutamicum* was transformed via electroporation (van der Rest et al., 1999) at 2.5 kV, 200 Ω , and 25 μF . All cloned DNA fragments were shown to be correct by sequencing.

DELETION OF CAROTENOGENIC GENES IN *C. GLUTAMICUM* MB001

For deletion of the carotenogenic genes *crtY_{eff}* and *crtEb*, encoding the C45/C50 carotenoid ϵ -cyclase and the lycopene elongase, respectively, the suicide vector pK19*mobsacB* was used (Schäfer et al., 1994). Genomic regions flanking the *crtYeb* cluster were amplified from genomic DNA of *C. glutamicum* WT using primer pairs *crtY-A/crtY-B* and *crtEb-C/crtEb-D* (Table 2), respectively. The PCR products were purified and linked by crossover PCR using the primer pair *crtY-A/crtEb-D* (Table 2). The purified PCR product was cloned into pK19*mobsacB* resulting in the construction of deletion vector pK19*mobsacB- δ crtYeb* (Table 1). The targeted deletion of *crtYeb* via two-step homologous recombination as well as the selection for the first and second recombination events were carried out as described previously (Eggeling and Bott, 2005). Deletion of *crtYeb* was verified by PCR analysis of the constructed mutant using primer pair *crtY-E/crtEb-F* (Table 2).

CONSTRUCT DESIGN OF THE SYNTHETIC MEP OPERONS AND THEIR INTEGRATION INTO THE GENOME OF *C. GLUTAMICUM* LYC3

The integration of the synthetic operons Op1 and Op2 was conducted by using the suicide vector pK19*mobsacB* (Schäfer et al., 1994). Op1 consists of the MEP-pathway genes *ispD*, *ispF*, and *ispE* under the control of the constitutive P_{tuf} promoter. *IspD* and *ispF* form a transcription unit and were amplified as such from genomic DNA from *C. glutamicum* WT using the oligonucleotides 5 and 6. The primer pair 7/8 was used to amplify *ispE* from *C. glutamicum* WT, introducing an artificial ribosome binding site (RBS) in front of the gene. The promoter region was amplified using the oligonucleotides 3 and 4. In Op2 *dxr*, *ispG* and *ispH* were combined, by amplification from the *C. glutamicum* WT genome using

Table 1 | Strains and plasmids used in this study.

Strain, plasmid	Relevant characteristics	Source or reference
C. glutamicum STRAINS		
WT	ATCC 13032	Abe et al. (1967)
MB001	Prophage-cured ATCC 13032; in-frame deletion of prophages <i>cgp1</i> (cg1507-cg1524), <i>cgp2</i> (cg1746-cg1752), and <i>cgp3</i> (cg1890-cg2071)	Baumgart et al. (2013)
LYC3	<i>crtY_eY_fEb</i> deletion mutant of <i>C. glutamicum</i> MB001	This work
LYC3-P _{tuf} <i>dxs</i>	LYC3 derivative with <i>dxs</i> (cg2083) under control of the P _{tuf} promoter integrated into the intergenic region of cg2083 and cg2084	This work
LYC3-Op1	LYC3 derivative with <i>ispD</i> (cg2945), <i>ispF</i> (cg2944), and <i>ispE</i> (cg1039) under control of the P _{tuf} promoter integrated into the <i>cgp2</i> cured region between cg1745 and cg1753	This work
LYC3-Op2	LYC3 derivative with <i>dxr</i> (cg2208), <i>ispG</i> (cg2206), and <i>ispH</i> (cg1164) under control of the P _{tuf} promoter integrated into the <i>cgp1</i> cured region between cg1506 and cg1525	This work
LYC3-Op1Op2	LYC3-Op2 derivative with <i>ispD</i> (cg2945), <i>ispF</i> (cg2944), and <i>ispE</i> (cg1039) under control of the P _{tuf} promoter integrated into the <i>cgp2</i> cured region between cg1745 and cg1753	This work
LYC3-MEP	LYC3-Op1Op2 derivative with <i>dxs</i> (cg2083) under control of the P _{tuf} promoter integrated into the intergenic region of cg2083 and cg2084	This work
OTHER STRAINS		
<i>E. coli</i> DH5 α	F ⁻ <i>thi-1 endA1 hsdR17(r⁻ m⁻) supE44 ΔlacU169 (ϕ80lacZΔM15) recA1 gyrA96 relA1</i>	Hanahan (1983)
<i>Pantoea ananatis</i>	ATCC 19321	Misawa et al. (1990)
<i>Brevundimonas aurantiaca</i>	ATCC 15266	Abraham et al. (1999)
PLASMIDS		
pK19 <i>mobsacB</i>	Km ^R ; <i>E. coli/C. glutamicum</i> shuttle vector for construction of insertion and deletion mutants in <i>C. glutamicum</i> (pK18 <i>oriV_{Ec} sacB lacZα</i>)	Schäfer et al. (1993)
pK19 <i>mobsacB</i> - Δ <i>crtYEB</i>	pK19 <i>mobsacB</i> with a <i>crtY_eY_fEb</i> deletion construct	Heider et al. (2014a)
pK19 <i>mobsacB</i> -P _{tuf} <i>dxs</i>	pK19 <i>mobsacB</i> derivative with a <i>tuf</i> promoter region (200 bp upstream of the coding sequence of the <i>tuf</i> gene (cg0587) construct for the promoter exchange of <i>dxs</i>	This work
pK19 <i>mobsacB</i> -Op1	pK19 <i>mobsacB</i> derivative containing the artificial operon <i>ispDFE</i> under the control of the P _{tuf} promoter with an additional ribosome binding site in front of <i>ispE</i> for integration in the <i>cgp2</i> cured region of <i>C. glutamicum</i> MB001	This work
pK19 <i>mobsacB</i> -Op2	pK19 <i>mobsacB</i> derivative containing the artificial operon <i>dxr_ispGH</i> under the control of the P _{tuf} promoter with addition ribosome binding sites in front of <i>ispG</i> and <i>ispH</i> for integration in the <i>cgp2</i> cured region of <i>C. glutamicum</i> MB001	This work
pVWEx1	Km ^R ; <i>E. coli/C. glutamicum</i> shuttle vector for regulated gene expression (P _{tac} , <i>lacI^q</i> , pCG1 <i>oriV_{Cg}</i>)	Peters-Wendisch et al. (2001)
pVWEx1- <i>crtEBI</i>	pVWEx1 derivative for IPTG-inducible expression of <i>crtE</i> and the cluster <i>crtBI</i> from <i>C. glutamicum</i> containing artificial ribosome binding sites each	Heider et al. (2014a)
pVWEx1- <i>dxs</i>	pVWEx1 derivative for IPTG-inducible overexpression of <i>dxs</i> (cg2083) containing an artificial ribosome binding site in front of the gene	This work
pVWEx1- <i>idi</i>	pVWEx1 derivative for IPTG-inducible overexpression of <i>idi</i> (cg2531) containing an artificial ribosome binding site in front of the gene	This work
pVWEx1- <i>glpFKD</i>	pVWEx1 derivative for IPTG-inducible overexpression of <i>glpF</i> , <i>glpK</i> , and <i>glpD</i> from <i>E. coli</i> MG1655	Rittmann et al. (2008)
pEKEx3	Spec ^R ; <i>E. coli/C. glutamicum</i> shuttle vector for regulated gene expression (P _{tac} , <i>lacI^q</i> , pBL1 <i>oriV_{Cg}</i>)	Stansen et al. (2005)

(Continued)

Table 1 | Continued

Strain, plasmid	Relevant characteristics	Source or reference
pEKEx3- <i>crtEbY</i>	pEKEx3 derivative for IPTG-inducible expression of <i>crtEb</i> and <i>crtY</i> from <i>C. glutamicum</i> containing artificial ribosome binding sites in front of each gene	Heider et al. (2014a)
pEKEx3- <i>crtY</i>	pEKEx3 derivative for IPTG-inducible expression of <i>crtY</i> from <i>P. ananatis</i> containing an artificial ribosome binding site in front of the gene	Heider et al. (2014a)
pEKEx3- <i>crtYZ</i>	pEKEx3 derivative for IPTG-inducible expression of <i>crtY</i> and <i>crtZ</i> from <i>P. ananatis</i> containing artificial ribosome binding sites in front of each gene	Heider et al. (2014a)
pEKEx3- <i>crtYZW</i>	pEKEx3 derivative for IPTG-inducible expression of <i>crtY</i> and <i>crtZ</i> from <i>P. ananatis</i> and <i>crtW</i> of <i>Brevundimonas aurantiaca</i> containing artificial ribosome binding sites in front of each gene	This work
pEKEx3- <i>crtZWY</i>	pEKEx3 derivative for IPTG-inducible expression of <i>crtY</i> and <i>crtZ</i> from <i>P. ananatis</i> and <i>crtW</i> of <i>Brevundimonas aurantiaca</i> containing artificial ribosome binding sites in front of each gene in the order as depicted by the name	This work
pEKEx3- <i>dxs</i>	pEKEx3 derivative for IPTG-inducible overexpression of <i>dxs</i> (cg2083) containing an artificial ribosome binding site in front of the gene	This work

the primer pairs 15/16, 17/18, and 19/20, respectively. An artificial RBS in front of *ispG* and *ispH* each was introduced by the oligonucleotides 17 and 19, respectively. Also the genes of Op2 were put under the control of the P_{tuf} promoter, amplified from genomic DNA using the primers 13 and 14. Genomic regions flanking the selected insertion region were amplified from genomic DNA of *C. glutamicum* LYC3 using primer pairs 1/2 and 9/10 for integration in the *cgp2* cured region in the case of Op1, or 11/12 and 20/22 for integration of Op2 in the *cgp1* cured region (Table 2), respectively. The purified PCR products were either linked by crossover PCR or were directly combined together with the plasmid by Gibson assembly (Gibson et al., 2009). The final assembly of the insert with linearized pK19*mobsacB* led to the construction of the respective integration vectors pK19*mobsacB*-Op1 and pK19*mobsacB*-Op2 (Table 1). The following integration of the operon by two-step homologous recombination was performed according to the deletion of genes. The integration of operon1 and 2 was verified by PCR using the primers 29/30 and 31/32, respectively.

PROMOTER EXCHANGE OF THE *dxs* GENE IN *C. GLUTAMICUM* LYC3

The plasmid pK19*mobsacB*- P_{tuf} *dxs* was constructed to replace the native *dxs* promoter with the *tuf* promoter region from *C. glutamicum* WT. For this purpose, the upstream region of *dxs* (483 bp), the 3' part of *dxs* and the *tuf* promoter region [200 bp upstream of the coding sequence of the *tuf* gene (cg0587)] were amplified from chromosomal DNA of *C. glutamicum* LYC3 using the oligonucleotide pairs 27/28, 23/24, and 25/26, respectively (Table 2). By crossover PCR, the *dxs* 3' fragment and the *tuf* promoter region were fused with oligonucleotides 23/26. Afterward, the *dxs* upstream region was fused to this 644 bp long fragment using oligonucleotides 27/26. The final purified PCR product was cloned into pK19*mobsacB* resulting in the vector pK19*mobsacB*- P_{tuf} *dxs* (Table 1). The following process for the promoter exchange by two-step homologous recombination was performed as described earlier for the deletion of genes. The promoter exchange was verified by PCR using the primers *dxs_E* and 33, and sequencing of the PCR product.

OVEREXPRESSION OF CAROTENOGENIC GENES

Plasmids harboring a carotenogenic gene (general abbreviation *crt*), pEKEx3-*crt* or pVWEx1-*crt* allowed an IPTG-inducible overexpression of *crt*. They were constructed on the basis of pEKEx3 (Stansen et al., 2005) or pVWEx1 (Peters-Wendisch et al., 2001), respectively. Amplification of *crt* by polymerase chain reaction (PCR) from genomic DNA of *C. glutamicum* WT, *P. ananatis* and *B. aurentiaca*, which was prepared as described (Eikmanns et al., 1995), was carried out using the respective primers (Table 2). The amplification of the *crt* genes from was based on genomic DNA as template. The amplified products were cloned into the appropriately restricted pEKEx3 or pVWEx1 plasmid DNA.

EXTRACTION ANALYSIS OF CAROTENOIDS

To extract carotenoids from the *C. glutamicum* strains 15 ml aliquots of the cell cultures were centrifuged at 10,000 × g for 15 min and the pellets were washed with deionized H₂O. The pigments were extracted with 10 ml methanol:acetone mixture (7:3) at 60°C for 30 min with thorough vortexing every 10 min. When necessary, several extraction cycles were performed to remove all visible colors from the cell pellet (Heider et al., 2012).

The extraction mixture was centrifuged 10,000 × g for 15 min and the supernatant was transferred to a new tube. The carotenoid content in the extracts was quantified through absorbance at 470 nm by HPLC analysis (see below) and the concentrations were calculated using a standard curve and appropriate dilutions. High performance liquid chromatography (HPLC) analyses of the *C. glutamicum* extracts were performed like described earlier (Heider et al., 2014a) on an Agilent 1200 series HPLC system (Agilent Technologies Sales & Services GmbH & Co., KG, Waldbronn), including a diode array detector (DAD) for UV/visible (Vis) spectrum recording. For separation, a column system consisting of a precolumn (10 mm × 4 mm MultoHigh 100 RP18-5, CS Chromatographie Service GmbH, Langerwehe, Germany) and a main column (ProntoSIL 200-5 C30, 250 mm × 4 mm, CS Chromatographie Service GmbH, Langerwehe, Germany) was used. Quantification of carotenoids was performed using the extracted wavelength

Table 2 | Oligonucleotides used in this study.

Oligonucleotide	Sequence (5' → 3')
<i>crtEb-A</i>	AAAACCCGGGACTACCACTCCCGAGGTT
<i>crtEb-B</i>	CCCATCCACTAAACTTAAACATAGAATTAGTCTTATTTTTCCATCAT
<i>crtEb-C</i>	TGTTTAAGTTTAGTGGATGGGACGATACTGCTAATAGCAATTCATCAGATATAA
<i>crtEb-D</i>	AAAACCCGGGATGTGTGGGAGGCTTCGC
<i>crtEb-E</i>	GGAGACTCAGCGTTTATGTC
<i>crtEb-F</i>	AAAACAATGCGCAGCGCA
<i>crtY-A</i>	AAAAGGATCCAGTCGGCTTCAGCATCC
<i>crtY-B</i>	CCCATCCACTAAACTTAAACATGAAATATCGATGATAGGGATCAA
<i>crtY-E</i>	TTGCACCTGCTGGATACGAA
<i>crtY-F</i>	ATCGCTGCTGAAGGAGATGT
1	GCAGGTCGACTCTAGAGGATCCCCGTGCTTCGCATCGTCTATGTC
2	CATTCGCAGGGTAACGGCCAATAGTTGGGGAAATTTATAAGGATTTG
3	CAAATCCTTATAAATCCCCCAACTATTGGCCGTTACCCTGCGAATG
4	GGGATTCGTGTAGACGACATTGTATGTCCTCCTGGACTTC
5	GAAGTCCAGGAGGACATACAATGTCGTCTACACGAATCCC
6	CGCCTTAGCGGTAATTTTCATCT GAAGGGCCTCCTTTCT TAAAGCCTTCCACACCACTGC
7	GCAGTGGTGTGGAAGGCTTAA GAAAGGAGGCCCTTCAG ATGAAAATTACCGCTAAGGCG
8	CTAATGGACGGTGAAGTATCATTTATGTTATGAAACAGTCAAATGTGTGC
9	GCACACATTTTACTGTTTCATAACATAAATGATACTTCACCGTCCATTAG
10	CCAGTGAATTCGAGCTCGGTACCCCGCCGTATGTAACAAGATTTG
11	GCAGTTCGACTCTAGAGGATCCCCAGTGAAGGATCGGTGCG
12	CATTCGCAGGGTAACGGCCACCTATCTGCTGGCCGGTG
13	CACCGGCCAGCAGATAGGTGGCCGTTACCCTGCGAATG
14	GATCTTTTATGTCACGACTCCCATTGTATGTCCTCCTGGACTTC
15	GAAGTCCAGGAGGACATACAATGGGAGTCGTGACTAAAAAGATC
16	CTAGAAAAGGAAGCCGCAT CTGAAGGGCCTCCTTTCT TACAAGTTGGTTGCCAACCG
17	CGGTTGGCAACCAACTTGTAA GAAAGGAGGCCCTTCAG ATGCGGCTTCCTTTCTAG
18	GCTGATAACAGGTGAGCTCAT CTGAAGGGCCTCCTTTCT TACTTGGTTACCTTCACCTCAG
19	CTGAAGTGAAGGTAACCAAGTAA GAAAGGAGGCCCTTCAG ATGAGCTCACCTGTTATCAGC
20	TCTTACTACTGCGCTAGGTACAGTAAATCTTGTGGCGCAGC
21	GCTGCGCCACAAGAATTAAGTGTACCTAGCGCAAGTAGTAAGA
22	CCAGTGAATTCGAGCTCGGTACCCCTGCTCATCCTTCAACAACGT
23	TGGCCGTTACCCTGCGAATG
24	TGTATGTCCTCCTGGACTTC
25	GAAGTCCAGGAGGACATACAATGGGAATTCTGAACAGTATTTTC
26	CCAGTGAATTCGAGCTCGGTACCCACACTATGCGTGGTATCG
27	GCAGGTCGACTCTAGAGGATCCCCGTGCACTTTCCACACTGGTC
28	CATTCGCAGGGTAACGGCCATGGCGCGAGTCAGACAC
29	TCGCACCATCTACGACAACC
30	CTACGAAGCTGACGCCGAAG
31	GTGGTGCTCGAGAACATAAG
32	CGGTCACCCGTAACAATCAG
33	CAGGATCTTATGCACATAGGACTG
<i>dxs_E</i>	CTGCGGCGTATTCAGAGTTC
<i>Pa_crtY-fw</i>	CTGCAGGTCGACTCTAGAG GAAAGGAGGCCCTTCAG ATGCAACCGCATTATGATCTG
<i>Pa_crtY-rv1</i>	CGGTACCCGGGGATCTTAACGATGAGTCGTGATAATGG
<i>Pa_crtY-rv2</i>	GGCATTCCAAATCCACAACATCTGAAGGGCCTCCTTTCTTAACGATGAGTCGTGATAATGG
<i>Pa_crtZ-fw2</i>	CCATTATGACGACTCATCGTTAA GAAAGGAGGCCCTTCAG ATGTTGTGGATTTGGAATGCC
<i>Pa_crtZ-rv</i>	CGGTACCCGGGGATCTTACTTCCCGGATGCGG
<i>crtW-fw 2</i>	CATCCGGGAAGTAAGATCCCC GAAAGGAGGCCCTTCAG ATGACCGCCGCCGTCG
<i>crtW-rv</i>	CGGTACCCGGGGATCTCAAGACTCGCCGCGCCAC

(Continued)

Table 2 | Continued

Oligonucleotide	Sequence (5' → 3')
A1	<i>CTGCAGGTCGACTCTAGAGGAAAGGAGGCCCTTCAG</i> ATGACCGCCGCCGTCG
A2	<i>CGGTACCCGGGGATCTCAAGACTCGCCGCGCCAC</i>
A3	<i>CAGATCATAATGCGGTTGCAT</i> CTGAAGGGCCTCCTTTCT CAAGACTCGCCGCGCCAC
A4	<i>GTGGCGCGGGGAGTCTTGAGAAAGGAGGCCCTTCAG</i> ATGCAACCGCATTATGATCTG
A6	<i>CTGCAGGTCGACTCTAGAGGAAAGGAGGCCCTTCAG</i> ATGTTGTGGATTTGGAATGCC
A7	<i>CGACGGCGGGGTCAT</i> CTGAAGGGCCTCCTTTCT TACTTCCCGGATGCGG
A8	<i>CCGCATCCGGGAAGTAA</i> GAAAGGAGGCCCTTCAG ATGACCGCCGCCGTCG
pVWEx-fw	CATCATAACGGTTCTGGC
pVWEx-rv	ATCTTCTCTCATCCGCCA
M13 fw	CACAGCGGGAGTGCCTATTGTTTTG
M13 rv	CAGCGATGATCACTTCTGGCTC

Sequence in bold: artificial ribosome binding site; sequence underlined: restriction site; sequence in italics: linker sequence for hybridization.

chromatogram at 470 nm for decaprenoxanthin and carotenoids with corresponding UV/Vis profiles as well as for lycopene and corresponding carotenoids. Lycopene from tomato (Sigma, Steinheim, Germany), astaxanthin (Ehrenstorfer GmbH, Augsburg, Germany), and β -carotene (Merck, Darmstadt, Germany) were used as standards. The carotenoids were dissolved in chloroform according to its solubility and diluted in methanol:acetone (7:3). Due to the lack of appropriate standards decaprenoxanthin and zeaxanthin quantification was calculated based on a β -carotene standard and reported as β -carotene equivalents. The HPLC protocol comprised a gradient elution for 10 min and a mobile phase composition of (A) methanol and (B) methanol/methyl tert-butyl ether/ethyl acetate (5:4:1) starting from 10 to 100% eluent B followed by 20 min of isocratic elution with 100% B. After that, the eluent composition is set back to 10% B for 3 min. The injection volume was 50 μ l and the flow rate was kept constant at 1.4 ml/min.

DXS ACTIVITY ASSAY

The DXS activity of *C. glutamicum* crude extracts was determined using an endpoint assay adopted from Xiang et al. (2007), which is based on the measurement of the remaining pyruvate level in the reaction mixture. The assays were carried out at 30°C in total volume of 1 ml containing 50 mM Tris (pH 7.5), 60 μ M pyruvate, 60 μ M GAP, 10 mM dithiothreitol (DTT), 5 mM MgCl₂, and 600 μ M TPP. Reactions were stopped after 5, 15, 30, and +60 min of incubation by heat inactivation (5 min at 95°C). Subsequent the leftover pyruvate was converted to lactate with lactate dehydrogenase and the concomitant consumption of NADH was determined by fluorescence. Therefore, the reaction was allowed to proceed for 60 min at room temperature. Then, 2.5 U ml⁻¹ lactate dehydrogenase and 0.1 mM NADH was added to the reaction mixture and incubated for 30 min at 37°C. The NADH diminution was determined photometrically at 340 nm.

RESULTS

OVEREXPRESSION OF *dxs* INCREASED LYCOPENE YIELD

The first and often rate-limiting reaction in the MEP pathway is the condensation of pyruvate and GAP to DXP catalyzed by Dxs (Harker and Bramley, 1999; Kim and Keasling, 2001). To test if Dxs is a bottleneck in carotenoid biosynthesis in *C. glutamicum*, *dxs* was

Table 3 | Influence of chromosomal promoter exchange of the 1-deoxy-d-xylulose 5-phosphate synthase gene *dxs* on Dxs activities, growth rates, and lycopene production.

<i>C. glutamicum</i> strain	Growth rate (h ⁻¹)	final OD (600 nm)	Dxs sp. act. (mU mg ⁻¹)	Lycopene production (mg g ⁻¹ DCW)
LYC3	0.45 ± 0.01	27 ± 1	9 ± 1	0.04 ± 0.01
LYC3-P _{tuf} <i>dxs</i>	0.44 ± 0.02	24 ± 2	16 ± 1	0.08 ± 0.01
LYC3(pEKEx3- <i>dxs</i>)	0.38 ± 0.01	22 ± 2	26 ± 3	0.06 ± 0.01

Cells were grown in glucose CGXII minimal medium for 24 h. Means and standard deviations of three cultivations are reported.

overexpressed in *C. glutamicum* LYC3, a mutant derived from the genome-reduced *C. glutamicum* strain MB001 (Baumgart et al., 2013) that accumulates lycopene due to deletion of the lycopene elongase and C45/C50 carotenoid ϵ -cyclase genes *crtEb* and *crtY_{eff}*. To exchange the native *dxs* promoter by the strong constitutive promoter of *tuf* (cg0587), which encodes for the elongation factor EF-Tu (Fukui et al., 2011), the replacement vector pK19mobsacB-P_{tuf}*dxs* was constructed and *C. glutamicum* LYC3-P_{tuf}*dxs* was obtained. Dxs activities measured in crude extracts were about twofold higher in *C. glutamicum* LYC3-P_{tuf}*dxs* (16 ± 1 mU mg⁻¹) than in the control strain *C. glutamicum* LYC3 (Table 3). As consequence of enhanced Dxs activity, lycopene production doubled (0.08 ± 0.01 mg g⁻¹ DCW as compared to 0.04 ± 0.01 mg g⁻¹ DCW) (Table 3). Thus, increased Dxs activity improved lycopene production by *C. glutamicum*. Increased specific Dxs activities were also observed when a plasmid-borne copy of *dxs* was overexpressed from an IPTG-inducible promoter in LYC3, but lycopene production was only slightly improved (Table 3). Hence, chromosomal overexpression proved better and was therefore chosen for subsequent metabolic engineering of the MEP pathway.

OVERPRODUCTION OF ENZYMES CONVERTING DXP TO IPP USING TWO SYNTHETIC OPERONS INTEGRATED INTO THE *C. GLUTAMICUM* CHROMOSOME

For overproduction of the six MEP pathway enzymes catalyzing the conversion of DXP to IPP, two synthetic operons were

constructed and integrated into the chromosome of *C. glutamicum* LYC3. Operon 1 was constructed to drive expression of *ispDF*, which are cotranscribed naturally, fused to *ispE* from P_{tuf} . The RBS of the *tuf* gene was inserted upstream of *ispD*, while the endogenous RBS of *ispF* and a perfect *C. glutamicum* RBS upstream of *ispE* were used. To construct operon 2, *dxr*, *ispG*, and *ispH* were fused for expression from P_{tuf} and perfect *C. glutamicum* RBS were inserted upstream of *ispG* and *ispH* while the RBS of the *tuf* gene was used upstream of *dxr*. Both operons were integrated by homologous recombination into the chromosome of *C. glutamicum* LYC3, which lacks prophages *cgp1* and *cgp2*. Operon 1 was integrated into the chromosome of *C. glutamicum* LYC3 between *cg1506* and *cg1525*, i.e., at the position that harbors prophage *cgp2* in the *C. glutamicum* WT, but which is absent from LYC3, and the resulting strain was named LYC3-Op1. Similarly, *C. glutamicum* LYC3-Op2 was obtained by integrating operon 2 into the chromosome of *C. glutamicum* LYC3 at the position (between *cg1745* and *cg1753*) that in *C. glutamicum* WT harbors prophage *cgp1*, but which is absent from LYC3. The constructed *C. glutamicum* strain LYC3-Op1Op2 contains both operons in the chromosome instead of prophages *cgp1* and *cgp2*. *C. glutamicum* LYC3-Op1 showed slightly higher lycopene accumulation than *C. glutamicum* strains LYC3 and LYC3-Op2. *C. glutamicum* LYC3-Op2 grew slower than LYC3 and LYC3-Op1. *C. glutamicum* LYC3-Op1Op2 that harbors both operons also grew slower, but accumulated almost threefold more lycopene than LYC3. Thus, overexpression of MEP pathway genes from two chromosomally integrated synthetic operons improved lycopene production (Figure 2).

IMPROVED IPP SUPPLY BY CHROMOSOME-BASED ENHANCEMENT OF MEP PATHWAY GENE EXPRESSION

To combine chromosome-based overexpression of the genes necessary for conversion of DXP to IPP with overproduction of Dxs, the first enzyme of the MEP pathway, the endogenous promoter of chromosomal *dxs* was exchanged by P_{tuf} in *C. glutamicum* LYC3-Op1Op2 and the resulting strain was named *C. glutamicum* LYC3-MEP. Surprisingly, LYC3-MEP showed slower growth on solid as well as in liquid medium. Poor growth in liquid glucose medium was accompanied by little lycopene production, although LYC3-MEP colonies appeared well pigmented on plates. Since the central carbon metabolites pyruvate and GAP are the immediate precursors of the MEP pathway, it was tested if lycopene production by *C. glutamicum* LYC3-MEP was affected by the carbon source. To this end, pyruvate and glycerol were tested as carbon sources. Since glycerol is no carbon source for *C. glutamicum* WT, *glpFKD* from *E. coli* encoding the enzyme for conversion of glycerol to GAP were expressed from plasmid pVWEx1-*glpFKD* (Rittmann et al., 2008) in *C. glutamicum* LYC3-MEP. Growth by *C. glutamicum* LYC3-MEP(pVWEx1-*glpFKD*) on glycerol, glycerol + glucose, or glycerol + pyruvate was still impaired, but about twofold more lycopene (around $0.07 \pm 0.01 \text{ mg g}^{-1} \text{ DCW}$) accumulated than with glucose as sole carbon source (Figure 3).

Since *IspH* synthesizes both IPP and DMAPP, but typically not in equimolar amounts (Rohdich et al., 2002; Gräwert et al., 2004; Xiao et al., 2008), it is possible that unbalanced biosynthesis of IPP and DMAPP in *C. glutamicum* LYC3-MEP impairs

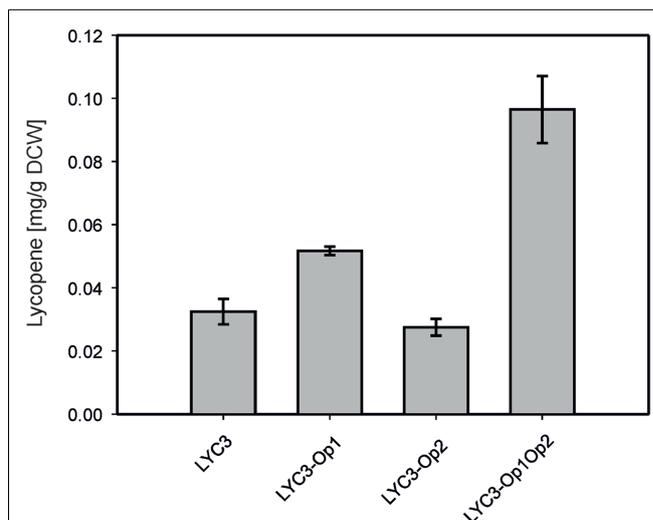


FIGURE 2 | Lycopene production by *C. glutamicum* LYC3 and derived strains expressing the synthetic *ispDFE* operon (LYC3-Op1), the synthetic *dxr-ispGH* operon (LYC3-Op2) or both operons (LYC3-Op1Op2) for overproduction of MEP pathway enzymes. Cells were grown in glucose CGXII minimal medium. Means and standard deviations of three cultivations are shown.

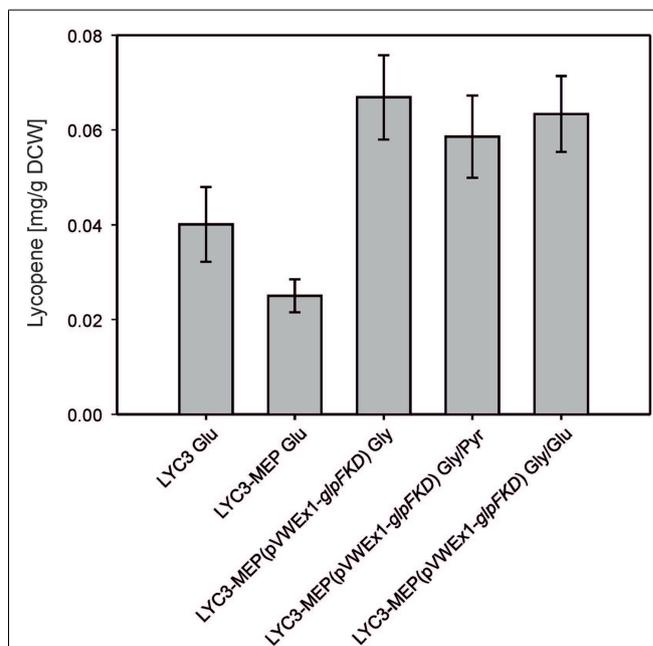


FIGURE 3 | Lycopene production by *C. glutamicum* LYC3-MEP(pVWEx1-*glpFKD*) on glycerol as sole and combined carbon source. LYC3-MEP (pVWEx1-*glpFKD*) cells were grown in CGXII minimal medium with 200 mM glycerol (Gly), 100 mM glycerol + 100 mM pyruvate (Gly/Pyr), or 100 mM glycerol + 50 mM glucose (Gly/Glu), respectively. Expression of *glpFKD* was induced by $50 \mu\text{M}$ IPTG. As reference, lycopene production of the strains LYC3 and LYC3-MEP grown in CGXII minimal medium with 100 mM glucose (Glu) is given. Means and standard deviations of three cultivations are reported.

growth and carotenogenesis. To test this hypothesis, isopentenyl pyrophosphate isomerase *Idi* was overproduced. Indeed, *C. glutamicum* LYC3-MEP(pVWEX1-*idi*)(pEKEx3) produced twofold more lycopene ($0.08 \pm 0.02 \text{ mg g}^{-1} \text{ DCW}$) than *C. glutamicum* strains LYC3, LYC3-MEP, and the empty vector control strain, but still showed impaired growth (Table 4). Thus, a lycopene producing *C. glutamicum* strain with improved IPP supply overexpressing all MEP pathway genes and *idi* could be constructed. However, lycopene production by this strain (Table 4) was comparable to that by *C. glutamicum* strains LYC3-*P_{turf}dxs* (Table 3) and LYC3-Op1Op2 (Figure 2) indicating that the positive effects did not act synergistically. This was also observed when the strains were grown in LB medium supplemented with 100 mM glucose; however, they grew faster (data not shown). Taken together, *C. glutamicum* strains with improved IPP and DMAPP supply showed higher lycopene production than the respective parental strains.

APPLICATION OF *C. GLUTAMICUM* WITH IMPROVED IPP SUPPLY FOR PRODUCTION OF DECAPRENOXANTHIN AND ASTAXANTHIN

To test if *C. glutamicum* LYC3-MEP overexpressing *idi* is suitable for production of the endogenous C50 carotenoid decaprenoxanthin, this strain was transformed with plasmid pEKEx3-*crtEbY*. Expression of lycopene elongase gene *crtEb* and of carotenoid ϵ -cyclase gene *crtY_{eff}* from this plasmid complements the lycopene producing *C. glutamicum* LYC3-MEP, which carries chromosomal *crtEb* and *crtY_{eff}* deletions allowing for decaprenoxanthin biosynthesis. The resulting strain LYC3-MEP(pVWEX1-*idi*)(pEKEx3-*crtEbY*) overproduces

all enzymes of endogenous carotenogenesis except *crtE*, *crtB*, and *crtI* (Figure 1). Although it grew slowly, LYC3-MEP(pVWEX1-*idi*)(pEKEx3-*crtEbY*) produced $0.35 \pm 0.02 \text{ mg g}^{-1} \text{ DCW}$ (Table 5) and, thus, is a genome-reduced strain with improved IPP supply suitable for the overproduction of the endogenous C50 carotenoid decaprenoxanthin.

C. glutamicum has previously been engineered for the production of the non-native C40 carotenoids β -carotene and zeaxanthin (Heider et al., 2014a). When *crtY_{pa}* (PANA_4160) encoding lycopene cyclase from *Pantoea ananatis* was expressed, β -carotene accumulated. Additional expression of *crtZ_{pa}* (PANA_4163), which encodes β -carotene hydroxylase, resulted in partial conversion of β -carotene to zeaxanthin (Heider et al., 2014a). To enable astaxanthin production, *crtW_{Ba}* encoding carotene C(4) oxygenase from *Brevundimonas aurantiaca*, which oxidizes zeaxanthin to yield astaxanthin, was expressed in addition to *crtY_{pa}* and *crtZ_{pa}*. The resulting plasmid pEKEx3-*crtZWY* was used to transform LYC3-MEP(pVWEX1-*idi*). *C. glutamicum* LYC3-MEP(pVWEX1-*idi*)(pEKEx3-*crtZWY*) produced $0.14 \pm 0.01 \text{ mg g}^{-1} \text{ DCW}$ astaxanthin and neither β -carotene nor zeaxanthin accumulated (Table 5). Thus, to the best of our knowledge, this is the documentation of astaxanthin production by recombinant *C. glutamicum*. Although levels were low, LYC3-MEP(pVWEX1-*idi*)(pEKEx3-*crtZWY*) produced astaxanthin as only carotenoid.

Based on our previous findings that overexpression of the genes *crtE*, *crtB*, and *crtI* (Figure 1) strongly increased lycopene production (Heider et al., 2012), as well as decaprenoxanthin production (Heider et al., 2014a); these genes were overexpressed from plasmid pVWEX3-*crtEBI*. The resulting strain *C. glutamicum* LYC3-MEP(pVWEX3-*crtEBI*)(pEKEx3-*crtZWY*) produced $2.1 \pm 1.3 \text{ mg g}^{-1} \text{ DCW}$ β -carotene and $1.2 \pm 0.2 \text{ mg g}^{-1} \text{ DCW}$ zeaxanthin (Table 5), but also ninefold more astaxanthin ($1.2 \pm 0.5 \text{ mg g}^{-1} \text{ DCW}$) than LYC3-MEP(pVWEX1-*idi*)(pEKEx3-*crtZWY*). Thus, it was shown that astaxanthin can be produced by recombinant *C. glutamicum* in the milligrams per gram DCW range.

DISCUSSION

Recently, *C. glutamicum* has been engineered for production of diverse lycopene-derived carotenoids (Heider et al., 2014a) and of a sesquiterpene (Frohwitter et al., 2014). There is an increasing demand for efficient, low-cost, and natural production of terpenoids (Zhu et al., 2014) as they have many applications,

Table 4 | Growth rates and lycopene production by prophage-cured, MEP pathway genes overexpressing *C. glutamicum* strain LYC3-MEP.

<i>C. glutamicum</i> strain	Growth rate (h ⁻¹)	final OD (600 nm)	Lycopene production (mg g ⁻¹ DCW)
LYC3	0.45 ± 0.01	27 ± 1	0.04 ± 0.01
LYC3-MEP	0.16 ± 0.01	23 ± 1	0.03 ± 0.01
LYC3-MEP(pVWEX1)(pEKEx3)	0.15 ± 0.00	23 ± 2	0.04 ± 0.02
LYC3-MEP(pVWEX1- <i>idi</i>)(pEKEx3)	0.13 ± 0.01	20 ± 1	0.08 ± 0.02

Cells were grown in glucose CGXII minimal medium and plasmid carrying strains were induced with 50 μM IPTG. Means and standard deviations of three cultivations are shown.

Table 5 | Astaxanthin and decaprenoxanthin production by recombinant *C. glutamicum* strains with improved IPP supply.

<i>C. glutamicum</i> strain	Production (mg g ⁻¹ DCW)			
	Decaprenoxanthin	β -Carotene	Zeaxanthin	Astaxanthin
LYC3-MEP(pVWEX1)(pEKEx3)	<0.01	<0.01	<0.01	<0.01
LYC3-MEP(pVWEX1- <i>idi</i>)(pEKEx3- <i>crtEbY</i>)	0.4 ± 0.1	<0.01	<0.01	<0.01
LYC3-MEP(pVWEX1- <i>idi</i>)(pEKEx3- <i>crtZWY</i>)	<0.01	<0.01	<0.01	0.1 ± 0.0
LYC3-MEP(pVWEX1- <i>crtEBI</i>)(pEKEx3- <i>crtZWY</i>)	<0.01	2.1 ± 1.3	1.2 ± 0.2	1.2 ± 0.5

Cells were grown in glucose CGXII minimal medium with 50 μM IPTG. Means and standard deviations of three cultivations are reported.

e.g., in the medicinal and nutraceutical industries or as fuels (Martin et al., 2003; Ajikumar et al., 2010; Peralta-Yahya et al., 2011). Besides terminal terpenoid pathway engineering, an efficient supply of the prenyl pyrophosphate precursors is important (Heider et al., 2014b). It could be shown here that MEP pathway engineering to improve IPP supply in *C. glutamicum* improved lycopene production. However, as observed in similar studies of MEP pathway engineering in other bacteria individual bottlenecks may be overcome, but the individual beneficial effects do not necessarily add up (Kim and Keasling, 2001; Martin et al., 2003; Rodriguez-Villalon et al., 2008). Overexpressing the initial MEP pathway gene, *dxs* improved lycopene production by *C. glutamicum* (see **Figure 1**) and by other bacteria (Harker and Bramley, 1999; Matthews and Wurtzel, 2000). However, optimal overexpression levels need to be established since, e.g., chromosomal overexpression proved better than overexpression from a multi-copy plasmid (Yuan et al., 2006). Similarly, when *dxs* was overexpressed in *C. glutamicum* by exchanging the native promoter of *dxs* with the strong constitutive *tuf* promoter more lycopene accumulated than when plasmid-borne *dxs* overexpression, which led to higher Dxs activities, was tested (**Table 3**). The complex interplay of MEP pathway enzymes is also reflected by the fact that overexpression of *dxr*, *ispG*, and *ispH* in LYC3-Op2 only improved lycopene accumulation when combined with overexpression of *ispDF* and *ispE* (Op1) (**Figure 2**). Although lycopene titers obtained with *C. glutamicum* LYC3-Op1Op2 were comparable to the *dxs* overexpressing strain LYC3-*P_{tuf}dxs* (**Figure 2** and **Table 3**), their combination in strain LYC3-MEP was not synergistic and even perturbed growth. This may be explained by accumulation of inhibitory MEP pathway intermediates as shown for *B. subtilis* (Sivý et al., 2011) and *E. coli* (Martin et al., 2003; Zou et al., 2013), from an excessive drain of central metabolic intermediates (Kim and Keasling, 2001) and/or from an imbalance between IPP and DMAPP (Kajiwara et al., 1997). In *C. glutamicum*, improved lycopene production as consequence of overexpression of IPP isomerase gene *idi* was observed in LYC3-MEP (**Table 4**). However, lycopene production by LYC3-MEP overexpressing *idi* was not higher than by LYC3-*P_{tuf}dxs* or by LYC3-Op1Op2. Moreover, when *dxs* was overexpressed in the WT-derived strain $\Delta crtEb$ lycopene production increased from about 0.04 to about 0.12 mg g⁻¹ DCW, but combined overexpression of *dxs* and *idi* did not further increase lycopene production (data not shown). Thus, the perturbed growth may not only be due to an imbalance between IPP and DMAPP.

It remains to be shown if combinatorial approaches to optimize multiple gene expression levels (Zelcbuch et al., 2013; Nowroozi et al., 2014) would improve the IPP precursor supply in *C. glutamicum*. Fine-tuning of gene expression in recombinant *C. glutamicum* by varying promoters (Holátko et al., 2009; van Ooyen et al., 2011; Schneider et al., 2012), RBSs (Schneider et al., 2012), translational start codons (Schneider et al., 2012), or translational stop codons (Jensen and Wendisch, 2013) improved production of amino acids and diamines. In addition, overexpression of heterologous instead of endogenous genes may be beneficial, e.g., as shown for improving isoprene production by *E. coli* via overexpression of two MEP pathway genes *dxs* and *dxr* from *B. subtilis* (Zhao et al., 2011) or by combining overexpression of *xylA* from *Xanthomonas*

campestris with endogenous *xylB* to accelerate xylose utilization of *C. glutamicum* (Meiswinkel et al., 2013).

Besides fine-tuning of MEP pathway gene overexpression, growth, and terpenoid production by recombinant *C. glutamicum* with increased IPP supply could be improved by metabolic pull, i.e., by overexpression of genes of the downstream terpenoid pathway (**Table 5**). Similarly, amorphadiene synthase overexpression prevented accumulation of inhibitory isoprenoid pathway intermediates in *E. coli* (Martin et al., 2003). Overcoming the toxicity of accumulating IPP and DMAPP was successfully used as screening method for the identification of genes that are involved in isoprenoid biosynthesis (Withers et al., 2007). Accumulation of the MEP pathway intermediate ME-cPP inhibits growth and isoprenoid production by recombinant *E. coli*. To abolish its accumulation overexpressing the two enzymes downstream of ME-cPP (*ispG* and *ispH*) needed to be combined with overexpressing an operon for iron-sulfur cluster assembly since both IspG and IspH are containing iron-sulfur clusters (Zou et al., 2013).

To the best of our knowledge, production of astaxanthin by recombinant *C. glutamicum* was shown here for the first time. Astaxanthin is the third most important carotenoid after β -carotene and lutein and its global market amounted to about 230 million US\$ in 2010 (BBC Research, 2011). The economically most significant application of astaxanthin is its use as feed additive in aquaculture industry (Lorenz and Cysewski, 2000; Higuera-Ciapara et al., 2006; Schmidt et al., 2011), but it also exhibits high potential as a nutraceutical and as an approved ingredient for cosmetics due to its remarkably high antioxidative activity (Miki, 1991; Schmidt et al., 2011). Astaxanthin is mainly produced by marine bacteria and microalgae, but only the green freshwater microalga *Haematococcus pluvialis* and the red yeasts *Xanthophyllomyces dendrohous/Phaffia rhodozyma* are established as hosts for commercial production (Bhosale and Bernstein, 2005; Rodriguez-Saiz et al., 2010). Algae-based production of astaxanthin is still more costly than chemical synthesis (Jackson et al., 2008), but markets more and more demand naturally produced carotenoids. The astaxanthin titers by recombinant *C. glutamicum* reported here are in the milligrams per gram DCW range and, thus, they are comparable to yields described for *P. rhodozyma* (ranging from 0.16 to 6.6 mg g⁻¹ DCW (Cruz and Parajo, 1998; Jacobson et al., 1999)). The highest product titer of 9.7 mg g⁻¹ DCW is reported for a *P. rhodozyma* strain improved by metabolic engineering and classical mutagenesis (Gassel et al., 2013), while the highest titer in a recombinant bacterium, i.e., *E. coli* strain was 5.8 mg g⁻¹ DCW astaxanthin (Zelcbuch et al., 2013). Thus, the astaxanthin titers reported for *C. glutamicum* are comparable and it is conceivable that they may be improved further by combining metabolic engineering with classical mutagenesis as in *P. rhodozyma* (Gassel et al., 2013), by combinatorial approaches to gene expression (Zelcbuch et al., 2013), or by high-cell density cultivation since biomass concentrations of up to 95 g DCW/l have been reported for *C. glutamicum* (Riesenberg and Guthke, 1999).

REFERENCES

- Abe, S., Takayama, K., and Kinoshita, S. (1967). Taxonomical studies on glutamic acid producing bacteria. *J. Gen. Appl. Microbiol.* 13, 279–301. doi:10.2323/jgam.13.279

- Abraham, W. R., Strömpl, C., Meyer, H., Lindholst, S., Moore, E. R., Christ, R., et al. (1999). Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *Int. J. Syst. Bacteriol.* 49(Pt 3), 1053–1073. doi:10.1099/00207713-49-3-1053
- Adam, P., Hecht, S., Eisenreich, W., Kaiser, J., Grawert, T., Arigoni, D., et al. (2002). Biosynthesis of terpenes: studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12108–12113. doi:10.1073/pnas.182412599
- Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70–74. doi:10.1126/science.1191652
- Baumgart, M., Unthan, S., Ruckert, C., Sivalingam, J., Grunberger, A., Kalinowski, J., et al. (2013). Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol.* 79, 6006–6015. doi:10.1128/AEM.01634-13
- BBC Research. (2011). *The Global Market for Carotenoids*. Available from: <http://www.bccresearch.com/market-research/food-and-beverage/carotenoids-global-market-fod025d.html>
- Begley, T. P., Downs, D. M., Ealick, S. E., Mclafferty, F. W., Van Loon, A. P., Taylor, S., et al. (1999). Thiamin biosynthesis in prokaryotes. *Arch. Microbiol.* 171, 293–300. doi:10.1007/s002030050713
- Bhosale, P., and Bernstein, P. S. (2005). Microbial xanthophylls. *Appl. Microbiol. Biotechnol.* 68, 445–455. doi:10.1007/s00253-005-0032-8
- Blombach, B., and Eikmanns, B. J. (2011). Current knowledge on isobutanol production with *Escherichia coli*, *Bacillus subtilis* and *Corynebacterium glutamicum*. *Bioeng. Bugs* 2, 346–350. doi:10.4161/bbug.2.6.17845
- Cruz, J. M., and Parajo, J. C. (1998). Improved astaxanthin production by *Xanthophyllomyces dendrorhous* growing on enzymatic wood hydrolysates containing glucose and cellobiose. *Food Chem.* 63, 479–484. doi:10.1016/S0308-8146(98)00061-2
- Cutzu, R., Coi, A., Rosso, F., Bardi, L., Ciani, M., Budroni, M., et al. (2013). From crude glycerol to carotenoids by using a *Rhodotorula glutinis* mutant. *World J. Microbiol. Biotechnol.* 29, 1009–1017. doi:10.1007/s11274-013-1264-x
- Das, A., Yoon, S. H., Lee, S. H., Kim, J. Y., Oh, D. K., and Kim, S. W. (2007). An update on microbial carotenoid production: application of recent metabolic engineering tools. *Appl. Microbiol. Biotechnol.* 77, 505–512. doi:10.1007/s00253-007-1206-3
- Downham, A., and Collins, P. (2000). Colouring our foods in the last and next millennium. *Int. J. Food Sci. Technol.* 35, 5–22. doi:10.1046/j.1365-2621.2000.00373.x
- Eggeling, L., and Bott, M. (eds) (2005). *Handbook of Corynebacterium glutamicum*. Boca Raton: CRC Press.
- Eggeling, L., and Reyes, O. (2005). “Experiments,” in *Handbook of Corynebacterium glutamicum*, eds L. Eggeling and M. Bott (Boca Raton: CRC Press), 3535–3566.
- Eikmanns, B. J., Rittmann, D., and Sahm, H. (1995). Cloning, sequence analysis, expression, and inactivation of the *Corynebacterium glutamicum icd* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. *J. Bacteriol.* 177, 774–782.
- Frohwitter, J., Heider, S., Peters-Wendisch, P., Beekwilder, J., and Wendisch, V. (2014). Production of the sesquiterpene (+)-valencene by metabolically engineered *Corynebacterium glutamicum*. *J. Biotechnol.* doi:10.1016/j.jbiotec.2014.05.032
- Fukui, K., Koseki, C., Yamamoto, Y., Nakamura, J., Sasahara, A., Yuji, R., et al. (2011). Identification of succinate exporter in *Corynebacterium glutamicum* and its physiological roles under anaerobic conditions. *J. Biotechnol.* 154, 25–34. doi:10.1016/j.jbiotec.2011.03.010
- Gassel, S., Schewe, H., Schmidt, I., Schrader, J., and Sandmann, G. (2013). Multiple improvement of astaxanthin biosynthesis in *Xanthophyllomyces dendrorhous* by a combination of conventional mutagenesis and metabolic pathway engineering. *Biotechnol. Lett.* 35, 565–569. doi:10.1007/s10529-012-1103-4
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318
- Gopinath, V., Meiswinkel, T. M., Wendisch, V. F., and Nampoothiri, K. M. (2011). Amino acid production from rice straw and wheat bran hydrolysates by recombinant pentose-utilizing *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 92, 985–996. doi:10.1007/s00253-011-3478-x
- Gräwert, T., Groll, M., Rohdich, F., Bacher, A., and Eisenreich, W. (2011). Biochemistry of the non-mevalonate isoprenoid pathway. *Cell. Mol. Life Sci.* 68, 3797–3814. doi:10.1007/s00018-011-0753-z
- Gräwert, T., Kaiser, J., Zepeck, F., Laupitz, R., Hecht, S., Amslinger, S., et al. (2004). IspH protein of *Escherichia coli*: studies on iron-sulfur cluster implementation and catalysis. *J. Am. Chem. Soc.* 126, 12847–12855. doi:10.1021/ja0471727
- Hahn, F. M., Hurlburt, A. P., and Poulter, C. D. (1999). *Escherichia coli* open reading frame 696 is *idi*, a nonessential gene encoding isopentenyl diphosphate isomerase. *J. Bacteriol.* 181, 4499–4504.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi:10.1016/S0022-2836(83)80284-8
- Harada, H., and Misawa, N. (2009). Novel approaches and achievements in biosynthesis of functional isoprenoids in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 84, 1021–1031. doi:10.1007/s00253-009-2166-6
- Harker, M., and Bramley, P. M. (1999). Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatase in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. *FEBS Lett.* 448, 115–119.
- Heider, S. A., Peters-Wendisch, P., Netzer, R., Stafnes, M., Brautaset, T., and Wendisch, V. F. (2014a). Production and glucosylation of C50 and C40 carotenoids by metabolically engineered *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 98, 1223–1235. doi:10.1007/s00253-013-5359-y
- Heider, S. A., Peters-Wendisch, P., Wendisch, V. F., Beekwilder, J., and Brautaset, T. (2014b). Metabolic engineering for the microbial production of carotenoids and related products with a focus on the rare C50 carotenoids. *Appl. Microbiol. Biotechnol.* 98, 4355–4368. doi:10.1007/s00253-014-5693-8
- Heider, S. A., Peters-Wendisch, P., and Wendisch, V. F. (2012). Carotenoid biosynthesis and overproduction in *Corynebacterium glutamicum*. *BMC Microbiol.* 12:198. doi:10.1186/1471-2180-12-198
- Higuera-Ciapara, I., Felix-Valenzuela, L., and Goycoolea, F. M. (2006). Astaxanthin: a review of its chemistry and applications. *Crit. Rev. Food Sci. Nutr.* 46, 185–196. doi:10.1080/10408690590957188
- Hill, R. E., Himmeldirk, K., Kennedy, I. A., Pauloski, R. M., Sayer, B. G., Wolf, E., et al. (1996). The biogenetic anatomy of vitamin B6. A 13C NMR investigation of the biosynthesis of pyridoxol in *Escherichia coli*. *J. Biol. Chem.* 271, 30426–30435. doi:10.1074/jbc.271.48.30426
- Holátko, J., Elisáková, V., Prouza, M., Sobotka, M., Nesvera, J., and Pátek, M. (2009). Metabolic engineering of the L-valine biosynthesis pathway in *Corynebacterium glutamicum* using promoter activity modulation. *J. Biotechnol.* 139, 203–210. doi:10.1016/j.jbiotec.2008.12.005
- Jackson, H., Braun, C. L., and Ernst, H. (2008). The chemistry of novel xanthophyll carotenoids. *Am. J. Cardiol.* 101, 50D–57D. doi:10.1016/j.amjcard.2008.02.008
- Jacobson, G., Jolly, S., Sedmak, J., Skatrud, T., and Wasileski, J., Inventors. (1999). *Astaxanthin Over-Producing Strains of Phaffia rhodozyma, Methods for Their Cultivation, and Their Use in Animal Feeds*. Assignee: Archer Daniels Midland Company, Decatur, IL. United States Patent Application 08/967,034.
- Jensen, J. V., and Wendisch, V. F. (2013). Ornithine cyclodeaminase-based proline production by *Corynebacterium glutamicum*. *Microb. Cell Fact.* 12, 63. doi:10.1186/1475-2859-12-63
- Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., et al. (1999). Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimarial drugs. *Science* 285, 1573–1576. doi:10.1126/science.285.5433.1573
- Julsing, M. K., Rijpkema, M., Woerdenbag, H. J., Quax, W. J., and Kayser, O. (2007). Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 75, 1377–1384. doi:10.1007/s00253-007-0953-5
- Kajiwara, S., Fraser, P. D., Kondo, K., and Misawa, N. (1997). Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochem. J.* 324(Pt 2), 421–426.
- Kim, S. W., and Keasling, J. D. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.* 72, 408–415. doi:10.1002/1097-0290(20000220)72:4<408::AID-BIT1003>3.0.CO;2-H
- Krubasik, P., Takaichi, S., Maoka, T., Kobayashi, M., Masamoto, K., and Sandmann, G. (2001). Detailed biosynthetic pathway to decaprenoxanthin diglucoside in *Corynebacterium glutamicum* and identification of novel intermediates. *Arch. Microbiol.* 176, 217–223. doi:10.1007/s002030100315

- Lange, B. M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13172–13177. doi:10.1073/pnas.240454797
- Lee, P. C., and Schmidt-Dannert, C. (2002). Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl. Microbiol. Biotechnol.* 60, 1–11. doi:10.1007/s00253-002-1101-x
- Lorenz, R. T., and Cysewski, G. R. (2000). Commercial potential for *Haematococcus microalgae* as a natural source of astaxanthin. *Trends Biotechnol.* 18, 160–167. doi:10.1016/S0167-7799(00)01433-5
- Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802. doi:10.1038/nbt833
- Matano, C., Uhde, A., Youn, J. W., Maeda, T., Clermont, L., Marin, K., et al. (2014). Engineering of *Corynebacterium glutamicum* for growth and L-lysine and lycopene production from N-acetyl-glucosamine. *Appl. Microbiol. Biotechnol.* 98, 5633–5643. doi:10.1007/s00253-014-5676-9
- Matthews, P. D., and Wurtzel, E. T. (2000). Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl. Microbiol. Biotechnol.* 53, 396–400. doi:10.1007/s002530051632
- Meiswinkel, T. M., Rittmann, D., Lindner, S. N., and Wendisch, V. F. (2013). Crude glycerol-based production of amino acids and putrescine by *Corynebacterium glutamicum*. *Bioresour. Technol.* 145, 254–258. doi:10.1016/j.biortech.2013.02.053
- Miki, W. (1991). Biological functions and activities of animal carotenoids. *Pure Appl. Chem.* 63, 141–146. doi:10.1351/pac199163010141
- Mimitsuka, T., Sawai, H., Hatsu, M., and Yamada, K. (2007). Metabolic engineering of *Corynebacterium glutamicum* for cadaverine fermentation. *Biosci. Biotechnol. Biochem.* 71, 2130–2135. doi:10.1271/bbb.60699
- Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., et al. (1990). Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J. Bacteriol.* 172, 6704–6712.
- Nowroozi, F. F., Baidoo, E. E., Ermakov, S., Redding-Johanson, A. M., Batth, T. S., Petzold, C. J., et al. (2014). Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. *Appl. Microbiol. Biotechnol.* 98, 1567–1581. doi:10.1007/s00253-013-5361-4
- Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* 2, 483. doi:10.1038/ncomms1494
- Peters-Wendisch, P. G., Schiel, B., Wendisch, V. F., Katsoulidis, E., Mockel, B., Sahm, H., et al. (2001). Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. *J. Mol. Microbiol. Biotechnol.* 3, 295–300.
- Puan, K. J., Wang, H., Dairi, T., Kuzuyama, T., and Morita, C. T. (2005). *fldA* is an essential gene required in the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *FEBS Lett.* 579, 3802–3806. doi:10.1016/j.febslet.2005.05.047
- Riesenberg, D., and Guthke, R. (1999). High-cell-density cultivation of microorganisms. *Appl. Microbiol. Biotechnol.* 51, 422–430. doi:10.1007/s002530051412
- Rittmann, D., Lindner, S. N., and Wendisch, V. F. (2008). Engineering of a glycerol utilization pathway for amino acid production by *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* 74, 6216–6222. doi:10.1128/AEM.00963-08
- Rodríguez-Concepcion, M., and Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130, 1079–1089. doi:10.1104/pp.007138
- Rodríguez-Saiz, M., De La Fuente, J. L., and Barredo, J. L. (2010). *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Appl. Microbiol. Biotechnol.* 88, 645–658. doi:10.1007/s00253-010-2814-x
- Rodríguez-Villalón, A., Perez-Gil, J., and Rodríguez-Concepcion, M. (2008). Carotenoid accumulation in bacteria with enhanced supply of isoprenoid precursors by upregulation of exogenous or endogenous pathways. *J. Biotechnol.* 135, 78–84. doi:10.1016/j.jbiotec.2008.02.023
- Rohdich, F., Bacher, A., and Eisenreich, W. (2004). Perspectives in anti-infective drug design. The late steps in the biosynthesis of the universal terpenoid precursors, isopentenyl diphosphate and dimethylallyl diphosphate. *Bioorg. Chem.* 32, 292–308. doi:10.1016/j.bioorg.2004.05.012
- Rohdich, F., Hecht, S., Gartner, K., Adam, P., Krieger, C., Amslinger, S., et al. (2002). Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1158–1163. doi:10.1073/pnas.032658999
- Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565–574. doi:10.1039/a709175c
- Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* 295(Pt 2), 517–524.
- Sakai, S., Tsuchida, Y., Okino, S., Ichihashi, O., Kawaguchi, H., Watanabe, T., et al. (2007). Effect of lignocellulose-derived inhibitors on growth of and ethanol production by growth-arrested *Corynebacterium glutamicum*. *R. Appl. Environ. Microbiol.* 73, 2349–2353. doi:10.1128/AEM.01858-07
- Sambrook, J., and Russell, D. (2001). *Molecular Cloning. A Laboratory Manual*, 3rd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sauret-Gueto, S., Ramos-Valdivia, A., Ibanez, E., Boronat, A., and Rodríguez-Concepcion, M. (2003). Identification of lethal mutations in *Escherichia coli* genes encoding enzymes of the methylerythritol phosphate pathway. *Biochem. Biophys. Res. Commun.* 307, 408–415. doi:10.1016/S0006-291X(03)01211-7
- Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Puhler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 145, 69–73. doi:10.1016/0378-1119(94)90324-7
- Schäfer, T., Selig, M., and Schönheit, P. (1993). Acetyl-CoA synthetase (ADP forming) in archaea, a novel enzyme involved in acetate formation and ATP synthesis. *Arch. Microbiol.* 159, 72–83. doi:10.1007/BF00244267
- Schmidt, I., Schewe, H., Gassel, S., Jin, C., Buckingham, J., Humbelin, M., et al. (2011). Biotechnological production of astaxanthin with *Phaffia rhodozyma/Xanthophyllomyces dendrorhous*. *Appl. Microbiol. Biotechnol.* 89, 555–571. doi:10.1007/s00253-010-2976-6
- Schneider, J., Eberhardt, D., and Wendisch, V. F. (2012). Improving putrescine production by *Corynebacterium glutamicum* by fine-tuning ornithine transcarbamoylase activity using a plasmid addiction system. *Appl. Microbiol. Biotechnol.* 95, 169–178. doi:10.1007/s00253-012-3956-9
- Schneider, J., and Wendisch, V. F. (2010). Putrescine production by engineered *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 88, 859–868. doi:10.1007/s00253-010-2778-x
- Seibold, G., Aucher, M., Berens, S., Kalinowski, J., and Eikmanns, B. J. (2006). Utilization of soluble starch by a recombinant *Corynebacterium glutamicum* strain: growth and lysine production. *J. Biotechnol.* 124, 381–391. doi:10.1016/j.jbiotec.2005.12.027
- Sivy, T. L., Fall, R., and Rosenstiel, T. N. (2011). Evidence of isoprenoid precursor toxicity in *Bacillus subtilis*. *Biosci. Biotechnol. Biochem.* 75, 2376–2383. doi:10.1271/bbb.110572
- Sprenger, G. A., Schorken, U., Wiegert, T., Grolle, S., De Graaf, A. A., Taylor, S. V., et al. (1997). Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12857–12862. doi:10.1073/pnas.94.24.12857
- Stansen, C., Uy, D., Delaunay, S., Eggeling, L., Goergen, J. L., and Wendisch, V. F. (2005). Characterization of a *Corynebacterium glutamicum* lactate utilization operon induced during temperature-triggered glutamate production. *Appl. Environ. Microbiol.* 71, 5920–5928. doi:10.1128/AEM.71.10.5920-5928.2005
- Steinbüchel, A. (2003). Production of rubber-like polymers by microorganisms. *Curr. Opin. Microbiol.* 6, 261–270. doi:10.1016/S1369-5274(03)00061-4
- Testa, C. A., and Brown, M. J. (2003). The methylerythritol phosphate pathway and its significance as a novel drug target. *Curr. Pharm. Biotechnol.* 4, 248–259. doi:10.2174/1389201033489784
- Tsuchidate, T., Tateno, T., Okai, N., Tanaka, T., Ogino, C., and Kondo, A. (2011). Glutamate production from beta-glucan using endoglucanase-secreting *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 90, 895–901. doi:10.1007/s00253-011-3116-7
- Uhde, A., Youn, J. W., Maeda, T., Clermont, L., Matano, C., Kramer, R., et al. (2013). Glucosamine as carbon source for amino acid-producing *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 97, 1679–1687. doi:10.1007/s00253-012-4313-8

- van der Rest, M. E., Lange, C., and Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl. Microbiol. Biotechnol.* 52, 541–545. doi:10.1007/s002530051557
- van Ooyen, J., Emer, D., Bussmann, M., Bott, M., Eikmanns, B. J., and Eggeling, L. (2011). Citrate synthase in *Corynebacterium glutamicum* is encoded by two *gltA* transcripts which are controlled by RamA, RamB, and GlxR. *J. Biotechnol.* 154, 140–148. doi:10.1016/j.jbiotec.2010.07.004
- Withers, S. T., Gottlieb, S. S., Lieu, B., Newman, J. D., and Keasling, J. D. (2007). Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. *Appl. Environ. Microbiol.* 73, 6277–6283. doi:10.1128/AEM.00861-07
- Xiang, S., Usunow, G., Lange, G., Busch, M., and Tong, L. (2007). Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* 282, 2676–2682. doi:10.1074/jbc.M610235200
- Xiao, Y., Zhao, Z. K., and Liu, P. (2008). Mechanistic studies of IspH in the deoxyxylulose phosphate pathway: heterolytic C-O bond cleavage at C4 position. *J. Am. Chem. Soc.* 130, 2164–2165. doi:10.1021/ja710245d
- Yuan, L. Z., Rouviere, P. E., Larossa, R. A., and Suh, W. (2006). Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* 8, 79–90. doi:10.1016/j.ymben.2005.08.005
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., et al. (2013). Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* 41, e98. doi:10.1093/nar/gkt151
- Zhao, Y., Yang, J., Qin, B., Li, Y., Sun, Y., Su, S., et al. (2011). Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. *Appl. Microbiol. Biotechnol.* 90, 1915–1922. doi:10.1007/s00253-011-3199-1
- Zhu, F., Zhong, X., Hu, M., Lu, L., Deng, Z., and Liu, T. (2014). In vitro reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in *Escherichia coli*. *Biotechnol. Bioeng.* 111, 1396–1405. doi:10.1002/bit.25198
- Zou, R., Zhou, K., Stephanopoulos, G., and Too, H. P. (2013). Combinatorial engineering of 1-deoxy-D-xylulose 5-phosphate pathway using cross-lapping in vitro assembly (CLIVA) method. *PLoS ONE* 8:e79557. doi:10.1371/journal.pone.0079557

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 May 2014; accepted: 31 July 2014; published online: 20 August 2014.

Citation: Heider SAE, Wolf N, Hofemeier A, Peters-Wendisch P and Wendisch VF (2014) Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum*. *Front. Bioeng. Biotechnol.* 2:28. doi: 10.3389/fbioe.2014.00028

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Heider, Wolf, Hofemeier, Peters-Wendisch and Wendisch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Engineering sugar utilization and microbial tolerance toward lignocellulose conversion

Lizbeth M. Nieves[†], Larry A. Panyon[†] and Xuan Wang^{*}

School of Life Sciences, Arizona State University, Tempe, AZ, USA

Edited by:

Pablo Carbonell, University of Evry, France

Reviewed by:

Weiwen Zhang, Tianjin University, China

Taek Soon Lee, Lawrence Berkeley National Laboratory, USA

*Correspondence:

Xuan Wang, School of Life Sciences, Arizona State University, 427 E. Tyler Mall, Tempe, AZ 85287, USA
e-mail: wangxuan@asu.edu

[†]Lizbeth M. Nieves and Larry A. Panyon have contributed equally to this work.

Production of fuels and chemicals through a fermentation-based manufacturing process that uses renewable feedstock such as lignocellulosic biomass is a desirable alternative to petrochemicals. Although it is still in its infancy, synthetic biology offers great potential to overcome the challenges associated with lignocellulose conversion. In this review, we will summarize the identification and optimization of synthetic biological parts used to enhance the utilization of lignocellulose-derived sugars and to increase the biocatalyst tolerance for lignocellulose-derived fermentation inhibitors. We will also discuss the ongoing efforts and future applications of synthetic integrated biological systems used to improve lignocellulose conversion.

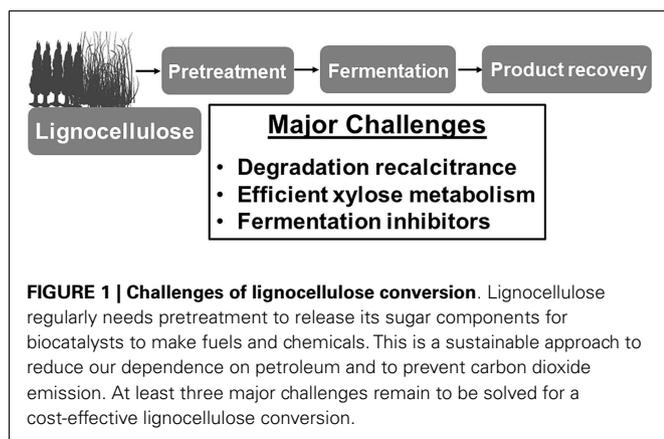
Keywords: synthetic biology, metabolic engineering, lignocellulose, xylose, furan aldehydes

INTRODUCTION

One of the daunting challenges faced by the modern world is our unsustainable dependence on petroleum as the primary source for transportation fuels and many chemical products including solvents, fertilizers, pesticides, and plastics (Service, 2007). To fulfill future societal needs, we have to find a sustainable supply of energy and chemicals. Synthetic biology has emerged as a young discipline with the great potential to construct a novel biological system to produce fuels and chemicals from renewable sources in a cost-effective manner, thus ultimately achieving energy self-sufficiency independent of petroleum. We will apply the synthetic biology definition of “the design and construction of new biological components, such as enzymes, genetic circuits, and cells, or the redesign of existing biological systems” throughout this review (Keasling, 2008). The engineered biological systems created by synthetic biology include enzymes with new functions, genetic circuits, and engineered cells with unique specifications (Cameron et al., 2014; Way et al., 2014). In many cases, the ultimate goal is to rationally manipulate organisms to facilitate novel functions, which do not exist in nature (Cameron et al., 2014; Way et al., 2014). Thus far, synthetic biology has contributed to many fields such as bio-based production (Keasling, 2008; Jarboe et al., 2010), tissue and plant engineering (Bacchus et al., 2012; Moses et al., 2013; Xu et al., 2013; Trantidou et al., 2014), and cell-free synthesis (Lee and Kim, 2013).

Plant biomass (lignocellulose) represents arguably the most important renewable feedstock on the planet. Lignocellulose is a complex matrix of various polysaccharides, phenolic polymers, and proteins that are present in the cell walls of woody plants (Saha, 2003; Girio et al., 2010). Conversion of non-food plant biomass, especially agricultural residues such as corn stover and sugarcane bagasse, avoids the many concerns about the production of fuels and chemicals derived from food sources (Lynd, 1990).

Additionally, non-food-based biofuels offer greater cost reduction in the longer term (Lynd, 1990). For numerous types of agricultural residues, the sugar content is comparable to corn (Saha, 2003). However, the conversion of these sugars from agricultural residues to fuels and chemicals in a cost-effective manner still remains challenging. There are at least three major challenges to be solved before lignocellulose bioconversion becomes financially feasible (Figure 1). First, in contrast to starch, which is easily degraded into fermentable sugar monomers, sugars in lignocellulose are locked into very stable polymeric structures including cellulose and hemicellulose (Saha, 2003; Girio et al., 2010). These polymers are designed by nature to resist deconstruction (Alvira et al., 2010). The crystalline-like fibers of cellulose are encased in a covalently linked mesh of lignin and hemicellulose. Cellulose (30–40% of biomass dry weight) is composed of only D-glucose linked by β -1,4 glycosidic bonds while a mixture of pentoses, especially D-xylose, and hexoses comprises the main component of hemicellulose (20–40% of biomass dry weight) (Saha, 2003). Lignin is not the saccharides polymer but a complex polymer of aromatic alcohols. Different types of lignocellulosic biomass vary in the composition of cellulose, hemicellulose, and lignin (Saha, 2003). Chemical pretreatment processes are commonly required for lignocellulose conversion. Steam pretreatment with dilute mineral acids is an efficient approach to depolymerize hemicellulose into sugar monomers and to increase the accessibility of cellulase enzymes to degrade cellulose (Saha, 2003; Sousa et al., 2009; Alvira et al., 2010). After pretreatment and cellulase digestion, most of the sugars in agricultural waste will be released into the broth and thus ready to be converted into fuels and chemicals if a suitable biocatalyst is applied. The cost of cellulase enzymes is currently still prohibitive to wide application of lignocellulose conversion. Continuing efforts of synthetic biologists from academic and industrial labs are improving cellulase enzymes or



enzyme complexes aiming to develop catalysts that are cost-effective enough to be suitable for commercialization. The recent advancements in cellulases have been extensively reviewed (Elkins et al., 2010; Garvey et al., 2013; Hasunuma et al., 2013; Bommarius et al., 2014) and therefore are not the scope of this review. Second, one of the major carbohydrates in the typical lignocellulosic biomass is D-xylose, a five-carbon aldose, which is difficult for many microbes to metabolize. For instance, common ethanol-producing industrial microbes such as *Saccharomyces cerevisiae* and *Zymomonas mobilis*, do not natively metabolize xylose (Saha, 2003). Although some microbes such as *Escherichia coli* and *Klebsiella pneumonia* have the native xylose metabolic pathway, it is not efficient and is commonly repressed by the presence of glucose (Saha, 2003). Third, side products that hinder cell growth and fermentation such as furfural, 5-hydroxymethylfurfural, formate, acetate, and soluble lignin products are formed during common chemical pretreatment processes (Saha, 2003; Mills et al., 2009). For example, furfural (dehydration product of pentose sugars) is widely regarded as one of the most potent inhibitors (Mills et al., 2009; Geddes et al., 2010a, 2011). It can completely inhibit cellular growth at low concentrations (Zaldivar et al., 1999; Liu and Blaschek, 2010). The concentration of furfural is correlated with the toxicity of dilute acid hydrolyzates (Martinez et al., 2000). Overliming to pH 10 with Ca(OH)₂ or active carbon filter reduces the level of furfural and toxicity, but increases the process complexity and operational cost, thus reducing economic viability (Martinez et al., 2000). There has been a growing interest to engineer industrially related strains to be more resistant to these inhibitors (Wang et al., 2012a,b; Zheng et al., 2012; Geddes et al., 2014; Xiao and Zhao, 2014). For example, beneficial genetic traits to increase host tolerance of furan aldehydes have been identified (Tahezadeh et al., 2000; Liu et al., 2004, 2005, 2008; Gorsich et al., 2006; Petersson et al., 2006; Almeida et al., 2008; Geddes et al., 2014; Glebes et al., 2014a,b; Luhe et al., 2014), knowledge about toxicity mechanisms has been accumulated (Lin et al., 2009a; Miller et al., 2009a,b; Ma and Liu, 2010; Glebes et al., 2014a,b), and thus the integrated synthetic detoxification systems have been constructed and proven effective in different biocatalysts (Wang et al., 2013).

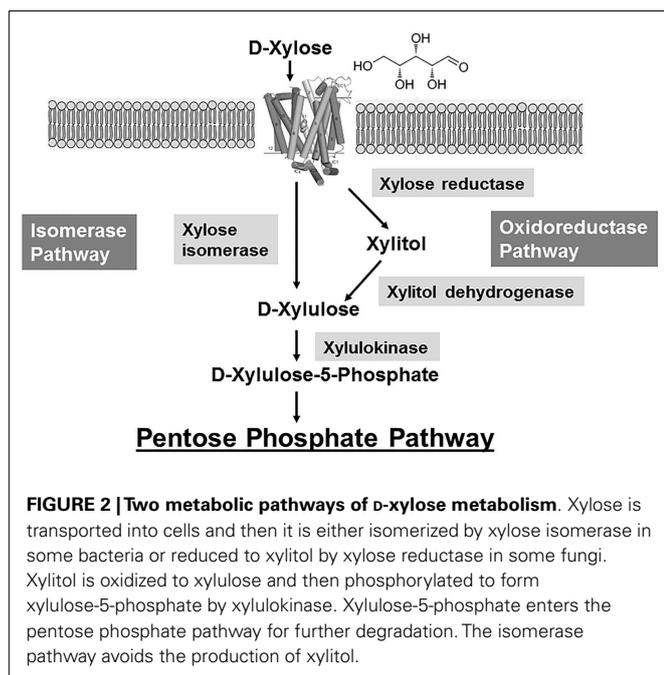
Despite government incentives and mandates, these grand challenges have prohibited the commercialization of lignocellulose

conversion into fuels and chemicals at low cost (Sheridan, 2013). Until now, most efforts for lignocellulose conversion have been devoted to microbial ethanol production. By pathway engineering and metabolic engineering, the microbial hosts can extend their metabolism to produce valuable chemicals other than ethanol from lignocellulose. This review focuses on engineering new biological components by synthetic biology to improve lignocellulose conversion. The past efforts, current status, and future challenges will be discussed.

GENETIC IMPROVEMENT OF UTILIZATION AND TRANSPORT OF MONOSACCHARIDES DERIVED FROM LIGNOCELLULOSE

Hydrolysis of hemicellulose and cellulose into five- and six-carbon sugars by pretreatments provides the mixture of sugars. Microorganisms tend to selectively utilize a preferred sugar, usually D-glucose, by a regulation mechanism called catabolite repression. Synthetic biology has the potential to re-design microbial biology to simultaneously use D-glucose and other pentoses efficiently. Lignocellulosic raw materials commonly contain much higher amounts of D-xylose compared to other pentoses, and therefore, improving xylose fermentation has become a priority (Girio et al., 2010). Xylose degradation is not universal for all microbes in spite of being the most abundant monosaccharide in hemicellulose. At the current stage, most related research still uses the trial-and-error approach to accelerate xylose transport and xylose metabolism. A more quantitative understanding of sugar catabolism is necessary before synthetic biologists are able to predict and design a biological system that efficiently transports and metabolizes sugars.

There are two major metabolic pathways to catabolize xylose: xylose isomerase pathway and oxidoreductase pathway used by bacteria and fungi, respectively (Figure 2). These pathways have been constructed and optimized in industrial biocatalysts such as *S. cerevisiae* and *Z. mobilis*, which cannot natively metabolize xylose. There are comprehensive reviews that excellently summarized this research topic (Jeffries and Jin, 2004; Chu and Lee, 2007; Matsushika et al., 2009; Young et al., 2010; Cai et al., 2012; Kim et al., 2013). Here, we only briefly review some of important past efforts. The xylose oxidoreductase pathway is commonly used by some ascomycetous yeasts such as *Pichia stipitis* (Figure 2). Although the *S. cerevisiae* chromosome has genes encoding xylose reductase, xylitol dehydrogenase, and xylulokinase, their native expression level is too low to support cellular growth when using xylose as the sole carbon source (Yang and Jeffries, 1997; Richard et al., 2000; Traff et al., 2002; Toivari et al., 2004). Anaerobic xylose fermentation by *S. cerevisiae* was first demonstrated by heterologous expression of *XYL1* (Rizzi et al., 1988) and *XYL2* (Rizzi et al., 1989) genes encoding xylose reductase and xylitol dehydrogenase from *P. stipitis* (Kotter et al., 1990; Tantirungkij et al., 1994). However, the xylitol is accumulated as a significant side product when genes *XYL1* and *XYL2* are overexpressed in the recombinant *S. cerevisiae*, which lowers the ethanol yield. The accumulation of xylitol is likely due to the cofactor imbalance of the first two steps in the oxidoreductase pathway (Figure 2). NADPH is the preferred cofactor for xylose reductase to reduce xylose, while NAD is used by xylitol dehydrogenase to oxidize xylitol, resulting in the formation of xylulose (Figure 2). Unlike many bacteria, *S. cerevisiae*



lacks pyridine nucleotide transhydrogenases, which catalyze the conversion between these two reducing cofactors, NADPH and NADH (Nissen et al., 2001). Therefore, this imbalance of cofactors caused by these two reactions will eventually lead to slow kinetics for xylose degradation and xylitol accumulation. Although overexpression of the xylose reductase and xylitol dehydrogenase genes has been shown to enable xylose metabolism in recombinant *S. cerevisiae* strains, overexpression of the xylulokinase gene is often required to create a complete functional heterologous pathway and to further reduce xylitol production (Ho et al., 1998; Jin et al., 2005; Bettiga et al., 2008). One successful example of engineering an efficient xylose-metabolizing yeast is the recombinant *Saccharomyces* sp. strain 1400(pLNH32) (Ho et al., 1998). In this strain, the *P. stipitis* xylose reductase, *P. stipitis* xylitol dehydrogenase, and *S. cerevisiae* xylulokinase genes under the control of the strong native glycolytic promoters were cloned into the plasmid pLNH32 to achieve high expression level. The aerobic conversion of xylose to ethanol has relatively high titer (23 g/L), yield (~0.45 g ethanol/g xylose, theoretic yield is ~0.5 g ethanol/g xylose for ethanol fermentation), and productivity (4 g/L/h) in a complex medium (Ho et al., 1998). Further improvements of ethanol titer and yields in several xylose-fermenting industrial yeast strains such as TMB 3400 and 424A(LNF-ST) have been achieved by utilizing the heterologous xylose oxidoreductase pathway and other genetic modifications that enhance the downstream pentose phosphate pathway (Matsushika et al., 2009). This demonstrates the potential of the xylose oxidoreductase pathway to improve xylose metabolism.

The xylose isomerase pathway, dominantly used by many bacteria including *E. coli* and *Bacillus subtilis*, has also been constructed in *S. cerevisiae* strains. In this pathway, xylose is directly converted to xylulose through a one-step reaction catalyzed by xylose isomerase or other aldose isomerases (Figure 2). This pathway does

not involve xylitol formation and it does not require a reducing cofactor. However, this isomerization reaction thermodynamically favors xylose over xylulose at equilibrium (Jeffries, 1983), which requires an alternative driving force such as efficient downstream reactions to promote the equilibrium moving toward the formation of xylulose (Figure 2). In addition, it has been shown that the expression of functional bacterial xylose isomerase genes often result in inefficient enzymatic activities and thus low xylose utilization (Sarthý et al., 1987; Gardonyi and Hahn-Hagerdal, 2003). The unsuccessful heterologous expression is probably due to the protein misfolding and post-transcriptional modification. Even though the successful synthesis of active xylose isomerases derived from different microbes including thermophilic bacterium *Thermus thermophilus* (Walfridsson et al., 1996), *Piromyces* sp.E2 (Kuyper et al., 2003), *Orpinomyces* (Madhavan et al., 2009), and *Clostridium phytofermentans* (Brat et al., 2009) has been achieved in *S. cerevisiae* at high levels, the rate of growth on xylose was still poor. It is possible that further optimization is needed to increase metabolic flux of downstream reactions, especially the pentose phosphate pathway. Ethanol yield is often higher in these recombinant *S. cerevisiae* using the xylose isomerase pathway than those using the heterologous xylose oxidoreductase pathway because xylitol production is avoided. However, the titer and productivity of *S. cerevisiae* using the xylose isomerase pathway are very low. The *Piromyces* sp. xylose isomerase has been extensively engineered to increase catalytic efficiency, and the *S. cerevisiae* BY4741-S1 derivatives expressing this mutant enzyme improved both its aerobic growth rate and ethanol production (Lee et al., 2012). However, in terms of xylose utilization and ethanol production, these optimized recombinant *S. cerevisiae* strains still perform more poorly with a final ethanol titer lower than 4 g/L. The heterologous xylose isomerase pathway has also been successfully constructed in other biocatalysts such as *Z. mobilis*, a bacterium notable for its bioethanol-producing capabilities, which has been used as a natural fermentative agent in alcoholic beverage production (Skotnicki et al., 1983). Similar to *S. cerevisiae*, *Z. mobilis* cannot metabolize xylose, which limits its application in lignocellulose conversion. In addition, *Z. mobilis* metabolizes glucose into pyruvate using the Entner–Doudoroff pathway instead of glycolysis (Embden–Meyerhof–Parnas pathway) and then converts pyruvate into ethanol and CO₂ (Conway, 1992). Even with the successful expression of the xylose isomerase and xylulokinase genes from *Xanthomonas campestris* or *Klebsiella pneumoniae*, *Z. mobilis* was still unable to grow using xylose as the sole carbon source (Liu et al., 1988; Feldmann et al., 1992). Interestingly, in addition to overexpression of the xylose isomerase and xylulokinase genes, overexpression of the transaldolase and transketolase genes (the main enzymes in the pentose phosphate pathway) resulted in a recombinant *Z. mobilis* with a functional xylose metabolism (Zhang et al., 1995). The resulting strain CP4 (pZB5) is able to convert xylose to ethanol with a higher titer (11 g/L) and yield (0.44 g/g xylose) compared to recombinant *S. cerevisiae* using the xylose isomerase pathway (Zhang et al., 1995). This excellent work strongly suggests a high flux of downstream metabolic reactions such as the pentose phosphate pathway is required for a functional xylose catabolism using the xylose isomerase pathway (Figure 2). A high performance

of xylose to ethanol conversion using a bacterial xylose isomerase pathway has been achieved in a wild-type *E. coli* strain (ATCC9637) after extensive metabolic engineering and adaptive laboratory evolution (Jarboe et al., 2007). The recombinant *E. coli* strain LY180 uses the native xylose isomerase pathway and the *Z. mobilis* ethanol-producing pathway to achieve the efficient conversion of xylose to ethanol with a high titer (45 g/L after 48 h) and yield (0.48 g/g xylose) using mineral salts medium (Miller et al., 2009b; Yomano et al., 2009). These successful examples of engineering *Z. mobilis* and *E. coli* suggest that the bacterial xylose isomerase pathway has the potential for efficient xylose conversion when the metabolic flux in downstream pathways is efficient.

Another challenge for the conversion of sugars derived from lignocellulose is the sequential metabolism of sugar mixtures, a phenomenon called catabolite repression. D-glucose represses the utilization of other sugars such as xylose in many industrial catalysts, thus impeding the rapid and complete utilization of sugar mixtures during fermentation. The mechanism of glucose repression is very complex and involves multiple levels of regulation. For example, *E. coli* has complex glucose repression mechanisms mainly through cyclic AMP, cyclic AMP-binding protein and enzymes of the phosphotransferase system (Kim et al., 2010). There are also other mechanisms involving the inhibition of transport of alternative sugars and a dual transcriptional regulator called Cra (Ramseier, 1996). Strains with the relaxed glucose repression should be able to simultaneously use a heterogeneous sugar mixture. However, genetic perturbation of glucose repression components can disrupt regular glucose metabolism and result in decreased glucose metabolism. It is challenging to engineer a biocatalyst with relaxed glucose repression while keeping a high glucose utilization rate. There are different engineering strategies developed to improve sugar co-utilization (Yomano et al., 2009; Chiang et al., 2013). In a recombinant *E. coli* strain, a combinatory engineering strategy has achieved efficient co-utilization of glucose and xylose (30 g/L for each) in 16 h (Chiang et al., 2013). This genetic engineering strategy includes (1) deletion of *ptsG* (the glucose permease in phosphotransferase system) to release catabolite repression; (2) overexpression of a glucose transporter from *Z. mobilis* to restore glucose transport and metabolism; (3) overexpression of genes *rpiA*, *tktA*, *rpe*, and *talB* to increase pentose phosphate pathway. Recently, a completely different approach to decrease glucose repression has been developed (Galazka et al., 2010; Ha et al., 2011). Cellodextrins are glucose polymers of varying length (two or more glucose monomers) resulting from degradation of cellulose. Wild-type *S. cerevisiae* cannot assimilate cellodextrin because it lacks both the cellodextrin transporter and β -glucosidase capable of hydrolyzing cellodextrin into glucose. By integrating efficient transporters, the complemented hydrolytic enzymes for cellodextrin and the xylose oxidoreductase pathway (Figure 2) into *S. cerevisiae*, this recombinant *S. cerevisiae* strain is able to simultaneously consume cellodextrin and xylose probably because the glucose concentration is never high enough to induce the catabolite repression phenotype (Ha et al., 2011). It is plausible that intracellular hydrolysis of cellodextrin minimizes glucose repression of xylose fermentation allowing this co-consumption (Galazka et al., 2010; Ha et al., 2011). This novel strategy has

the potential to enable efficient co-utilization of sugar mixtures derived from lignocellulose.

Successful lignocellulose conversion requires efficient transport of the mixture of sugars into the cells. The transport of xylose is less efficient than the transport of glucose and often inhibited by D-glucose, which suggests xylose transport is a limiting factor for lignocellulose conversion (Jeffries and Jin, 2004; Luo et al., 2014). Overexpression of homologous and heterologous sugar transporters enables recombinant strains to transport xylose, but have very limited positive effect on xylose fermentation and growth (Weierstall et al., 1999; Hamacher et al., 2002; Gardonyi et al., 2003; Sedlak and Ho, 2004; Saloheimo et al., 2007; Hector et al., 2008; Runquist et al., 2009). To improve xylose transporters, the substrate affinities for xylose of different yeast hexose transporters were altered and selected through mutagenesis and screening approaches (Young et al., 2012, 2014; Farwick et al., 2014). These efforts identified regions and motifs of the hexose transporters as the engineering targets for reprogramming transporter properties (Farwick et al., 2014; Young et al., 2014). However, whether the transport of xylose is the limiting factor for xylose fermentation requires more characterization. Theoretically, xylose uptake becomes a limiting step only when the rate of xylose fermentation is higher than xylose uptake (Cai et al., 2012). The wild-type *S. cerevisiae* CEN.PK2-1C with its native hexose transporter Hxt was reported to be able to take up 0.14 g xylose/h/g dry cell weight in the presence of 50 mM xylose, which exceeds the xylose consumption rate in most recombinant *S. cerevisiae* strains (Hamacher et al., 2002; Cai et al., 2012). Without optimization of sugar transporters, engineered yeast strains already achieved relatively high performance of xylose fermentation using native hexose sugar transporters for xylose uptake (Ho et al., 1998; Sonderegger et al., 2004). The potential beneficial effect of these improved xylose transporters in the recombinant yeast strains with high xylose metabolism remains to be tested.

ENGINEERING BIOCATALYSTS RESISTANT TO LIGNOCELLULOSE INHIBITORS

Pretreatments such as dilute acid at elevated temperature are effective for the hydrolysis of pentose polymers in hemicellulose and also increase the access of cellulase enzymes to cellulose fibers. However, the fermentation of the resulting syrups, called hydrolyzates, is hindered by minor reaction products such as furan aldehydes including furfural and 5-hydroxymethylfurfural (5-HMF), organic acids, and phenolic compounds (Saha, 2003). Furfural and 5-HMF are formed by the dehydration of sugars (pentoses and hexoses, respectively) during pretreatment and more furfural than 5-HMF is present in most hemicellulose hydrolyzates (Saha, 2003; Geddes et al., 2010a,b, 2013). Furfural is of particular importance as a fermentation inhibitor because of its abundance and toxicity (Saha, 2003; Almeida et al., 2009; Mills et al., 2009; Geddes et al., 2010b, 2011). Furfural is more toxic than 5-HMF to industrial catalysts such as *E. coli* and *S. cerevisiae* (Zaldivar et al., 1999; Gorsich et al., 2006). In model studies with various hydrolyzate inhibitors, furfural was unique in potentiating the toxicity of other compounds (Zaldivar et al., 1999). The advancement of engineering tolerance to organic acids and phenolic compounds has been excellently summarized in recent reviews

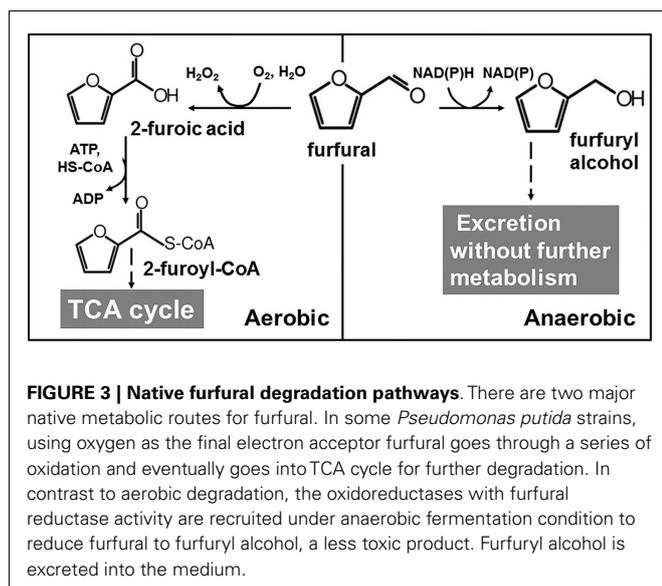
(Mills et al., 2009; Laluce et al., 2012). This review mainly focuses on furan aldehydes as important lignocellulose inhibitors.

A significant amount of effort has been contributed to the identification and optimization of biological components to increase the resistance to furan aldehydes, especially furfural (Table 1). The toxicity mode of furan aldehydes is complex and involves multiple factors (Almeida et al., 2009; Lin et al., 2009a,b; Mills et al., 2009). Cellular growth is arrested in the presence of furan aldehydes and growth resumes after the complete reduction of furfural. This furan-induced delay in growth was observed in both *E. coli* and *S. cerevisiae* (Taherzadeh et al., 2000; Miller et al., 2009b; Wang et al., 2012b). There are two major metabolic pathways to metabolize or reduce furan aldehydes in nature (Figure 3). Some bacteria such as *Cupriavidus basilensis* HMF14 can catabolize furan aldehyde as a sole carbon source when growing aerobically (Koopman et al., 2010). Furan aldehydes such as furfural are firstly oxidized into 2-furoic acid and then further metabolized to 2-oxoglutaric acid that eventually enters the TCA cycle to provide energy and biosynthetic building block (Trudgill, 1969; Koenig and Andreesen, 1990; Koopman et al., 2010) (Figure 3). The key step of this furfural degradation is dependent on oxygen thus limiting its application for anaerobic fermentative production

(Koopman et al., 2010; Ran et al., 2014). *E. coli* and *S. cerevisiae* do not have furan aldehydes oxidative degradation pathways. Under anaerobic fermentation conditions, these microbes use their native oxidoreductases to reduce furan aldehydes to furan alcohol, which is much less toxic (Zaldivar et al., 1999, 2000). Furan alcohols are secreted outside of cells and remain in the fermentation broth without further degradation (Liu and Blaschek, 2010; Wang et al., 2012b). Cells do not grow until furfural or 5-HMF is reduced to a low threshold concentration (~5 mM) (Liu and Blaschek, 2010; Wang et al., 2012b; Ran et al., 2014) (Figure 3). This native detoxification approach has been strengthened in *S. cerevisiae* strains by overexpression of the native oxidoreductase genes such as *ADH1* (Laadan et al., 2008), *ADH6* (Petersson et al., 2006; Almeida et al., 2008; Liu et al., 2008), and *ADH7* (Liu et al., 2008) encoding the enzymes with activities to reduce furan aldehydes (Table 1). Overexpression of these oxidoreductase genes increases the 5-HMF reduction rate and shortens the lag time of cell growth. Interestingly, this native detoxification response causes the growth arrest in *E. coli*. The presence of furfural activates the expression of the *yqhD* gene encoding an oxidoreductase able to reduce furfural to furfuryl alcohol using NADPH as the reducing cofactor (Miller et al., 2009b; Turner et al., 2010). However, NADPH is essential

Table 1 | Beneficial genetic traits for furan aldehydes degradation and tolerance.

Beneficial genetic traits	Microbial host	Proposed detoxification mechanism	Reference
<i>yqhD</i> deletion	<i>E. coli</i>	Avoid the competition for NADPH	Miller et al. (2009b)
<i>pntAB</i> overexpression	<i>E. coli</i>	Increase NADPH level	Miller et al. (2009a)
<i>fucO</i> overexpression	<i>E. coli</i>	Reduce furfural to furfuryl alcohol	Wang et al. (2011)
Mutation of <i>irrE</i>	<i>E. coli</i>	Stress related global regulator	Wang et al. (2012a)
<i>ucpA</i> overexpression	<i>E. coli</i>	Unknown	Wang et al. (2012b)
<i>thyA</i> overexpression	<i>E. coli</i>	Increase the availability of dTMP for DNA repair	Zheng et al. (2012)
<i>fucO</i> missense mutations	<i>E. coli</i>	Improve furfural reductase activity	Zheng et al. (2013)
<i>potE</i> overexpression	<i>E. coli</i>	Polyamine binding to negatively charged cellular constituents	Geddes et al. (2014)
<i>puuP</i> overexpression	<i>E. coli</i>	Polyamine binding to negatively charged cellular constituents	Geddes et al. (2014)
<i>plaP</i> overexpression	<i>E. coli</i>	Polyamine binding to negatively charged cellular constituents	Geddes et al. (2014)
<i>potABCD</i> overexpression	<i>E. coli</i>	Polyamine binding to negatively charged cellular constituents	Geddes et al. (2014)
<i>lpcA</i> overexpression	<i>E. coli</i>	Strengthen cell wall or indirectly increase NADPH availability	Glebes et al. (2014b)
<i>groESL</i> overexpression	<i>E. coli</i>	Possibly related to solvent stress response	Glebes et al. (2014b)
<i>ahpC</i> overexpression	<i>E. coli</i>	Unknown	Glebes et al. (2014a)
<i>yhiH</i> overexpression	<i>E. coli</i>	Unknown	Glebes et al. (2014a)
<i>rna</i> overexpression	<i>E. coli</i>	Unknown	Glebes et al. (2014a)
<i>dicA</i> overexpression	<i>E. coli</i>	Unknown	Glebes et al. (2014a)
<i>ZWF1</i> overexpression	<i>S. cerevisiae</i>	Maintain NADPH levels needed for furan oxidoreductases	Gorsich et al. (2006)
<i>ADH6</i> overexpression	<i>S. cerevisiae</i>	Reduce HMF to alcohol form	Petersson et al. (2006)
<i>ADH1</i> missense mutations	<i>S. cerevisiae</i>	S109P, L116S, and Y294C increase affinity to NADH	Almeida et al. (2008), Laadan et al. (2008)
<i>ADH7</i> overexpression	<i>S. cerevisiae</i>	Reduce HMF to alcohol form	Liu et al. (2008)
<i>YAP1</i> overexpression	<i>S. cerevisiae</i>	Mitigate oxidative stress	Ma and Liu, 2010, Kim and Hahn, 2013
Inactivation of <i>SIZ1</i>	<i>S. cerevisiae</i>	Likely related to oxidative stress	Xiao and Zhao, 2014
Aerobic HMF degradation	<i>C. basilensis</i> HMF14	Oxidize HMF	Koopman et al. (2010)
Aerobic furfural degradation	<i>C. basilensis</i> HMF14 <i>P. putida</i> Fu1, <i>C. basilensis</i> HMF14	Oxidize furfural	Trudgill, 1969, Koopman et al. (2010)

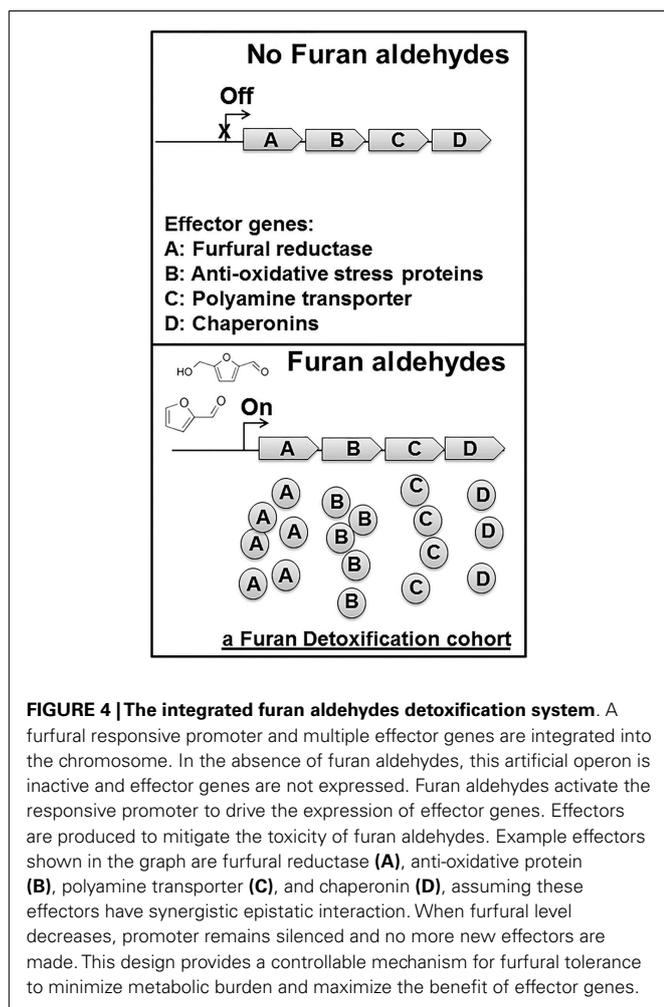


for biosynthesis but is very limited under anaerobic xylose fermentation (Frick and Wittmann, 2005; Miller et al., 2009a). It is this depletion of NADPH by YqhD that has been proposed as the mechanism for growth inhibition in *E. coli* (Miller et al., 2009a,b; Turner et al., 2010). The NADPH-intensive pathway for sulfate assimilation was identified as a sensitive site that may be responsible for growth inhibition (Miller et al., 2009a). Addition of cysteine, deletion of *yqhD*, or increased expression of *pntAB* (transhydrogenase for interconversion of NADH and NADPH) conferred the tolerance to furan aldehydes including furfural and 5-HMF in *E. coli* (Miller et al., 2009a,b, 2010). To accelerate the furfural reduction but avoid using NADPH as the reducing cofactor, an alternative NADH-dependent furfural reductase is desired. A native oxidoreductase, FucO, was identified to have such properties and its overexpression did increase furfural tolerance in different *E. coli* biocatalysts (Wang et al., 2011). FucO normally functions in fucose metabolism and its catalytic efficiency for furfural reduction is low (Wang et al., 2011). The enzyme properties of FucO as a furfural reductase were improved by site-saturated mutagenesis and growth-based selection (Zheng et al., 2013). Overall, optimization of NADH-dependent furfural reductase has potential to shorten the lag phase and to increase tolerance of biocatalysts under fermentation conditions.

A variety of genomic and transcriptomic approaches have yielded many beneficial genetic traits related to furan aldehydes tolerance (Table 1). *S. cerevisiae* gene disruption library was screened for mutants with growth deficiencies in the presence of furfural and *ZWF1* was found to relate to furfural tolerance (Gorsich et al., 2006). Overexpression of *ZWF1* increased furfural tolerance (Gorsich et al., 2006). *ZWF1* encodes glucose-6-phosphate dehydrogenase, which catalyzes the first step of the pentose phosphate pathway, the major pathway providing NADPH when utilizing glucose as the carbon source. A similar approach using genome-wide RNAi screen showed that inactivation of the *SIZ1* gene increased furfural tolerance (Xiao and Zhao, 2014). *SIZ1* encodes E3 SUMO-protein ligase and inactivation of *SIZ1*

increases the tolerance to oxidative stress besides furfural (Xiao and Zhao, 2014). At least part of the toxicity mechanism induced by furfural is suggested to be associated with oxidative stress (Mills et al., 2009). Furfural was shown to induce the accumulation of reactive oxygen species inside of the *S. cerevisiae* cells and to cause damage to mitochondria, vacuole membranes, and cytoskeletons (Allen et al., 2010). Furan aldehydes were also reported to act as thiol-reactive electrophiles, to directly activate Yap1 transcription factor and to deplete glutathione (Kim and Hahn, 2013). Overexpression of either wild-type *YAP1* or its target genes *CTA1* and *CTT1* encoding catalases increased tolerance to furan aldehydes (Kim and Hahn, 2013). Interestingly, furan aldehydes do not induce oxidative responses in *E. coli*. The expression of the genes in major oxidative regulons such as OxyR and SoxRS regulons is not activated by the presence of furfural (Miller et al., 2009a). This strain difference adds another layer of complexity to engineering tolerance of furan aldehydes. In *E. coli*, an oxidoreductase UcpA with an undefined function was found to be associated with furfural tolerance by a transcriptomic analysis and its overexpression increased furan aldehyde tolerance (Wang et al., 2012b). Genomic libraries from three different bacteria were screened for genes that conferred furfural resistance to *E. coli* on plates. Beneficial plasmids containing the *thyA* gene were recovered from all three genomic libraries. The *thyA* gene encodes thymidylate synthase, important for dTMP biosynthesis, suggesting furfural toxicity is possibly related to DNA damage (Zheng et al., 2012). The microarray studies and whole genome sequencing of furfural resistant *E. coli* mutants led to the discovery of some polyamine transporters including PotE, PuuP, PlaP, and PotABCD with a beneficial role for furfural tolerance (Geddes et al., 2014). The detoxification mechanism was proposed to relate to the protection role of polyamine for important cellular constituents such as DNA (Geddes et al., 2014). Other advanced genomic tools such as multiScale Analysis of Library Enrichments (SCALE) (Lynch et al., 2007) and trackable multiplex recombineering (TRMR) (Warner et al., 2010) have been used to identify more furfural related genetic traits in *E. coli* (Glebes et al., 2014a,b). These experiments showed the *lpcA*, *groESL*, *ahpC*, *yhiH*, *rna*, and *dicA* genes are associated with furfural tolerance although the overexpression of these genes individually only showed limited positive effect (Glebes et al., 2014a,b). Another interesting approach is to select a mutant form of the stress-related exogenous regulator IrrE, which confers *E. coli* the tolerance to furan aldehydes (Wang et al., 2012a). Considering the complexity of the toxicity mode induced by furfural, it is not surprising to identify multiple biological parts beneficial for furan tolerance (Table 1). However, all these individual beneficial genetic traits discussed above only provide limited improvement for furan aldehyde tolerance. How to combine multiple beneficial genetic traits to achieve a significant increase of tolerance is a great challenge for synthetic biologists. An ideal synthetic detoxification system should contain a furfural responsive promoter driving the expression of the optimal combinations of different effector genes to minimize metabolic burden and maximize the benefit of effector genes (Figure 4).

There are at least two major challenges for designing such an integrated detoxification system. First, most epistatic interactions between beneficial genetic traits are not predictable and



the experimental search for the optimal combination of multiple effector genes is time-consuming and labor-intensive (Sandoval et al., 2012b). Negative epistatic interactions are present for different beneficial genetic traits for furan aldehyde tolerance. For example, the combination of two beneficial traits, the increased expression of *pntAB* and the deletion of the *yqhD* gene together, made cells less tolerance to furfural than the cells with either one of these two beneficial genetic traits alone (Wang et al., 2013). Further characterization of the beneficial traits in a high-throughput manner is desired to eventually construct an optimal combination of multiple effector genes. Second, the technical challenges to achieve optimal expression of effector genes at the chromosomal level remain to be solved. The effector genes are normally expressed from an expression vector with expensive inducers and antibiotics or other selective conditions. The application of a plasmid-based expression system is undesired in large-scale bio-based production conditions due to the genetic instability, metabolic burden, and the costs (Keasling, 2008; Jarboe et al., 2010). Integration of furan aldehydes detoxification systems into the chromosome is desired. However, it is challenging to achieve the optimal expression of target genes at the chromosomal level, especially when high expression is needed.

CONCLUSION AND FUTURE PERSPECTIVES

Efficient xylose metabolism and tolerance to furan aldehydes are desired features of microbial catalysts used in lignocellulose conversion. Past efforts of synthetic biology focused on identification and optimization of individual biological parts needed for a successful lignocellulose conversion. We have gradually accumulated much knowledge about xylose metabolism and transport, glucose repression, and furan aldehyde toxicity. Limited success of lignocellulose conversion has been achieved using these individual optimized parts (Sandoval et al., 2012a; Wang et al., 2013). Instead of taking a reductionist approach, we are reaching a new phase to characterize the epistatic interactions and to integrate the optimal combinations of different biological parts. This development is dependent on the modular high-throughput approach for epistasis characterization and large-scale genome editing. With the new development of high-throughput techniques and genome editing tools such as CRISPR/Cas9 technology (Doench et al., 2014; Harrison et al., 2014; Sampson and Weiss, 2014), constructing an effective platform strain for lignocellulose conversion is in the scope. The platform strains with high efficiency of sugar co-utilization and tolerance to chemical insult can be used to produce a variety of fuels and chemicals from lignocellulosic biomass by metabolic engineering. These common platforms can also be tuned to different types of biomass by laboratory adaptive evolution.

ACKNOWLEDGMENTS

This work was supported by start-up fund from Arizona State University.

REFERENCES

- Allen, S. A., Clark, W., McCaffery, J. M., Cai, Z., Lanctot, A., Slininger, P. J., et al. (2010). Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 3, 2. doi:10.1186/1754-6834-3-2
- Almeida, J. R., Bertilsson, M., Gorwa-Grauslund, M. F., Gorsich, S., and Liden, G. (2009). Metabolic effects of furfuraldehydes and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* 82, 625–638. doi:10.1007/s00253-009-1875-1
- Almeida, J. R., Roder, A., Modig, T., Laadan, B., Liden, G., and Gorwa-Grauslund, M. F. (2008). NADH- vs NADPH-coupled reduction of 5-hydroxymethyl furfural (HMF) and its implications on product distribution in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 78, 939–945. doi:10.1007/s00253-008-1364-y
- Alvira, P., Tomas-Pejo, E., Ballesteros, M., and Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresour. Technol.* 101, 4851–4861. doi:10.1016/j.biortech.2009.11.093
- Bacchus, W., Lang, M., El-Baba, M. D., Weber, W., Stelling, J., and Fussenegger M. (2012). Synthetic two-way communication between mammalian cells. *Nat. Biotechnol.* 30, 991–996. doi:10.1038/nbt.2351
- Bettiga, M., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. F. (2008). Comparing the xylose reductase/xyloxylic acid dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting *Saccharomyces cerevisiae* strains. *Biotechnol. Biofuels* 1, 16. doi:10.1186/1754-6834-1-16
- Bommarius, A. S., Sohn, M., Kang, Y., Lee, J. H., and Realf, M. J. (2014). Protein engineering of cellulases. *Curr. Opin. Biotechnol.* 29C, 139–145. doi:10.1016/j.copbio.2014.04.007
- Brat, D., Boles, E., and Wiedemann, B. (2009). Functional expression of a bacterial xylose isomerase in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 75, 2304–2311. doi:10.1128/AEM.02522-08
- Cai, Z., Zhang, B., and Li, Y. (2012). Engineering *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: reflections and perspectives. *Biotechnol. J.* 34–46. doi:10.1002/biot.201100053

- Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12, 381–390. doi:10.1038/nrmicro3239
- Chiang, C. J., Lee, H. M., Guo, H. J., Wang, Z. W., Lin, L. J., and Chao, Y. P. (2013). Systematic approach to engineer *Escherichia coli* pathways for co-utilization of a glucose-xylose mixture. *J. Agric. Food Chem.* 61, 7583–7590. doi:10.1021/jf401230r
- Chu, B. C., and Lee, H. (2007). Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. *Biotechnol. Adv.* 25, 425–441. doi:10.1016/j.biotechadv.2007.04.001
- Conway, T. (1992). The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol. Rev.* 9, 1–27. doi:10.1111/j.1574-6968.1992.tb05822.x
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. doi:10.1038/nbt.3026
- Elkins, J. G., Raman, B., and Keller, M. (2010). Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Curr. Opin. Biotechnol.* 21, 657–662. doi:10.1016/j.copbio.2010.05.008
- Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., and Boles, E. (2014). Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5159–5164. doi:10.1073/pnas.1323464111
- Feldmann, S. D., Sahm, H., and Sprenger, G. A. (1992). Pentose metabolism in *Zymomonas mobilis* wild-type and recombinant strains. *Appl. Microbiol. Biotechnol.* 38, 354–361. doi:10.1007/BF00170086
- Frick, O., and Wittmann, C. (2005). Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative ¹³C flux analysis. *Microb. Cell Fact.* 4, 30. doi:10.1186/1475-2859-4-30
- Galazka, J. M., Tian, C. G., Beeson, W. T., Martinez, B., Glass, N. L., and Cate, J. H. D. (2010). Celodextrin transport in yeast for improved biofuel production. *Science* 330, 84–86. doi:10.1126/science.1192838
- Gardonyi, M., and Hahn-Hagerdal, B. (2003). The *Streptomyces rubiginosus* xylose isomerase is misfolded when expressed in *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* 32, 252–259. doi:10.1016/S0141-0229(02)00285-5
- Gardonyi, M., Jeppsson, M., Liden, G., Gorwa-Grausland, M. F., and Hahn-Hagerdal, B. (2003). Control of xylose consumption by xylose transport in recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 82, 818–824. doi:10.1002/bit.10631
- Garvey, M., Klose, H., Fischer, R., Lambert, C., and Commandeur, U. (2013). Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol.* 31, 581–593. doi:10.1016/j.tibtech.2013.06.006
- Geddes, C. C., Mullinnix, M. T., Nieves, I. U., Hoffman, R. W., Sagues, W. J., York, S. W., et al. (2013). Seed train development for the fermentation of bagasse from sweet sorghum and sugarcane using a simplified fermentation process. *Bioresour. Technol.* 128, 716–724. doi:10.1016/j.biortech.2012.09.121
- Geddes, C. C., Mullinnix, M. T., Nieves, I. U., Peterson, J. J., Hoffman, R. W., York, S. W., et al. (2010a). Simplified process for ethanol production from sugarcane bagasse using hydrolytase-resistant *Escherichia coli* strain MM160. *Bioresour. Technol.* 102, 2702–2711. doi:10.1016/j.biortech.2010.10.143
- Geddes, C. C., Peterson, J. J., Roslander, C., Zacchi, G., Mullinnix, M. T., Shanmugam, K. T., et al. (2010b). Optimizing the saccharification of sugar cane bagasse using dilute phosphoric acid followed by fungal cellulases. *Bioresour. Technol.* 101, 1851–1857. doi:10.1016/j.biortech.2009.09.070
- Geddes, C. C., Nieves, I. U., and Ingram, L. O. (2011). Advances in ethanol production. *Curr. Opin. Biotechnol.* 22, 312–319. doi:10.1016/j.copbio.2011.04.012
- Geddes, R. D., Wang, X., Yomano, L. P., Miller, E. N., Zheng, H., Shanmugam, K. T., et al. (2014). Polyamine transporters and polyamines increase furfural tolerance during xylose fermentation with ethanologenic *Escherichia coli* strain LY180. *Appl. Environ. Microbiol.* 80, 5955–5964. doi:10.1128/AEM.01913-14
- Girio, F. M., Fonseca, C., Carvalho, F., Duarte, L. C., Marques, S., and Bogel-Lukasik, R. (2010). Hemicelluloses for fuel ethanol: a review. *Bioresour. Technol.* 101, 4775–4800. doi:10.1016/j.biortech.2010.01.088
- Glebes, T. Y., Sandoval, N. R., Gillis, J. H., and Gill, R. T. (2014a). Comparison of genome-wide selection strategies to identify furfural tolerance genes in *Escherichia coli*. *Biotechnol. Bioeng.* 112, 129–140. doi:10.1002/bit.25325
- Glebes, T. Y., Sandoval, N. R., Reeder, P. J., Schilling, K. D., Zhang, M., and Gill, R. T. (2014b). Genome-wide mapping of furfural tolerance genes in *Escherichia coli*. *PLoS ONE* 9:e87540. doi:10.1371/journal.pone.0087540
- Gorsich, S. W., Dien, B. S., Nichols, N. N., Slininger, P. J., Liu, Z. L., and Skory, C. D. (2006). Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 71, 339–349. doi:10.1007/s00253-005-0142-3
- Ha, S. J., Galazka, J. M., Kim, S. R., Choi, J. H., Yang, X. M., Seo, J. H., et al. (2011). Engineered *Saccharomyces cerevisiae* capable of simultaneous cellulose and xylose fermentation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 504–509. doi:10.1073/pnas.1010456108
- Hamacher, T., Becker, J., Gardonyi, M., Hahn-Hagerdal, B., and Boles, E. (2002). Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. *Microbiology* 148, 2783–2788.
- Harrison, M. M., Jenkins, B. V., O'Connor-Giles, K. M., and Wildonger, J. (2014). A CRISPR view of development. *Genes Dev.* 28, 1859–1872. doi:10.1101/gad.248252.114
- Hasunuma, T., Okazaki, F., Okai, N., Hara, K. Y., Ishii, J., and Kondo, A. (2013). A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresour. Technol.* 135, 513–522. doi:10.1016/j.biortech.2012.10.047
- Hector, R. E., Qureshi, N., Hughes, S. R., and Cotta, M. A. (2008). Expression of a heterologous xylose transporter in a *Saccharomyces cerevisiae* strain engineered to utilize xylose improves aerobic xylose consumption. *Appl. Microbiol. Biotechnol.* 80, 675–684. doi:10.1007/s00253-008-1583-2
- Ho, N. W., Chen, Z., and Brainard, A. P. (1998). Genetically engineered *Saccharomyces* yeast capable of effective cofermentation of glucose and xylose. *Appl. Environ. Microbiol.* 64, 1852–1859.
- Jarboe, L. R., Grabar, T. B., Yomano, L. P., Shanmugam, K. T., and Ingram, L. O. (2007). Development of ethanologenic bacteria. *Adv. Biochem. Eng. Biotechnol.* 108, 237–261.
- Jarboe, L. R., Zhang, X., Wang, X., Moore, J. C., Shanmugam, K. T., and Ingram, L. O. (2010). Metabolic engineering for production of bio-renewable fuels and chemicals: contributions of synthetic biology. *J. Biomed. Biotechnol.* 2010, 761042. doi:10.1155/2010/761042
- Jeffries, T. W. (1983). Utilization of xylose by bacteria, yeasts, and fungi. *Adv. Biochem. Eng. Biotechnol.* 27, 1–32.
- Jeffries, T. W., and Jin, Y. S. (2004). Metabolic engineering for improved fermentation of pentoses by yeasts. *Appl. Microbiol. Biotechnol.* 63, 495–509. doi:10.1007/s00253-003-1450-0
- Jin, Y. S., Alper, H., Yang, Y. T., and Stephanopoulos, G. (2005). Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach. *Appl. Environ. Microbiol.* 71, 8249–8256. doi:10.1128/AEM.71.12.8249-8256.2005
- Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* 3, 64–76. doi:10.1021/cb700243a
- Kim, D., and Hahn, J. S. (2013). Roles of the Yap1 transcription factor and antioxidants in *Saccharomyces cerevisiae*'s tolerance to furfural and 5-hydroxymethylfurfural, which function as thiol-reactive electrophiles generating oxidative stress. *Appl. Environ. Microbiol.* 79, 5069–5077. doi:10.1128/AEM.00643-13
- Kim, J. H., Block, D. E., and Mills, D. A. (2010). Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. *Appl. Microbiol. Biotechnol.* 88, 1077–1085. doi:10.1007/s00253-010-2839-1
- Kim, S. R., Park, Y. C., Jin, Y. S., and Seo, J. H. (2013). Strain engineering of *Saccharomyces cerevisiae* for enhanced xylose metabolism. *Biotechnol. Adv.* 31, 851–861. doi:10.1016/j.biotechadv.2013.03.004
- Koenig, K., and Andreesen, J. R. (1990). Xanthine dehydrogenase and 2-furoyl-coenzyme-A dehydrogenase from *Pseudomonas putida* Ful – 2 molybdenum-containing dehydrogenases of novel structural composition. *J. Bacteriol.* 172, 5999–6009.
- Koopman, F., Wierckx, N., De Winde, J. H., and Ruijsenaars, H. J. (2010). Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of *Cupriavidus basilensis* HMF14. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4919–4924. doi:10.1073/pnas.0913039107
- Kotter, P., Amore, R., Hollenberg, C. P., and 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. doi:10.1007/BF00327019
- Kuyper, M., Harhangi, H. R., Stave, A. K., Winkler, A. A., Jetten, M. S., De Laat, W. T., et al. (2003). High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*? *FEMS Yeast Res.* 4, 69–78. doi:10.1016/S1567-1356(03)00141-7

- Laadan, B., Almeida, J. R., Radstrom, P., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. (2008). Identification of an NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Yeast* 25, 191–198. doi:10.1002/yea.1578
- Laluce, C., Schenberg, A. C. G., Gallardo, J. C. M., Coradello, L. F. C., and Pombeiro-Sponchiado, S. R. (2012). Advances and developments in strategies to improve strains of *Saccharomyces cerevisiae* and processes to obtain the lignocellulosic ethanol – a review. *Appl. Biochem. Biotechnol.* 166, 1908–1926. doi:10.1007/s12010-012-9619-6
- Lee, K. H., and Kim, D. M. (2013). Applications of cell-free protein synthesis in synthetic biology: interfacing bio-machinery with synthetic environments. *Biotechnol. J.* 8, 1292–1300. doi:10.1002/biot.201200385
- Lee, S. M., Jellison, T., and Alper, H. S. (2012). Directed evolution of xylose isomerase for improved xylose catabolism and fermentation in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 78, 5708–5716. doi:10.1128/AEM.01419-12
- Lin, F. M., Qiao, B., and Yuan, Y. J. (2009a). Comparative proteomic analysis of tolerance and adaptation of ethanologenic *Saccharomyces cerevisiae* to furfural, a lignocellulosic inhibitory compound. *Appl. Environ. Microbiol.* 75, 3765–3776. doi:10.1128/AEM.02594-08
- Lin, F. M., Tan, Y., and Yuan, Y. J. (2009b). Temporal quantitative proteomics of *Saccharomyces cerevisiae* in response to a nonlethal concentration of furfural. *Proteomics* 9, 5471–5483. doi:10.1002/pmic.200900100
- Liu, C. Q., Goodman, A. E., and Dunn, N. W. (1988). Expression of cloned xanthomonas D-xylose catabolic genes in *Zymomonas mobilis*. *J. Biotechnol.* 7, 61–70. doi:10.1016/0168-1656(88)90035-1
- Liu, Z. L., and Blaschek, H. P. (2010). “Biomass conversion inhibitors and in situ detoxification,” in *Biomass to Biofuels: Strategies for Global Industries*, eds O. A. Vertès, H. P. Blaschek, and H. Yukawa (West Sussex: John Wiley and Sons), 233–259.
- Liu, Z. L., Moon, J., Andersh, B. J., Slininger, P. J., and Weber, S. (2008). Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 81, 743–753. doi:10.1007/s00253-008-1702-0
- Liu, Z. L., Slininger, P. J., Dien, B. S., Berhow, M. A., Kurtzman, C. P., and Gorsich, S. W. (2004). Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. *J. Ind. Microbiol. Biotechnol.* 31, 345–352.
- Liu, Z. L., Slininger, P. J., and Gorsich, S. W. (2005). Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. *Appl. Biochem. Biotechnol.* 12, 451–460. doi:10.1385/ABAB:121:1-3:0451
- Luhe, A. L., Lim, C. Y., Gerken, H., Wu, J., and Zhao, H. (2014). Furfural and hydroxymethylfurfural tolerance in *Escherichia coli* DeltaacrR regulatory mutants. *Biotechnol. Appl. Biochem.* doi:10.1002/bab.1232
- Luo, Y., Zhang, T., and Wu, H. (2014). The transport and mediation mechanisms of the common sugars in *Escherichia coli*. *Biotechnol. Adv.* 32, 905–919. doi:10.1016/j.biotechadv.2014.04.009
- Lynch, M. D., Warnecke, T., and Gill, R. T. (2007). SCALEs: multiscale analysis of library enrichment. *Nat. Methods* 4, 87–93. doi:10.1038/nmeth946
- Lynd, L. R. (1990). Large-scale fuel ethanol from lignocellulose – potential, economics, and research priorities. *Appl. Biochem. Biotechnol.* 2, 695–719. doi:10.1007/BF02920289
- Ma, M., and Liu, Z. L. (2010). Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for *Saccharomyces cerevisiae*. *BMC Genomics* 11:660. doi:10.1186/1471-2164-11-660
- Madhavan, A., Tamalampudi, S., Ushida, K., Kanai, D., Katahira, S., Srivastava, A., et al. (2009). Xylose isomerase from polycentric fungus *Orpinomyces*: gene sequencing, cloning, and expression in *Saccharomyces cerevisiae* for bio-conversion of xylose to ethanol. *Appl. Microbiol. Biotechnol.* 82, 1067–1078. doi:10.1007/s00253-008-1794-6
- Martinez, A., Rodriguez, M. E., York, S. W., Preston, J. F., and Ingram, L. O. (2000). Effects of Ca(OH)(2) treatments (“overliming”) on the composition and toxicity of bagasse hemicellulose hydrolysates. *Biotechnol. Bioeng.* 69, 526–536. doi:10.1002/1097-0290(20000905)69:5<526::AID-BIT7>3.0.CO;2-E
- Matsushika, A., Inoue, H., Kodaki, T., and Sawayama, S. (2009). Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl. Microbiol. Biotechnol.* 84, 37–53. doi:10.1007/s00253-009-2101-x
- Miller, E. N., Jarboe, L. R., Turner, P. C., Pharkya, P., Yomano, L. P., York, S. W., et al. (2009a). Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. *Appl. Environ. Microbiol.* 75, 6132–6141. doi:10.1128/AEM.01187-09
- Miller, E. N., Jarboe, L. R., Yomano, L. P., York, S. W., Shanmugam, K. T., and Ingram, L. O. (2009b). Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 4315–4323. doi:10.1128/AEM.00567-09
- Miller, E. N., Turner, P. C., Jarboe, L. R., and Ingram, L. O. (2010). Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing *Escherichia coli* LY180. *Biotechnol. Lett.* 32, 661–667. doi:10.1007/s10529-010-0209-9
- Mills, T. Y., Sandoval, N. R., and Gill, R. T. (2009). Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnol. Biofuels* 2, 26. doi:10.1186/1754-6834-2-26
- Moses, T., Pollier, J., Thevelein, J. M., and Goossens, A. (2013). Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and invitro. *New Phytol.* 200, 27–43. doi:10.1111/nph.12325
- Nissen, T. L., Anderlund, M., Nielsen, J., Villadsen, J., and Kielland-Brandt, M. C. (2001). Expression of a cytoplasmic transhydrogenase in *Saccharomyces cerevisiae* results in formation of 2-oxoglutarate due to depletion of the NADPH pool. *Yeast* 18, 19–32. doi:10.1002/1097-0061(200101)18:1<19::AID-YEA650>3.3.CO;2-X
- Pettersson, A., Almeida, J. R. M., Modig, T., Karhumaa, K., Hahn-Hagerdal, B., Gorwa-Grauslund, M. F., et al. (2006). A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast* 23, 455–464. doi:10.1002/yea.1370
- Ramseier, T. M. (1996). Cra and the control of carbon flux via metabolic pathways. *Res. Microbiol.* 147, 489–493. doi:10.1016/0923-2508(96)84003-4
- Ran, H., Zhang, J., Gao, Q. Q., Lin, Z. L., and Bao, J. (2014). Analysis of biodegradation performance of furfural and 5-hydroxymethylfurfural by *Amorphotheca resiniae* ZN1. *Biotechnol. Biofuels* 7, 51. doi:10.1186/1754-6834-7-51
- Richard, P., Toivari, M. H., and Penttila, M. (2000). The role of xylulokinase in *Saccharomyces cerevisiae* xylose catabolism. *FEMS Microbiol. Lett.* 190, 39–43. doi:10.1111/j.1574-6968.2000.tb09259.x
- Rizzi, M., Erlemann, P., Buithanh, N. A., and Dellweg, H. (1988). Xylose fermentation by yeasts: 4. Purification and kinetic studies of xylose reductase from *Pichia stipitis*. *Appl. Microbiol. Biotechnol.* 29, 148–154. doi:10.1007/BF00939299
- Rizzi, M., Klein, C., Schulze, C., Bui-Thanh, N. A., and Dellweg, H. (1989). Xylose fermentation by yeasts. 5. Use of ATP balances for modeling oxygen-limited growth and fermentation of yeast *Pichia stipitis* with xylose as carbon source. *Biotechnol. Bioeng.* 34, 509–514. doi:10.1002/bit.260340411
- Runquist, D., Fonseca, C., Radstrom, P., Spencer-Martins, L., and Hahn-Hagerdal, B. (2009). Expression of the Gxf1 transporter from *Candida intermedia* improves fermentation performance in recombinant xylose-utilizing *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 82, 123–130. doi:10.1007/s00253-008-1773-y
- Saha, B. C. (2003). Hemicellulose bioconversion. *J. Ind. Microbiol. Biotechnol.* 30, 279–291. doi:10.1007/s10295-003-0049-x
- Saloheimo, A., Rauta, J., Stasyk, O. V., Sibirny, A. A., Penttila, M., and Ruohonen, L. (2007). Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases. *Appl. Microbiol. Biotechnol.* 74, 1041–1052. doi:10.1007/s00253-006-0747-1
- Sampson, T. R., and Weiss, D. S. (2014). Exploiting CRISPR/Cas systems for biotechnology. *Bioessays* 36, 34–38. doi:10.1002/bies.201300135
- Sandoval, N. R., Kim, J. Y., Glebes, T. Y., Reeder, P. J., Aucoin, H. R., Warner, J. R., et al. (2012a). Strategy for directing combinatorial genome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10540–10545. doi:10.1073/pnas.1206299109
- Sandoval, N. R., Kim, J. Y. H., Glebes, T. Y., Reeder, P. J., Aucoin, H. R., Warner, J. R., et al. (2012b). Strategy for directing combinatorial genome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10540–10545. doi:10.1073/pnas.1206299109
- Sarthy, A. V., McConaughy, B. L., Lobo, Z., Sundstrom, J. A., Furlong, C. E., and Hall, B. D. (1987). Expression of the *Escherichia coli* xylose isomerase gene in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 53, 1996–2000.
- Sedlak, M., and Ho, N. W. Y. (2004). Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant *Saccharomyces* yeast. *Yeast* 21, 671–684. doi:10.1002/yea.1060

- Service, R. F. (2007). Cellulosic ethanol – biofuel researchers prepare to reap a new harvest. *Science* 315, 1488–1491. doi:10.1126/science.315.5818.1488
- Sheridan, C. (2013). Big oil turns on biofuels. *Nat. Biotechnol.* 31, 870–873. doi:10.1038/nbt.2704
- Skotnicki, M. L., Warr, R. G., Goodman, A. E., Lee, K. J., and Rogers, P. L. (1983). High-productivity alcohol fermentations using *Zymomonas mobilis*. *Biochem. Soc. Symp.* 48, 53–86.
- Sonderegger, M., Jeppsson, M., Larsson, C., Gorwa-Grauslund, M. F., Boles, E., Olsson, L., et al. (2004). Fermentation performance of engineered and evolved xylose-fermenting *Saccharomyces cerevisiae* strains. *Biotechnol. Bioeng.* 87, 90–98. doi:10.1002/bit.20094
- Sousa, L. D., Chundawat, S. P. S., Balan, V., and Dale, B. E. (2009). ‘Cradle-to-grave’ assessment of existing lignocellulose pretreatment technologies. *Curr. Opin. Biotechnol.* 20, 339–347. doi:10.1016/j.copbio.2009.05.003
- Taherzadeh, M. J., Gustafsson, L., Niklasson, C., and Liden, G. (2000). Physiological effects of 5-hydroxymethylfurfural on *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 53, 701–708. doi:10.1007/s002530000328
- Tantirungkij, M., Seki, T., and Yoshida, T. (1994). Genetic improvement of *Saccharomyces cerevisiae* for ethanol production from xylose. *Ann. N. Y. Acad. Sci.* 721, 138–147. doi:10.1111/j.1749-6632.1994.tb47386.x
- Toivari, M. H., Salusjarvi, L., Ruohonen, L., and Penttila, M. (2004). Endogenous xylose pathway in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 70, 3681–3686. doi:10.1128/AEM.70.6.3681-3686.2004
- Traff, K. L., Jonsson, L. J., and Hahn-Hagerdal, B. (2002). Putative xylose and arabinose reductases in *Saccharomyces cerevisiae*. *Yeast* 19, 1233–1241. doi:10.1002/yea.913
- Trantidou, T., Rao, C., Barrett, H., Camelliti, P., Pinto, K., Yacoub, M. H., et al. (2014). Selective hydrophilic modification of Parylene C films: a new approach to cell micro-patterning for synthetic biology applications. *Biofabrication* 6:025004. doi:10.1088/1758-5082/6/2/025004
- Trudgill, P. W. (1969). Metabolism of 2-furoic acid by *Pseudomonas* F2. *Biochem. J.* 113, 577–587.
- Turner, P. C., Miller, E. N., Jarboe, L. R., Baggett, C. L., Shanmugam, K. T., and Ingram, L. O. (2010). YqhC regulates transcription of the adjacent *Escherichia coli* genes yqhD and dkgA that are involved in furfural tolerance. *J. Ind. Microbiol. Biotechnol.* 38, 431–439. doi:10.1007/s10295-10010-10787-10295
- Walfridsson, M., Bao, X. M., Anderlund, M., Lilius, G., Bulow, L., and Hahn-Hagerdal, B. (1996). Ethanolic fermentation of xylose with *Saccharomyces cerevisiae* harboring the *Thermus thermophilus* xylA gene, which expresses an active xylose (glucose) isomerase. *Appl. Environ. Microbiol.* 62, 4648–4651.
- Wang, J. Q., Zhang, Y., Chen, Y. L., Lin, M., and Lin, Z. L. (2012a). Global regulator engineering significantly improved *Escherichia coli* tolerances toward inhibitors of lignocellulosic hydrolysates. *Biotechnol. Bioeng.* 109, 3133–3142. doi:10.1002/bit.24574
- Wang, X., Miller, E. N., Yomano, L. P., Shanmugam, K. T., and Ingram, L. O. (2012b). Increased furan tolerance in *Escherichia coli* due to a cryptic ucpA gene. *Appl. Environ. Microbiol.* 78, 2452–2455. doi:10.1128/AEM.07783-11
- Wang, X., Miller, E. N., Yomano, L. P., Zhang, X., Shanmugam, K. T., and Ingram, L. O. (2011). Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. *Appl. Environ. Microbiol.* 77, 5132–5140. doi:10.1128/AEM.05008-11
- Wang, X., Yomano, L. P., Lee, J. Y., York, S. W., Zheng, H., Mullinnix, M. T., et al. (2013). Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4021–4026. doi:10.1073/pnas.1217958110
- Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., and Gill, R. T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* 28, 856–U138. doi:10.1038/nbt.1653
- Way, J. C., Collins, J. J., Keasling, J. D., and Silver, P. A. (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
- Weierstall, T., Hollenberg, C. P., and Boles, E. (1999). Cloning and characterization of three genes (SUT1-3) encoding glucose transporters of the yeast *Pichia stipitis*. *Mol. Microbiol.* 31, 871–883. doi:10.1046/j.1365-2958.1999.01224.x
- Xiao, H., and Zhao, H. (2014). Genome-wide RNAi screen reveals the E3 SUMO-protein ligase gene SIZ1 as a novel determinant of furfural tolerance in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 7, 78. doi:10.1186/1754-6834-7-78
- Xu, P., Bhan, N., and Koffas, M. A. G. (2013). Engineering plant metabolism into microbes: from systems biology to synthetic biology. *Curr. Opin. Biotechnol.* 24, 291–299. doi:10.1016/j.copbio.2012.08.010
- Yang, V. W., and Jeffries, T. W. (1997). Regulation of phosphotransferases in glucose- and xylose-fermenting yeasts. *Appl. Biochem. Biotechnol.* 6, 97–108. doi:10.1007/BF02920416
- Yomano, L. P., York, S. W., Shanmugam, K. T., and Ingram, L. O. (2009). Deletion of methylglyoxal synthase gene (mgsA) increased sugar co-metabolism in ethanol-producing *Escherichia coli*. *Biotechnol. Lett.* 31, 1389–1398. doi:10.1007/s10529-009-0011-8
- Young, E., Lee, S. M., and Alper, H. (2010). Optimizing pentose utilization in yeast: the need for novel tools and approaches. *Biotechnol. Biofuels* 3, 24. doi:10.1186/1754-6834-3-24
- Young, E. M., Comer, A. D., Huang, H. S., and Alper, H. S. (2012). A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. *Metab. Eng.* 14, 401–411. doi:10.1016/j.ymben.2012.03.004
- Young, E. M., Tong, A., Bui, H., Spofford, C., and Alper, H. S. (2014). Rewiring yeast sugar transporter preference through modifying a conserved protein motif. *Proc. Natl. Acad. Sci. U.S.A.* 111, 131–136. doi:10.1073/pnas.1311970111
- Zaldivar, J., Martinez, A., and Ingram, L. O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. *Biotechnol. Bioeng.* 65, 24–33. doi:10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2
- Zaldivar, J., Martinez, A., and Ingram, L. O. (2000). Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic *Escherichia coli*. *Biotechnol. Bioeng.* 68, 524–530. doi:10.1002/(SICI)1097-0290(20000605)68:5<524::AID-BIT6>3.3.CO;2-K
- Zhang, M., Eddy, C., Deanda, K., Finkstein, M., and Picataggio, S. (1995). Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* 267, 240–243. doi:10.1126/science.267.5195.240
- Zheng, H., Wang, X., Yomano, L. P., Geddes, R. D., Shanmugam, K. T., and Ingram, L. O. (2013). Improving *Escherichia coli* FucO for furfural tolerance by saturation mutagenesis of individual amino acid positions. *Appl. Environ. Microbiol.* 79, 3202–3208. doi:10.1128/AEM.00149-13
- Zheng, H., Wang, X., Yomano, L. P., Shanmugam, K. T., and Ingram, L. O. (2012). Increase in furfural tolerance in ethanologenic *Escherichia coli* LY180 by plasmid-based expression of thyA. *Appl. Environ. Microbiol.* 78, 4346–4352. doi:10.1128/AEM.00356-12

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 November 2014; accepted: 04 February 2015; published online: 18 February 2015.

Citation: Nieves LM, Panyon LA and Wang X (2015) Engineering sugar utilization and microbial tolerance toward lignocellulose conversion. *Front. Bioeng. Biotechnol.* 3:17. doi: 10.3389/fbioe.2015.00017

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Nieves, Panyon and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cofactor engineering for enhancing the flux of metabolic pathways

M. Kalim Akhtar^{1*} and Patrik R. Jones^{2*}

¹ Department of Biochemical Engineering, University College London, London, UK

² Department of Life Sciences, Imperial College London, London, UK

Edited by:

Pablo Carbonell, University of Evry, France

Reviewed by:

Juan Manuel Pedraza, Universidad de los Andes, Colombia

Gary Sawers, Martin-Luther University Halle-Wittenberg, Germany

*Correspondence:

M. Kalim Akhtar, Department of Biochemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK

e-mail: kalim.akhtar@ucl.ac.uk;

Patrik R. Jones, Department of Life Sciences, Imperial College London, Sir Alexander Fleming building, London, SW7 2AZ, UK

e-mail: p.jones@imperial.ac.uk

The manufacture of a diverse array of chemicals is now possible with biologically engineered strains, an approach that is greatly facilitated by the emergence of synthetic biology. This is principally achieved through pathway engineering in which enzyme activities are coordinated within a genetically amenable host to generate the product of interest. A great deal of attention is typically given to the quantitative levels of the enzymes with little regard to their overall qualitative states. This highly constrained approach fails to consider other factors that may be necessary for enzyme functionality. In particular, enzymes with physically bound cofactors, otherwise known as holoenzymes, require careful evaluation. Herein, we discuss the importance of cofactors for biocatalytic processes and show with empirical examples why the synthesis and integration of cofactors for the formation of holoenzymes warrant a great deal of attention within the context of pathway engineering.

Keywords: cofactors, metabolic pathway engineering, Fe-S clusters, enzymatic activity, synthetic biology

INTRODUCTION

Synthetic biology permits the engineering of biological devices or systems with novel or enhanced functions (Church et al., 2014). Such an approach has numerous applications, most notably in the manufacture of a diverse number of molecules including household chemicals, biofuels, and pharmaceutical drugs (Keasling, 2010). This is principally achieved through pathway engineering in which enzyme activities are carefully coordinated to generate the product of interest. Approaches based on the activities of isolated enzymes (*in vitro*) or whole cells (*in vivo*) can be employed for this purpose (Guterl et al., 2012; Stephanopoulos, 2012). The latter approach in particular offers a significant benefit with respect to complex, multi-step pathways that rely on secondary cellular factors, as well as additional pathway processing. Since the enzymes serve as the workhorse components of these biocatalytic systems, both their quantitative levels and qualitative states are important parameters to consider for pathway engineering. A great deal of focus is typically placed on the quantitative levels of the enzyme (Zelcbuch et al., 2013) with little attention given to their overall qualitative states. Thus, the assumption is usually made that these enzyme components are functioning at full capacity. However, this may not always be the case given that a large subset of enzymes depend on cofactors for functionality.

SIGNIFICANCE OF COFACTORS IN BIOLOGY

All biological organisms possess a network of pathways that lead to the production of metabolites with an array of cellular functions (Feist et al., 2009). The synthesis and interconversion of these metabolites is made possible by the catalytic activities of countless enzymes. By lowering the activation energy barrier, enzymes

catalyze reactions at considerably faster rates than their chemical counterparts. This characteristic property along with the relatively high degree of substrate selectivity, permit the use of enzymes as catalysts for industrial purposes. For those enzymes, which rely solely on amino acids for catalysis, the types of reactions are extremely narrow in scope and, for the most part, restricted to acid/base and electrophilic/nucleophilic reactions (Broderick, 2001). To further extend, the range of reactions, enzymes are commonly associated with non-protein moieties known as cofactors (Broderick, 2001).

In the broadest sense of the term, cofactors are thought to be associated with well over half of known proteins (Fischer et al., 2010a). However, for this article, the term “cofactor” will refer specifically to those moieties, either organic or inorganic, which remain physically associated with the enzyme throughout the catalytic cycle (Fischer et al., 2010a). This excludes dissociable cosubstrates such as NADPH and glutathione. Additionally, since the primary focus of this article is on pathway engineering, only those cofactors that require *de novo* pathways for syntheses will be mentioned and sole metal entities such as calcium and selenium will also be excluded. Using this strict definition, a selection of common cofactors are listed in **Table 1**. These are categorized into two types: organic and inorganic (Rees, 2002; Fischer et al., 2010b). Members of the organic group of cofactors tend to be derivatives of vitamins and undertake numerous types of reactions, while the inorganic group is usually based on various arrangements of iron-sulfur (Fe-S) clusters.

In its cofactor-bound state, enzymes are referred to as holoenzymes while in the unbound state, they are known as apoenzymes (**Figure 1A**). Two discrete structural parts are required for the

Table 1 | Examples of enzyme bound cofactors.

Enzyme bound cofactor	Type of reaction catalyzed	Example of a cofactor-containing enzyme	Associated pathway
ORGANIC COFACTORS			
Biotin	Carbon dioxide addition	Acetyl CoA carboxylase	Fatty acid biosynthesis
Factor F430	Methyl transfer	Methyl coenzyme M reductase	Methanogenesis
Flavin mononucleotide	Electron transfer	Cytochrome P450 reductase	Detoxification
Heme	Electron transfer	Cytochrome P450	Detoxification
Lipoic acid	Acyl/methyl amine transfer	2-oxoacid dehydrogenase	Citric acid cycle
MIO cofactor	Carbon–hydrogen bond activation	Phenylalanine ammonia-lyase	Polyphenol biosynthesis
Molybdopterin	Electron transfer	Xanthine oxidase	Purine catabolism
Phosphopantetheine	Acyl carrier	Carboxylic acid reductase	Fatty acid metabolism
Pyridoxal 5' phosphate	Transamination	Glycogen phosphorylase	Glycogenesis
Pyrrroloquinoline quinone	Electron transfer	Methanol dehydrogenase	Methane metabolism
Thiamine pyrophosphate	Carbon dioxide removal	Pyruvate ferredoxin/ flavodoxin reductase	Pyruvate decarboxylation
Topaquinone	Amine oxidation	Amine oxidase	Urea cycle
INORGANIC COFACTORS			
Fe–S	Electron transfer	Ferredoxin	Iron–sulfur cluster biogenesis
H-cluster	Hydrogen activation	Fe–Fe hydrogenase	Hydrogen metabolism
Fe–Moco	Nitrogen reduction	Nitrogenase	Nitrogen fixation
C-cluster	Carbon monoxide oxidation	Carbon monoxide dehydrogenase	Carbon monoxide metabolism
P-cluster	Electron transfer	Nitrogenase	Nitrogen fixation

For a more comprehensive list of organic cofactors refer Fischer et al. (2010b).

synthesis of a holoenzyme: a polypeptide chain and a cofactor moiety. The former is generated by the ubiquitous translational machinery while the latter, depending upon the cofactor, is synthesized by a defined metabolic pathway. In certain cases, subtle variations of the cofactor pathway may exist. For example, in animals, fungi, and α -proteobacteria, heme synthesis is initiated by 5-aminolevulinic synthase, via condensation of glycine and succinyl CoA, while in photosynthetic eukaryotes and some species of the α -proteobacterial group, it depends on glutamate, via the concerted actions of three enzymes (Layer et al., 2010). Once synthesized, the cofactor is integrated with the apoenzyme, either in a co-translational or post-translational manner, to form the holoenzyme. The nature of the association may be covalent and, in such cases, linkages are typically formed with serine, threonine, histidine, tyrosine, and lysine residues and catalyzed by a distinct maturation system. As an example, the heme *c* in cytochrome *c* is attached to a cysteine residue, via the vinyl group with the aid of Ccm (cytochrome *c* maturation) factors (Sanders et al., 2010). Alternatively, the interaction may be tight but non-covalent as in the case of the flavin-containing enzyme, acyl CoA dehydrogenase (Thorpe and Kim, 1995).

IMPORTANCE OF COFACTOR SYNTHESIS FOR BOTH ENZYME AND PATHWAY FUNCTIONALITY

As integral components of numerous holoenzymes, cofactors are required for the majority of metabolic pathways. To name but a few, these include: FAD in succinate dehydrogenase for the Krebs cycle; TPP in transketolase for the pentose phosphate pathway; pyridoxal phosphate in glycogen phosphorylase for glycogenolysis; H-clusters in hydrogenases for hydrogen metabolism; Fe–MoCo in nitrogenases for nitrogen fixation; biotin in acetyl CoA carboxylase for fatty acid biosynthesis; and haem in cytochrome P450 for

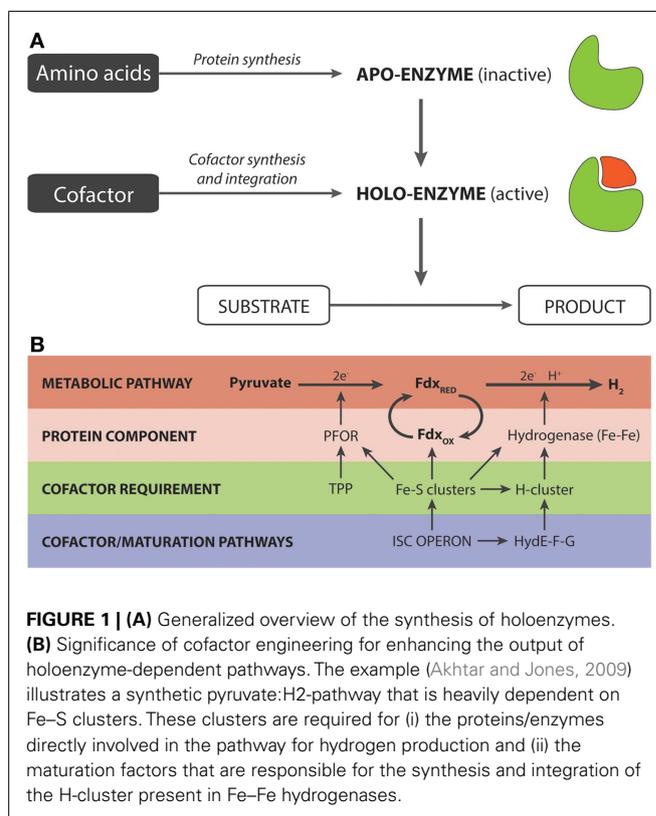


FIGURE 1 | (A) Generalized overview of the synthesis of holoenzymes. **(B)** Significance of cofactor engineering for enhancing the output of holoenzyme-dependent pathways. The example (Akhtar and Jones, 2009) illustrates a synthetic pyruvate:H₂-pathway that is heavily dependent on Fe–S clusters. These clusters are required for (i) the proteins/enzymes directly involved in the pathway for hydrogen production and (ii) the maturation factors that are responsible for the synthesis and integration of the H-cluster present in Fe–Fe hydrogenases.

various detoxification pathways. The most crucial point to consider is that the functional output of holoenzymes can only arise if the apoenzyme is correctly folded with its cofactor. Without the

cofactor, such enzymes would be rendered inoperable and the associated pathways would become redundant. This situation, though undesirable, is likely to be encountered in a bottom-up approach to microbial engineering in which a genetically amenable and well-characterized organism, such as *Escherichia coli* and *Saccharomyces cerevisiae* may be completely devoid of cofactors necessary for heterologous enzyme activity. Alternatively, the capability of the host to synthesize the required cofactor exists but may be insufficient, resulting in a pool of enzymes with a high ratio of apo to holo form (see next section). The most obvious strategy for resolving these issues would be to resort to “cofactor engineering,” by genetically modifying the host so that the cofactor assembly system is present or that the level of native cofactor assembly is in sufficient supply.

Consider the expression of the clostridial Fe–Fe hydrogenase. This enzyme depends on an H-cluster that essentially is a di-iron arrangement with three carbon monoxide, two cyanide ligands, and one dithiolate bridge (Mulder et al., 2011). The formation of this cluster is catalyzed by three maturation enzymes; namely HydE, HydF, and HydG (Posewitz et al., 2004). Since *E. coli* is not naturally endowed with the Hyd maturation enzymes, *E. coli* invariably produces a non-functional Fe–Fe hydrogenase. This can be circumvented by simply complementing the expression of Fe–Fe hydrogenases with the maturation pathway for the H-cluster in order to form the active Fe–Fe hydrogenase (Posewitz et al., 2004; Akhtar and Jones, 2008b). Pyrroloquinoline (PQQ) is another example of a cofactor, which is not naturally synthesized in *E. coli* (Matsushita et al., 1997). This cofactor, present in a family of quinoproteins, has potential uses in biofuel cells, bioremediation, and biosensing (Matsushita et al., 2002). By incorporating the *pqqABCDE* gene cluster from *Gluconobacter oxydans*, Yang et al. (2010) were able to successfully demonstrate the activity of a PQQ-requiring D-glucose dehydrogenase in *E. coli*. They noted, however, that the gene cluster was most likely complemented by the native *tldD* gene. A final example is tetrahydrobiopterin, which in itself is a desirable commodity for the treatment of mild and moderate forms of phenylketonuria (Perez-Duenas et al., 2004). It can be synthesized *in vivo* from GTP via a three-step pathway comprising GTP cyclohydrolase I, 6-pyruvoyl-tetrahydropterin synthase, and sepiapterin reductase (Yamamoto et al., 2003). By augmenting the pathway with expression of a GTP cyclohydrolase I sourced from *Bacillus subtilis*, a 1.5-fold improvement was observed with titers reaching as high as 4 g biopterin per liter of culture (Yamamoto et al., 2003).

COFACTOR INSERTION IS KEY TO MAXIMIZING TOTAL AND SPECIFIC HOLOENZYME ACTIVITY

To maximize the specific activity of a holoenzyme, cofactor synthesis would need to be complemented and/or coupled with cofactor insertion. For enzymes, which bind to cofactors in a non-covalent fashion, the process of cofactor insertion is somewhat of an enigma. It is still not known, even for the well-known heme *b* cofactor, whether this process is facilitated by dedicated *in vivo* components or is a spontaneous process (Thöny-Meyer, 2009). Though overwhelming *in vitro* data show that heme *b* insertion can be a spontaneous event, recent evidence with *in vivo* model systems have implicated the involvement of cellular factors that have yet to be elucidated (Waheed et al., 2010; Correia et al., 2011). For

those enzymes that have covalently attached cofactors, specialized maturation systems have evolved to catalyze both insertion and covalent linkage of the cofactor. Induction of the maturation system can increase holoenzyme activity, as in the case of a carboxylic acid reductase (CAR), which was recently employed for the production of a broad range of chemical commodities (Akhtar et al., 2013). Venkitasubramanian et al. (2007) had verified that CAR requires a cofactor known as phosphopantetheine. This cofactor, during its synthesis, is concomitantly integrated with the enzyme, via a phosphodiester bond, by a maturation enzyme known as phosphopantetheinyl transferase. The sole expression of CAR in *E. coli* leads to an observable, but exceedingly poor, activity. However, with coexpression of the phosphopantetheinyl transferase Sfp from *Bacillus subtilis*, the specific activity of CAR can be enhanced several-fold to a level that is on par with one that is purified from the native organism.

In addition to stimulating the specific enzyme activity, increasing the intracellular levels of cofactors is also known to improve the overall production levels of holoenzymes, suggesting a relationship between holo/apo-forms and degradation. This is a phenomenon that has been frequently observed for hemoproteins (Harnastai et al., 2006; Lu et al., 2010, 2013; Michener et al., 2012). By elevating heme levels, via supplementation of the media with δ -aminolevulinic acid, the expression levels of hemoglobin and cytochrome *b₅* can be significantly improved (Gallagher et al., 1992; Liu et al., 2014). Likewise, increasing Fe–S levels, via overexpression of the *isc* (Fe–S cluster) operon, also leads to similar effects (Nakamura et al., 1999; Akhtar and Jones, 2008a). An explanation for the increased holoenzyme levels may be gleaned from studies of proteins, which utilize divalent metal ions as cofactors (Wilson et al., 2004; Bushmarina et al., 2006; Palm-Espling et al., 2012). Evidence from these published reports suggests that cofactors may aid in the folding of the polypeptide chain and, in turn, accelerate the formation of a functional protein (Goedken et al., 2000). In the case of ribonuclease HI, metal cofactor integration was found to impart a greater degree of rigidity on the final native conformational state of the protein, in addition to improving the refolding rate of the protein (Wittung-Stafshede, 2002). Further insights on the importance of cofactors in protein folding can be gained with the S-adenosylmethionine-containing biotin synthase. This enzyme relies on an intact Fe–S cluster for the addition of sulfur to dethiobiotin to form the biotin thiophane ring. Reyda et al. (2008) noticed that the loss of the Fe–S cluster destabilized the protein, which led to transient unfolding of specific regions, as well as increased proteolysis. Proteolytic degradation was found to proceed by an apparent ATP-dependent proteolysis mechanism, via sequential cleavage of small C-terminal fragments. Interestingly, it was also speculated that since the activity of the protein is generally well maintained under high-iron conditions, a repair process, possibly mediated by the Isc and/or Suf (Sulfur mobilization) machinery, may be active under conditions of destabilization (Reyda et al., 2008).

STIMULATING COFACTOR SYNTHESIS CAN ENHANCE THE FLUX OF SYNTHETIC PATHWAYS

With regard to the actual impact that cofactor engineering can have on the metabolic performance of synthetic pathways, two

studies are particularly worthy of mention. In the first study relating to the production of the vitamin C precursor, 2-keto-L-gulonic acid, Gao et al. (2013) utilized two PQQ-dependent dehydrogenases with D-sorbitol as the starting substrate. The authors noted that induced expression did not improve titers beyond a certain threshold and hypothesized that PQQ was the limiting factor. This was proven to be correct since induction of a pathway for PQQ synthesis resulted in a 20% increase in overall titer. A later refinement of the work in which the two pathway enzymes were incorporated as a fusion protein in *Ketogulonigenium vulgare* also resulted in a similar improvement in titer (Gao et al., 2014).

The second study relates to a synthetic pathway for hydrogen production consisting of a pyruvate:ferredoxin oxidoreductase (PFOR, also known as YdbK in the literature), Fdx and Fe–Fe hydrogenase (Akhtar and Jones, 2009). The design of this pathway was based on the observation that pyruvate, rather than NAD(P)H, was a metabolically superior source of electrons for hydrogen synthesis (Veit et al., 2008). Initial structural analysis of the PFOR-based pathway, in addition to the maturation enzymes, revealed that each protein component was associated with at least one Fe–S cluster that essentially provides the route of electron transfer from pyruvate to the H-cluster of the Fe–Fe hydrogenase (Figure 1B). Up to a total of 12 Fe–S clusters were found to be required for the pathway. In *E. coli*, the formation of Fe–S clusters is undertaken by the *isc* operon, though an analogous *suf* operon is also present (Fontecave et al., 2005; Roche et al., 2013). Since the *isc* operon is controlled by the negative IscR transcriptional regulator, our initial work using a Δ *iscR* background strain had shown that the levels of holo-ferredoxin and the *in vitro* hydrogenase activity could be improved two and threefold, respectively, relative to the wild-type strain (Akhtar and Jones, 2008a). Based on this insight and given the heavy demand for Fe–S clusters, we reasoned that the Δ *iscR* strain may improve the pathway flux toward hydrogen production by increasing the availability of Fe–S clusters. In accordance with our prediction, we observed a twofold improvement in hydrogen yield relative to the wild-type, resulting in an overall yield of 1.5 moles of H₂ per mole of glucose. Even more remarkably with addition of TPP, which serves as a cofactor for the PFOR component, the hydrogen yield was further increased to give a final yield of 1.9, out of a theoretical yield of two moles per mole of glucose (Akhtar and Jones, 2009). Given also that both the specific and total *in vitro* activity of PFOR was improved, this suggests that the availability of the TPP cofactor may well be another potential limiting cofactor for hydrogen production (Akhtar and Jones, 2009).

Although cofactor biosynthesis and integration under native control can be limiting and unresponsive to high-expression levels of the apofactor, presumably to conserve cellular resources, cofactor engineering can quite clearly be advantageous for the assembly of metabolic pathways that depend on the activity of heterologous holoenzymes. This benefit presumably arises from the improved holoenzyme activity, via increased structural stability and/or folding rates in conjunction with reduced protein degradation and/or protein unfolding (explained in the previous section). Interestingly, data from our work on the *in vitro*

activity of Fe–S proteins also suggests that, under certain conditions, a steady and constant supply of cofactors may well aid the restoration of damaged or inactivated enzymes (Akhtar and Jones, 2008a).

CONCLUDING REMARKS

Achieving a balanced production of polypeptide and cofactor for optimal holoenzyme activity would be an iterative process involving the (i) modulated induction of genes, (ii) monitoring of cofactor levels, (iii) evaluation of enzyme activity, and (iv) evaluation of whole-system productivity. Genetic modulation can be controlled at the transcriptional and translational levels by varying the strengths of promoter and ribosomal binding sites, while cofactor levels can be monitored and quantified using suitable analytical methods, e.g., mass spectrometry, high-performance liquid chromatography (HPLC). Combining this with information relating to specific enzyme activity should allow provide profound insights into the intracellular cofactor levels required to achieve optimal holoenzyme activity. Furthermore, if a product reporter system is available, it may also be possible to screen for optimal cofactor metabolism, for example using RBS-variation libraries (Zelcbuch et al., 2013).

Given the importance of cofactor synthesis and integration for holoenzyme activity, a few key points need to be considered when assembling pathways involving holoenzymes. Firstly, holoenzyme activity will only be possible within a host that is metabolically equipped to synthesize the necessary cofactor, otherwise a complementary pathway for cofactor production would also need to be implemented. This is particularly relevant for synthetic pathways that employ holoenzymes from diverse origins. Secondly, to ensure maximal activity of the holoenzyme, the apoenzyme needs to be sufficiently coupled with the synthesis and insertion of its respective cofactor. An imbalance between the two will lead to poor enzyme activity, one that would most likely be inadequate for catalytic purposes. Even native cofactor biosynthesis may not be optimally tuned or responsive to demand from an over-expressed apoenzyme, as would be required in a biocatalytic system where the objective function has shifted from biomass to metabolite producer. Thirdly, since cofactor stimulation is known to improve the stability and activity of holoenzymes, cofactor engineering is likely to be a useful strategy for enhancing the total activity of all holoenzymes in the engineered pathway, and maximizing chances for high flux toward the product of interest. As it currently stands, the impact of cofactor engineering on the activity of holoenzymes is still very much a poorly studied area and certainly warrants more attention given its potential impact on the success of engineered biocatalytic systems.

ACKNOWLEDGMENTS

This work that spurred our views, reflected within this perspective, was supported in part by the Academy of Finland project no. 25369 and European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/European Research Council Grant Agreement 260661 (Patrik R. Jones).

REFERENCES

- Akhtar, M., and Jones, P. (2008a). Deletion of *iscR* stimulates recombinant clostridial Fe-Fe hydrogenase activity and H₂-accumulation in *Escherichia coli* BL21(DE3). *Appl. Microbiol. Biotechnol.* 78, 853–862. doi:10.1007/s00253-008-1377-6
- Akhtar, M., and Jones, P. (2008b). Engineering of a synthetic *hydF-hydE-hydG-hydA* operon for biohydrogen production. *Anal. Biochem.* 373, 170–172. doi:10.1016/j.ab.2007.10.018
- Akhtar, M., and Jones, P. (2009). Construction of a synthetic *YdbK*-dependent pyruvate: H₂ pathway in *Escherichia coli* BL21(DE3). *Metab. Eng.* 11, 139–147. doi:10.1016/j.ymben.2009.01.002
- Akhtar, M. K., Turner, N. J., and Jones, P. R. (2013). Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci. U.S.A.* 110, 87–92. doi:10.1073/pnas.1216516110
- Broderick, J. B. (2001). *Coenzyme and Cofactors*. eLS.
- Bushmarina, N. A., Blanchet, C. E., Vernier, G., and Forge, V. (2006). Cofactor effects on the protein folding reaction: acceleration of alpha-lactalbumin refolding by metal ions. *Protein Sci.* 15, 659–671. doi:10.1110/ps.051904206
- Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., and Weiss, R. (2014). Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* 15, 289–294. doi:10.1038/nrm3767
- Correia, M. A., Sinclair, P. R., and De Matteis, F. (2011). Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug Metab. Rev.* 43, 1–26. doi:10.3109/03602532.2010.515222
- Feist, A., Herrgård, M., Thiele, I., Reed, J., and Palsson, B. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143. doi:10.1038/nrmicro1949
- Fischer, J. D., Holliday, G. L., Rahman, S. A., and Thornton, J. M. (2010a). The structures and physicochemical properties of organic cofactors in biocatalysis. *J. Mol. Biol.* 403, 803–824. doi:10.1016/j.jmb.2010.09.018
- Fischer, J. D., Holliday, G. L., and Thornton, J. M. (2010b). The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics* 26, 2496–2497. doi:10.1093/bioinformatics/btq442
- Fontecave, M., Choudens, S. O., Py, B., and Barras, F. (2005). Mechanisms of iron-sulfur cluster assembly: the SUF machinery. *J. Biol. Inorg. Chem.* 10, 713–721. doi:10.1007/s00775-005-0025-1
- Gallagher, J., Kaderbhai, N., and Kaderbhai, M. A. (1992). Gene-dose-dependent expression of soluble mammalian cytochrome b5 in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 38, 77–83. doi:10.1007/BF00169423
- Gao, L., Du, G., Zhou, J., Chen, J., and 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, 1398–1404. doi:10.1002/btpr.1803
- Gao, L., Hu, Y., Liu, J., Du, G., Zhou, J., and 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. doi:10.1016/j.ymben.2014.04.003
- Goedken, E. R., Keck, J. L., Berger, J. M., and Marqusee, S. (2000). Divalent metal cofactor binding in the kinetic folding trajectory of *Escherichia coli* ribonuclease HI. *Protein Sci.* 9, 1914–1921. doi:10.1110/ps.9.10.1914
- Guterl, J. K., Garbe, D., Carsten, J., Steffler, F., Sommer, B., Reisse, S., et al. (2012). Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* 5, 2165–2172. doi:10.1002/cssc.201200365
- Harnastai, I. N., Gilep, A. A., and Usanov, S. A. (2006). The development of an efficient system for heterologous expression of cytochrome P450s in *Escherichia coli* using *hemA* gene co-expression. *Protein Expr. Purif.* 46, 47–55. doi:10.1016/j.pep.2005.07.006
- Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science* 330, 1355–1358. doi:10.1126/science.1193990
- Layer, G., Reichelt, J., Jahn, D., and Heinz, D. W. (2010). Structure and function of enzymes in heme biosynthesis. *Protein Sci.* 19, 1137–1161. doi:10.1002/pro.405
- Liu, L., Martinez, J. L., Liu, Z., Petranovic, D., and Nielsen, J. (2014). Balanced globin protein expression and heme biosynthesis improve production of human hemoglobin in *Saccharomyces cerevisiae*. *Metab. Eng.* 21, 9–16. doi:10.1016/j.ymben.2013.10.010
- Lu, H., Ma, J., Liu, N., and Wang, S. (2010). Effects of heme precursors on CYP1A2 and POR expression in the baculovirus/*Spodoptera frugiperda* system. *J. Biomed. Res.* 24, 242–249. doi:10.1016/S1674-8301(10)60034-6
- Lu, H. Y., Qiu, L. L., Yang, X. J., Zhang, X. M., Zhang, Z., and Wang, S. L. (2013). Optimization of heme precursors for the expression of human cytochrome P450 2A13 and its co-expression with oxidoreductase in baculovirus/sf9 system. *J. Biochem.* 153, 555–563. doi:10.1093/jb/mvt018
- Matsushita, K., Arents, J. C., Bader, R., Yamada, M., Adachi, O., and Postma, P. W. (1997). *Escherichia coli* is unable to produce pyrroloquinoline quinone (PQQ). *Microbiology* 143, 3149–3156. doi:10.1099/00221287-143-10-3149
- Matsushita, K., Toyama, H., Yamada, M., and Adachi, O. (2002). Quinoproteins: structure, function, and biotechnological applications. *Appl. Microbiol. Biotechnol.* 58, 13–22. doi:10.1007/s00253-001-0851-1
- Michener, J. K., Nielsen, J., and Smolke, C. D. (2012). Identification and treatment of heme depletion attributed to overexpression of a lineage of evolved P450 monooxygenases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19504–19509. doi:10.1073/pnas.1212287109
- Mulder, D. W., Shepard, E. M., Meuser, J. E., Joshi, N., King, P. W., Posewitz, M. C., et al. (2011). Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure* 19, 1038–1052. doi:10.1016/j.str.2011.06.008
- Nakamura, M., Saeki, K., and Takahashi, Y. (1999). Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-*iscS-iscU-iscA-hscB-hscA-fdx-ORF3* gene cluster. *J. Biochem.* 126, 10–18. doi:10.1093/oxfordjournals.jbchem.a022409
- Palm-Espling, M. E., Niemiec, M. S., and Wittung-Stafshede, P. (2012). Role of metal in folding and stability of copper proteins in vitro. *Biochim. Biophys. Acta* 1823, 1594–1603. doi:10.1016/j.bbamcr.2012.01.013
- Perez-Duenas, B., Vilaseca, M. A., Mas, A., Lambruschini, N., Artuch, R., Gomez, L., et al. (2004). Tetrahydrobiopterin responsiveness in patients with phenylketonuria. *Clin. Biochem.* 37, 1083–1090. doi:10.1016/j.clinbiochem.2004.09.005
- Posewitz, M. C., King, P. W., Smolinski, S. L., Zhang, L., Seibert, M., and Ghirardi, M. L. (2004). Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. *J. Biol. Chem.* 279, 25711–25720. doi:10.1074/jbc.M403206200
- Rees, D. C. (2002). Great metalloclusters in enzymology. *Annu. Rev. Biochem.* 71, 221–246. doi:10.1146/annurev.biochem.71.110601.135406
- Reyda, M. R., Dippold, R., Dotson, M. E., and Jarrett, J. T. (2008). Loss of iron-sulfur clusters from biotin synthase as a result of catalysis promotes unfolding and degradation. *Arch. Biochem. Biophys.* 471, 32–41. doi:10.1016/j.abb.2007.12.001
- Roche, B., Aussel, L., Ezratty, B., Mandin, P., Py, B., and Barras, F. (2013). Iron/sulfur proteins biogenesis in prokaryotes: formation, regulation and diversity. *Biochim. Biophys. Acta* 1827, 455–469. doi:10.1016/j.bbabi.2012.12.010
- Sanders, C., Turkarslan, S., Lee, D. W., and Daldal, F. (2010). Cytochrome c biogenesis: the Ccm system. *Trends Microbiol.* 18, 266–274. doi:10.1016/j.tim.2010.03.006
- Stephanopoulos, G. (2012). Synthetic biology and metabolic engineering. *ACS Synth. Biol.* 1, 514–525. doi:10.1021/sb300094q
- Thöny-Meyer, L. (2009). “Heme transport and incorporation into proteins,” in *Tetrapyrroles: Birth, Life and Death*, eds M. J. Warren and A. G. Smith (Landes Bioscience and Springer Science+Business Media), 149–159.
- Thorpe, C., and Kim, J. J. (1995). Structure and mechanism of action of the acyl-CoA dehydrogenases. *FASEB J.* 9, 718–725.
- Veit, A., Akhtar, M., Mizutani, T., and Jones, P. (2008). Constructing and testing the thermodynamic limits of synthetic NAD(P)H:H₂ pathways. *Microb. Biotechnol.* 1, 382–394. doi:10.1111/j.1751-7915.2008.00033.x
- Venkatasubramanian, P., Daniels, L., and Rosazza, J. P. (2007). Reduction of carboxylic acids by *Nocardia* aldehyde oxidoreductase requires a phosphopantetheinylated enzyme. *J. Biol. Chem.* 282, 478–485. doi:10.1074/jbc.M607980200
- Waheed, S. M., Ghosh, A., Chakravarti, R., Biswas, A., Haque, M. M., Panda, K., et al. (2010). Nitric oxide blocks cellular heme insertion into a broad range of heme proteins. *Free Radic. Biol. Med.* 48, 1548–1558. doi:10.1016/j.freeradbiomed.2010.02.038
- Wilson, C. J., Apiyo, D., and Wittung-Stafshede, P. (2004). Role of cofactors in metalloprotein folding. *Q. Rev. Biophys.* 37, 285–314. doi:10.1017/S00335835000404X
- Wittung-Stafshede, P. (2002). Role of cofactors in protein folding. *Acc. Chem. Res.* 35, 201–208. doi:10.1021/ar010106e
- Yamamoto, K., Kataoka, E., Miyamoto, N., Furukawa, K., Ohsuye, K., and Yabuta, M. (2003). Genetic engineering of *Escherichia coli* for production of tetrahydrobiopterin. *Metab. Eng.* 5, 246–254. doi:10.1016/S1096-7176(03)00046-6

- Yang, X. P., Zhong, G. F., Lin, J. P., Mao, D. B., and Wei, D. Z. (2010). Pyrroloquinoline quinone biosynthesis in *Escherichia coli* through expression of the gluconobacter oxydans pqqABCDE gene cluster. *J. Ind. Microbiol. Biotechnol.* 37, 575–580. doi:10.1007/s10295-010-0703-z
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., et al. (2013). Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucl. Acids Res.* 41:e98. doi:10.1093/nar/gkt151

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 June 2014; accepted: 12 August 2014; published online: 28 August 2014.
Citation: Akhtar MK and Jones PR (2014) Cofactor engineering for enhancing the flux of metabolic pathways. *Front. Bioeng. Biotechnol.* 2:30. doi: 10.3389/fbioe.2014.00030
This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.
Copyright © 2014 Akhtar and Jones. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Can the natural diversity of quorum-sensing advance synthetic biology?

René Michele Davis^{1,2*}, Ryan Yue Muller^{3,4} and Karmella Ann Haynes¹

¹ Ira A. Fulton School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA

² Biological Design Graduate Program, Arizona State University, Tempe, AZ, USA

³ Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ, USA

⁴ School of Life Sciences, Arizona State University, Tempe, AZ, USA

Edited by:

Zhanglin Lin, Tsinghua University, China

Reviewed by:

Baojun Wang, University of Edinburgh, UK

Jesus Picó, Universitat Politècnica de Valencia, Spain

*Correspondence:

René Michele Davis, Ira A. Fulton School of Biological and Health Systems Engineering, Arizona State University, 501 E Tyler Mall, 9709, Tempe, AZ 85287, USA
e-mail: rene.davis@asu.edu

Quorum-sensing networks enable bacteria to sense and respond to chemical signals produced by neighboring bacteria. They are widespread: over 100 morphologically and genetically distinct species of eubacteria are known to use quorum sensing to control gene expression. This diversity suggests the potential to use natural protein variants to engineer parallel, input-specific, cell–cell communication pathways. However, only three distinct signaling pathways, Lux, Las, and Rhl, have been adapted for and broadly used in engineered systems. The paucity of unique quorum-sensing systems and their propensity for crosstalk limits the usefulness of our current quorum-sensing toolkit. This review discusses the need for more signaling pathways, roadblocks to using multiple pathways in parallel, and strategies for expanding the quorum-sensing toolbox for synthetic biology.

Keywords: quorum sensing, homoserine lactone, crosstalk, orthogonal, genetic wire, synthetic gene circuit

MODULES FROM NATURAL QUORUM-SENSING NETWORKS CAN BE DECOUPLED AND INTEGRATED INTO SYNTHETIC SYSTEMS

Scientists first explored the genetic circuitry of a quorum-sensing system through basic research of *Vibrio fischeri*, a symbiotic microbe that populates the light organ of the bobtail squid, *Euprymna scolopes*. Researchers identified an operon called “Lux” that allowed individual *V. fischeri* cells to produce a glowing phenotype by expressing Luciferase specifically in dense bacterial populations (Ruby and Nealson, 1976). Explorations of other microbial genomes revealed dozens of Lux homologs that are collectively known as homoserine lactone (HSL) quorum-sensing networks (Fuqua et al., 1996; Williams et al., 2007; Dickschat, 2010). In addition to bioluminescence, they found that these bacteria use quorum sensing to couple population density with the onset of group behaviors such as virulence, biofilm formation, sporulation, competence, and disruption of neighboring bacterial biofilms (Eberl, 1999).

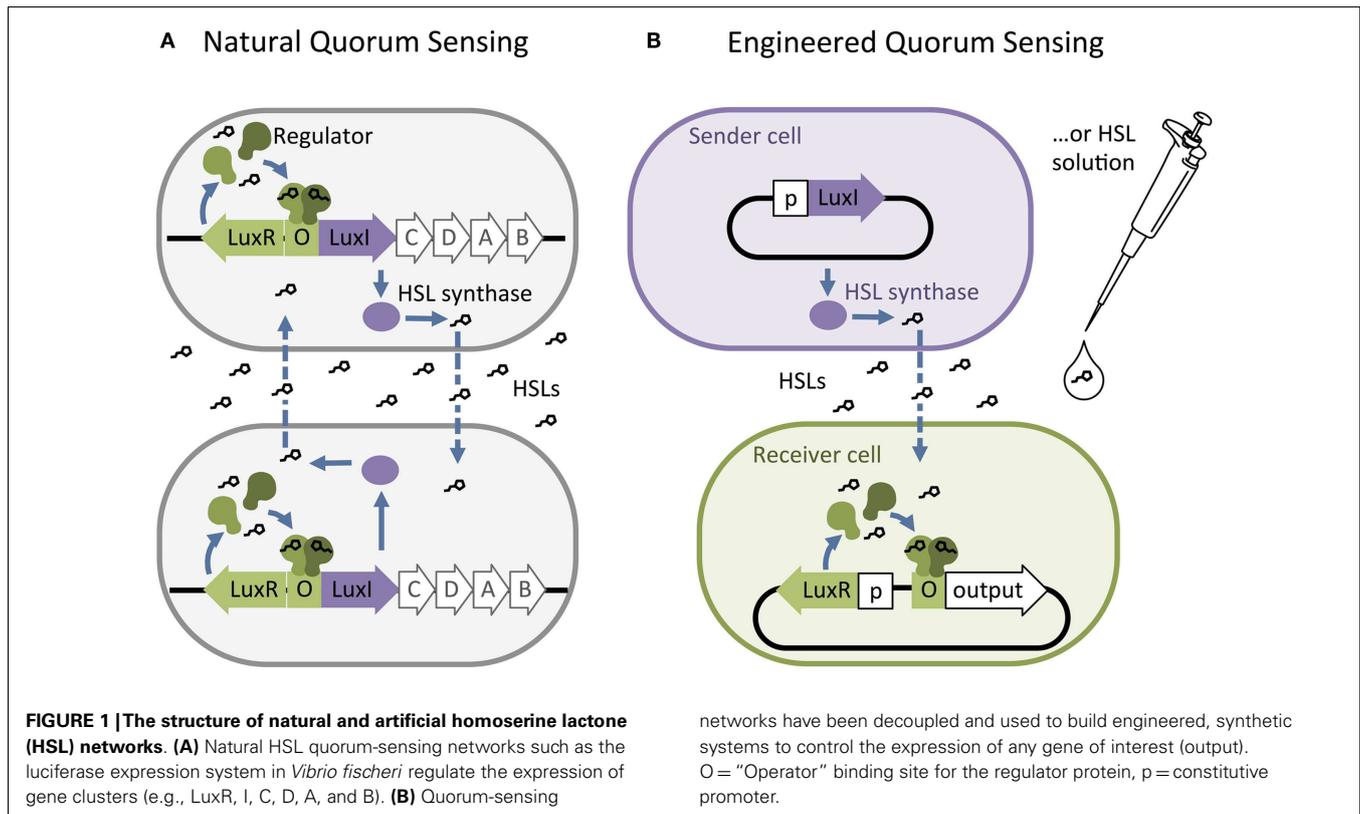
Homoserine lactones networks are more commonly known as *N*-acyl homoserine lactone (AHL) quorum-sensing networks. However, our discussion includes LuxI-like synthases that produce compounds with a homoserine lactone ring but groups other than the acyl tail. In this review, we will consider homoserine lactone, HSL, to include AHLs as well as non-acyl tail compounds. We will refer to HSL with an acyl tail as acyl-HSL.

Homoserine lactones quorum-sensing networks generally consist of an HSL synthase LuxI-like protein, an HSL-binding LuxR-like regulator, and promoters that are regulated by LuxR-like/HSL complexes. The LuxI-like HSL synthase enzyme produces chemical signals called HSLs (Engebrecht and Silverman, 1984; Kaplan and Greenberg, 1987). Most HSLs diffuse passively across the cell

membrane, while some require active transport. Quorum sensing is triggered when high external HSL concentrations drive net influx, allowing HSLs to bind and activate a LuxR-like regulator. The activated LuxR-like/HSL complex binds to a 20 base pair inverted repeat known as a Lux-like-box and regulates expression of downstream genes (Figure 1A). Synthases from various species of bacteria produce different HSL signals, and their corresponding regulators generally bind their cognate HSL with varying levels of specificity.

Researchers have also identified two other families of cell–cell communication networks: autoinducer-2 (AI-2) networks (Vendeville et al., 2005; Reading and Sperandio, 2006) and autoinducing peptides (AIPs), also called peptide pheromone networks (Kleerebezem et al., 1997; Reading and Sperandio, 2006). While AI-2 and AIP networks may be used in engineered systems, the molecular components of HSL networks are simpler, more diverse, and require little modification to function as expected when they are transferred into new host cells. These characteristics of the HSL family of quorum-sensing networks are well suited for building sophisticated, multi-component, synthetic systems. Therefore, we focus primarily on the HSL networks in this review.

Synthetic biologists and genetic engineers often use HSL quorum-sensing pathways to engineer novel behaviors in prokaryotic microorganisms. In these engineered systems, quorum-sensing pathways are used as a set of decoupled components where the HSL synthase is the “Sender” component and the regulator and promoter are collectively the “Receiver” component (Figure 1B) (Miller and Bassler, 2001). They can be employed as “genetic wires” linking the functional elements of multi-component biological systems (Tamsir et al., 2011; Goñi-Moreno et al., 2013). The wires connect circuit components within a cell or across a population of



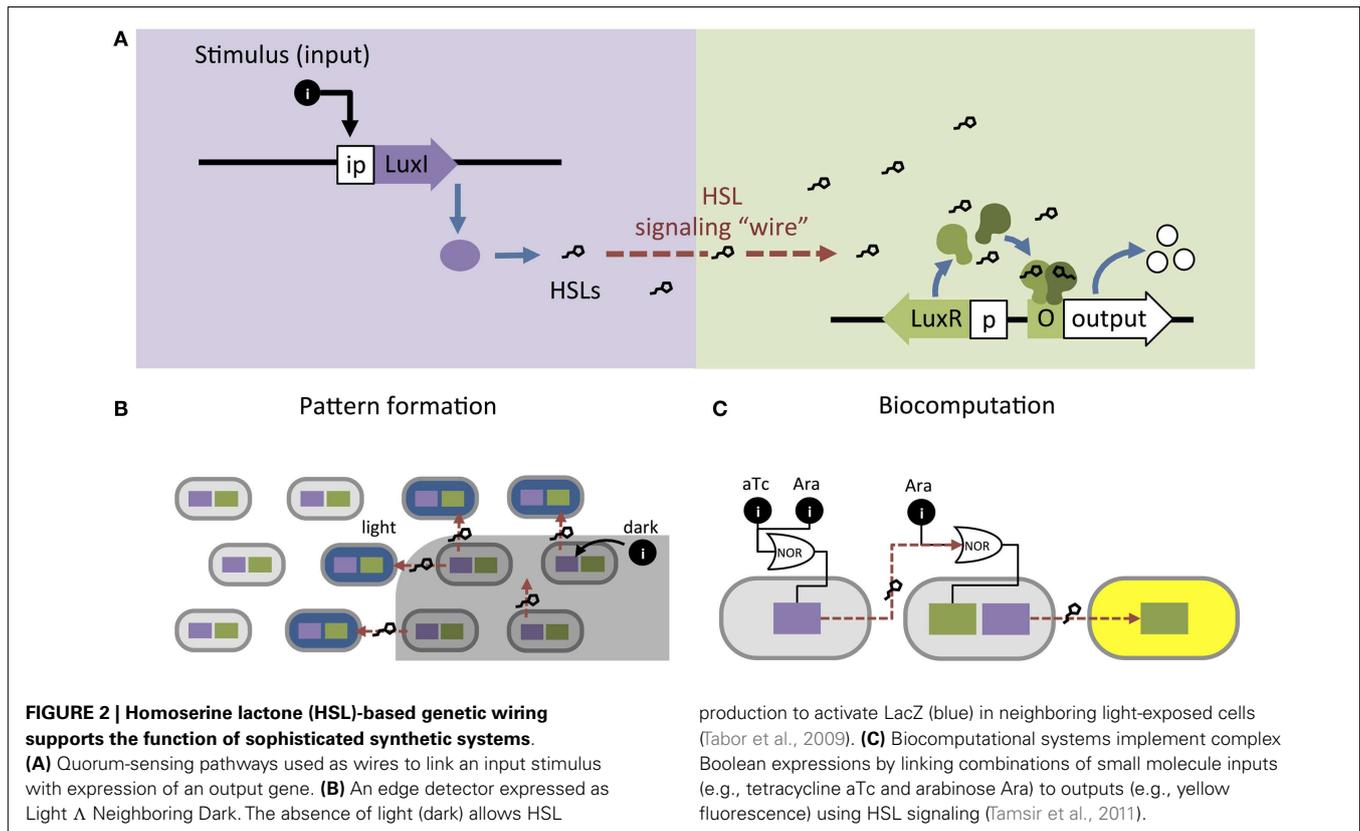
single or multiple strains. The Sender converts an input stimulus into a transmittable signal, the HSL, which activates the Receiver. The Receiver modulates expression of an output as designated by the designer (Figure 2A). This input stimulus may be anything that activates a promoter, including heavy metals (Prindle et al., 2012a; Wang et al., 2013), specific wavelengths of light (Tabor et al., 2009), biochemical signals secreted by pathogens (Saeidi et al., 2011; Gupta et al., 2013), the hypoxic microenvironment surrounding a tumor (Anderson et al., 2006), and HSLs from tandem quorum-sensing networks (Tamsir et al., 2011). The output may be any gene controlled by a Lux-like promoter, such as a visible reporter (Canton et al., 2008; Tabor et al., 2009; Tamsir et al., 2011), cell motility (Liu et al., 2011), antimicrobial proteins (Saeidi et al., 2011; Gupta et al., 2013), and anti-cancer drugs (Anderson et al., 2006).

The simplicity of these networks allows researchers to model how quorum-sensing-controlled gene expression is regulated in response to HSL signal concentration (McMillen et al., 2002; Pai and You, 2009; Pai et al., 2014). These models can inform how a quorum-sensing network should be implemented in a synthetic circuit to achieve the desired behavior (McMillen et al., 2002; Pai and You, 2009). Furthermore, dry lab researchers have used modeling to demonstrate how quorum-sensing systems control group response in the presence of noisy signal concentrations, supporting their use in synthetic biology as robust circuit components (Koseska et al., 2009; Weber and Buceta, 2013).

Incorporating quorum-sensing networks into production strains has advanced the field of metabolic engineering. Quorum

sensing has been used to synchronize gene expression across a population to reduce cell-to-cell variability and to increase yields in engineered strains (Danino et al., 2010; Prindle et al., 2012a,b; Anesiadis et al., 2013). For example, by linking a Lux-based genetic oscillator with a gas phase signal oscillator, researchers coordinated gene expression among 2.5 million cells across 5 mm of space with minimal noise (Danino et al., 2010; Prindle et al., 2012a,b). Anesiadis et al. (2013) employed this type of circuit in a production strain, where they engineered a cell-density-dependent switch using the Lux system to control production of serine in an *Escherichia coli* knockout strain. Group-controlled gene expression implemented by an HSL quorum-sensing network leads to overall higher serine production.

Quorum-sensing networks are also used in genetic circuits to perform computation. Tabor et al. (2009) took advantage of the diffusibility of HSL through agar to build a bacterial edge detector using the Lux network. They demonstrated that stationary physical spacing of bacteria relative to different inputs drives controlled expression of an output. The circuit was designed such that bacteria exposed to darkness expressed HSLs but no output (LacZ). The circuit allowed only bacteria that were both adjacent to HSL-producers and exposed to light to express LacZ, which resulted in a pigmented outline at the edges of a light-masked region (Figure 2B). While most biocomputation is digital, Daniel et al. (2013) showed the versatility of quorum-sensing networks by demonstrating analog computation using the Lux network; their circuit converts logarithmic HSL input into linear fluorescent output over a large range of HSL concentrations. Thus far, engineered



biocomputation has used monolayers of cells. Three-dimensional (3D) colony-printing techniques will increase the sophistication of these systems (Connell et al., 2013). Controlled spacing of colonies based on HSL-diffusion rates could allow engineering a temporal element into a split circuit.

In the preceding examples, the cells in each system are expressing the same circuit. However, engineers may also coordinate gene circuits distributed among multiple populations. Brenner et al. (2007) used the Rhl and Las networks from *Pseudomonas aeruginosa* to build two strains of *E. coli* that form a biofilm together once both populations reach a threshold density. Balagaddé et al. (2008) used components from the Lux and Las networks to engineer a predator–prey relationship between *E. coli* strains. High predator population density induces cell death in the prey strain, while high prey population density supports survival of the predator strain. You et al. (2004) placed a cell death gene under the control of the Lux promoter and built a bistable system that maintains a population density of a defined range. At high cell density, the Lux network activates a cell death. At decreased cell density, the cell death gene is inactive and the population begins to grow again. Computation may be split across multiple strains, distributing the energy demands of a complex computation that are too great for a single cell (Ji et al., 2013; Payne and You, 2014) (Figure 2A). Wang et al. (2013) built a two-strain, three-input biosensor in *E. coli* that produces RFP only in the presence of three heavy metal contaminants: arsenic, mercury, and copper. Cell 1 produces LuxI after exposure to arsenic and mercury; Cell 2 expresses RFP in response to 3O-C6-HSL produced by Cell 1 and copper. Tamsir et al. (2011)

linked circuits expressed in multiple cell populations using two quorum-sensing networks derived from *P. aeruginosa*, Rhl and Las (Figure 2C). They implemented complex Boolean expressions using different spatial arrangements on agar plates. Their system is built with the functionally completed NOR operator and can implement any Boolean expression.

CROSSTALK BETWEEN QUORUM-SENSING PATHWAYS CHALLENGES THE DEVELOPMENT OF SYNTHETIC GENETIC CIRCUITS

Attempts to isolate, study, and apply quorum-sensing pathways for bioengineering is often thwarted by unexpected crosstalk. Quorum sensing is a popular tool among synthetic biologists for designing multicellular systems, but widely utilized HSL quorum-sensing networks are currently limited to three pathways: Lux, Las, and Rhl. These networks all exhibit crosstalk with each other, complicating the design of complex genetic systems implemented with quorum-sensing networks.

For instance, a single regulator can be activated by multiple acyl-HSL-class molecules, resulting in cross-activation of regulators from different species of bacteria. This phenomenon was observed in a proof-of-concept system designed by Canton et al. (2008) wherein an output gene for green fluorescent protein (GFP) was placed under the control of a LuxR receiver module (Wu et al., 2014). Four chemically distinct acyl-HSLs, C6-HSL, C7-HSL, 3O-C8-HSL and, at higher concentrations, C8-HSL, all activated expression of GFP at levels comparable to or even greater than the cognate LuxI acyl-HSL, 3O-C6-HSL (Canton et al., 2008). Many

different HSL synthases, including EsaI, ExpI, and AhII, produce the same major cognate acyl-HSL as LuxI, suggesting that these pathways would have high levels of crosstalk if built into the same network (Miller and Bassler, 2001; Pöllumaa et al., 2012). In the report of their predator–prey, two-strain system, Balagaddé et al. (2008) discussed low-level crosstalk between LuxI and LasR, which was recently confirmed by observing LuxI and LasR interactions in a single-strain system (Wu et al., 2014). However, Balagaddé's system functioned such that crosstalk was apparently below the threshold for altering intended behavior. Interestingly, this type of crosstalk is also observed in nature (Fuqua et al., 1996). For example, two opportunistic pathogens, *Burkholderia cepacia* and *P. aeruginosa*, are known to co-infect patients with cystic fibrosis (Lewenza et al., 2002). Each pathogen's quorum-sensing regulators respond to the other's HSLs, resulting in coordination of virulence-gene expression.

Crosstalk can also occur at the level of the target, or “output,” gene; similarities between promoter sequences and the DNA-binding domains within the regulator proteins contribute to crosstalk between quorum-sensing pathways. The acyl-HSL-activated LuxR regulator stimulates transcription at its cognate promoter as well as the Esa promoter, while acyl-HSL-activated LasR, EsaR, and ExpR regulators are also capable of initiating transcription at the Lux promoter (von Bodman et al., 2003; Saeidi et al., 2011; Shong et al., 2013). While this type of crosstalk can be avoided by using only one regulator per strain, they will not behave as two orthogonal wires within a single cell.

EXPANDING THE SET OF ORTHOGONAL QUORUM-SENSING PATHWAYS ENABLES DESIGN OF COMPLEX GENETIC CIRCUITS

Synthetic circuits may be engineered to detect specific combinations of input signals so long as each sensing pathway functions independently (orthogonally) without undesired intercommunication (crosstalk). Genetic circuits designed to respond to complex combinations of environmental conditions must distinguish and integrate multiple distinct input signals. Orthogonal quorum-sensing pathways are necessary to implement complex circuits that respond to signals produced by living cells, rather than requiring synthetic, exogenous inputs. Engineered division of labor is a major research area in metabolic engineering (Bernstein et al., 2012; Vinuselvi and Lee, 2012); orthogonal quorum-sensing modules will enable further development of cell-autonomous metabolic regulation in multi-strain bioreactor systems. Quorum-sensing circuits could be used to engineer multi-strain, self-monitoring microbial populations that perform energetically expensive metabolic processes in a single culture. Multiple co-cultured strains could be designed to monitor and maintain a target population ratio, or steps in a metabolic process could be timed for accumulation of precursors (Tamsir et al., 2011).

Circuit sophistication is limited by metabolic capacity, transcription and translation resources, and crosstalk within the cell. Moon et al. (2012) pushed the bounds of single-cell computational capability by building a four input AND gate in *E. coli*. Their complex logic gate allows living bacterial cells to express GFP in the presence of four exogenous compounds and no fewer.

Transcription activator complexes were decoupled and placed under the control of distinct inducible promoters that respond to the presence of soluble compounds (arabinose, IPTG, tetracycline, and the acyl-HSL 3O-C6-homoserine lactone) in the cell culture medium. More complex circuits could be implemented by replacing the exogenous inputs in the Moon et al. system with quorum-sensing wires linking cells performing independent computation. Scaling can be achieved through modularity by building complex computational systems with simple independent components. By designing the components in separate strains and connecting them with orthogonal quorum-sensing wires, computational steps can be performed independently without exhausting cellular resources.

Connecting complex circuits requires orthogonal HSL networks to independently signal each strain's computation. However, using even two quorum-sensing networks in parallel is constrained by crosstalk. To our knowledge, there is no published demonstration of three or more orthogonal quorum-sensing networks in a single system. The complexity of multi-input integration circuits remains constrained by reliance on exogenous signals and by the limited number of orthogonal input–output pathways.

STRATEGIES FOR MINIMIZING CROSSTALK

Promiscuous interactions between HSLs and regulators, as well as between regulators and promoters, prevent many quorum-sensing systems from operating independently and in parallel. Some have used gene-network engineering approaches to mitigate crosstalk (Brenner et al., 2007; Balagaddé et al., 2008; Tamsir et al., 2011; Wu et al., 2014). For instance, Brenner et al. (2007) engineered their system to avoid crosstalk between the Las and Rhl networks. They split the networks between two strains to eliminate promoter–regulator crosstalk and controlled HSL synthase production via a positive feedback loop to achieve a two-strain, biofilm-forming consortium.

Another approach to eliminate crosstalk between signaling pathways is using quorum-sensing pathways from distinct families (the aforementioned HSL, AI-2, and AIP pathways). Significant variance in the chemistry of the signaling molecules suggests that cross-reactivity is unlikely: HSLs contain a lactone ring with a hydrocarbon acyl or aryl tail, AI-2 is a furanosyl borate diester composed of two five-membered rings stabilized by a boron atom, and AIPs are relatively large circular peptides composed of amino acids (Chen et al., 2002; Marchand and Collins, 2013). However, this approach may be limited in its flexibility since both AI-2 and AIP require active transport and multiple proteins to generate and detect the signals. With a few exceptions, HSL networks require only two proteins and one promoter. While AI-2 quorum sensing is limited to only one signaling molecule, multiple AIP pathways may exist that do not have cross-reactivity. Marchand and Collins (2013) recently demonstrated modularity and orthogonality of two AIP signals from *Staphylococcus aureus*. In their system, *E. coli* was the AIP sender, producing and secreting two AIPs, and two engineered strains of *Bacillus megaterium* each received one of the signals but not the other. While the ability to use two AIPs in a single cell was not explored, this is a promising result and further research could demonstrate orthogonality between AIPs and HSL quorum sensing.

Directed evolution could also be used to generate regulator proteins that specifically respond to any desired HSL. Mutational analyses and 3D protein structure data have helped to identify key amino acid residues that govern the interaction between regulators and acyl-HSL ligands. Using positive and negative selection, Collins et al. (2006) generated a LuxR mutant that no longer responds to the cognate 3O-C6-HSL but gained responsiveness to C10-HSL, to which wild-type LuxR does not respond. They then demonstrated the orthogonality of LuxR wild type versus the mutant. However, directed evolution of regulator proteins to generate novel orthogonality is technically daunting and only generates mutants with minor changes to the wild-type binding pocket, limiting the range of possible novel behaviors. Furthermore, they bind to and activate the same promoter and, while this feature could be leveraged to build OR gates, they cannot be used as orthogonal networks in the same cell without further mutagenesis to alter promoter-binding specificity.

Finally, scientists could explore other microbial genomes for quorum-sensing homologs that have not yet been exploited for synthetic biology. Comparative genomics has identified dozens of HSL family (Lux-like) homologs in divergent species (Case et al., 2008). A major advantage of exploring wild-type homologs over directed evolution is that natural evolution has already “discovered” functional regulators in a very broad exploration space of amino acid sequences. Evolution has selected for regulator proteins of significantly different sizes, as opposed to artificial selection techniques that, due to practical constraints, do not deviate significantly from pre-existing primary structures.

THE BASIS OF SPECIFICITY IN THE HSL SIGNALING FAMILY

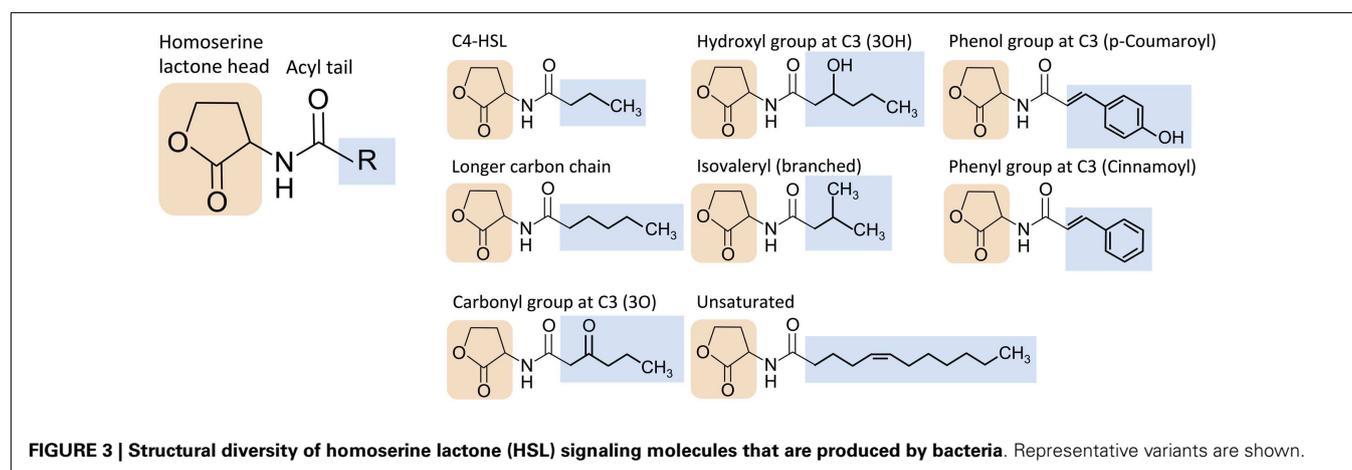
Investigations of microbial quorum-sensing pathways have revealed molecular characteristics that underlie the diversity of HSL signaling pathways in different species. These signaling pathways have been distinguished on the basis of the operator binding sites at promoters elsewhere (Vannini et al., 2002). In this review, we focus on diversity in the geometries of HSL signaling molecules and the HSL-binding pockets within the regulator proteins.

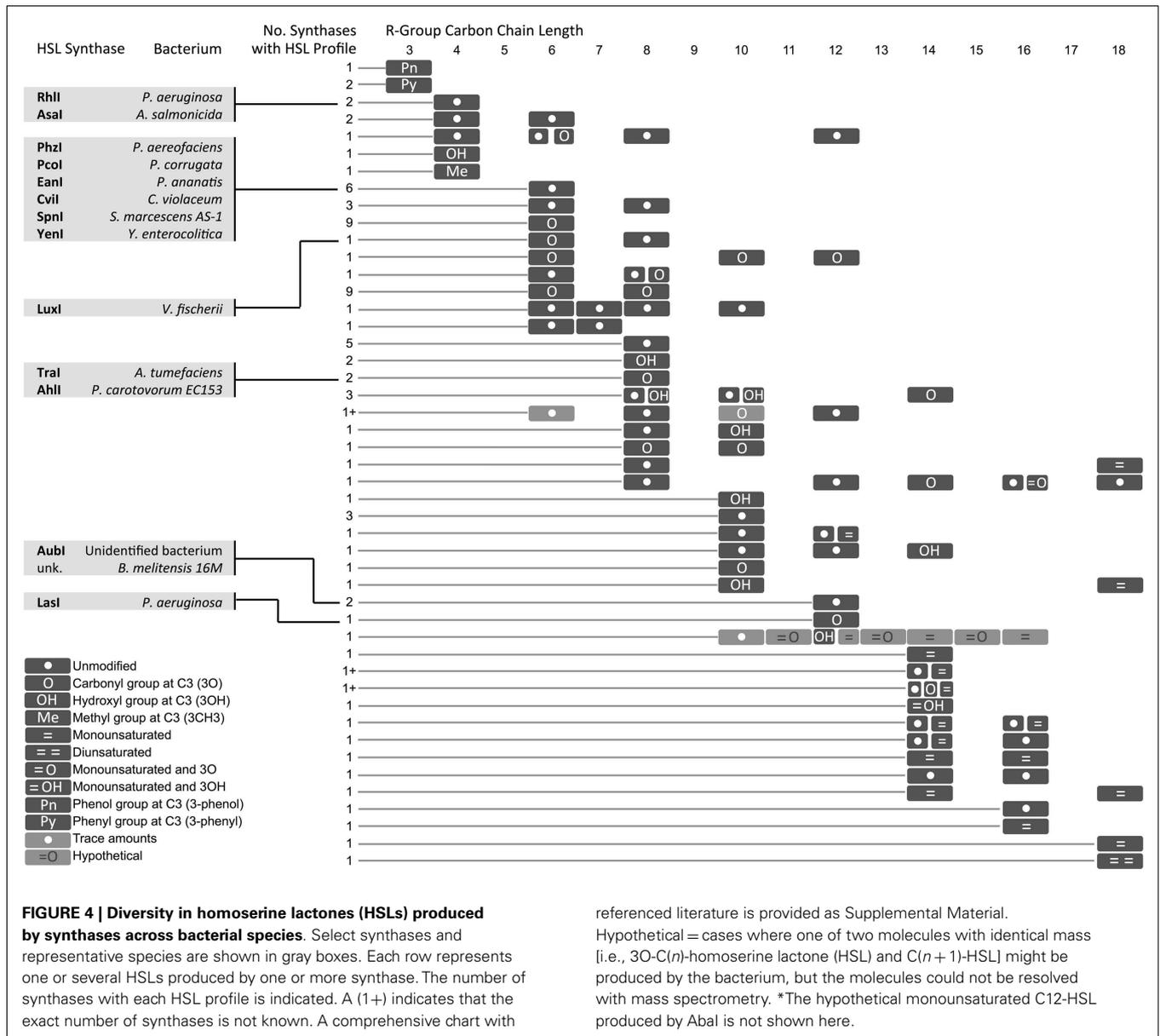
The extensive molecular diversity of naturally occurring HSL signaling molecules suggests that many functionally distinct HSLs, and thus orthogonal pathways, may exist. HSLs vary in the R-group, an acyl or aryl tail that extends from a

homoserine lactone head (Dickschat, 2010) (**Figure 3**). HSL synthases have been reported to generate HSLs of varying carbon chain lengths, branching functional groups, and hydrocarbon saturation (**Figure 4**). Straight-chain acyl R-groups vary by chain lengths (e.g., C4-HSL versus C6-HSL in **Figure 3**) from 4 to 18 carbon atoms. Some acyl R-groups carry side-group replacements at the third or fourth carbon in the chain: a carbonyl group at C3 (e.g., 3O-C6-HSL), a hydroxyl group at C4 (e.g., 3OH-C6-HSL), or a methyl group at C3 (e.g., branched-chain isovaleryl-HSL). Aryl R-groups have a phenol group at C4 (p-Coumaroyl-HSL), or a phenyl group at C4 (Cinnamoyl-HSL). R-groups also differ by the degree of saturation in the carbon chain (e.g., monounsaturated 3OH-C14:1); unsaturation results in a carbon-carbon double bond, which changes the shape of the acyl tail, compared to the saturated form.

In some instances, a single synthase can produce two or more HSL variants. This variety arises from the species-specific combination of acyl tails that are carried into the HSL synthesis pathway by acyl-carrier proteins (ACPs) or aryl-CoA (Lindemann et al., 2011). Some HSL synthases display promiscuity in ACP or CoA-binding affinity and can catalyze formation of several different HSL molecules. Other HSL molecules show no overlap across species, suggesting that the cognate regulators may have evolved to respond specifically to certain HSLs, and orthogonal quorum-sensing systems may exist in nature.

Regulator proteins from the HSL quorum-sensing family (LuxR homologs) consist of two major domains: an N-terminal autoinducer (HSL) binding region and a C-terminal region that binds DNA (**Figure 5**). To visually compare the topologies of functional regions in different regulators, we have generated scaled protein domain maps using descriptions from the literature (Egland and Greenberg, 2001; Zhang et al., 2002; Bottomley et al., 2007; Chen et al., 2011) and annotations from protein domain-scanning databases Uniprot (UniProt Consortium, 2014), Prosite (Sigrist et al., 2013), InterPro (Mitchell et al., 2015), and the Protein Data Bank (Berman et al., 2000) (**Figure 5**). Autoinducer-binding regions (InterPro IPR005143) contain roughly six alpha helices and six beta strands. Published 3D structures for TraR (PDB 1L3L) (Zhang et al., 2002), LasR (PDB et al., 2UV0) (Bottomley et al., 2007), CviR (PDB 3QP6) (Chen et al., 2011), and SdiA (Yao et al., 2006) reveal that a five-strand beta sheet is sandwiched



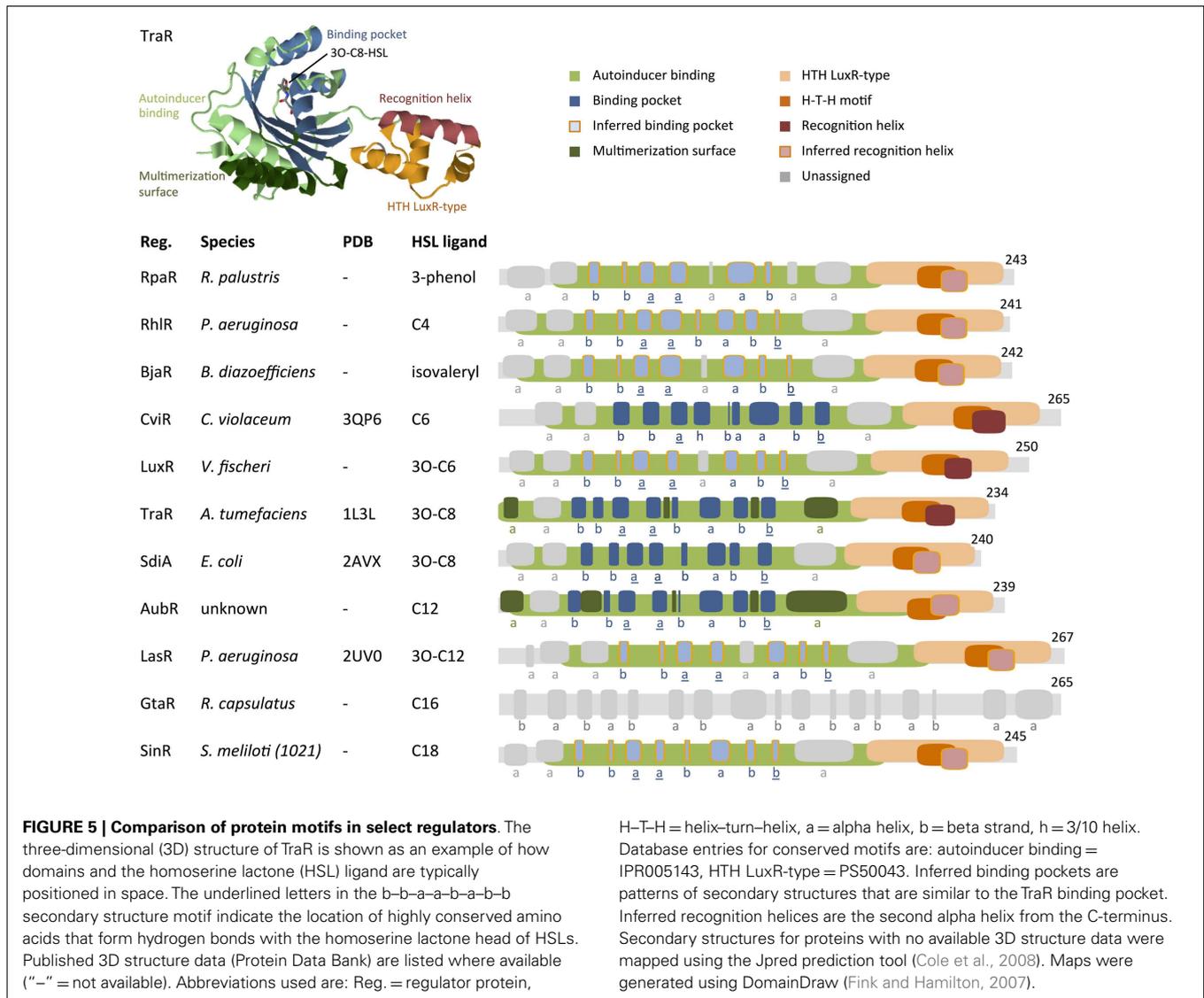


between two three-helix bundles. The C-terminal DNA-binding domains are characterized as “helix–turn–helix” (HTH) regions (Prosite PS50043) that consist of four alpha helices. The second and third helices within the HTH region are often identified as a conserved H–T–H motif (Prosite PRU00411); the third helix has been characterized as the DNA recognition helix in TraR (Zhang et al., 2002). When HSL molecules bind to their corresponding quorum-sensing regulator, they often induce multimerization of regulator proteins. This multimeric state is the active form, capable of binding an inverted DNA sequence repeat at the target promoter and inducing transcription of downstream genes.

Analysis of the HSL-regulator binding pockets suggests that the shape, size, hydrophobicity, and functionalization determine the binding affinity of a regulator for a specific HSL. This implies that

comparison of known HSL–regulator interactions may identify likely candidates for orthogonal quorum-sensing networks. For example, it has been hypothesized that quorum-sensing systems that produce long, straight-chain acyl-HSLs have regulators with longer binding pockets; likewise, a system that uses acyl-HSL molecules with branching functional groups will have regulators with binding pockets that accommodate the branches (Bottomley et al., 2007). Thus, taking sterics into account, a quorum-sensing system that uses HSL molecules with a relatively short hydrocarbon tail and bulky functional groups may be orthogonal to a system that uses long-chain, non-branched HSL molecules.

Hydrophobic interactions between the HSL tail and amino acid residues within the binding pocket suggest that these binding interactions are dominated by van der Waals forces (Bottomley et al., 2007) (Figure 3). Because the HSL tail is buried within the binding



pocket, the hydrophobicity of each component also determines the entropic stability, with a predominantly hydrophobic tail pairing stably with a predominantly hydrophobic binding pocket and less hydrophobic tail pairing stably with a less hydrophobic binding pocket. Therefore, HSL tail and binding pocket hydrophobicity may be a predictor of orthogonality between quorum-sensing pathways.

Pharmacophore models for HSL-regulator binding developed by Geske et al. (2008) support the idea that functionalization of the HSL molecule underlies binding pocket selectivity. Their models are based on the response of Tra, Las, and Lux regulators to libraries of HSLs and synthetic analogs in a system that used beta-galactosidase as the output gene. Comparison of the atomic geometries of ligands reveals three general properties linked with HSL efficacy: spacing of hydrophobic regions, hydrogen bond donor regions, and hydrogen bond acceptor regions within the R-group attached to the lactone ring. For instance, TraR shows the greatest response to a group

of ligands in which the acyl tail contains one hydrogen bond donor region followed by two hydrogen bond acceptor regions arranged in *trans* and ended in a hydrophobic region (Geske et al., 2008).

Conservation and divergence in the conformation of regulator N-terminal HSL-binding regions support the idea that variation in HSL R-groups coordinates selective regulator–ligand interactions. Here, we explore whether motifs in the protein structures of regulators provide insight into the underlying mechanism of HSL-binding selectivity. Primary sequence alignments show 10–25% identity in regulator homologs (Bottomley et al., 2007) and therefore provide very limited information. We have attempted a more coarse-grained approach on a select set of well-characterized regulators by annotating secondary structures that correspond to the TraR binding pocket. For regulators that lack published 3D structure data, we have annotated secondary structures as hypothetical HSL-binding pockets (Figure 5).

The autoinducer-binding region contains two functional domains in its tertiary structure: the multimerization surface and the HSL-binding pocket. The multimerization surface of the TraR homodimer consists primarily of alpha helices $\alpha 1$ and $\alpha 6$ (Bottomley et al., 2007), plus other residues within loops that link helices and beta strands (Figure 5). The HSL-binding pocket binds a single HSL molecule in the space between a five-strand antiparallel beta sheet and a three-helix bundle (Bottomley et al., 2007). In the primary structure of TraR, these secondary structures are arranged in the order of b–b–a–a–b–a–b–b. The first and second alpha helix and the last beta strand of this motif (underlined in Figure 5) contain the amino acids that form hydrogen bonds with the homoserine lactone head of the HSL ligand. These residues are highly conserved in LuxR protein homologs, reflecting a common binding mechanism at the non-variable head regions of HSL molecules. In contrast, the variable acyl tail extends into the region of the binding pocket that is formed by residues that show less conservation in LuxR homologs, suggesting a mechanism for HSL selectivity (Bottomley et al., 2007).

TraR and SdiA are most responsive to the ligand 3O-C8-HSL (Michael et al., 2001; Geske et al., 2008). These regulators contain the same b–b–a–a–b–a–b–b pattern of secondary structures in their HSL-binding pocket domains (Figure 5). This pair of regulators fits the attractive idea that binding pockets with similar secondary structures may prefer the same HSL ligands. However, comparisons of other regulators challenge this idea. While some regulators that respond to HSLs with smaller or larger R-groups deviate from the b–b–a–a–b–a–b–b motif, there are others, i.e., RhIR, LasR, and SinR, which contain the same motif yet respond to different ligands: C4-HSL, 3O-C12-HSL, and C18-HSL, respectively (Llamas et al., 2004; Kumari et al., 2006; Geske et al., 2008). It is possible that variations in specific residues in RhIR, LasR, and SinR account for their preferences for different ligands. AubR contains a substitution of the third beta strand with an alpha helix in the b–b–a–a–b–a–b–b motif, similar to LuxR and BjaR. LuxR and BjaR respond to 3O-C6-HSL (Canton et al., 2008) and isovaleryl-HSL (Lindemann et al., 2011), respectively. Assuming that the ligand for AubR is C12-HSL [produced by AubI (Nasuno et al., 2012)], AubR, LuxR, and BjaR represent another set of regulators where similarities in secondary structures do not appear to correspond to similar ligands. Exploration of the range of HSL-responsiveness of these regulators may provide more insight into their structure–function relationships.

Interestingly, no HSL-regulator protein-related motifs appear in GtaR. Leung et al. (2012) reported that GtaR regulates its target promoter (a Lux promoter homolog) in response to C16-HSL and cell-free growth medium collected from HSL-producing strains. GatR shows sequence conservation with the TatD family of deoxyribonuclease proteins. Like the LuxR homologs, TatD proteins contain interspersed beta strands and alpha helices. Here, we have annotated predicted secondary structures within GtaR; these domains are inferred from comparisons of GatR with closely related TatD proteins that have published 3D structures. Given its distinct arrangement of secondary structures, GtaR might represent a unique set of HSL-responsive regulator proteins.

CONCLUSION AND DISCUSSION

The information we present here on the diversity of HSL molecules and regulator proteins is insufficient to conclude that the structures of regulator binding pockets and the atomic geometries of the HSL ligands imply orthogonality. Regulators that respond to distinct HSL ligands show different protein folding patterns in some cases but similar structures in others. With limited data on regulator promiscuity, secondary structure alone cannot predict HSL ligand preference; thus, interaction between specific amino acid residues and atoms within the HSL molecule may need to be considered. This investigation is limited by the lack of 3D structure data for LuxR homologs.

Many gaps in knowledge remain in understanding the extent of orthogonality or interaction between the homologous pathways in living cells. To date, the published functional studies of the quorum-sensing homologs in synthetic circuits (HSL synthases, regulators, and promoters) include just three homologous quorum-sensing pathways, or they use purified compounds to stimulate one or a few regulators. Furthermore, the available 3D structure data for regulator proteins is sparse compared to the total number of putative regulator homologs that have been identified via metagenomic analysis (Nasuno et al., 2012). More comprehensive analyses to study the responses of regulator proteins to different HSLs, such as that of Geske et al. (2008), may enable us to predict and select orthogonal pathways for use in complex synthetic systems. For instance, *E. coli* could be used as a universal host to carry dozens of decoupled sender and receiver components (Figure 1), derived from the genomes of various bacterial species. Culture media from sender strains could be used to stimulate receiver strains carrying a reporter driven by a receiver system (regulator protein and its corresponding promoter).

The discovery of novel orthogonal quorum-sensing pathways will provide metabolic engineers and synthetic biologists with HSL signaling wires that do not cross-react. Using these insulated, independently functioning pathways, synthetic circuits could be designed to detect distinct combinations of multiple input signals and scale simple single-cell components to sophisticated multi-strain circuits. It is imperative to continue research on quorum-sensing pathway behavior across multiple disciplines, including crystallography, molecular biology, microbiology, metabolic engineering, and synthetic biology, to fill critical gaps in knowledge that have prevented us from engineering highly sophisticated biological systems.

ACKNOWLEDGMENTS

RD is supported by Achievement Rewards for College Students (ARCS). RM is supported by the ASU School of Life Sciences Undergraduate Research Program (SOLUR). RD and KH are supported by the ASU Foundation: Women and Philanthropy. The authors thank F. Wu, W. Alexander, D. Nyer, S. Hays, J. Kemper, and K. Breeden for constructive criticism and help in finalizing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fbioe.2015.00030/abstract>

REFERENCES

- Anderson, J. C., Clarke, E. J., Arkin, A. P., and Voigt, C. A. (2006). Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* 355, 619–627. doi:10.1016/j.jmb.2005.10.076
- Anesiadis, N., Kobayashi, H., Cluett, W. R., and Mahadevan, R. (2013). Analysis and design of a genetic circuit for dynamic metabolic engineering. *ACS Synth. Biol.* 2, 442–452. doi:10.1021/sb300129j
- Balagaddé, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., et al. (2008). A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.* 4, 187. doi:10.1038/msb.2008.24
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank www.rcsb.org. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bernstein, H., Paulson, S., and Carlson, R. (2012). Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. *J. Biotechnol.* 157, 159–166. doi:10.1016/j.jbiotec.2011.10.001
- Bottomley, M. J., Muraglia, E., Bazzo, R., and Carfi, A. (2007). Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. *J. Biol. Chem.* 282, 13592–13600. doi:10.1074/jbc.M700556200
- Brenner, K., Karig, D. K., Weiss, R., and Arnold, F. H. (2007). Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17300–17304. doi:10.1073/pnas.0704256104
- Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26, 787–793. doi:10.1038/nbt1413
- Case, R. J., Labbate, M., and Kjelleberg, S. (2008). AHL-driven quorum-sensing circuits: their frequency and function among the proteobacteria. *ISME J.* 2, 345–349. doi:10.1038/ismej.2008.13
- Chen, G., Swem, L. R., Swem, D. L., Stauff, D. L., O'Loughlin, C. T., Jeffrey, P. D., et al. (2011). A strategy for antagonizing quorum sensing. *Mol. Cell* 42, 199–209. doi:10.1016/j.molcel.2011.04.003
- Chen, X., Schauder, S., Potier, N., Van Dorselaer, A., Pelcer, I., Bassler, B., et al. (2002). Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* 415, 545–549. doi:10.1038/415545a
- Cole, C., Barber, J., and Barton, G. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 1, W197–W201. doi:10.1093/nar/gkn238
- Collins, C. H., Leadbetter, J. R., and Arnold, F. H. (2006). Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* 24, 708–712. doi:10.1038/nbt0806-1033c
- Connell, J. L., Ritschdorff, E. T., Whiteley, M., and Shear, J. B. (2013). 3D printing of microscopic bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18380–18385. doi:10.1073/pnas.1309729110
- Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature* 497, 619–623. doi:10.1038/nature12148
- Danino, T., Mondragón-Palomino, O., Tsimring, L., and Hasty, J. (2010). A synchronized quorum of genetic clocks. *Nature* 463, 326–330. doi:10.1038/nature08753
- Dickschat, J. S. (2010). Quorum sensing and bacterial biofilms. *Nat. Prod. Rep.* 27, 343–369. doi:10.1039/b804469b
- Eberl, L. (1999). N-acyl homoserinylactone-mediated gene regulation in gram-negative bacteria. *Syst. Appl. Microbiol.* 22, 493–506. doi:10.1016/S0723-2020(99)80001-0
- Egland, K., and Greenberg, E. (2001). Quorum sensing in *Vibrio fischeri* – analysis of the LuxR DNA binding region by alanine-scanning mutagenesis. *J. Bacteriol.* 183, 382–386. doi:10.1128/JB.183.1.382-386.2001
- Engelbrecht, J., and Silverman, M. (1984). Identification of genes and gene products necessary for bacterial bioluminescence. *Proc. Natl. Acad. Sci. U.S.A.* 81, 4154–4158. doi:10.1073/pnas.81.13.4154
- Fink, J., and Hamilton, N. (2007). DomainDraw: a macromolecular feature drawing program. *In silico Biol.* 7, 145–150.
- Fuqua, C., Winans, S. C., and Greenberg, E. P. (1996). Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu. Rev. Microbiol.* 50, 727–751. doi:10.1146/annurev.micro.50.1.727
- Geske, G., Mattmann, M., and Blackwell, H. (2008). Evaluation of a focused library of N-aryl L-homoserine lactones reveals a new set of potent quorum sensing modulators. *Bioorg. Med. Chem. Lett.* 18, 5978–5981. doi:10.1016/j.bmcl.2008.07.089
- Goñi-Moreno, A., Amos, M., and de la Cruz, F. (2013). Multicellular computing using conjugation for wiring. *PLoS ONE* 8:e65986. doi:10.1371/journal.pone.0065986
- Gupta, S., Bram, E. E., and Weiss, R. (2013). Genetically programmable pathogen sense and destroy. *ACS Synth. Biol.* 2, 715–723. doi:10.1021/sb4000417
- Ji, W., Shi, H., Zhang, H., Sun, R., Xi, J., Wen, D., et al. (2013). A formalized design process for bacterial consortia that perform logic computing. *PLoS ONE* 8:e57482. doi:10.1371/journal.pone.0057482
- Kaplan, H., and Greenberg, E. P. (1987). Overproduction and purification of the luxR gene product: transcriptional activator of the *Vibrio fischeri* luminescence system. *Proc. Natl. Acad. Sci. U.S.A.* 84, 6639–6643. doi:10.1073/pnas.84.19.6639
- Kleerebezem, M., Quadri, L. E. N., Kuipers, O., and de Vos, W. (1997). Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Mol. Microbiol.* 24, 895–904. doi:10.1046/j.1365-2958.1997.4251782.x
- Koseska, A., Zaikin, A., Kurths, J., and García-Ojalvo, J. (2009). Timing cellular decision making under noise via cell-cell communication. *PLoS ONE* 4:e4872. doi:10.1371/journal.pone.0004872
- Kumari, A., Pasini, P., Deo, S. K., Flomenhoft, D., Shashidhar, H., and Daunert, S. (2006). Biosensing systems for the detection of bacterial quorum signaling molecules. *Anal. Chem.* 78, 7603–7609. doi:10.1021/ac061421n
- Leung, M., Brimacombe, C., Spiegelman, G., and Beatty, J. (2012). The GtaR protein negatively regulates transcription of the gtaRI operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* 83, 759–774. doi:10.1111/j.1365-2958.2011.07963.x
- Lewenza, S., Visser, M. B., and Sokol, P. A. (2002). Interspecies communication between *Burkholderia cepacia* and *Pseudomonas aeruginosa*. *Can. J. Microbiol.* 48, 707–716. doi:10.1139/w02-079
- Lindemann, A., Pessi, G., Schaefer, A., Mattmann, M., Christensen, Q., Kessler, A., et al. (2011). Isovaleryl-homoserine lactone, an unusual branched-chain quorum-sensing signal from the soybean symbiont *Bradyrhizobium japonicum*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16765–16770. doi:10.1073/pnas.1114125108
- Liu, C., Fu, X., Liu, L., Ren, X., Chau, C. K. L., Li, S., et al. (2011). Sequential establishment of stripe patterns in an expanding cell population. *Science* 334, 238–241. doi:10.1126/science.1209042
- Llamas, I., Keshavan, N., and González, J. (2004). Use of *Sinorhizobium meliloti* as an indicator for specific detection of long-chain N-acyl homoserine lactones. *Appl. Environ. Microbiol.* 70, 3715–3723. doi:10.1128/AEM.70.6.3715-3723.2004
- Marchand, N., and Collins, C. H. (2013). Peptide-based communication system enables *Escherichia coli* to *Bacillus megaterium* interspecies signaling. *Biotechnol. Bioeng.* 110, 3003–3012. doi:10.1002/bit.24975
- McMillen, D., Kopell, N., Hasty, J., and Collins, J. J. (2002). Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl. Acad. Sci. U.S.A.* 99, 679–684. doi:10.1073/pnas.022642999
- Michael, B., Smith, J., Swift, S., Heffron, F., and Ahmer, B. (2001). SdiA of *Salmonella enterica* is a LuxR homolog that detects mixed microbial communities. *J. Bacteriol.* 183, 5733–5742. doi:10.1128/JB.183.19.5733-5742.2001
- Miller, M. B., and Bassler, B. L. (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55, 165–199. doi:10.1146/annurev.micro.55.1.165
- Mitchell, A., Chang, H., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The Interpro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, D213–D221. doi:10.1093/nar/gku1241
- Moon, T., Lou, C., Tamsir, A., Stanton, B., and Voigt, C. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature* 491, 249–253. doi:10.1038/nature11516
- Nasuno, E., Kimura, N., Fujita, M. J., Nakatsu, C. H., Kamagata, Y., and Hanada, S. (2012). Phylogenetically novel LuxI/LuxR-type quorum sensing systems isolated using a metagenomic approach. *Appl. Environ. Microbiol.* 78, 8067–8074. doi:10.1128/AEM.01442-12
- Pai, A., Srimani, J. K., Tanouchi, Y., and You, L. (2014). Generic metric to quantify quorum sensing activation dynamics. *ACS Synth. Biol.* 3, 220–227. doi:10.1021/sb400069w
- Pai, A., and You, L. (2009). Optimal tuning of bacterial sensing potential. *Mol. Syst. Biol.* 5, 286. doi:10.1038/msb.2009.43
- Payne, S., and You, L. (2014). Engineered cell-cell communication and its applications. *Adv. Biochem. Eng. Biotechnol.* 146, 97–121. doi:10.1007/10_2013_249

- Pöllumaa, L., Alamäe, T., and Mäe, A. (2012). Quorum sensing and expression of virulence in *Pectobacterium*. *Sensors* 12, 3327–3349. doi:10.3390/s120303327
- Prindle, A., Samayoa, P., Razinkov, I., Danino, T., Tsimring, L. S., and Hasty, J. (2012a). A sensing array of radically coupled genetic “biopixels.” *Nature* 481, 39–44. doi:10.1038/nature10722
- Prindle, A., Selimkhanov, J., Danino, T., Samayoa, P., Goldberg, A., Bhatia, S., et al. (2012b). Genetic circuits in *Salmonella typhimurium*. *ACS Synth. Biol.* 1, 458–464. doi:10.1021/sb300060e
- Reading, N. C., and Sperandio, V. (2006). Quorum sensing: the many languages of bacteria. *FEMS Microbiol. Lett.* 254, 1–11. doi:10.1111/j.1574-6968.2005.00001.x
- Ruby, E. G., and Neelson, K. H. (1976). Symbiotic association of *Photobacterium fischeri* with the marine luminous fish *Monocentris japonica*: a model of symbiosis based on bacterial studies. *Biol. Bull.* 151, 574–586. doi:10.2307/1540507
- Saeidi, N., Wong, C. K., Lo, T.-M., Nguyen, H. X., Ling, H., Leong, S. S. J., et al. (2011). Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol.* 7, 1–11. doi:10.1038/msb.2011.55
- Shong, J., Huang, Y.-M., Bystroff, C., and Collins, C. H. (2013). Directed evolution of the quorum-sensing regulator EsaR for increased signal sensitivity. *ACS Chem. Biol.* 8, 789–795. doi:10.1021/cb3006402
- Sigrist, C., de Castro, E., Cerutti, L., Cucho, B., Hulo, N., Bridge, A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi:10.1093/nar/gks1067
- Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A., Levskaya, A., Marcotte, E. M., et al. (2009). A synthetic genetic edge detection program. *Cell* 137, 1272–1281. doi:10.1016/j.cell.2009.04.048
- Tamsir, A., Tabor, J., and Voigt, C. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical “wires.” *Nature* 469, 212–215. doi:10.1038/nature09565
- UniProt Consortium. (2014). Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* 42, D191–D198. doi:10.1093/nar/gkt1140
- Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., de Francesco, R., et al. (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* 21, 4393–4401. doi:10.1093/emboj/cdf459
- Vendeville, A., Winzer, K., Heurlier, K., Tang, C. M., and Hardie, K. R. (2005). Making “sense” of metabolism: autoinducer-2, LuxS and pathogenic bacteria. *Nat. Rev. Microbiol.* 3, 383–396. doi:10.1038/nrmicro1146
- Vinuselvi, P., and Lee, S. K. (2012). Engineered *Escherichia coli* capable of utilization of cellobiose and xylose. *Enzyme Microb. Technol.* 50, 1–4. doi:10.1016/j.enzmictec.2011.10.001
- von Bodman, S., Ball, J., Faini, M., Herrera, C., Minogue, T., Urbanowski, M., et al. (2003). The quorum sensing negative regulators EsaR and ExpREcc, homologues within the LuxR family, retain the ability to function as activators of transcription. *J. Bacteriol.* 185, 7001–7007. doi:10.1128/JB.185.23.7001-7007.2003
- Wang, B., Barahona, M., and Buck, M. (2013). A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosens. Bioelectron.* 40, 368–376. doi:10.1016/j.bios.2012.08.011
- Weber, M., and Buceta, J. (2013). Dynamics of the quorum sensing switch: stochastic and non-stationary effects. *BMC Syst. Biol.* 7:6. doi:10.1186/1752-0509-7-6
- Williams, P., Winzer, K., Chan, W. C., and Cámara, M. (2007). Look who’s talking: communication and quorum sensing in the bacterial world. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1119–1134. doi:10.1098/rstb.2007.2039
- Wu, F., Menn, D. J., and Wang, X. (2014). Article circuits: from unimodality to trimodality. *Chem. Biol.* 21, 1629–1638. doi:10.1016/j.chembiol.2014.10.008
- Yao, Y., Martinez-Yamout, M., Dickerson, T. J., Brogan, A. P., Wright, P. E., and Dyson, H. J. (2006). Structure of the *Escherichia coli* quorum sensing protein SdiA: activation of the folding switch by acyl homoserine lactones. *J. Mol. Biol.* 355, 262–273. doi:10.1016/j.jmb.2005.10.041
- You, L., Cox, R., Weiss, R., and Arnold, F. (2004). Programmed population control by cell–cell communication and regulated killing. *Nature* 428, 868–871. doi:10.1038/nature02491
- Zhang, R., Pappas, K. M., Brace, J. L., Miller, P. C., Oulmassov, T., Molyneux, J. M., et al. (2002). Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature* 417, 971–974. doi:10.1038/nature00833

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 December 2014; accepted: 21 February 2015; published online: 10 March 2015.

Citation: Davis RM, Muller RY and Haynes KA (2015) Can the natural diversity of quorum-sensing advance synthetic biology? *Front. Bioeng. Biotechnol.* 3:30. doi: 10.3389/fbioe.2015.00030

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Davis, Muller and Haynes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Signal-to-noise ratio measures efficacy of biological computing devices and circuits

Jacob Beal*

Raytheon BBN Technologies, Cambridge, MA, USA

Engineering biological cells to perform computations has a broad range of important potential applications, including precision medical therapies, biosynthesis process control, and environmental sensing. Implementing predictable and effective computation, however, has been extremely difficult to date, due to a combination of poor composability of available parts and of insufficient characterization of parts and their interactions with the complex environment in which they operate. In this paper, the author argues that this situation can be improved by quantitative signal-to-noise analysis of the relationship between computational abstractions and the variation and uncertainty endemic in biological organisms. This analysis takes the form of a $\Delta\text{SNR}_{\text{dB}}$ function for each computational device, which can be computed from measurements of a device's input/output curve and expression noise. These functions can then be combined to predict how well a circuit will implement an intended computation, as well as evaluating the general suitability of biological devices for engineering computational circuits. Applying signal-to-noise analysis to current repressor libraries shows that no library is currently sufficient for general circuit engineering, but also indicates key targets to remedy this situation and vastly improve the range of computations that can be used effectively in the implementation of biological applications.

Keywords: synthetic biology, controls, signals, digital circuits, Boolean logic, analysis

1. Introduction

Engineering biological cells to perform computations has been one of the major goals of synthetic biology from its inception (Knight and Sussman, 1998; Elowitz and Leibler, 2000; Gardner et al., 2000; Weiss, 2001). The complexity of computations that have actually been implemented, however, has been quite small (Purnick and Weiss, 2009), only quite recently rising as high as a 3-layer logic circuit comprising 6 regulatory devices (Moon et al., 2012). A number of well-known obstacles have contributed to the difficulty of building multi-element logic circuits, including insufficient numbers of strong regulatory elements for building circuits, undesirable interactions between genetic elements, difficulties in constructing and delivering large genetic constructs, and difficulty in modeling and predicting circuit behavior. A number of ongoing efforts are showing progress toward decreasing these problems in circuit engineering, promising to soon deliver many more strong regulatory elements [e.g., Bonnet et al. (2013), Kiani et al. (2014), and Stanton et al. (2014)], improved isolation between components [e.g., Lou et al. (2012) and Mutalik et al. (2013)], fast and easy construction and delivery [e.g., Weber et al. (2011) and Linshiz et al. (2012)], and better predictive circuit models [e.g., Davidsohn et al. (2015) and Beal et al. (2015)].

OPEN ACCESS

Edited by:

Karmella Ann Haynes,
Arizona State University, USA

Reviewed by:

Ilias Tagkopoulos,
University of California Davis, USA
Linh Huynh,
University of California Davis, USA (in
collaboration with Ilias Tagkopoulos)
Naglis Malys,
University of Warwick, UK

*Correspondence:

Jacob Beal,
Raytheon BBN Technologies, 10
Moulton Street, Cambridge, MA
02138, USA
jakebeal@bbn.com

Specialty section:

This article was submitted to
Synthetic Biology, a section of the
journal *Frontiers in Bioengineering and
Biotechnology*

Received: 19 January 2015

Accepted: 15 June 2015

Published: 30 June 2015

Citation:

Beal J (2015) Signal-to-noise ratio
measures efficacy of biological
computing devices and circuits.
Front. Bioeng. Biotechnol. 3:93.
doi: 10.3389/fbioe.2015.00093

Among all of these improvements in our ability to engineer computational circuits, however, there are two critical and surprisingly unresolved questions:

- Just how good an implementation of computation is provided by the biological circuits and genetic elements that are currently available?
- How much better do they need to be in order to realize various applications?

A number of efforts have been made toward providing a clear definition for biological computational devices [e.g., Knight and Sussman (1998) and Weiss (2001)] and toward characterizing their performance [e.g., Canton et al. (2008), Ellis et al. (2009), Kelly et al. (2009), and Beal et al. (2012)]. None of these efforts to date, however, has provided a practical method for quantifying the performance of real devices and circuits that can be implemented with readily obtainable information about biological devices.

This paper aims to provide such a method, based on the mathematical foundation of a signal-to-noise ratio (SNR). The basic idea is this: although biological computation is defined in the platonic realm of abstract numbers and symbols, it must be realized in the noisier physical reality of quantities like chemical concentration. Such reality is never perfect, and a signal-to-noise ratio quantifies how much of a problem the noise is with respect to the intended representation. In electronics, signal-to-noise analysis is a foundational tool for the engineering of computation and communication; this paper now adapts this tool to the engineering of biological computation circuits.

To this end, Section 2.1 of this paper thus begins by reviewing these foundational concepts and adjusting their application to be suitable for biological circuits. Section 2.2 applies this to computing devices, analyzing them in terms of the degree to which they enhance or degrade signal strength under various conditions. Section 2.3 shows how SNR analysis of individual devices can be used to predict the behavior of circuits, and Section 3.1 follows the implications of these methods to develop a new framework for engineering biological circuits based on SNR analysis. Building on this framework, Section 3.2 applies SNR analysis to existing libraries of biological computational devices, finding that none are yet suitable for large-scale circuit engineering and identifying targets for improvement that may remedy this situation. Finally, Section 4 summarizes and considers future directions.

2. Materials and Methods

2.1. Boolean Biochemical Signals

For any biochemical implementation of Boolean values, we need to choose what physical phenomena will be interpreted as the abstract values “true” and “false.” In this paper, we will focus on one of the earliest proposed (Knight and Sussman, 1998; Weiss, 2001) and most commonly used representations, in which Boolean values are represented by the concentration of particular chemical species within a cell.

Many other biological phenomena have also been proposed or used to represent Boolean values, including extracellular concentration of chemicals [e.g., Danino et al. (2009)], rate of transcription by DNA polymerase or translation by ribosomes [e.g., Canton

et al. (2008)], presence, absence, or inversion of a given DNA sequence [e.g., Bonnet et al. (2013)], epigenetic markings on a DNA sequence [e.g., Keung et al. (2014)], fluorescence or light emission [e.g., Kim and Lin (2013)], and trans-membrane voltage [e.g., Adams and Levin (2013)]. For nearly all such mechanisms of biological computation, however, at some point the coupling between elements in the computation is regulated by the concentration of some chemical species. Thus, for many of these alternative representations of Boolean values, it is possible to identify an equivalent chemical representation to which the signal analysis developed in this paper can be applied.

We can evaluate the quality of a chemical concentration representation of Boolean values by comparing the distribution of concentrations per cell produced when the chemical should be in the “true” state with the distribution of concentrations per cell when the chemical should be in the “false” state. The more that these two distributions overlap, the harder it is to distinguish between them, and therefore the worse the quality of the signal and the more difficult it is to engineer an effective computation. Likewise, the more that the two distributions they are separated, the better the quality and the easier it is to engineer.

In electromagnetic systems, this notion of quality is typically quantified as a signal-to-noise ratio (SNR).¹ Signal-to-noise ratio is normally measured on a logarithmic scale of decibels, which can be computed using the standard definition:

$$\text{SNR}_{\text{dB}} = 20 \log_{10} \frac{A_{\text{signal}}}{A_{\text{noise}}} \quad (1)$$

where A is the root-mean-square (RMS) amplitude of the signal and noise waveforms, respectively (Oppenheim and Willsky, 1997). Applied to a general Boolean signal, this becomes:

$$\text{SNR}_{\text{dB}} = 20 \log_{10} \frac{|\mu_{\text{true}} - \mu_{\text{false}}|}{2\sigma} \quad (2)$$

computing expected signal amplitude as half the difference between mean “true” value and the mean “false” value (i.e., approximated by the RMS amplitude of a square wave), and noise amplitude as σ , the mean standard deviation for true and false states (i.e., the RMS amplitude for the waveform remaining when the intended Boolean signal is subtracted).

Superficially, it seems that the same analysis should be applicable to biochemical systems. In fact, however, this is not the case. The problem is that strong cellular expression of chemicals typically exhibits a log-normal distribution of concentration per cell [see, e.g., Friedman et al. (2006), Beal et al. (2012), Bonnet et al. (2013), Davidsohn et al. (2015), and Stanton et al. (2014)] – even if output might be population-level, computation is currently typically carried out within individual cells, because there are currently many more intracellular than intercellular devices available. This means that both signal and noise are generally

¹Note that the comparison of signal and noise distinguishes this discussion from prior investigations of gene expression noise in cells [e.g., Elowitz et al. (2002), Ozbudak et al. (2002), Rosenfeld et al. (2005), Bar-Even et al. (2006), and Friedman et al. (2006)]: these prior investigations characterize the characteristics of noise, but we cannot analyze the efficacy of computation without comparing such noise to an intended signal.

better represented using geometric statistics, implying that the signal-to-noise ratio calculation becomes:

$$\text{SNR}_{\text{dB}} = 20 \log_{10} \frac{|\log_{10}(\mu_{g,\text{true}}/\mu_{g,\text{false}})|}{2 \cdot \log_{10}(\sigma_g)} \quad (3)$$

where the μ_g variables are the geometric means of the true and false values and σ_g is the geometric standard deviation for both states (i.e., variation expressed in fold times/divide rather than value plus-minus).

The SNR that is actually required depends on the application. For example, if the goal is simply to detect that a computation is followed a specified truth table, this can be accomplished even when the signal is significantly less than the noise. For example, achieving a twofold difference in signal levels in a system with twofold SD of noise requires only an SNR_{dB} of only $20 \log_{10} \frac{|\log_{10}(2)|}{2 \cdot \log_{10}(2)} = -6.0\text{dB}$. For another example, controlling cells in an industrial fermenter, such that they select the most efficient of two modes of operation based on changing local conditions, are still a fairly permissive application, since individual cells selecting the wrong choice is likely to have only a minor effect on the overall batch process, and thus might require a fairly low SNR, in the 0–5 dB range. At the opposite end of the scale, a system intended to identify and kill cancer cells inside of a human patient likely needs to have a much higher SNR, perhaps in the range of 20–30 dB, since even a small fraction of cells erroneously killing healthy cells may have a major adverse impact on the patient's health.

For an example of such an SNR calculation, consider the simulated distributions shown in **Figure 1**. These distributions are generated from a log-normal process using values drawn from within the typical range of expression in mammalian cells, as based on the experimental data in (Beal et al., 2012, 2015; Kiani et al., 2014; Davidsohn et al., 2015): $\mu_{g,\text{true}} = 10^6$ molecules of equivalent fluorescein (MEFL),² $\mu_{g,\text{false}} = 10^4$ MEFL, and $\sigma_g = 3.2$ -fold. Each distribution shown is a 10 bins/decade histogram of expressed fluorescence from 50,000 simulated cells.

²MEFL units will be the unit of choice throughout this paper, as population histograms can be readily obtained experimentally in MEFL using protocols such as Beal et al. (2012), whereas other units such as concentration or number of molecules are much more difficult to obtain for large numbers of single cells at present. Using MEFL thus makes a simpler path for validation and application in the laboratory of the results presented.

Here, **Figure 1A** shows high expression representing a true state, while **Figure 1B** shows low expression representing a false state. The geometric means of these two distributions are nicely separated, with an approximately 100-fold ratio between the true and false levels. The cell-to-cell variation, however, is also fairly strong, with a σ_g of more than threefold, resulting in an overall SNR of only 6.2 dB.

Notice that the SNR value here is not very high, due to the high degree of cell-to-cell variation. Such relatively low SNRs are unfortunately rather typical for biological systems, and are an important factor in the difficulty of engineering reliable biological computations. The consequence is a low margin for error in design, putting even more importance on the quality of computing elements.

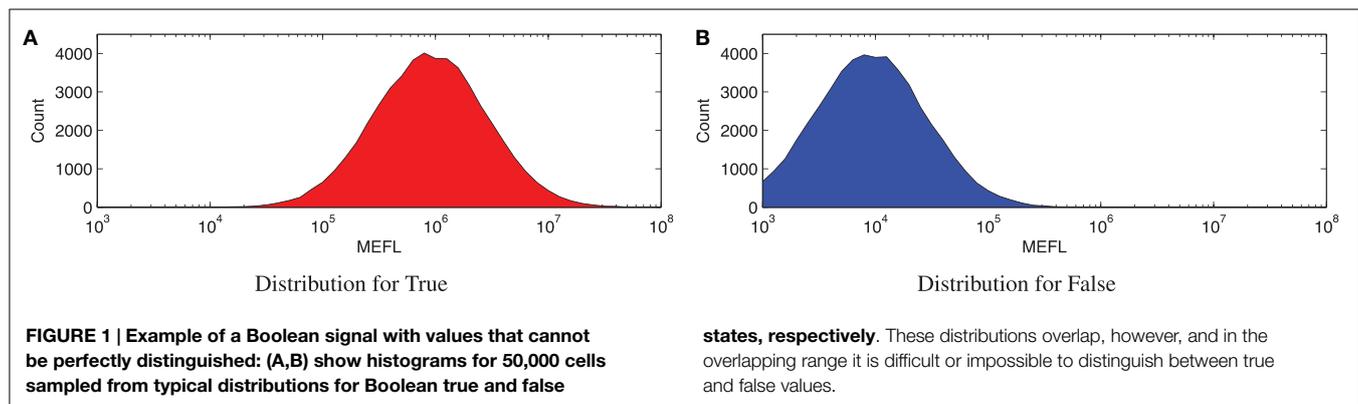
2.2. Effects of Computation on Signal Strength

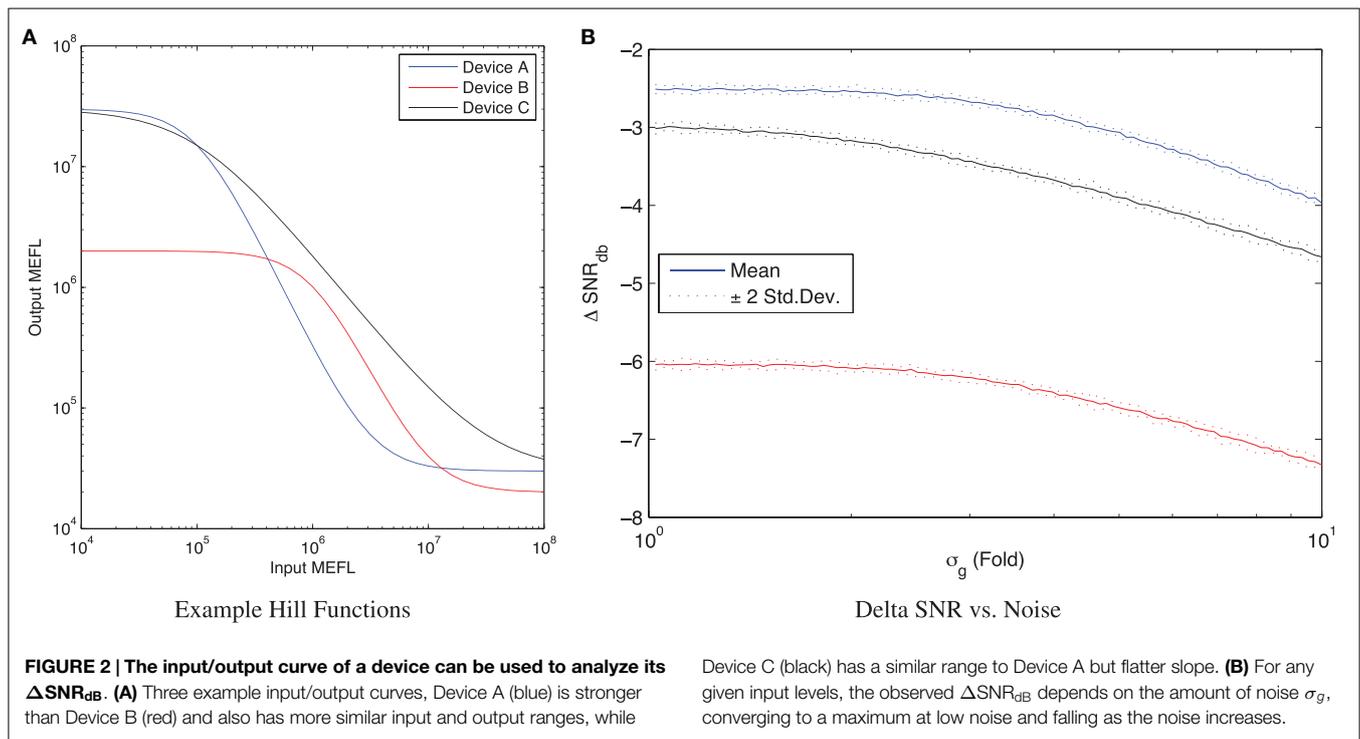
Each computational element in a biological circuit, in addition to performing its intended purpose, also affects the signal-to-noise characteristics of the signals passing through it. An element with strong amplification and inputs that are well-matched to its range of operation will produce *true* and *false* outputs that are more distinct than the inputs, i.e., with an increased SNR. An element with poorly matched inputs or poor amplification, on the other hand, will produce outputs that are less distinct than the inputs, i.e., with a decreased SNR. We may thus summarize the “quality” of a computational element in terms of the difference between input SNR and output SNR across the various combinations of inputs with which it may be supplied:

$$\Delta \text{SNR}_{\text{dB}} = \text{SNR}_{\text{dB,output}} - \text{SNR}_{\text{dB,input}} \quad (4)$$

Under this definition, the higher the $\Delta \text{SNR}_{\text{dB}}$, the better the biological element is at implementing a computation.

In general, the effect of a computational element is on SNR is not uniform, but depends on the circumstances of its use. This fact is independent of any additional biological effects of context, such as metabolic competition, toxicity, or translational read-through. Rather, it is an inherent characteristic of the non-linear relationships between input and output found in most computational elements: different combinations of input levels and noise environment ($\mu_{g,\text{true}}$, $\mu_{g,\text{false}}$, and σ_g) produce different output SNRs. The output SNR is also affected by the dynamics of a signal, e.g., how often the value of the input changes. For the





analysis in this paper, however, we will focus only on converged behavior in response to a stable input.³

The $\Delta\text{SNR}_{\text{dB}}$ for a computational element can be computed from an input/output curve, i.e., a function measuring the outputs observed across a range of input levels. **Figure 2A** shows three simulated examples of such curves for repressor devices with one input and one output (note that to allow easy visualization, all examples in this paper will be restricted to one input and one output, but the methods presented work for multiple inputs and multiple outputs as well). The three example devices have input/output curves f_i generated using Hill equations (Hill, 1910)⁴ of the form:

$$\text{Out} = \alpha \cdot \frac{1 + K^{-1} \left(\frac{In}{D}\right)^H}{1 + \left(\frac{In}{D}\right)^H}$$

with parameters selected to place the curve within the typical observed range for current repressor devices (Bonnet et al., 2013; Kiani et al., 2014; Stanton et al., 2014; Davidsohn et al., 2015), and both In and Out concentrations expressed in MEFL. In particular, Device A (blue) uses $K = 10^3$, $D = 10^5$, $H = 2$, $\alpha = 3 \times 10^7$, while Device B (red) uses $K = 10^2$, $D = 10^6$, $H = 2$, $\alpha = 2 \times 10^6$, and Device C (black) uses $K = 10^3$, $D = 10^5$, $H = 1.2$, $\alpha = 3 \times 10^7$. As can be seen in **Figure 2A**, Device A has a larger range than Device B, though their slope is similar in the transition between values; Device A also has a more similar range for its outputs and inputs. Device C, meanwhile, has similar input and output ranges to Device A, but a significantly flatter slope.

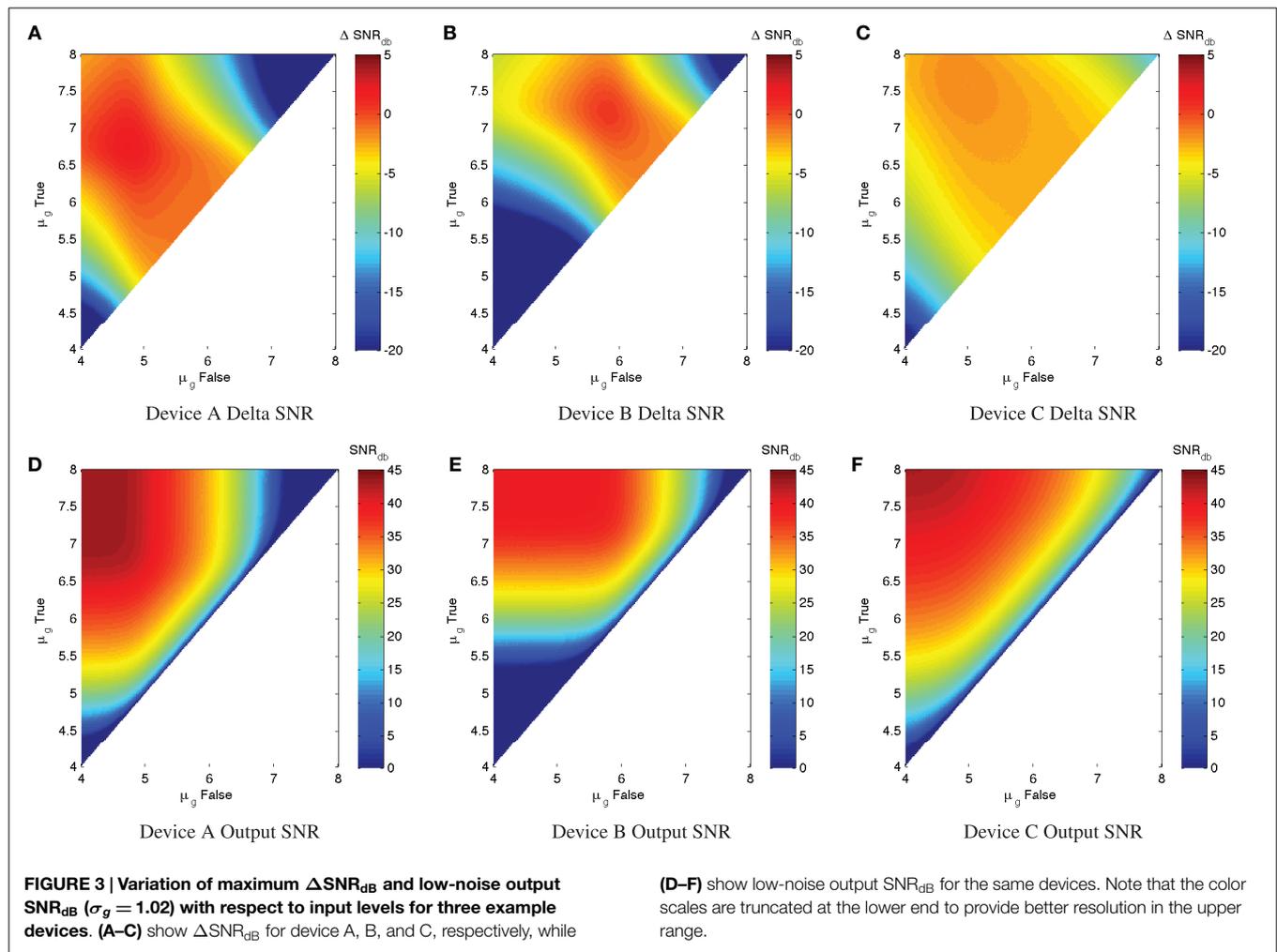
³Such static analysis is expected to be a reasonable first approximation of cellular behavior with strong signals, given the apparent dominance of extrinsic vs. intrinsic noise under such conditions per Elowitz et al. (2002) and Rosenfeld et al. (2005).

⁴Hill equations simulate regulated production, not concentration, but in steady state (as we consider here) the two are linearly related by a constant.

When the expression noise σ_g is low, the $\Delta\text{SNR}_{\text{dB}}$ for a computational element is converged to a maximum determined by the difference between input range and output range. At the opposite extreme, as σ_g continues to increase, the $\Delta\text{SNR}_{\text{dB}}$ decreases, eventually converging to a linear slope entirely dominated by noise. For example, **Figure 2B** shows $\Delta\text{SNR}_{\text{dB}}$ for the three example devices as a function of uncorrelated noise⁵ for inputs with $\mu_{g,\text{true}} = 10^8$ MEFL and $\mu_{g,\text{false}} = 10^4$ MEFL, simulating 10 samples of 50,000 cells per sample. Notice that as expression noise decreases toward the minimal-noise limit of $\sigma_g = 1$, the $\Delta\text{SNR}_{\text{dB}}$ converges to an upper limit of around -2.5 dB for Device A, -6 dB for Device B, and -3 dB for Device C. As the expression noise increases, the SNR degrades as the distributions become less separated. By $\sigma_g = 3$, the noise is having a noticeable effect on all three devices, and by $\sigma_g = 10$ it degrades device performance by around 1.5 dB for all three devices.

A good upper limit on the computational quality of a device can thus be estimated by considering the noise-free limit of its performance for various input levels. **Figure 3** illustrates this with a simulation parameter scan of $\mu_{g,\text{true}}$ and $\mu_{g,\text{false}}$ for each example device. In specific, the scan simulates all combinations of $\mu_{g,\text{true}} > \mu_{g,\text{false}}$ in the range of 10^4 – 10^8 MEFL in logarithmic steps at 50 levels/decade, for each combination running one sample of 50,000 cells at a very low noise $\sigma_g = 1.02$. Given the input/output functions for each of the example devices, the maximum output SNR at this noise level is 43.5 dB for Device A, 40 dB for Device B, and 43.1 dB for Device C. For each device, the strongest output SNR is found for input signals in the saturated regions of the device input/output curves: for Device A roughly $\mu_{g,\text{true}} > 10^{6.5}$ and

⁵Correlations in expression noise can shift the results slightly, but the overall trends remain the same.



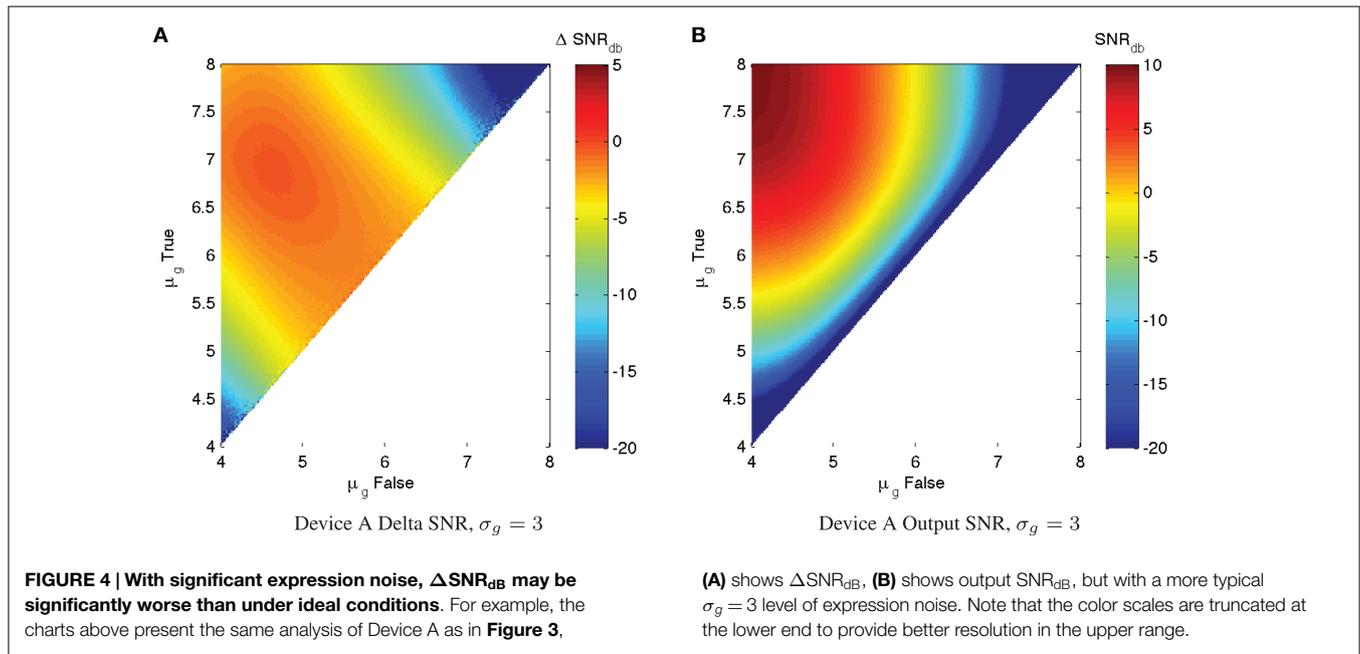
$\mu_{g,\text{false}} < 10^5$, for Device B roughly $\mu_{g,\text{true}} > 10^7$ and $\mu_{g,\text{false}} < 10^6$, and for Device C roughly $\mu_{g,\text{true}} > 10^{7.5}$ and $\mu_{g,\text{false}} < 10^5$. At the boundary of this region, the strong slope of the Devices A and B allows some minor signal restoration, but outside of a relatively small “sweet spot” the output SNR degrades badly with respect to the input SNR. Device C has a similar “sweet spot” pattern, but its lower input/output curve slope means that even its best possible performance still sees a significant signal degradation $\Delta\text{SNR}_{\text{dB}} = -1.6\text{dB}$.

Such a $\Delta\text{SNR}_{\text{dB}}$ chart can provide a good first analysis of the efficacy and operating range of a device. For example, with our three example devices, Device A has a decent range of potential use, while Device B is much narrower, and Device C, although has a very strong on/off ratio, significantly degrades signal strength even under ideal conditions of usage.

In practice, of course, there is typically a significant level of expression noise, which further degrades the SNR characteristics of a device. With measurements of the expected σ_g for a device (which can be readily obtained through high-throughput per-cell assays such as flow cytometry or microscopy with automated image analysis), we can apply the same SNR analysis to estimate the actual expected $\Delta\text{SNR}_{\text{dB}}$, which will always be overall worse (more negative) than in the ideal minimal-noise

condition. **Figure 4** shows an example of such an analysis for Device A with $\sigma_g = 3$, a typical level of observed expression noise (simulated using the same parameters as before). Notice that the essential character of the chart is not changed, meaning that the conclusions drawn from the maximum SNR analysis still apply. All of the features of the SNR chart, however, are more “blurred,” degrading the regions of high performance. Ironically, this also somewhat mitigates the regions of worst performance, but performance in these regions is still generally too poor to be useful.

Computation of $\Delta\text{SNR}_{\text{dB}}$ can be used as a first stage of triage in analyzing whether a given biological device will be useful in attempting to realize digital computations. First, a device cannot be used at all unless it has both a $\Delta\text{SNR}_{\text{dB}}$ that is sufficient to meet application SNR requirements, and also can achieve that $\Delta\text{SNR}_{\text{dB}}$ requirement in a range matched with its inputs [such evaluation has the obvious pre-requisite of characterizing the input/output relation using SI units rather than relative units, e.g., by means of the protocols in Beal et al. (2012, 2015), Davidsohn et al. (2015), and Kiani et al. (2014)]. Beyond that, the wider the region of good $\Delta\text{SNR}_{\text{dB}}$, the easier it will be to match a device with others to form a circuit and the more tolerant a device will be of other types of perturbations inflicted by its context of deployment.



2.3. Multi-Device Computational Circuits

Just as the computational efficacy of a single biological device may be analyzed in terms of its signal-to-noise characteristics, so can the same approach be applied to analyzing the computational efficacy of an entire computational circuit. The complete circuit can, after all, be viewed as just a more complicated device, and the SNR for its inputs and outputs be computed in the same way as for a single device.

The converged SNR characteristics of a circuit with no feedback loops can be predicted using the single-device SNR charts presented in the previous section. As has recently been demonstrated (Davidsohn et al., 2015), the mean and expression variation of such circuits can be predicted with high accuracy. Given such predictions, the maximum possible $\Delta\text{SNR}_{\text{dB}}$ can be predicted directly from the input signal levels, using the input-output curves and $\Delta\text{SNR}_{\text{dB}}$ analyses for the individual devices. For example, consider a chain of repressors, acting as logical inverters. For the i th inverter in the chain, its output is given by its input/output function f_i , producing the input for the next stage. In the minimal-noise case, the SNR changes at each step are independent, meaning that they add linearly. The $\Delta\text{SNR}_{\text{dB}}$ for the circuit can thus be computed by composing together input-output curves to predict the inputs for each device, then summing for each device i the device $\Delta\text{SNR}_{\text{dB},i}$ along the path from input to output. This produces a total end-to-end change of:

$$\begin{aligned} \Delta \text{SNR}_{\text{dB},\text{total}} = & \Delta \text{SNR}_{\text{dB},1}(\mu_{g,\text{true}}, \mu_{g,\text{false}}) \\ & + \Delta \text{SNR}_{\text{dB},2}(f_1(\mu_{g,\text{false}}), f_1(\mu_{g,\text{true}})) \\ & + \Delta \text{SNR}_{\text{dB},3}(f_2 \cdot f_1(\mu_{g,\text{true}}), f_2 \cdot f_1(\mu_{g,\text{false}})) + \dots \quad (5) \end{aligned}$$

The overall efficacy of a circuit is thus a function of both the SNR properties of individual devices and how well signal levels are matched between devices. As seen in the previous section, positive

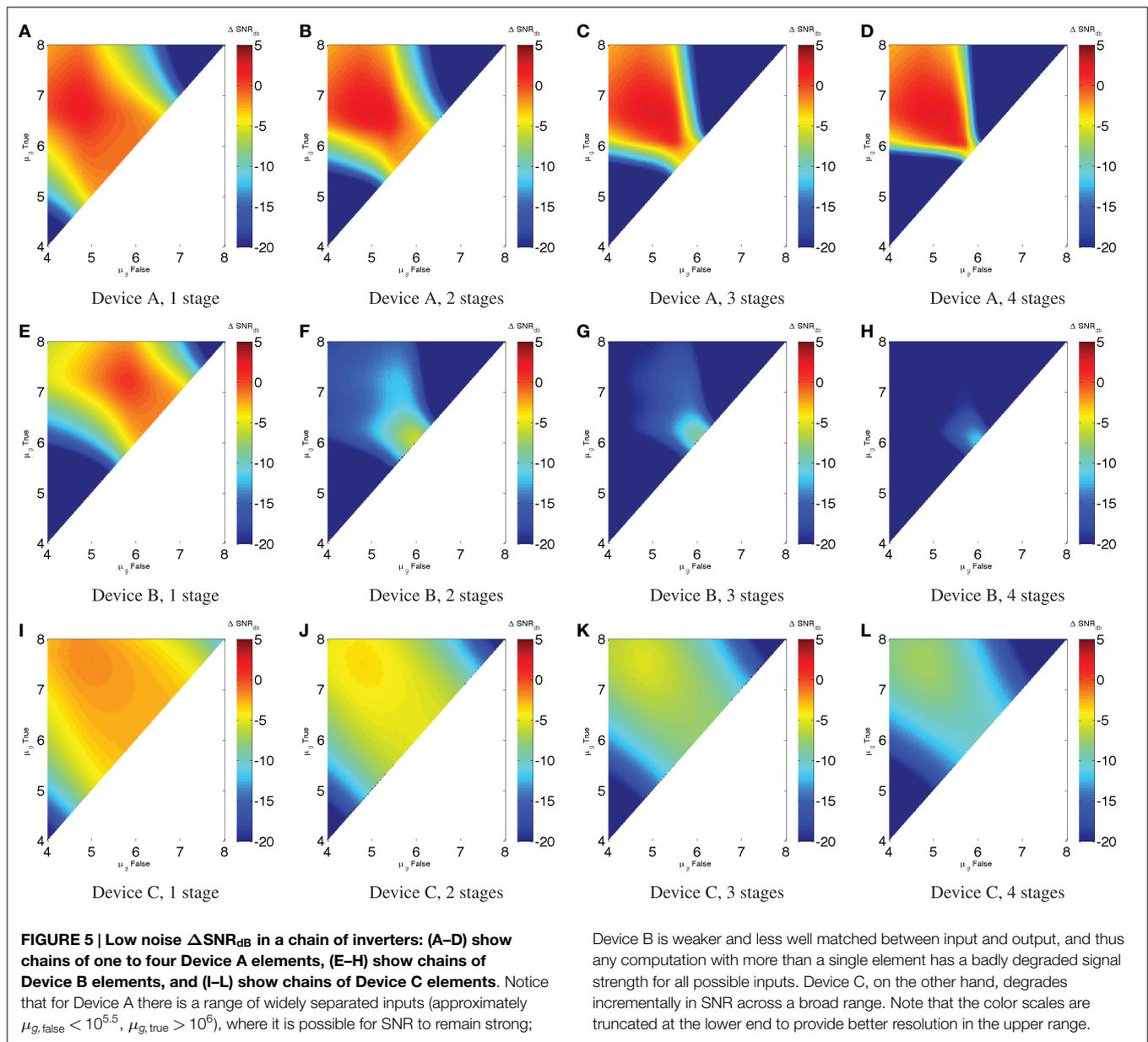
SNR ranges may often be quite narrow, and even a relatively small mismatch can be disastrous for the efficacy of a computation.

For example, **Figures 5A–D** show the $\Delta\text{SNR}_{\text{dB}}$ for chains of one to four repressors, each with the characteristics of Device A. This is a nice example of (potentially) effective digital computation: Device A is strong enough to restore signal and a fairly good match between its input and output ranges. As a result, any input starting in a fairly broad region of the upper left can maintain a strong SNR over multiple stages of computation – in fact, an unbounded number in the absence of noise. Inputs falling outside of this good operating range, however, quickly degrade away to very low SNR.

This presages the problems that occur when the output and input levels of devices are not as well matched (or less strong, which makes for a smaller “sweet spot” and more difficulty in matching levels). For example, **Figures 5E–H** show chains of one to four repressors with the characteristics of Device B. Although its performance characteristics are not much worse than Device A for a single repressor (as seen in the previous section), the poor match with a narrow high-SNR “sweet spot” means that $\Delta\text{SNR}_{\text{dB}}$ collapses when a second repressor is added – much worse than the twice the original $\Delta\text{SNR}_{\text{dB}}$ – and continues to degrade thereafter. Indeed, the “least bad” region is where the high and low inputs hold almost the same value to begin with, meaning there is little signal to be lost in the first place.

With a good match between signal levels but not a steep enough slope of the input/output curve, there is a third mode of behavior. This is exemplified by **Figures 5I–L**, which show chains of one to four repressors with the characteristics of Device C. Without a region of positive $\Delta\text{SNR}_{\text{dB}}$, the signal cannot be sustained, but degrades incrementally. With devices of this sort, it is impossible to implement many-layer computations, but computations with only a few devices between any input and output are viable.

As with individual devices, of course, the minimal-noise model gives only a best-case evaluation of the computational efficacy of a circuit. This is still valuable, because it can be used to eliminate



Device B is weaker and less well matched between input and output, and thus any computation with more than a single element has a badly degraded signal strength for all possible inputs. Device C, on the other hand, degrades incrementally in SNR across a broad range. Note that the color scales are truncated at the lower end to provide better resolution in the upper range.

many non-viable options and to triage viable options based on the difficulty of attaining the (SNR_{dB} required for an application).

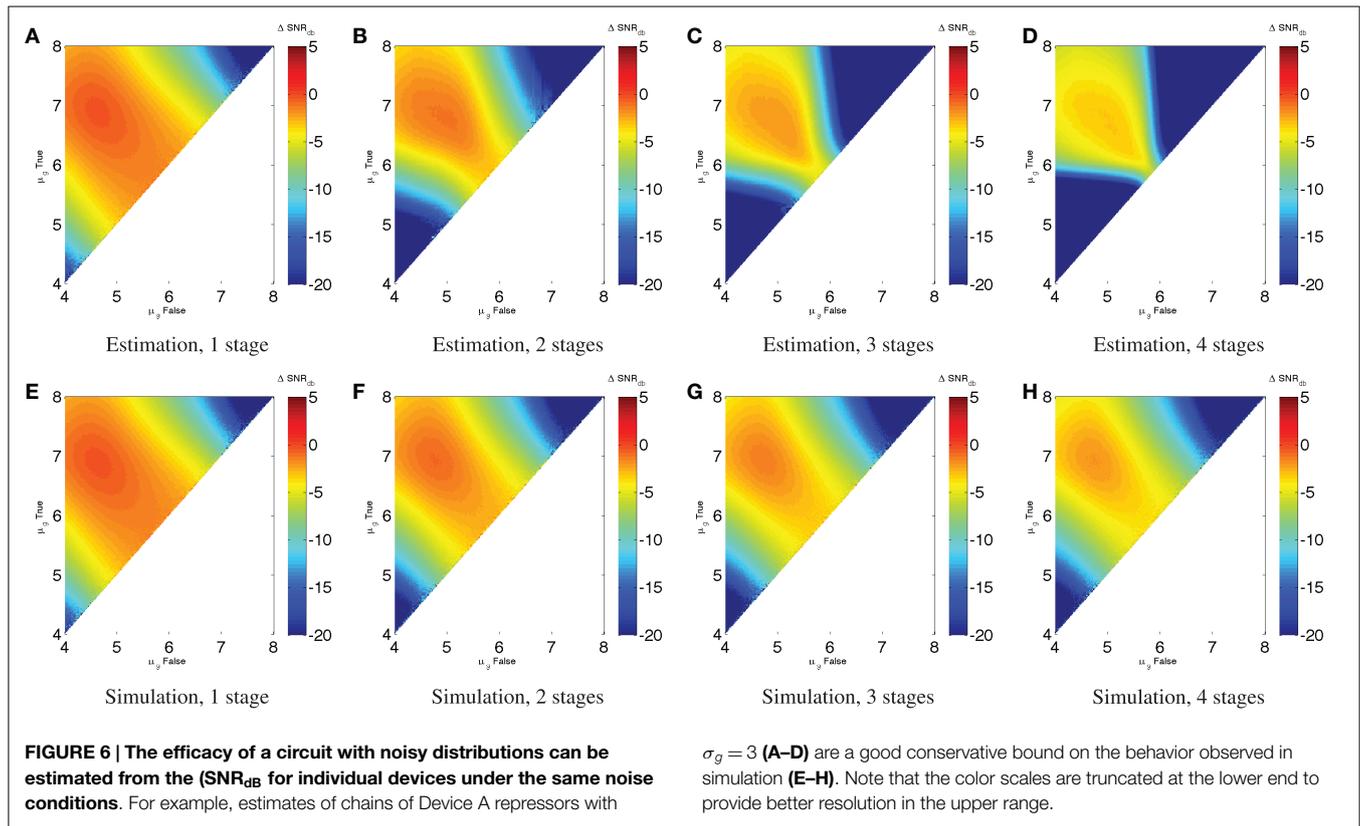
Just as with individual devices, however, we can use the same signal-to-noise models to estimate the performance of a circuit with higher σ_g . As before, the best (SNR_{dB} is expected to be less than can be achieved in a minimal-noise circuit, though some of the worst performance can be mitigated. Unlike the minimal-noise case, however, we cannot precisely predict performance of the circuit by adding single-device SNR losses. At higher levels of expression noise, SNR losses are not independent because the operation of each device affects the effective σ_g observed by the devices that consume its output. We can, however, estimate a conservative lower bound on performance by adding single-device SNR losses. For example, **Figure 6** shows (SNR_{dB} for the Device A repressor chains with $\sigma_g = 3$. **Figures 6A–D** estimate the value from the (SNR_{dB} of Device A with $\sigma_g = 3$ shown in

Figure 4A, while **Figures 6E–H** simulate chains of Device A using the same parameters as in Section 2.2 ($K = 10^3$, $D = 10^5$, $H = 2$, $\alpha = 3 \times 10^7$, $\sigma_g = 3,50,000$ cells per sample). As expected, these show that the estimate from individual devices is a good lower bound on the performance that can be attained from the device under conditions of noise, with the actual simulated performance being somewhere above that and below the minimal-noise performance.

3. Results

3.1. Implications for Biological Circuit Engineering

Let us now consider how the engineering of biological circuits can be assisted by these models. We must, however, remember that having a strong predicted SNR for a circuit will not ensure



that a biological circuit computes effectively, any more than using standard TTL components will ensure that an electronic circuit computes effectively: there are many other types of problems that also might interfere with the desired behavior. Importantly, though, having a strong SNR (both the overall circuit and the (SNR_{dB} at individual devices) does mean there is more margin for error in dealing with these other aspects of circuit engineering. Complementarily, an insufficient predicted SNR is a virtual guarantee that the circuit will not work. SNR analysis may thus be expected to be a useful tool for discriminating between possible circuit design alternatives.

As seen in the previous sections, in order to apply SNR analysis, it is necessary to have the following characterization data for each computational device:

- An input/output curve measured in compatible SI units, e.g., by means of the protocols in Beal et al. (2012, 2015), Davidsohn et al. (2015), and Kiani et al. (2014).⁶
- σ_g for the device's output in the (non-circuit) context in which the circuit is expected to operate.

To apply SNR analysis to a circuit requires the following additional information:

- The topology of the circuit, specifying the interconnections between device inputs and outputs.

⁶Technically, relative units could be also used [e.g., RFUs, per Kelly et al. (2009)], but the lack of SI measurements means that it is much more difficult to debug any problems that arise, particularly with regards to differences between practitioners or laboratories.

- Input signal levels σ_g and expression noise σ_g (also implying input SNR).
- SNR requirements for the circuit output.

Given a library of characterized devices and a circuit specification, it is then possible to search for good candidate circuits. The best candidates should go beyond satisfying output SNR requirements and maximize output SNR, in order to have the most margin for dealing with other engineering difficulties. With a homogeneous library of devices with very similar behavior [e.g., as CRISPR-based repressors appear likely to provide (Kiani et al., 2014)], circuit viability can be determined by a straightforward application of the SNR analysis presented in the prior section and devices assigned arbitrarily. With a more heterogeneous library, [e.g., the TetR homolog library in Stanton et al. (2014)], different combinations of devices will have different properties, but the design problem should still be susceptible to efficient search with any number of well-established constrained-search methods (Russell and Norvig, 2003).

More important to the success of circuit engineering is the SNR characteristics of the devices in the library. The three circuit examples in the previous section are characteristic of three general qualitatively different “phases” of expected difficult in engineering biological circuits. These phases are predicted by considering the selection of devices from a library as a search process proposed by Beal (2014). The behavior of such a search process is critically affected by degree of coupling between design choices (i.e., the likelihood that two independent choices are incompatible), as has been well-established in complexity theory (Cheeseman et al., 1991; Hogg et al., 1996) and statistical physics

(Krzakala and Kurchan, 2007; Dall’Asta et al., 2008; Zdeborová, 2008). In this case, the degree of coupling is determined by the likelihood of two devices having an output/input match with a high (SNR_{dB} , leading to three qualitatively different expected engineering environments:

- *Difficult circuits*: when most biological devices in a library are either weak or poorly matched (e.g., having characteristics like Device B), it is difficult to discover a working combination of components even in the best circumstances.

Engineering computational circuits using such devices is expected to be characterized by extensive and lengthy “tuning” and many failed attempts, since even small perturbations in device characteristics (e.g., from the biological operating context) can result in massive SNR losses.

- *Shallow circuits*: when many biological devices in a library have a large region of small negative SNR (e.g., having characteristics like Device C), it is easy to find acceptable matches, but there is still significant signal loss at every device.

Engineering computational circuits using such devices is expected to be relatively simple for circuits up to a certain depth, because there is tolerance for small perturbations and many good candidates for working circuits. When the circuit requires more depth than can be readily attained while maintaining sufficient SNR, however, this strain raises the effective coupling and it will be extremely difficult to engineer an effective circuit of such depth, just as in the prior case.

- *Deep circuits*: when many biological devices in a library have a well-matched region with positive SNR (e.g., having characteristics like Device A), it is easy to find combinations of devices where signals do not degrade from layer to layer.

Engineering computational circuits using such devices is expected to no longer be constrained by issues of computational efficacy: in principle, circuits of any depth and complexity can be readily engineered, and limits instead come from other aspects of the biological implementation, such as circuit delivery and demand on cellular resources.

Analysis of circuit and library SNR characteristics can determine which of these engineering environments we are operating in. Note, however, that there are no “hard” boundaries between phases: rather, as SNR characteristics improve, there is a gradual shift in the dominating engineering constraint from signal matching to signal degradation to non-signal constraints (with concomitant conclusions that can be drawn about the likely difficulty of circuit engineering). Unfortunately, knowing for certain if we are in trouble, while useful, does not actually make it any easier to engineer circuits. Quantification of SNR characteristics can, however, point to what target properties need to be achieved in order to move to a better engineering environment.

3.2. Prospects for Deep Circuit Libraries

Given the widely observed difficulty of engineering biological systems [e.g., Kwok (2010)], it seems intuitive to guess that synthetic biology is currently operating in the “difficult circuits” regime. By applying SNR analysis to current high-efficacy device families, we can verify that this is actually the case. More importantly, however,

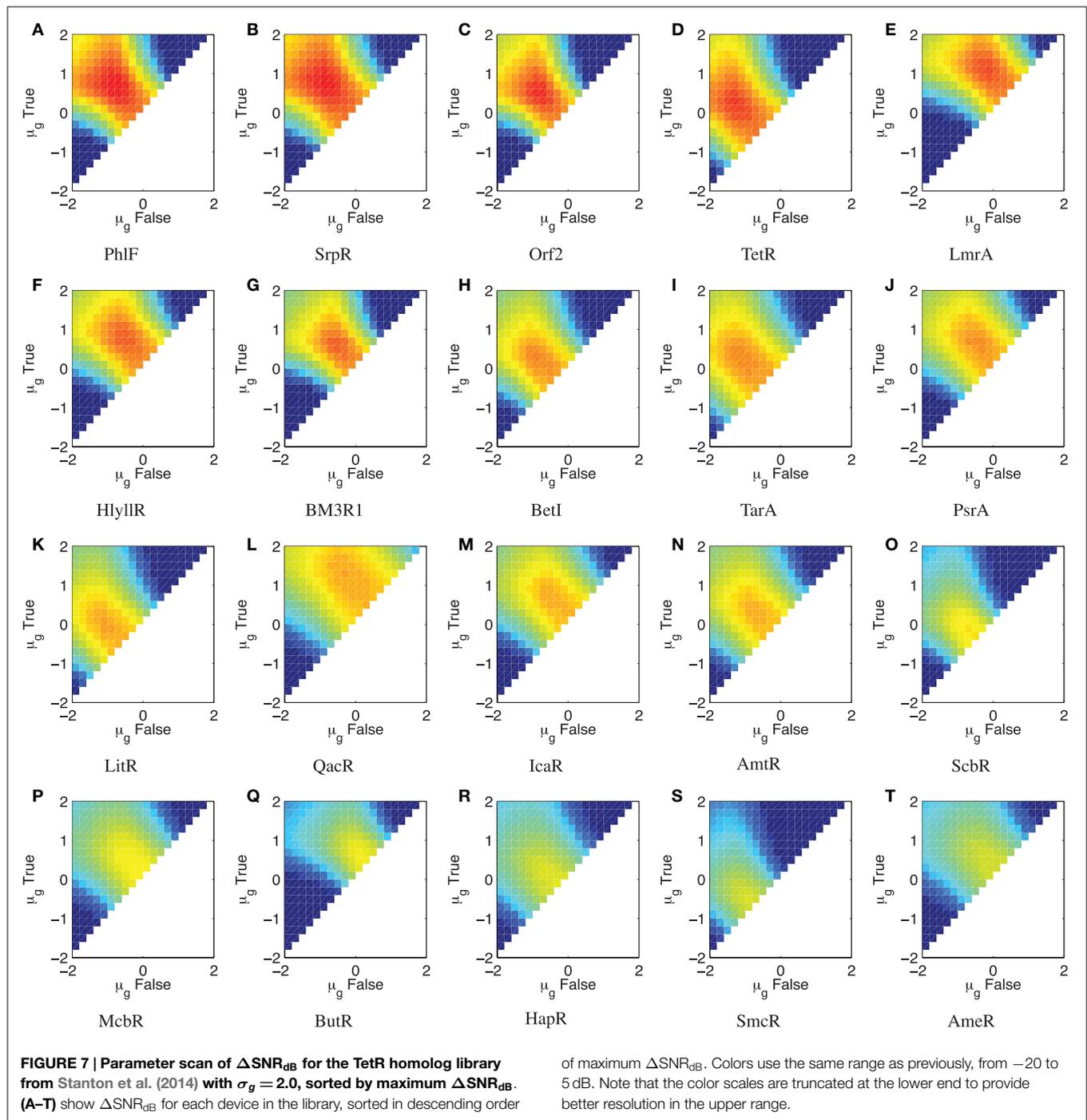
we can also estimate approximately how far these device families are from the “shallow circuit” or “deep circuit” regimes, and what changes would be likely to allow them to attain those goals. When analyzing some properties of some device families, the relevant device characteristics are well enough known to allow rough quantification of requirements; in other cases, only qualitative conclusions can be drawn at present.

At present, there are several families of biological computational devices with the prospect of producing large numbers of universal logic devices with a high differential between output signal levels. The strongest current candidates are homolog mining, integrase logic, TALE and zinc finger repressors, and CRISPR-based repressors, each of which we discuss in detail. Other promising candidates include miRNA, aptamers, RNA-binding proteins, riboregulators, and protein/protein regulation, but all of these currently face various obstacles that mean they appear to be significantly farther away from providing large families of strong universal logic devices. As these technologies continue to mature, however, the same type of analysis presented in this section can be applied to them as well.

3.2.1. Homolog Mining

The TetR repressor is a naturally occurring strong repressor that has been used successfully in many systems. Genomic mining for TetR homologs has produced a library of 20 orthogonal repressors, many of them with quite strong on/off ratios (Stanton et al., 2014). Each repressor has also been characterized with a high-resolution input/output curve (though only in relative units), and the models for these input/output curves are published in Stanton et al. (2014). **Figure 7** shows parameter scans of (SNR_{dB}) for a wide range of input level combinations for all 20 devices, using $\sigma_g = 2.0$ as a conservatively low estimate of a typical value of bacterial expression noise, as estimated from the histograms reported in Stanton et al. (2014) and the noise values reported in Ozbudak et al. (2002). Parameters scans are performed as in Section 2.2 except shifted to the relative unit range of the devices (10^{-2} – 10^2 relative units) and more coarsely, at five levels/decade. A summary of the results is given in **Figure 8A**, which lists the maximum (SNR_{dB}) for each device, along with the on/off ratio reported in Stanton et al. (2014). Of the 20 reported gates, only 4 have the positive (SNR_{dB}) needed that is a pre-requisite for deep circuits. Somewhere around another 5–10 are likely have sufficiently strong (SNR_{dB}) for shallow circuits, given their relatively high amplification and moderate signal loss. Since the library is highly heterogeneous (**Figure 8B**), signal matching must be done on a circuit-by-circuit basis. One significant challenge that is clear from the input/output functions, however, is that few devices have an output $\sigma_{g,\text{false}}$ low enough to achieve the optimal (SNR_{dB}) input; the mismatches between devices can thus be expected to lower the effective (SNR_{dB}) that can be achieved for any circuit.

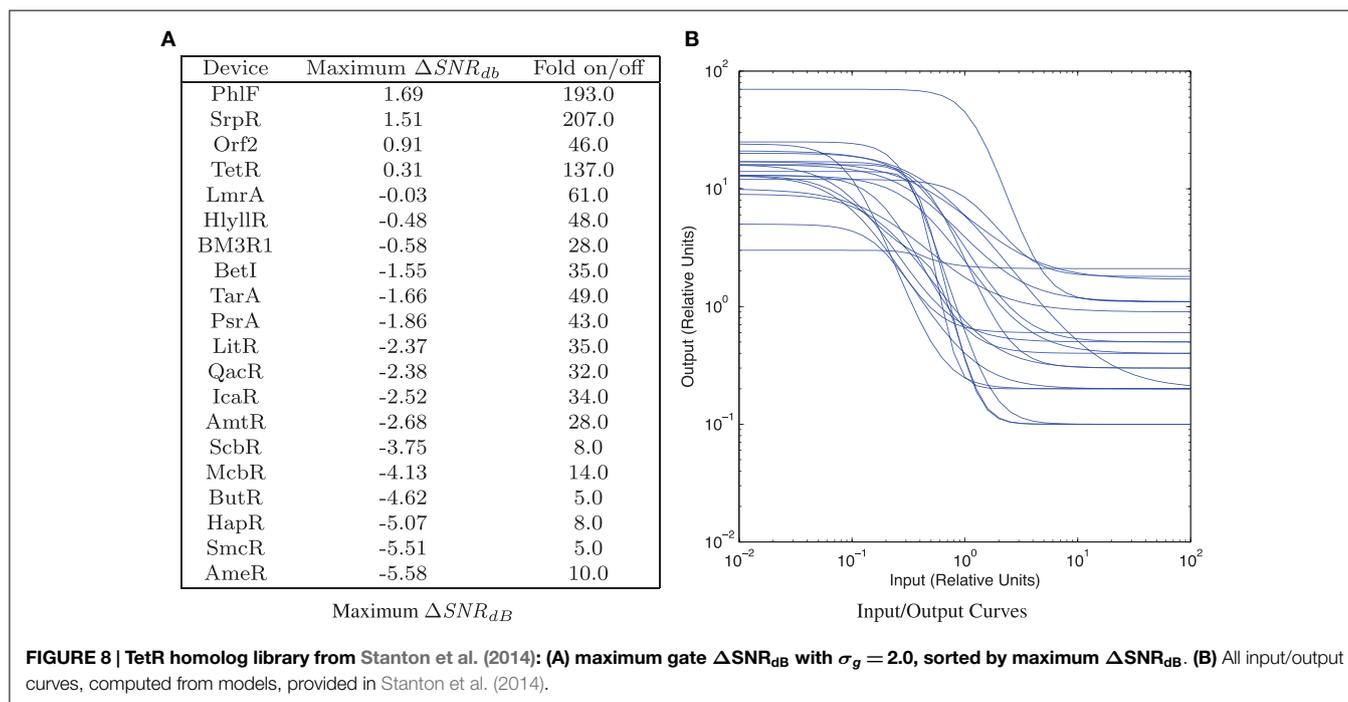
Nevertheless, this library is the closest currently in existence to supporting deep circuits. Key targets for developing that capability are to further expand the library by additional mining, to calibrate the input/output curves to SI units, and to adjust the signal levels to better match, likely by decreasing output expression via 5’UTR modifications.



3.2.2. Integrase Logic

Integrase logic gates, which operate by inverting segments of DNA, have been demonstrated to produce input/output curves with a very high amplification in their transition between high and low output (Bonnet et al., 2013). No model parameters were included in the publication, but the very steep slope of the curves makes it clear that these devices should have a high maximum (SNR_{dB}). This is tempered, however, by a significant number of cells that do not change state, leading to a (SNR_{dB} that appears to be net negative rather than net positive.

At present, however, these integrase logic gates have quite poorly matched input and output signal levels. In addition, to date, very few have been demonstrated: it is reasonable to expect that many more might be discovered through homolog mining, though the availability of usable naturally occurring orthogonal integrases is not yet clear. Key targets for expanding this technology into a library capable of deep computation are genomic mining to expand the number of devices, calibration of the input/output curves to SI units, and adjustment the signal levels to better match, likely by decreasing output expression via 5'UTR modifications.



3.2.3. TALE and Zinc Finger Repressors

TALE proteins are a modular DNA-binding protein that can be engineered to bind to specific sequences with high specificity. Coupled with appropriately designed promoters, TALE proteins have been used to implement extensible libraries of strong promoters (Garg et al., 2012; Davidsohn et al., 2015; Li et al., 2015). TALE repressors can produce remarkably strong repression [measured at a maximum of nearly 5000-fold repression in Garg et al. (2012)]. Detailed input/output curves taken for TALE repressors in Davidsohn et al. (2015) and Li et al. (2015), however, have found a poor slope and uncertain match between input and output levels, implying a poor (SNR_{dB} for composed TALEs – consistent with the low input/output differential observed in the composite circuits investigated in that paper.

At present, TALEs are thus viable only for implementing very shallow circuits with low SNR. One likely path for increasing their potential depth is to increase repression strength by adjusting the synthetic promoter architectures used for TALE repressors. Given the level of deamplification observed in circuits in Davidsohn et al. (2015) and Li et al. (2015), an approximately 10-fold increase would likely be sufficient and may be attainable through this approach. Another possibility might be to heighten cooperativity (steepening the input/output curve) by changing the TALE to a fusion protein. Furthermore, the characterization in Davidsohn et al. (2015) was of transient rather than converged operation (i.e., fluorescence levels were still changing over time, rather than having reached a stable level of expression), and it is possible that TALE repressors may have a significantly steeper input/output curve when converged.

Zinc finger repressors are a very similar modular protein technology, which has also been demonstrated to produce strong orthogonal repressors [e.g., Khalil et al. (2012) and Lohmueller et al. (2012)]. No detailed input/output curves of these strong

repressors have been produced to date, so obtaining input/output curves in SI units is the first key step to evaluating the viability of zinc finger repressors as a library. Given the similarity in promoter architectures used in the two technologies, however, it seems likely that they will face similar challenges to TALE repressors.

3.2.4. CRISPR-Based Repressors

CRISPR-based repressors are a recent addition to the set of candidate libraries (Kiani et al., 2014), based on a protein that can be targeted with high specificity by a separately expressed sequence of guide RNA (gRNA). Like TALE and zinc finger repressors, they have showed very high repression strength, and may be significantly more homogeneous and easier to engineer with since the sequences are much shorter and do not involve any protein design. They have not yet had detailed input/output curves measured, however, and what characterization has been done to date has been of transient rather than converged behavior, as with TALE repressors.

For this family, the clear next step toward deep computation is to determine the SNR characteristics of the components, though this is complicated by their current use of a Pol III promoter to express gRNA, which is not compatible with the fluorescent proteins typically used for characterization. If the CRISPR-based repressors prove to have a steep slope in their converged behavior, then their SNR may already be sufficient for deep circuits; otherwise, they will likely require similar promoter engineering to TALE and zinc finger devices.

All told, we see that the current situation of synthetic biology is one of difficult circuit engineering. Even though some devices provide good SNR, there are not enough and there is not enough compatibility to reliably support engineering of either shallow or deep circuits. Other devices may also provide good SNR, but require characterization before this can be determined and, if true,

effectively exploited. For all of these families of devices, however, the SNR approach identifies key targets for improvement that appear to be reasonable to aim for and that offer the prospect of enabling deep circuit engineering and the transformative capabilities that would imply.

4. Discussion of Contributions

This paper has developed methods for characterizing the efficacy of biological computing devices and circuits based on signal-to-noise ratio. This approach has the advantage of being firmly mathematically grounded in the fundamental definition of a signal, and can be applied using readily obtainable characterization data. This paper has also illustrated the use of SNR methods by applying them to analyze individual devices and predict the behavior of circuits in simulation, as well to develop a framework for SNR-based circuit engineering. Finally, a SNR-based analysis of current device libraries indicates that, while no library is yet sufficient to support deep biological circuits, several may be able to if particular targeted improvements can be realized.

References

- Adams, D. S., and Levin, M. (2013). Endogenous voltage gradients as mediators of cell-cell communication: strategies for investigating bioelectrical signals during pattern formation. *Cell Tissue Res.* 352, 95–122. doi:10.1007/s00441-012-1329-4
- Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., et al. (2006). Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38, 636–643. doi:10.1038/ng1807
- Beal, J. (2014). Bridging the gap: a roadmap to breaking the biological design barrier. *Front. Bioeng. Biotechnol.* 2:87. doi:10.3389/fbioe.2014.00087
- Beal, J., Lu, T., and Weiss, R. (2011). Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks. *PLoS ONE* 6:e22490. doi:10.1371/journal.pone.0022490
- Beal, J., Wagner, T. E., Kitada, T., Azizgolshani, O., Parker, J. M., Densmore, D., et al. (2015). Model-driven engineering of gene expression from RNA replicons. *ACS Synth. Biol.* 4, 48–56. doi:10.1021/sb500173f
- Beal, J., Weiss, R., Yaman, F., Davidsohn, N., and Adler, A. (2012). *A Method for Fast, High-Precision Characterization of Synthetic Biology Devices*. Technical Report MIT-CSAIL-TR-2012-008: MIT. Available at: <http://hdl.handle.net/1721.1/69973>.
- Bilitchenko, L., Liu, A., Cheung, S., Weeding, E., Xia, B., Leguia, M., et al. (2011). Eugene: a domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PLoS ONE* 6:e18882. doi:10.1371/journal.pone.0018882
- Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013). Amplifying genetic logic gates. *Science* 340, 599–603. doi:10.1126/science.1232758
- Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26, 787–793. doi:10.1038/nbt1413
- Cheeseman, P., Kanefsky, B., and Taylor, W. M. (1991). "Where the really hard problems are," in *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Vol. 1 (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 131–137.
- Dall'Asta, L., Ramezanpour, A., and Zecchina, R. (2008). Entropy landscape and non-gibbs solutions in constraint satisfaction problems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77, 031118. doi:10.1103/PhysRevE.77.031118
- Danino, T., Mondragon-Palomino, O., Tsimring, L., and Hasty, J. (2009). A synchronized quorum of genetic clocks. *Nature* 463, 326–330. doi:10.1038/nature08753
- Davidsohn, N., Beal, J., Kiani, S., Adler, A., Yaman, F., Li, Y., et al. (2015). Accurate predictions of genetic circuit behavior from part characterization and modular composition. *ACS Synth. Biol.* 4, 673–681. doi:10.1021/sb500263b
- Ellis, T., Wang, X., and Collins, J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* 27, 465–471. doi:10.1038/nbt.1536
- Elowitz, M., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi:10.1038/35002125
- Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi:10.1126/science.1070919
- Friedman, N., Cai, L., and Xie, X. S. (2006). Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97, 168302. doi:10.1103/PhysRevLett.97.168302
- Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi:10.1038/35002131
- Garg, A., Lohmueller, J. J., Silver, P. A., and Armel, T. Z. (2012). Engineering synthetic tal effectors with orthogonal target sites. *Nucleic Acids Res.* 40, 7584–7595. doi:10.1093/nar/gks404
- Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* 40, 4–7. doi:10.1371/journal.pone.0041098
- Hogg, T., Huberman, B. A., and Williams, C. P. (1996). Phase transitions and the search problem. *Artif. Intell.* 81, 1–15. doi:10.1016/0004-3702(95)00050-X
- Huynh, L., Tsoukalas, A., Köppe, M., and Tagkopoulou, I. (2013). Sbrome: a scalable optimization and module matching framework for automated biosystems design. *ACS Synth. Biol.* 2, 263–273. doi:10.1021/sb300095m
- Kelly, J. R., Ruben, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., et al. (2009). Measuring the activity of biobrick promoters using an in vivo reference standard. *J. Biol. Eng.* 3, 4. doi:10.1186/1754-1611-3-4
- Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J., and Khalil, A. S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* 158, 110–120. doi:10.1016/j.cell.2014.04.047
- Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., et al. (2012). A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150, 647–658. doi:10.1016/j.cell.2012.05.045
- Kiani, S., Beal, J., Ebrahimkhani, M. R., Huh, J., Hall, R. N., Xie, Z., et al. (2014). Crispr transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* 11, 723–726. doi:10.1038/nmeth.2969

- Kim, B., and Lin, M. Z. (2013). Optobiology: optical control of biological processes via protein engineering. *Biochem. Soc. Trans.* 41, 1183–1188. doi:10.1042/BST20130150
- Knight, T. F. Jr., and Sussman, G. J. (1998). “Cellular gate technology,” in *Unconventional Models of Computation* (Berlin: Springer-Verlag), 257–272.
- Krzakala, F., and Kurchan, J. (2007). Landscape analysis of constraint satisfaction problems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 021122. doi:10.1103/PhysRevE.76.021122
- Kwok, R. (2010). Five hard truths for synthetic biology. *Nature* 463, 288–290. doi:10.1038/463288a
- Li, Y., Jiang, Y., Chen, H., Liao, W., Li, Z., Weiss, R., et al. (2015). Modular construction of mammalian gene circuits using tale transcriptional repressors. *Nat. Chem. Biol.* 11, 207–213. doi:10.1038/nchembio.1736
- Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J. D., and Hillson, N. J. (2012). Par-par laboratory automation platform. *ACS Synth. Biol.* 2, 216–222. doi:10.1021/sb300075t
- Lohmueller, J. J., Armel, T. Z., and Silver, P. A. (2012). A tunable zinc finger-based framework for boolean logic computation in mammalian cells. *Nucleic Acids Res.* 40, 5180–5187. doi:10.1093/nar/gks142
- Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. (2012). Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* 30, 1137–1142. doi:10.1038/nbt.2401
- Marchisio, M. A., and Stelling, J. (2011). Automatic design of digital synthetic gene circuits. *PLoS Comput. Biol.* 7:e1001083. doi:10.1371/journal.pcbi.1001083
- Moon, T., Lou, C., Tamsir, A., Stanton, B., and Voigt, C. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature* 491, 249–253. doi:10.1038/nature11516
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., et al. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360. doi:10.1038/nmeth.2404
- Myers, C. J., Barker, N., Jones, K., Kuwahara, H., Madsen, C., and Nguyen, N.-P. D. (2009). ibiosim: a tool for the analysis and design of genetic circuits. *Bioinformatics* 25, 2848–2849. doi:10.1093/bioinformatics/btp457
- Oppenheim, A. V., and Willsky, A. S. (1997). *Signals and Systems*. Upper Saddle River, NJ: Prentice-Hall.
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73. doi:10.1038/ng869
- Purnick, P. E. M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422. doi:10.1038/nrm2698
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science* 307, 1962–1965. doi:10.1126/science.1106914
- Russell, S. J., and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, 2 Edn. New York, NY: Pearson Education.
- Stanton, B., Nielsen, A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* 10, 99–105. doi:10.1038/nchembio.1411
- Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* 6:e16765. doi:10.1371/journal.pone.0016765
- Weiss, R. (2001). *Cellular Computation and Communications using Engineered Genetic Regulatory Networks*. Ph.D. thesis, MIT, Cambridge, MA.
- Yaman, F., Bhatia, S., Adler, A., Densmore, D., and Beal, J. (2012). Automated selection of synthetic biology parts for genetic regulatory networks. *ACS Synth. Biol.* 1, 332–344. doi:10.1021/sb300032y
- Zdeborová, L. (2008). *Statistical Physics of Hard Optimization Problems*. arXiv. arXiv:0806.4112. Available at: <http://arxiv.org/abs/0806.4112>

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Beal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Obsolescence and intervention: on synthetic-biological entities

Andrés Moya^{1,2,3*}

¹ Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain

² Valencian Region Foundation for the Promotion of Health and Biomedical Research (FISABIO), Valencia, Spain

³ Consortium for Biomedical Research in Epidemiology and Public Health (CIBEResp), Madrid, Spain

*Correspondence: andres.moya@uv.es

Edited by:

Pablo Carbonell, University of Evry, France

Reviewed by:

Daniel James Nicholson, University of Exeter, UK

Keywords: natural entities, artificial entities, metabolism, evolution, machines

Oftentimes, topics that might fall outside of science's remit seem to end up becoming a part of it, sooner or later. This appears to be the case of synthetic biology, a new biological science (although some maintain that it is a form of engineering, or treat it as such; *Endy, 2005*), which seems to have become essential to the understanding of living beings and their extreme manipulation. I believe it to be a new form of biology. In truth, synthetic biology has a long history and, conceptually speaking, may well have formed part of the interests and research efforts of our illustrious predecessors throughout the first half of the twentieth century and even earlier. In any event, and broadly speaking, it may be asserted that such efforts were premature and that the state of the science, at that time, did not allow for progress, in terms of the modification, creation, or recreation of organisms or parts thereof that we have today. Now, it has rather come of age, which is why I think of synthetic biology as a young or new biological science (*Moya, 2014*).

To a great extent, when you talk about synthetic biology, what you are doing, first and foremost, is making a statement of intent, if not expressing your concern, wish, or hope that any biological organism which this discipline might examine should be completely under control, and that it should not deviate even the slightest bit from the role that has been ascribed to it. Of course, if some achievements had not already been made in the field of synthetic biology, clinging to a mere statement of intent would do little to further its chances of survival in the future. If some sciences

make headway, it is because some of their initial achievements thrust fame, awareness, and recognition upon them, in the eyes of the scientific community and the rest of society. Here is where we start to get to grips with wishes and hopes for these new sciences and their potential. This is what is happening with synthetic biology.

Fundamentally, the synthetic biologist aspires to make manufactured biological entities behave just as a car might, once it has been put together on an assembly line. The metaphor of the car is as valid as that of any other mechanical entity, or any other type of entity, with perhaps the only condition being that they should all be put together on an assembly line. A car, effectively, begins to take shape on an assembly line and ends up having a specific physical form. It comprises many different components and each one has a specific function, as designated by the manufacturer. Together, its components combine to form a mechanical entity that functions in a pre-determined way because of prior knowledge of the manner in which each component works, and the manufacturer puts them all together following a pre-determined plan, so that the whole may function as desired. Attempting to make a manufactured biological entity function in the same way as a car leads us to make two relevant observations. To begin with, there is the obsolescence of the entity, then there is intervention in the entity itself. The nature of these properties differs in cells when they are compared to cars.

First, let us take a look at the question of obsolescence. As everyone knows, a car has a limited lifespan. Its constituent

parts, in particular those that are essential in order for it to run, become worn down, inevitably, until the materials [from which they are made] break down and become altered through the contact of some parts with others, or on account of the various reasons for which the car might cease to run. I refer not to the failure of the car to run, on account of damage incurred following an accident, but quite simply to wear and tear, rendering it unable to perform the purpose for which it has been intended. To what degree might this metaphor apply to the biological entity, in general, and the synthetic-biological entity, in particular? In truth, the following question might be asked, equally, of the biological entity and the car. Might it suffer some type of wear and tear that causes it either to stop functioning or to function in a different way before, for example, reaching the culminating point of its division or reproduction? In other words, before it can reproduce itself? This is, in effect, the case. Let us take, for example, an individual biological entity; a microorganism or a cell from a multicellular organism which, on account of its specialization, could be a germ cell – the purpose of which is the reproduction of the organism – or a non-germ cell, which will subdivide through mitosis, creating copies of itself. In the case of the microorganism, or the non-germ cells, there comes a point at which they divide, producing two genetic copies of themselves; however, up until that point, they have undergone transformation processes, ending with metabolism, that have altered them in relation to what they might have

become in the moment that they were generated by their respective parent cells. It is generally believed that it is only at the point of division (reproduction), when genetic changes occur, that they are at their most relevant, when considering any possible transformation. These changes manifest themselves once they start to exhibit genetic differences from their progenitors; however, what I am talking about here are previous changes in the metabolic machinery. The cell is transformed, grows, and, before multiplying, changes. Consequently, this is not just about genetic changes; it is also about changes in the cell's metabolism (de Lorenzo, 2014). Contrary to cars, that cannot avoid wear and tear, cells are undergoing metabolic changes to avoid obsolescence up to a point. A microbial cell is usually said to be immortal, in that before it ceases to exist in that form, it has already divided. And in a multicellular organism, with specialized reproductive cells, it is usually these cells themselves that join with the reproductive cells of other organisms, having previously undergone meiosis, in order to reproduce. Yet, all of these different types of cells – the immortal microbial cells, the germ cells that have the capacity to reproduce, and all the other non-germ cells that undergo mitosis – transform metabolically, before reaching their respective points of division and reproduction. And these changes, which are quite noticeable and dramatic, can be crucial. So much so that they may end the very life of the corresponding entities before division or reproduction can take place. All of these cells have a lifespan that has been optimized, through evolution, to ensure that they reach the point of division or reproduction, whichever is relevant. We can, it seems, only take the metaphor of the cell as car so far. There is something that does not quite fit, because eventually cars become obsolete and end up in the scrapyard, whereas cells, even though they have a limited lifespan, and undergo transformation and decomposition eventually, are able to reproduce. They arrive, it would seem, at these stages, with a certain degree of autonomy. The higher autonomy displayed by living systems has its basis in their internal organizational dynamics, that is, in the fact that cells (and organisms in general) are self-maintaining, self-organizing, self-repairing, and self-reproducing systems

(Nicholson, 2013). It is these properties that confer organisms a far greater degree of functional autonomy when compared to machines. The metaphor of the car would only be valid if it were possible to extend the lifespan of its component parts and if, by means of systematic repair, the car continued to be the same car that rolled off the assembly line. In order for the car to remain the same, intervention must take place. And here is where we come to our second observation, the biological entity – in order to remain the same and reach the point of division or reproduction – effects its own intervention, regulating itself by means of its metabolism. All of evolution – more than 3000 million years of ceaseless effort – has conspired to ensure that this self-driven intervention of single-cell organisms, which divide, and multicellular organisms, which have cells that divide and others that reproduce, is effective; in other words, ensuring that the transformations that take place as a result of these organisms being in permanent interaction with their environment do not do so to such an extent that they cause them to decompose or degenerate before they can divide or reproduce (Danchin, 2009). The obsolescence of a cell is remedied by self-intervention, whereas the wear and tear of a car is externally remedied. The crucial difference with regards to intervention seems to be that cars require external intervention to remain viable and operational whereas cells have internal means of repairing damaged parts and compensating against external perturbations. This observation, which I consider to be crucial, and which I have formulated with regard to biological entities, must be kept in mind when considering synthetic-biological entities. Because even when cars become obsolescent, and their obsolescence can be remedied through intervention, biological entities more or less remedy their own obsolescence by effecting their own intervention.

If a synthetic-biological entity is, by definition, a biological entity, it will resist obsolescence autonomously, or much more autonomously than a car might (Nicholson, 2013). It would only fail to do so if, focusing on the synthetic part of the entity, we were to introduce some types of control that might prevent this natural dynamic. I cannot, however, conceive of another way of controlling the biological

entity that is not based on absolute knowledge of the entity. In other words, intervention into the biological entity or, failing that, manufacturing a biological entity from biological components, i.e., creating a synthetic-biological entity, differs fundamentally from the case of the entity we refer to as a car. Every aspect of the latter is designed, right from the start. By contrast, the biological entity is not designed, and our knowledge of how it functions is in no way complete. This being the case, we find ourselves in uncharted territory, not that I wish to suggest that we are in completely uncharted territory, or that there is no way to control what we are dealing with. There exist various controls of and external interventions into the entity itself, which may render said entity controllable, or perhaps, I should say, increasingly controllable, rather than totally controllable. The synthetic-biological entity is preceded by a biological entity which, as an autonomous system in the process of evolution, requires a certain level of knowledge in order for its self-effected interventions to be controlled (Serrano, 2007).

ACKNOWLEDGMENTS

This work was supported by projects SAF2012-31187 from the Ministerio de Ciencia e Innovación, PrometeoII/2014/065 from the Generalitat Valenciana, Spain, and ST-FLOW from the European Commission. The author acknowledges comments of the reviewer to improve the manuscript.

REFERENCES

- Danchin, A. (2009). Bacteria as computers making computers. *FEMS Microbiol. Rev.* 33, 3–26. doi:10.1111/j.1574-6976.2008.00137.x
- de Lorenzo, V. (2014). From the *selfish gene* to *selfish metabolism*: revisiting the central dogma. *Bioessays* 36, 226–235. doi:10.1002/bies.201300153
- Endy, D. (2005). Foundations for engineering biology. *Nature* 438, 449–453. doi:10.1038/nature04342
- Moya, A. (2014). *The Calculus of Life*. New York: Springer Verlag (in press).
- Nicholson, D. J. (2013). Organisms ≠ machines. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 669–678. doi:10.1016/j.shpsc.2013.05.014
- Serrano, L. (2007). Synthetic biology: promises and challenges (editorial). *Mol. Syst. Biol.* 3, 158. doi:10.1038/msb4100202

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 October 2014; accepted: 11 November 2014; published online: 26 November 2014.

Citation: Moya A (2014) Obsolescence and intervention: on synthetic-biological entities. *Front. Bioeng. Biotechnol.* 2:59. doi:10.3389/fbioe.2014.00059

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Moya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or

reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*

Tamás Fehér^{1,2}, Vincent Libis^{1,3}, Pablo Carbonell^{1,4,5} and Jean-Loup Faulon^{1,5}*

¹ Institute of Systems and Synthetic Biology, University of Evry Val d'Essonne, Evry, France

² Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary

³ Paris Diderot University, Paris, France

⁴ Research Programme on Biomedical Informatics (GRIB), Department of Experimental and Health Sciences, Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain

⁵ SYNBIOCHEM Center, Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, UK

Edited by:

Jean Marie François, CNRS, France

Reviewed by:

Mattheos Koffas, Rensselaer

Polytechnic Institute, USA

Shota Isogai, Technology Research

Association of Highly Efficient Gene

Design, Japan

*Correspondence:

Jean-Loup Faulon, 5 rue Henri

Desbruères, Genavenir 6, F-91030,

Evry Cedex, France

e-mail: jean-loup.faulon@issb.

genopole.fr

Production of value-added chemicals in microorganisms is regarded as a viable alternative to chemical synthesis. In the past decade, several engineered pathways producing such chemicals, including plant secondary metabolites in microorganisms have been reported; upscaling their production yields, however, was often challenging. Here, we analyze a modular device designed for sensing malonyl-CoA, a common precursor for both fatty acid and flavonoid biosynthesis. The sensor can be used either for high-throughput pathway screening in synthetic biology applications or for introducing a feedback circuit to regulate production of the desired chemical. Here, we used the sensor to compare the performance of several predicted malonyl-CoA-producing pathways, and validated the utility of malonyl-CoA reductase and malonate-CoA transferase for malonyl-CoA biosynthesis. We generated a second-order dynamic linear model describing the relation of the fluorescence generated by the sensor to the biomass of the host cell representing a filter/amplifier with a gain that correlates with the level of induction. We found the time constants describing filter dynamics to be independent of the level of induction but distinctively clustered for each of the production pathways, indicating the robustness of the sensor. Moreover, by monitoring the effect of the copy-number of the production plasmid on the dose–response curve of the sensor, we managed to coarse-tune the level of pathway expression to maximize malonyl-CoA synthesis. In addition, we provide an example of the sensor's use in analyzing the effect of inducer or substrate concentrations on production levels. The rational development of models describing sensors, supplemented with the power of high-throughput optimization provide a promising potential for engineering feedback loops regulating enzyme levels to maximize productivity yields of synthetic metabolic pathways.

Keywords: malonyl-CoA, dynamic pathway regulation, high-throughput screening, synthetic regulatory circuit, fluorescent reporter circuit, sensor-actuator circuit

INTRODUCTION

Natural metabolic pathways are highly regulated and can adapt dynamically to changes in the levels of chemicals within and around the cell (Chubukov et al., 2014). For this purpose, cells have developed a high number of sensors to monitor and control their own metabolic state. This stringent regulatory process is impaired or lost when metabolic engineers insert heterologous enzymes into artificial pathways, usually at non-optimal levels of activities, leading to accumulation of intermediate metabolites and reduced growth rate (Holtz and Keasling, 2010). To address this issue, several projects have attempted to mimic natural regulation by adding such sensors to engineered metabolic pathways. The advantages of using natural sensors such as transcription factors to implement synthetic dynamic regulation in heterologous pathways have been demonstrated in several studies: following pioneering work on lycopene synthesis regulation (Farmer and Liao, 2000), this strategy also led recently to higher yields in fatty

acid synthesis with the help of fatty acids (Zhang et al., 2012) and malonyl-CoA (Xu et al., 2014; Liu et al., 2015) biosensors.

Such biosensor-based approaches allow the bacterial factory to artificially monitor its own level of metabolites and to modify expression levels of certain enzymes in order to balance pathway function. They can also be used to devise external monitoring systems to aid the selection of cells displaying improved biological functions. A major application for this monitoring is the screening of libraries of variant strains in directed evolution of enzymes (Michener and Smolke, 2014). Screening of producers is still a major bottleneck in metabolic engineering as quantification of the compound of interest usually relies on low-throughput measurement techniques such as liquid or gas chromatography. However, the typical throughput necessary for directed evolution applications or combinatorial libraries' screening is several orders of magnitude higher (Dietrich et al., 2010). Therefore, *in vivo* biosensing offers a relevant non-destructive, cost-effective,

and high-throughput alternative for monitoring production levels (Schallmeyer et al., 2014). This strategy has proven successful in identifying high-producing strains, via the coupling of biosensor to reporter genes such as fluorescent proteins or fitness-related proteins. The screening can either be done at the colony level, in microtiter plates or by fluorescence-activated cell sorting (FACS). For instance, screening of *Escherichia coli* colonies expressing a transcription factor evolved to detect mevalonate allowed successful identification of high producers of mevalonate in a library of mevalonate-synthesis variants (Tang and Cirino, 2011). Theoretically, FACS may allow even higher throughputs and was notably used to evolve the caffeine demethylase enzyme in yeast cells harboring riboswitches able to detect theophylline, the enzyme's product (Smolke and Michener, 2012). FACS was also used in combination with transcription factor-based sensors to measure L-lysine production in *C. glutamicum* (Binder et al., 2012) and detect production of Kaempferol in engineered *E. coli* (Siedler et al., 2014). Alternatively, growth rate can be coupled to the production of the product of interest via transcription factor-dependent expression of antibiotic resistance proteins. This strategy has been used in a proof of principle experiment to enrich a population of *E. coli* for cells harboring the pathway for 1-butanol-production (Dietrich et al., 2013). Later, the power of multiplex automated genome engineering was combined with this approach to optimize the production of naringenin and glucaric acid in *E. coli* (Raman et al., 2014).

Because both pathway dynamic regulation and screening rely on the same type of biosensors, a synergy exists between these two approaches. Illustrating such synergy, here we adapted a malonyl-CoA biosensor previously developed for a dynamic regulation purpose (Liu et al., 2015) to a new screening goal. Transcription factor-based biosensing is particularly interesting for monitoring malonyl-CoA as it is difficult and time consuming to quantify this molecule *in vivo* via standard LC/MS. In addition to being an important step for fatty acid synthesis in *E. coli*, malonyl-CoA production is also a limiting step for the production of plant flavonoids in engineered *E. coli*. Our group is working on the development of computer-assisted design of metabolic pathways and is seeking high-throughput techniques to test predicted pathways and enzymes for continual feedback-optimization of the prediction algorithm. In our earlier proof of concept work, we used our software RetroPath (Carbonell et al., 2011) to identify and rank pathways capable of producing the high-value flavonoid pinocembrin in *E. coli*. During the implementation of multiple constructs, Malonyl-CoA, was identified as a bottleneck, preventing high-yield production. Here, we used the biosensor based on transcription factor FapR to analyze the malonyl-CoA-producing pathways that were proposed by our software to remove the bottleneck. We identified optimal conditions for malonyl-CoA monitoring by fine-tuning levels of expression of the transcription factor and we fitted our experimental data to a model giving insights on proper interpretation of output profiles encountered in such screening protocols. This approach helped us identify the best candidate pathways, validate two novel routes of malonyl-CoA synthesis, and optimize enzyme-expression levels through modification of plasmid copy-number.

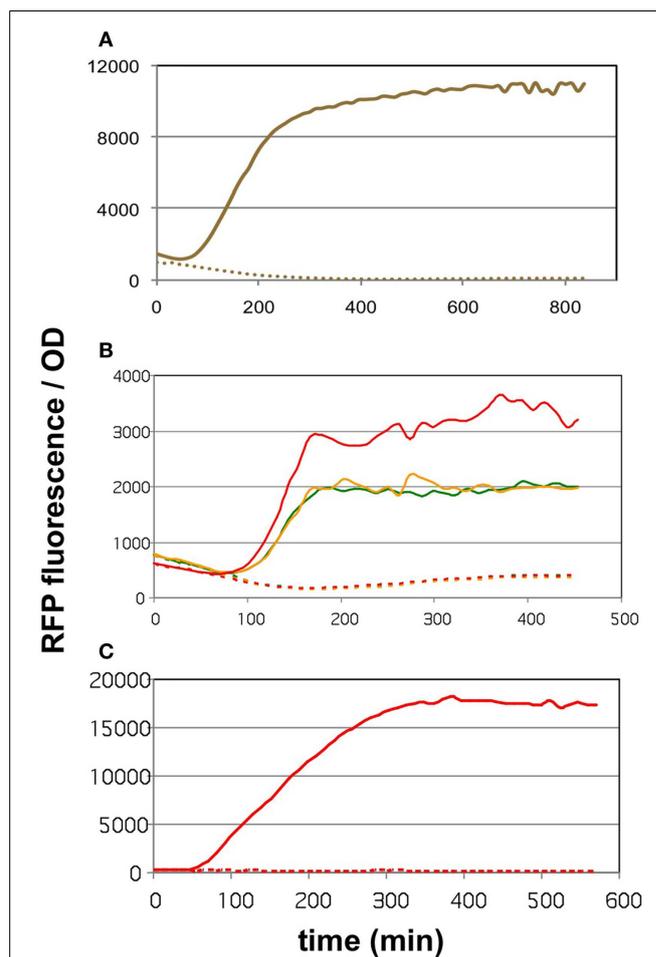
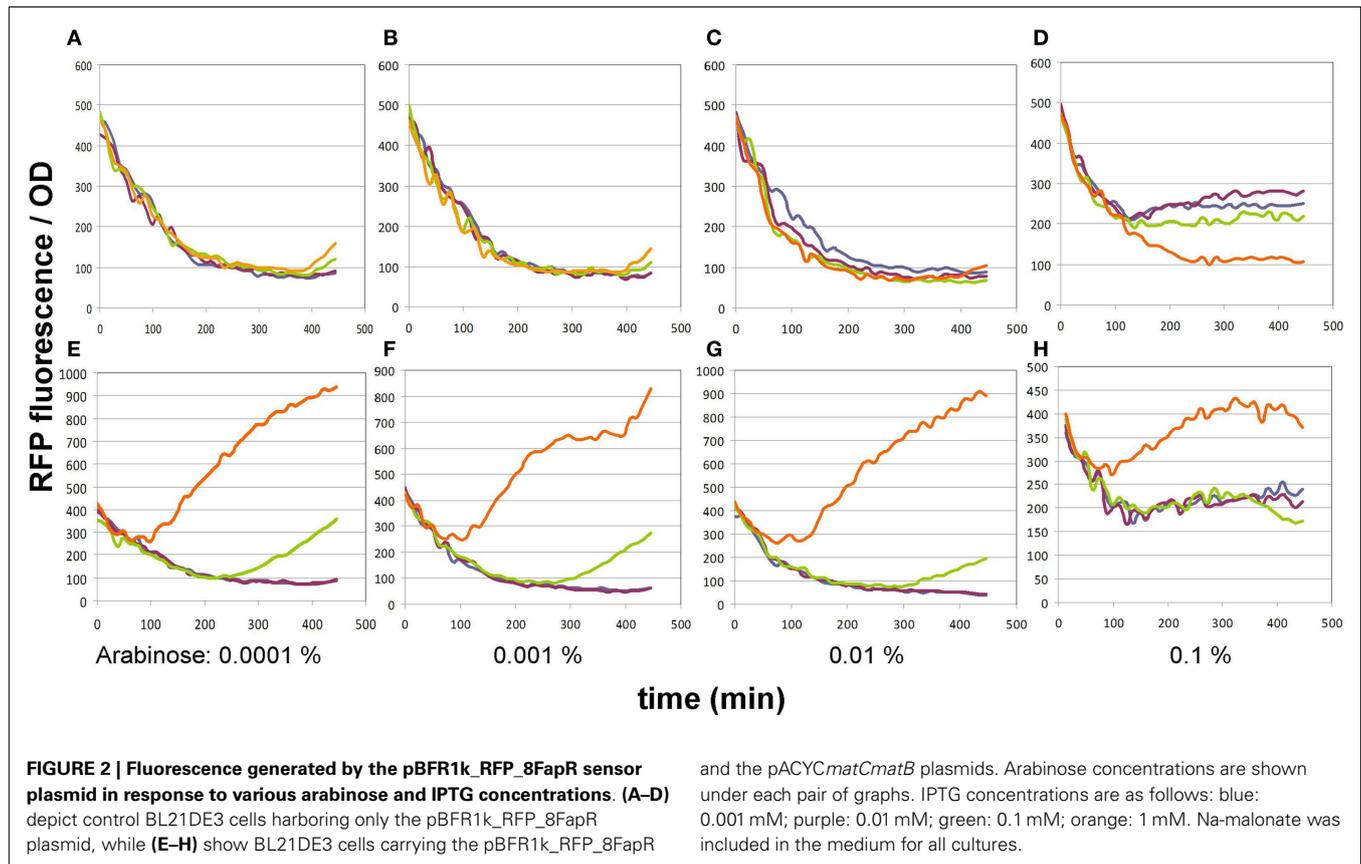


FIGURE 1 | The effect of cerulenin on fluorescence generated by the sensor circuits. On all three graphs, the solid lines show the cultures challenged with cerulenin, while the segmented lines depict the corresponding untreated cultures. **(A)** DH5 α cells harboring the pBFR1k_RFP_8FapR sensor plasmid in mineral salts medium. **(B)** DH5 α cells harboring the pCFR sensor plasmid in LB medium. Three cultures were tested, originating from three colonies obtained after transformation of the pCFR ligate. **(C)** DH5 α cells harboring the pCFR sensor plasmid from the best performer colony seen in **(B)**, cultured in mineral salts medium.

MATERIALS AND METHODS

STRAINS AND PLASMIDS

Plasmid construction was done in *E. coli* strain DH5 α , and all malonyl-CoA sensing experiments were carried out in strain BL21DE3. Plasmid pBFR1k_RFP_8FapR (Liu et al., 2015) was a kind gift of Dr. Fuzhong Zhang (Washington University, St. Louis, MO, USA). To construct plasmid pCFR, pBFR1k_RFP_8FapR was PCR-amplified using primers pBFR-FW (ACTGTCGACGAAACGATCCTCCTCCTG) and pBFR-Rev (GCTCTAGATTCTTCTGAGCGGGACTCTG), and the Cm-resistance marker of pSG76-CS (GeneBank: AF402780.1) was amplified with primers pSGCSH-FW (GCTCTAGAGTGAGGCA CCAATAACTG) and pSGCSH-Rev (ACTGTCGACGATCGGCA CGTAAGAGGTTTC). Both PCR products were double-digested



with *Xba*I and *Sal*I, and were ligated using T4 ligase. Plasmid pMSD8 was a kind gift of Prof. John Cronan (University of Illinois, Urbana, IL, USA), and plasmids pETM6-MaccABCD and pETM6-PaccABCD (Xu et al., 2013) were kindly provided by Prof. Mattheos Koffas (Rensselaer Polytechnic Institute, Troy, NY, USA). The construction of plasmids pRSFmatCmatB, pRSFmmsA, pRSFcagg1256, pRSFMaccABCD, and pRSFmatCatoDA was described elsewhere (Fehér et al., 2014). Plasmid pACYCmatCmatB (Wu et al., 2013) was a kind gift of Dr. Jingwen Zhou (LBBE, Jiangnan, China). pACYCMaccABCD was constructed by double-digesting pRSFMaccABCD with *Bam*HI and *Avr*II, and ligating the 4855 bp-long fragment with the similarly digested pACYCmatCmatB. pACYCmatCatoDA was engineered by double-digesting pRSFmatCatoDA with *Nde*I and *Xho*I, and ligating the 1408 bp fragment with the 5273 bp fragment of the similarly digested pACYCmatCmatB. pACYCmmsA and pACYCcagg1256 were constructed by double-digesting pRSFmmsA and pRSFcagg1256, respectively, both with *Eco*RI and *Xho*I, and ligating the respective 1685 and 3848 bp fragments with the similarly digested pACYCmatCmatB. Restriction endonucleases and T4 ligase were obtained from Thermo Fisher Scientific (Waltham, MA, USA), Q5 DNA polymerase for PCR was from New England Biolabs.

FLUORESCENCE MEASUREMENT OF BACTERIAL CULTURES

Each bacterial strain to be measured was grown overnight in mineral salts medium (Hall, 1998) supplemented with 0.2% glucose

and the appropriate antibiotics. On the day of the measurement, the cultures were diluted 20-fold in fresh medium, and dispensed into clear-bottomed black-walled 96-well plates (Costar Ref. 3603, Corning, NY, USA). Growth and measurement of the cultures took place in a TECAN Infinite 500 fluorescent reader, shaken and incubated at 37°C. Optical density was determined at 600 nm. Fluorescence intensity was recorded using an excitation wavelength of 580 ± 10 nm, and an emission wavelength of 610 ± 5 nm, with a gain set to 40. Readings were taken every 12 min. Antibiotics were used at the following concentrations: ampicillin (Ap): 50 μ g/ml, chloramphenicol (Cm): 25 μ g/ml, kanamycin (Km): 30 μ g/ml. Na-malonate was administered at 2 mg/ml, cerulenin was used at 20 μ g/ml, and β -alanine at 3 mM end concentrations. For characterization of the sensors' response to malonyl-CoA, IPTG was administered in a concentration-series of 0.01, 0.1, 0.3, 0.6, 1, and 10 mM. Arabinose was used in a log10 series from 10^{-4} to 10^{-1} % during optimization and at 10^{-2} % for comparison of alternate malonyl-CoA producer constructs. All chemicals were obtained from SIGMA (St. Louis, MO, USA).

IDENTIFICATION OF THE DYNAMIC MODEL OF THE SENSOR

In order to determine the dynamic response of the sensor, we consider three approximate dynamic models.

Zero-pole model

This model considers that the dynamics between biomass $X(t)$ and fluorescence $R(t)$ can be approximately described by an integral

time constant τ_p , a derivative time constant τ_z , and a gain K :

$$\tau_p \frac{dR(t)}{dt} = -R(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right)$$

The transfer function between the input and the output expressed through the Laplace transform notation is:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{1 + \tau_p s}$$

1 zero-2 poles model

In this model, we consider an additional integral time constant τ_{p2} between the input $X(t)$ and the output $R(t)$, represented in the following equation through an auxiliary state variable $X_1(t)$:

$$\begin{aligned} \tau_{p1} \frac{dX_1(t)}{dt} &= -X_1(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right) \\ \tau_{p2} \frac{dR(t)}{dt} &= -R(t) + X_1(t) \end{aligned}$$

In Laplace transform notation:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{(1 + \tau_{p1}s)(1 + \tau_{p2}s)}$$

1 zero-3 poles model

In this model, we consider an additional integral time constant τ_{p3} between the input $X(t)$ and the output $R(t)$, represented in the following equation through an additional state variable $X_2(t)$:

$$\begin{aligned} \tau_{p1} \frac{dX_1(t)}{dt} &= -X_1(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right) \\ \tau_{p2} \frac{dX_2(t)}{dt} &= -X_2(t) + X_1(t) \\ \tau_{p3} \frac{dR(t)}{dt} &= -R(t) + X_2(t) \end{aligned}$$

In Laplace transform notation:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{(1 + \tau_{p1}s)(1 + \tau_{p2}s)(1 + \tau_{p3}s)}$$

In order to identify the model from the samples, we used an approximate discrete model based on the bilinear transform (Oppenheim and Schaffer, 2010):

$$s \leftarrow \frac{2}{T_s} \frac{z - 1}{z + 1}$$

where z^{-1} corresponds to a pure delay in the sample and T_s is the sampling frequency. Parameters of the discrete model were fitted by linear regression using the R package with the dynlm package (Zeileis, 2014), which was also employed to perform the simulations.

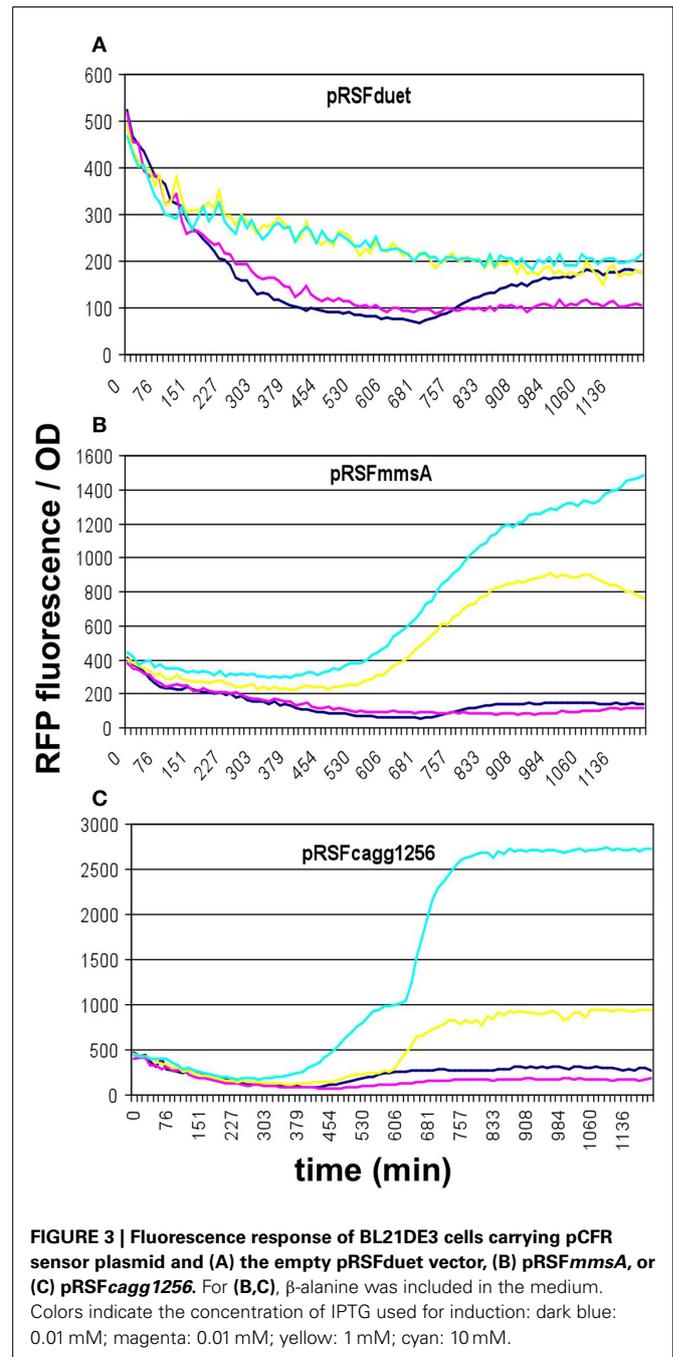
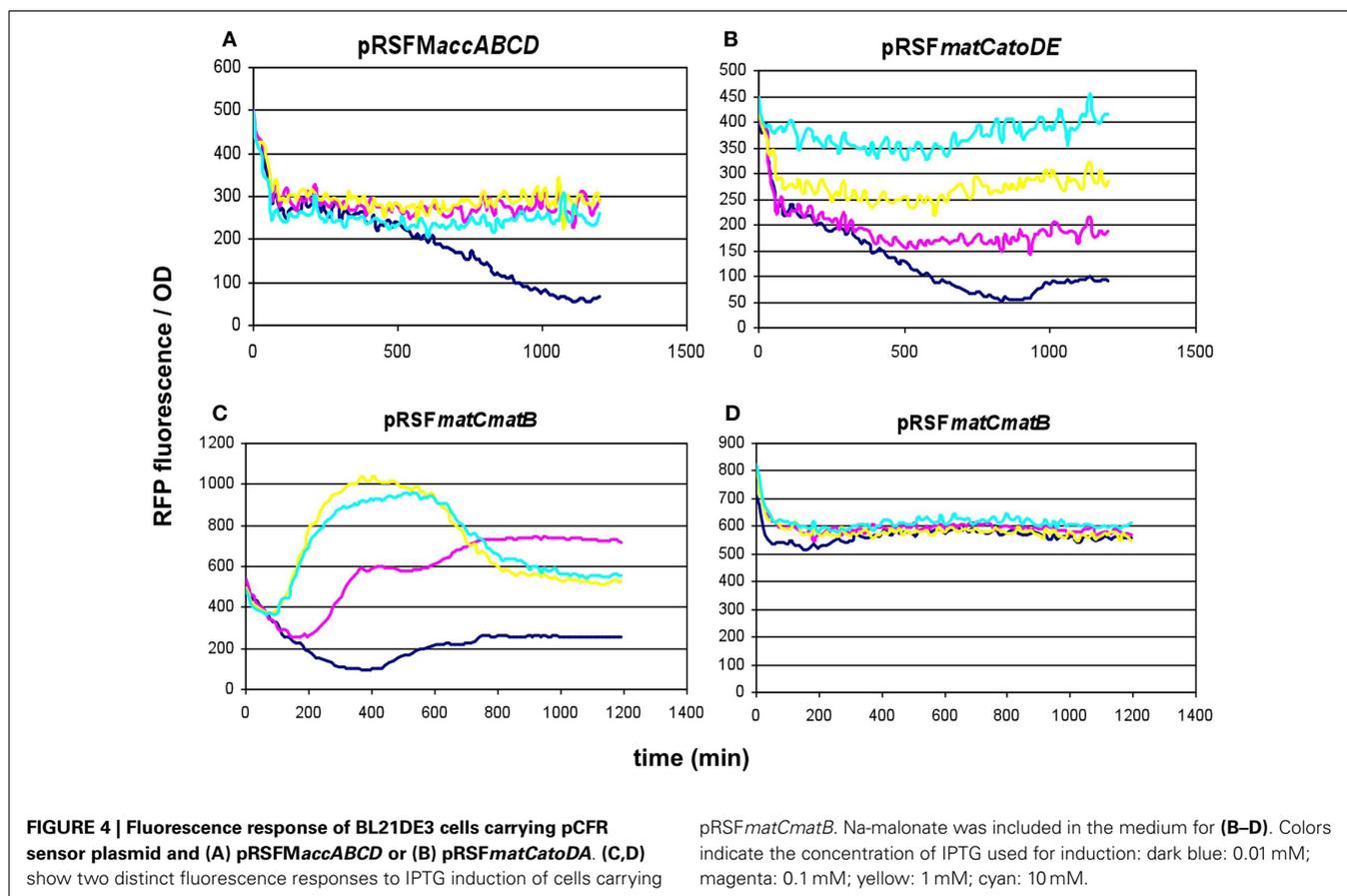


FIGURE 3 | Fluorescence response of BL21DE3 cells carrying pCFR sensor plasmid and (A) the empty pRSFduet vector, (B) pRSFmmsA, or (C) pRSFcagg1256. For (B,C), β -alanine was included in the medium. Colors indicate the concentration of IPTG used for induction: dark blue: 0.01 mM; magenta: 0.01 mM; yellow: 1 mM; cyan: 10 mM.

RESULTS

EXPERIMENTAL ASSESSMENT OF MALONYL-CoA SENSING

Malonyl-CoA is an important building block from the aspect of metabolic engineering, for it is required for the biosynthesis of fatty acids, polyketides, flavonoids, and other compounds (Fowler et al., 2009; Xu et al., 2011). In an earlier work, we have predicted and tested several alternative pathways that yield malonyl-CoA in order to boost pinocembrin production in *E. coli* (Fehér et al., 2014). Briefly, one candidate for malonyl-CoA synthase (*matB*, EC 6.2.1), one for malonate-CoA transferase (*atoDA*, EC 2.8.3.3),



two for malonyl-CoA reductase (*mmsA* and *cagg1256*, EC 1.2.1.75) as well as one for the acetyl-CoA carboxylase complex (*accABCD* from *E. coli*, EC 6.4.1.2) were expressed together with the complete pinocembrin pathway, and their efficiencies were deduced from the resulting normalized pinocembrin titers. Here, we used the molecular sensor as a more direct method to investigate the levels of malonyl-CoA produced by the alternative enzymes.

Verifying the function of the malonyl-CoA sensor

The core of the malonyl-CoA sensor is an RFP gene driven by the synthetic pFR1 promoter (Liu et al., 2015). pFR1 is a combination of P_{A1} promoter from phage T7 (Deuschle et al., 1986) and two FapR-binding segments flanking the –10 region. The FapR protein, originating from *Bacillus subtilis*, dissociates from the DNA upon the binding of malonyl-CoA (Schujman et al., 2003), thereby allowing the *E. coli* RNA polymerase to transcribe the downstream sequences. The transcription of FapR is driven by a pAra promoter (controlled by the AraC protein), pFR1 is therefore increasingly repressible by elevating the L-arabinose levels. The AraC and FapR transcription factors, as well as RFP are all encoded on pBFR1k_RFP_8FapR, a kanamycin-resistant plasmid constructed by the workgroup of Fuzhong Zhang (Liu et al., 2015). The authors validated the quantitative nature of malonyl-CoA sensing by generating a calibration curve that correlated the fluorescence levels to LC-MS measurements of intracellular malonyl-CoA. In order to use this sensor to compare the malonyl-CoA productivity of our

construct collection available on the kanamycin-resistant pRSF vector, we generated a chloramphenicol-resistant version of the plasmid, which we call pCFR (plasmids used in this study are listed in Table S1 in Supplementary Material). The functionality of pCFR, as well as that of the original sensor plasmid was benchmarked with the help of cerulenin. This is an inhibitor of 3-ketoacyl-ACP synthase I and II (D'Agnolo et al., 1973), which participate in the committing step of free fatty acid biosynthesis. The inhibition of these enzymes results in a substantial intracellular accumulation of malonyl-CoA (Davis et al., 2000), and should therefore cause an increase in RFP fluorescence. Indeed, *E. coli* cells carrying either pBFR1k_RFP_8FapR or pCFR responded to cerulenin with a sharp increase in OD-normalized fluorescence (Figures 1A,B). As a matter of fact, a small scale screening of four pCFR-carrying clones allowed the identification of the most effectively functioning plasmid, which was chosen for all downstream experiments (Figure 1B). As expected, the ratio of RFP/OD values of cerulenin-treated to control cultures was higher in minimal medium than in LB broth (Figure 1C), due to the relatively high background-fluorescence of LB medium (Xu et al., 1999). For this reason, all further measurements were carried out in minimal medium.

Optimizing sensitivity toward malonyl-CoA detection

In its original publication by Liu et al. (2015), the malonyl-CoA sensor was characterized in a split form: the FapR and RFP

genes were carried by two separate plasmids. Therefore, we re-characterized pBFR1k_RFP_8FapR in order to find the optimal conditions of comparing various malonyl-CoA producer enzymes using the single-plasmid sensor construct. The expression level of the FapR repressor was varied by administering a dilution series of arabinose, while the expression level of the malonate transporter (*matC*) and the *matB* from plasmid pACYC*matCmatB* was controlled by altering the IPTG concentration. A 4 × 4 concentration matrix was thus applied to cells carrying either both plasmids or just the sensor plasmid. As apparent from the OD-normalized fluorescence values shown on **Figure 2**, an arabinose concentration of 0.01% proved to be optimal, taking both specificity and sensitivity into account, similarly to the originally published sensor system. At lower arabinose concentrations, specificity deteriorated, since the sensor plasmid produced slight increases in fluorescence in response to IPTG even in the absence of the malonyl-CoA producer plasmid. This could indicate the promoter's non-specific binding to LacI, which is competitively inhibited by sufficient quantities of the FapR protein. Increasing arabinose levels to 0.1% strongly reduced fluorescence levels of the sample, leading to a decrease of sensitivity and of the signal to noise ratio, possibly resulting from an over-repression by FapR. The initial decrease in fluorescence/OD values, seen on all graphs is likely due to the delay of RFP expression compared to culture growth.

Comparison of malonyl-CoA-producing construct collection

To compare the malonyl-CoA production efficiencies of alternative malonyl-CoA-producing pathways, at least one representative enzyme for each pathway was cloned, as described earlier (Fehér et al., 2014). Briefly, the *matB*, *atoDA*, *mmsA*, and *cagg1256* genes, as well as the *accABCD* gene complex were expressed from the high-copy plasmid pRSFduet. To provide sufficient amounts of substrate, a MatC was co-expressed with *matB* and *atoDA*, and the medium was supplemented with Na-malonate. To elevate the substrate levels for *mmsA* and *cagg1256*, cultures were supplemented with β-alanine, which is converted to 3-oxopropanoate (malonate semialdehyde) by the cell's endogenous 4-aminobutyrate aminotransferase. In every case, the fluorescence/OD values were monitored over time after growing the cells in various IPTG concentrations. Cells carrying the empty pRSFduet plasmid (besides the sensor plasmid) were used as a negative control. Based on their response, the strains carrying the various constructs fell into either one of the two following categories: (i) the cells exhibited growth and a dose-dependent fluorescence/OD response to IPTG (pRSF*mmsA* and pRSF*cagg*) (**Figure 3**) or (ii) the cells exhibited absolutely no growth upon IPTG induction in minimal medium (pRSF*accABCD* and pRSF*atoDA*) (**Figures 4A,B**). It is important to note that certain cells of the latter category also gave a seemingly dose-dependent response to IPTG (such as pRSF*atoDA*), but their fluorescence and OD values were practically unchanged during the course of the experiment, most probably indicating that the trend in their ratio is an artifact (Figure S1 in Supplementary Material). Interestingly, both types of responses were seen with pRSF*matCmatB*-carrying cells, in a quite irreproducible manner (see Discussion and **Figures 4C,D** and Figure S2 in Supplementary Material). The growth curves obtained by inducing strains

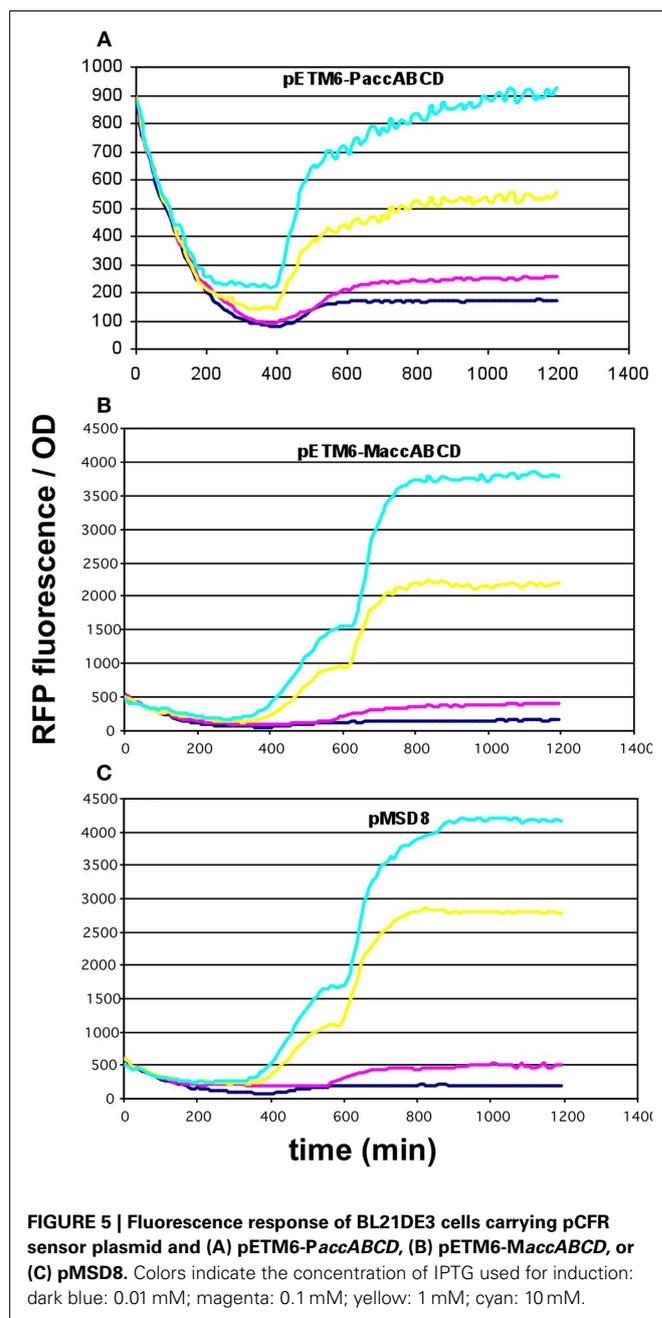
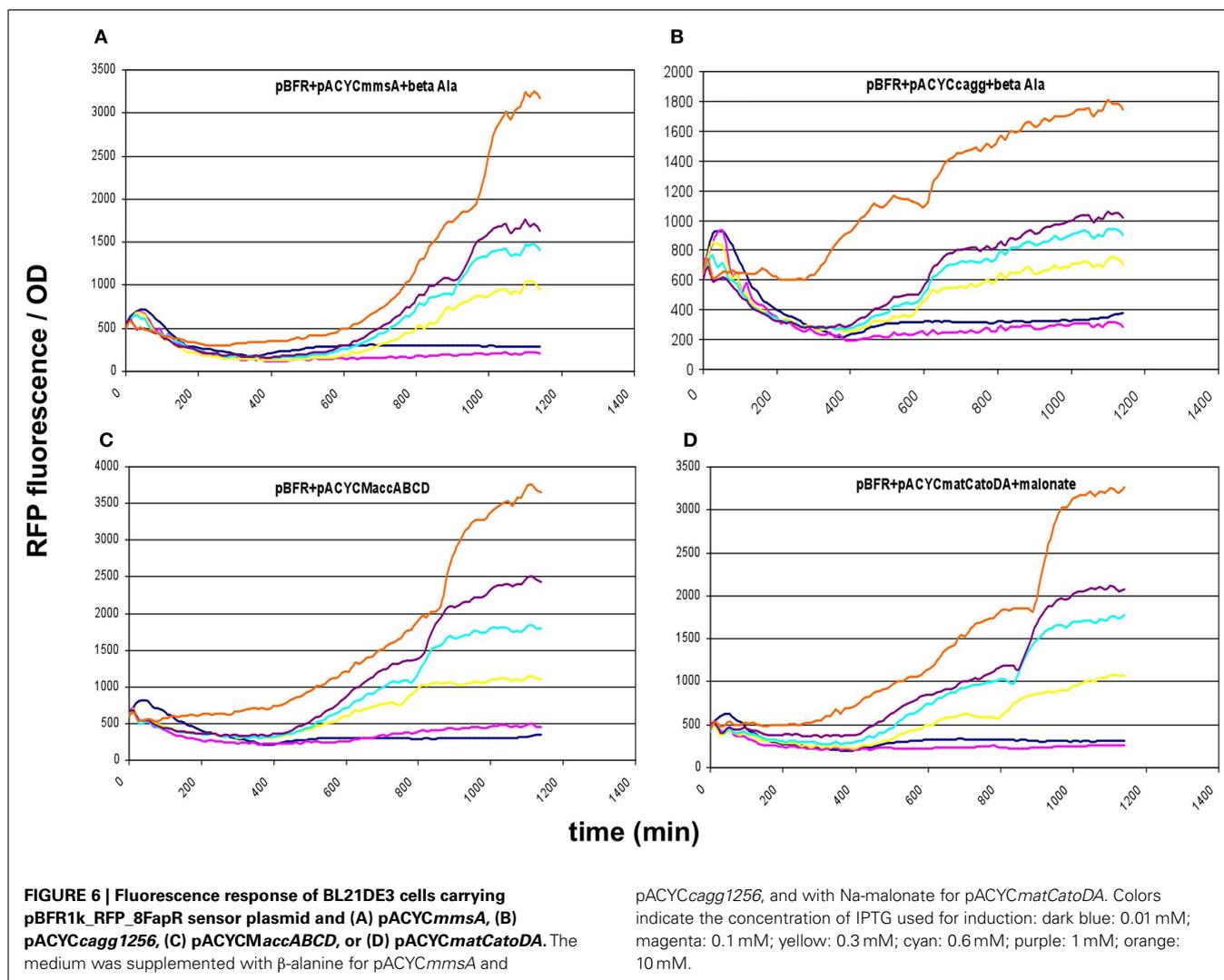


FIGURE 5 | Fluorescence response of BL21DE3 cells carrying pCFR sensor plasmid and (A) pETM6-P*accABCD*, (B) pETM6-M*accABCD*, or (C) pMSD8. Colors indicate the concentration of IPTG used for induction: dark blue: 0.01 mM; magenta: 0.1 mM; yellow: 1 mM; cyan: 10 mM.

carrying various producer plasmids are summarized on Figure S3 in Supplementary Material.

Investigating lower-copy-number alternative constructs

Since the overexpression of certain members of the *accABCD* complex (Davis et al., 2000), as well as protein overproduction in general (Flores et al., 2004) have been described to have a toxic effect on the cells, the overexpression of certain constructs was repeated using lower-copy vectors. For the *accABCD* complex, pMSD8, an ultra-low-copy variant was available expressing the four genes as a T7-driven operon. In addition, two medium-copy alternatives, pETM6-P*accABCD* and



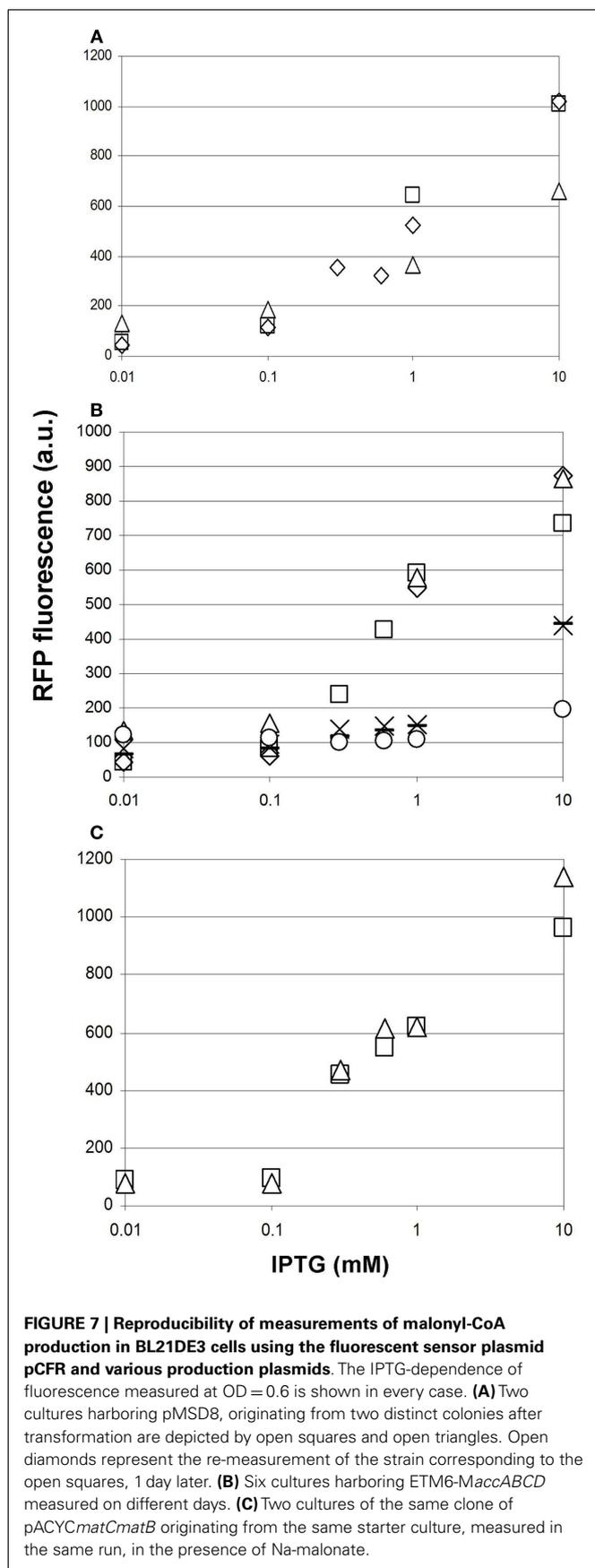
pETM6-*MaccABCD*, carrying the genes as a pseudo-operon (separate promoters) and in monocistronic form (separate promoters and terminators), respectively, were also obtained and tested (Xu et al., 2012, 2013). All three plasmids displayed a clear dose-dependent fluorescence/OD pattern, well above the levels of the negative control (Figures 5A–C). For *matCmatB*, pACYC*matCmatB* was readily available as a low-copy alternative that functioned in a quite well-reproducible fashion (Figure 2), when tested with the pBFR1k_RFP_8FapR sensor plasmid.

These results prompted us to subclone our pathway collection into a pACYC backbone, thereby reducing the copy-number of all constructs to ~ 15 . When expressed this way, all five constructs (*matCmatB*, *mmsA*, *cagg1256*, *M.accABCD*, and *matCatoDA*) allowed cell growth and displayed a scalable, IPTG-dependent fluorescence (Figures 2 and 6). As it turned out, the concise nature of the dataset obtained with the pACYC-collection allowed us to fit various models describing the response of the sensor to malonyl-CoA production (see below).

Quantification and reproducibility of malonyl-CoA levels

Since *E. coli* strain BL21DE3, the host used in our experiments is not deleted for the genes of arabinose catabolism, it breaks down arabinose when the medium is depleted for glucose. This likely causes a decrease in FapR expression, and a consequent derepression of RFP at the end of growth. As a result, a second, sudden increase in fluorescence was seen in many samples upon the transition to stationary phase, as apparent from juxtaposing the temporal evolution of fluorescence and OD (Figure S4 in Supplementary Material). To avoid this effect, we analyzed fluorescence levels corresponding to the log-phase ($OD = 0.6$), a physiological state when arabinose catabolism is still inhibited by the catabolite repression of *E. coli*. Figure S5 in Supplementary Material shows some of the fluorescence vs. OD plots generated to obtain such readings.

Prior to comparing the performance of various malonyl-CoA-producing constructs, we investigated the reproducibility of our measurements at three different levels. First, we compared the fluorescence values of parallel clones, resulting from



the transformation of a given producer construct. These displayed a strong variation in their response to IPTG, which was not unexpected taking into account their heterologous colony morphologies (Figure 7A). Second, we re-tested the best performing clones after their storage in the form of glycerol stocks. Reproducibility was substantially better, being the best for the lowest copy-number pMSD8 plasmid (Figure 7A). For other constructs, the performance seemed to fall into two distinct clusters: it either matched that of earlier measurements, or deteriorated to a lower, near-zero value, as demonstrated by the dose–response curves of pETM6-MaccABCD (Figure 7B). This phenomenon was reminiscent of a burden-relieving mutation that swept through the population early after induction. The third level of reproducibility corresponds to comparing the performance of a given strain in the same microtiter plate in the same measurement-run, starting from the same overnight seed-culture (Figure 7C). The high reproducibility of such measurements was sufficient to draw conclusions on the effect of providing the substrates for malonyl-CoA production (see below).

Investigating the effect of administering the substrate

To test whether providing the substrates of malonyl-CoA synthesis in the growth medium offers a measurable advantage in the product levels, we repeated the experiment for, pACYCmatCmatB and pACYCmatCatoDA with and without Na-malonate as well as for pRSFmmsA and pRSFcagg with and without β -alanine. For pACYCmatCmatB, we observed significantly boosted production at 0.3 and 0.6 mM IPTG (Figure 8A). However, the presence of Na-malonate significantly hindered the performance of pACYCmatCatoDA at IPTG concentrations of 0.3 mM and above (Figure 8B). Similarly, supplementing β -alanine to pRSFmmsA seemed to have a negative effect at 0.6 mM IPTG induction (Figure 8C). No significant effect was seen for adding the substrate to pRSFcagg (not shown).

Investigating the efficiencies of malonyl-CoA production

The four alternative pathways for malonyl-CoA production, as well as the top-scoring enzymes that catalyze the required steps were sought for by RetroPath, our retrosynthetic biology tool developed for metabolic engineering (Carbonell et al., 2011). As opposed to our initial plans, the growth-inhibitory effect of certain pRSF-based constructs prohibited us to validate the ranking of the alternative pathways by direct comparison of their measured fluorescence values. Indeed, the toxicity of overproducing malonyl-CoA by overexpression of acetyl-CoA carboxylase has been described before (Davis et al., 2000). The performance of the readily available low-copy variants of accABCD, however (pMSD8, pETM6-PaccABCD, and pETM6-MaccABCD), indicated the importance of using lower-copy expression vectors. After recloning our gene-set into pACYC-derived plasmids, matCatoDA, MaccABCD, mmsA, cagg1256, and matCmatB all permitted growth when expressed. Importantly, based on their fluorescence values, these constructs all produced significantly more malonyl-CoA when induced with 1 mM of IPTG, than the control cell harboring no production (see below). Ranking the alternative malonyl-CoA-producing pathways was not feasible due to the high variation of the data. Important practical conclusions could be nevertheless made, listed in the Section “Discussion.”

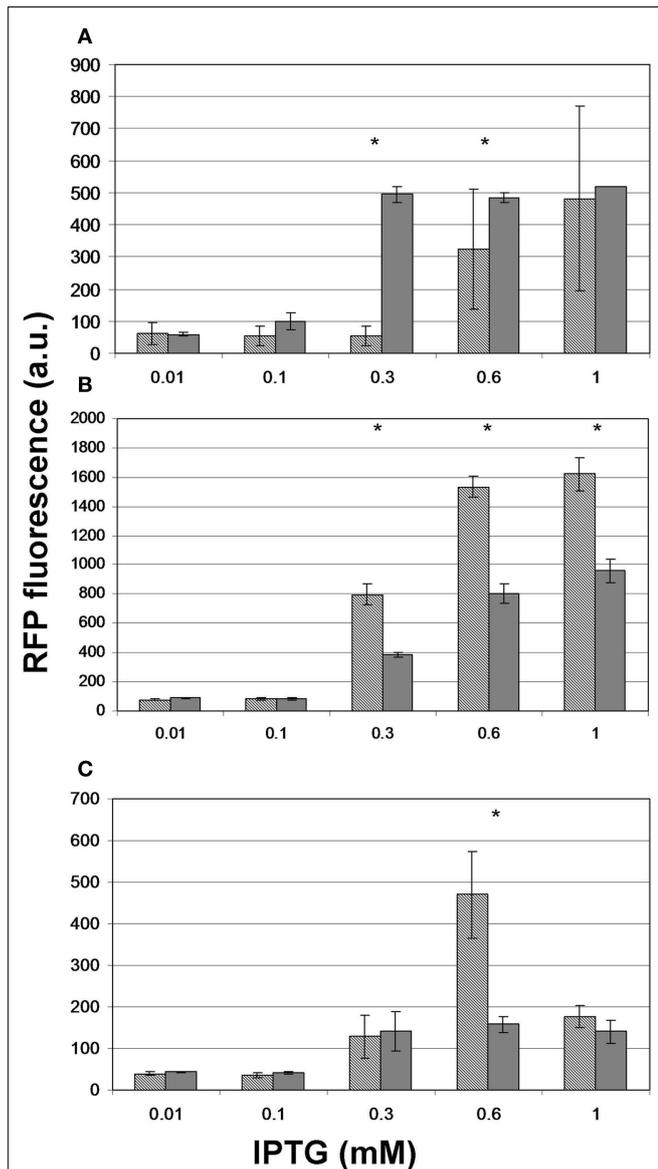


FIGURE 8 | The effect of administering the substrate on the level of fluorescence obtained upon induction of malonyl-CoA production. Hashed bars represent the effect of IPTG on the fluorescence obtained at OD = 0.6 with constructs **(A)** pACYCmatCmatB, **(B)** pACYCmatCatoDA, and **(C)** pRSFmmsA. Gray bars depict the peak fluorescence detected when inducing the same constructs in the presence of their substrate: Na-malonate for *matCmatB* and *matCatoDA* and β -alanine for *mmsA*. The sensor plasmid was pBFR1k_RFP_8FapR for **(A,B)** and pCFR for **(C)**. Asterisks indicate a significantly different fluorescence caused by the substrate ($p < 0.05$).

MODELING THE MALONYL-CoA SENSOR

Characterization of the sensor response

In order to characterize the sensor, we modeled the sensor response to the production of malonyl-CoA, its measured signal. Ideally, the sensor once calibrated should display a good reproducibility as well as a stable sensitivity for its use in applications such as screening or feedback regulation. For that purpose, it is desirable to keep the

dynamic response of the sensor uncoupled from any parameter variation so that the relationship with the measured signal and its output corresponds basically to a linear one. The time of response of such linear sensor, on the other hand, should be kept below some reasonable limit in order to avoid large lags introducing potential instability into the system. Here, as we are basing our model on the observations of OD and RFP, we need to consider in our model the relation between these variables with respect to the malonyl-CoA concentration, the signal that is measured by the sensor. In this study, the model necessarily will be kept simple, as our purpose is to provide a set of parameters that can be easily measured in order to calibrate the sensor.

Similarly to other proposed models (Anesiadis et al., 2008), we consider that the signal undergoes several transformation stages in cascade-mode going from the biomass to the fluorescence levels:

1. We assume that the variation on concentration of malonyl-CoA $M(t)$ depends on its instant production, given by the rate v_m , which is specific to each selected malonyl-CoA-producing construct or pathway, multiplied by the biomass $X(t)$ minus the malonyl-CoA that is consumed for growth defined as a constant rate v_g that multiplies the variation in time of biomass and another term for degradation, defined by a rate constant γ_m :

$$\frac{dM(t)}{dt} = v_m X(t) - v_g \frac{dX(t)}{dt} - \gamma_m M(t) \quad (1)$$

During the exponential growth phase, this previous equation is simplified as:

$$\frac{dM(t)}{dt} = (v_m - \mu v_g) X(t) - \gamma_m M(t) \quad (2)$$

where μ is the specific growth rate.

2. Malonyl-CoA binds to transcription factor FapR to form a complex whose concentration $C(t)$ depends on a dissociation constant K_d [2.4 μ M according to Schujman et al. (2006)]. The concentration of FapR $F(t)$ is considered constant depending on the concentration levels of the inducer L-arabinose:

$$F + M \xrightleftharpoons[k_r]{k_f} C$$

$$\frac{dC(t)}{dt} = k_f FM(t) - k_r C(t) \quad (3)$$

3. The change on concentration of RFP $R(t)$, depends on the concentration of the complex $C(t)$ through a constant κ associated to the promoter strength (Oyarzún and Stan, 2013) and a decay constant γ_r :

$$\frac{dR(t)}{dt} = \kappa C(t) - \gamma_r R(t) \quad (4)$$

Under the assumptions given by the model described by Eqs 1–4, the response of the sensor can be approximated as a cascade of filters as shown in **Figure 9**. The advantage of using such approximation is that we can assume that the dynamics

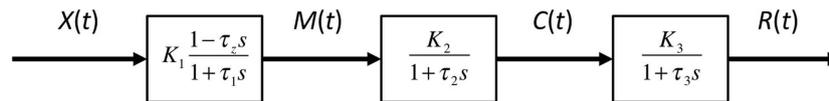


FIGURE 9 | An approximate model of the response behavior of the sensor based on a cascade of first-order filters. The biomass $X(t)$ is assumed to be approximately related to the concentration of malonyl-CoA $M(t)$ through a first-order pole-zero filter. The sensor complex malonyl-CoA-FapR concentration $C(t)$ is then related to $M(t)$ through a first-order filter and finally, the RFP concentration $R(t)$ is again related to $C(t)$ through another first-order filter.

due to the time constants of these three filters occur at different time scales and therefore, under the appropriate conditions, time-scale separations can be performed. Based on this principle, we can approximate the model by considering main time constants present in the model, i.e., by estimating parameters of a cascade of first-order models (see Materials and Methods).

Sensor model fitting

In our sensor model, we are assuming that under same inducer and regulatory conditions on a strain population, malonyl-CoA levels should change basically because of the change in the rate of production of this metabolite by the strain (v_m in Eq. 1), which would correspond to a different gain of the sensor, i.e., a different measured RFP and biomass X in steady state.

Such gain according to the sensor model in Eqs 1–4, the gain or relationship in steady state between RFP and X is given by the following relationship between model parameters:

$$\frac{R}{X} = K_d \frac{v_m}{\gamma_m \gamma_r} \quad (5)$$

Regarding model dynamics, we considered three simplified models for the response dynamics depending on the time constants that are considered (see Materials and Methods): (1) a first-order with integral and derivative terms; (2) a second-order model consisting of a first-order model with integral and derivative terms and an additional first-order integral term; and (3) a third-order model with the integral and derivative terms and two additional first-order integral terms.

As shown in **Figure 10**, measured responses were in general fitted with an increasing level of accuracy to each of the three approximate sensor models. For some constructs, however, measurements could not be successfully fitted to model 3 (see Figure S4 in Supplementary Material). Therefore, we decided to use model 2 (the one considering one derivative and two integral terms) as the reference model in order to characterize the dynamics of the sensor. **Table 1** provides a list of estimated parameters for different constructs. Time constants ranged between approximately 1 and 10 h.

Dependence of sensor parameters on induction

According to our simplified model of the sensor consisting on a second-order transfer function, gain between biomass and RFP should depend on the level of induction of the malonyl-CoA-producing enzyme. Time constants, on the contrary, should not depend on the level of induction, but should be constitutive parameters of the sensor that can be calibrated for each construct. As shown in **Figure 11**, we found effectively a good

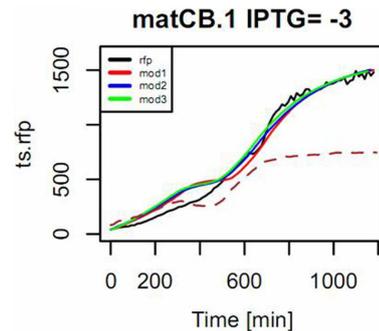


FIGURE 10 | Example of the mode of fitting models to describe the response of the pBFR1k_RFP_8FapR sensor in the presence of pACYCmatCmatB, induced with 1 mM of IPTG. Na-malonate was included in the medium. The input signal is the measured OD (shown in dotted line and scaled to 50% of maximum in the plot) and the simulated response of the model for the observed output RFP (in black) is shown for a fitting to a first-order model with input derivative (red), a second-order with input derivative (blue), and a third-order with input derivative (green).

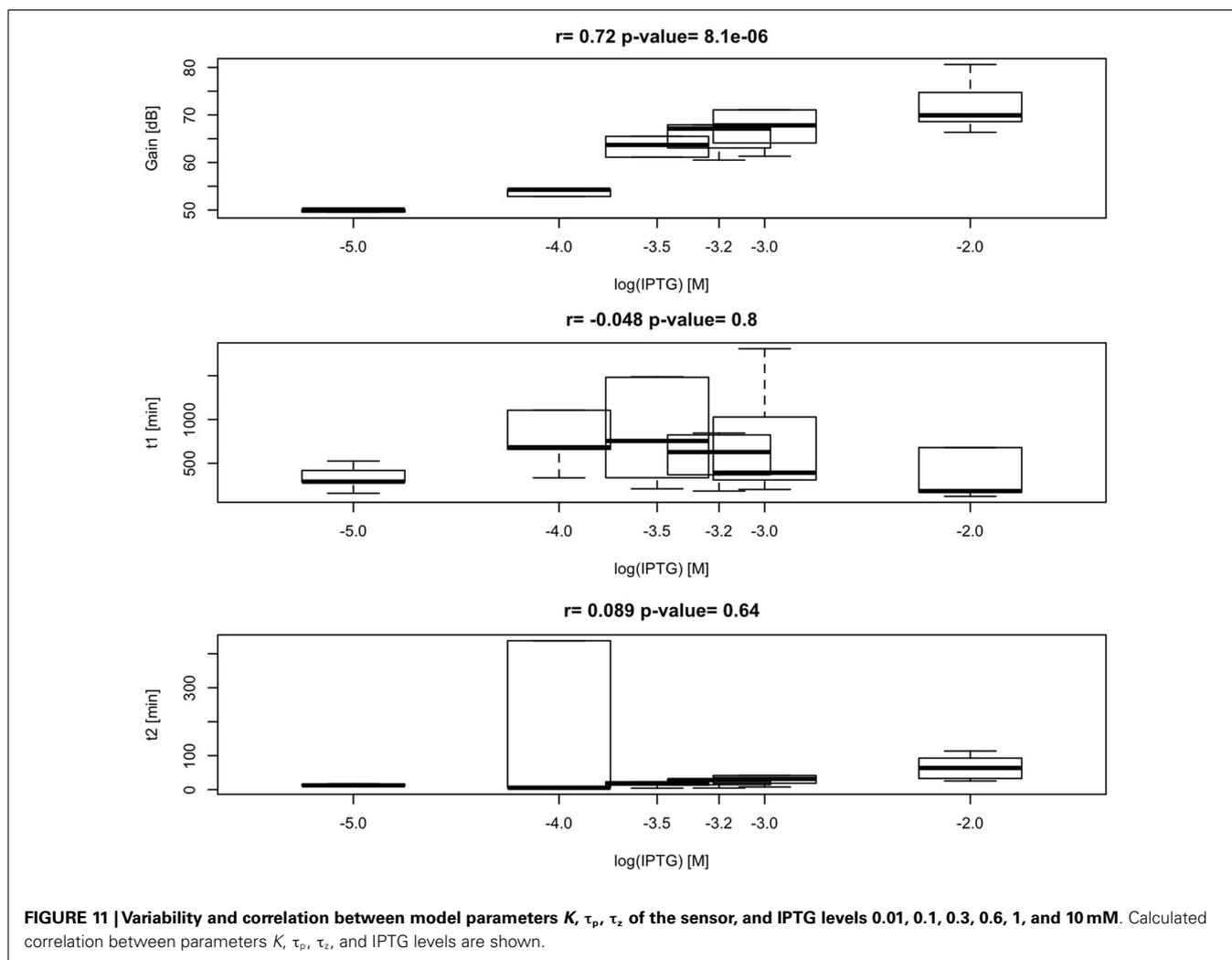
Table 1 | Gain and integral time constants of the sensor fitted to model 2 estimated for different constructs for IPTG = 1×10^{-5} .

Construct	K	τ_1	τ_2
matCmatB	250.84	290.11	12.69
mmsA	300.53	288.31	15.67
cagg	362.82	525.90	17.39
acc	320.20	417.92	11.49
matCatoDA	321.12	156.95	11.14

correlation between IPTG levels and sensor gain ($r = 0.72$, p -value = 8.1×10^{-6}), indicating that the sensor is responsive to changes in induction of the enzyme. Time constants, as expected, were not significantly affected by changes in the IPTG concentration. These time constants, on the other hand, appear as distinctively clustered for each construct, showing the robustness of the sensor (**Figure 12**).

DISCUSSION

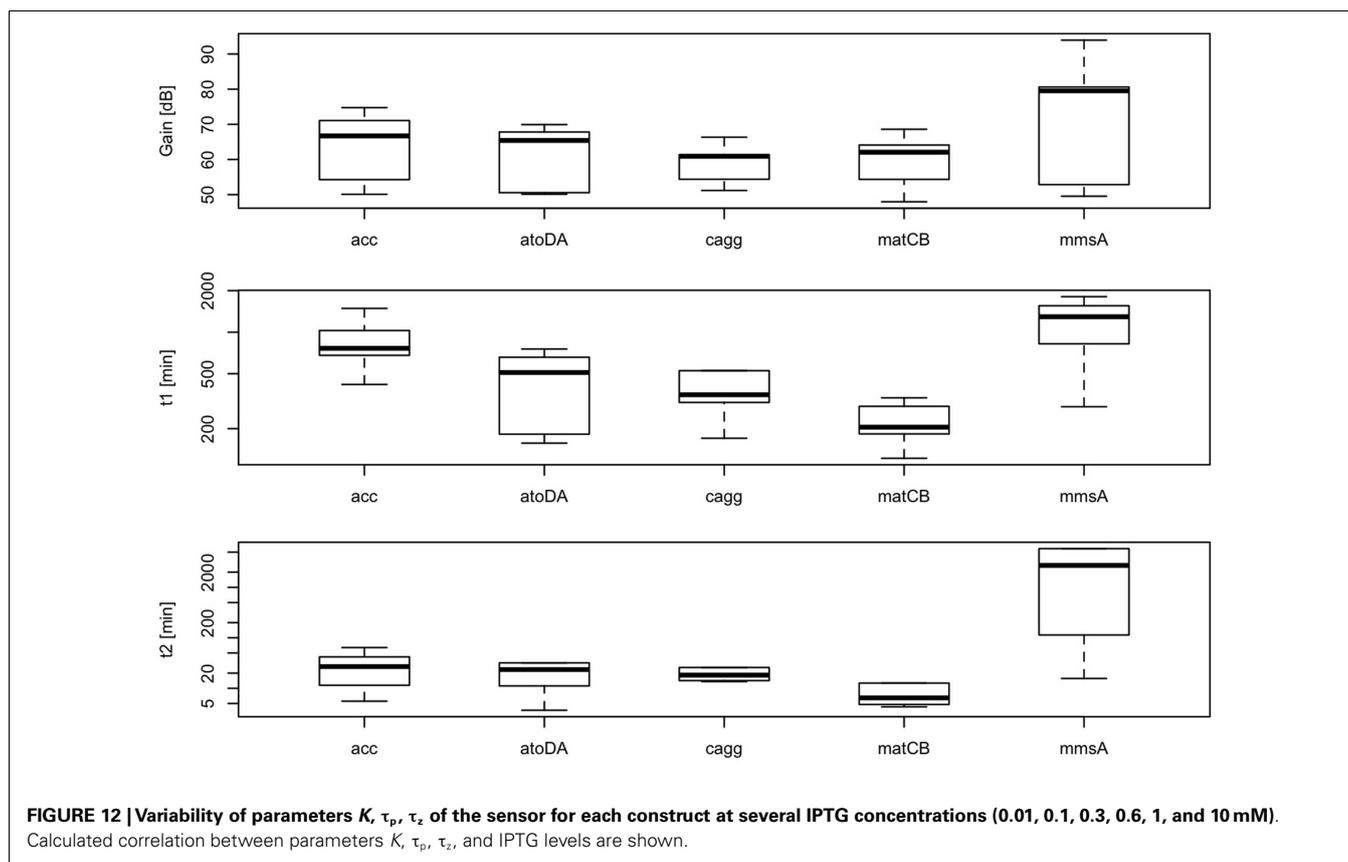
In this work, we have successfully implemented a malonyl-CoA sensor for the partial optimization of malonyl-CoA production in *E. coli* BL21DE3. Initially, we screened several conditions to be used for the measurements, and found the optimal one to match that described earlier (Liu et al., 2015), despite the possible changes in relative gene dosage caused by the single-plasmid vs.



dual-plasmid layout. We constructed a chloramphenicol-resistant version of the plasmid as well, to compare our construct collection encoding four alternative pathways for malonyl-CoA production, predicted by RetroPath. Besides their comparison, our aim was to use malonyl-CoA sensing to seek optimal conditions for efficient malonyl-CoA production, previously seen to be hindered by growth retardation (Fehér et al., 2014). Our current results also indicated, that the viability of the cells can be, in many cases strongly impaired by the high expression levels of the original pRSF-based constructs (Figure S3 in Supplementary Material). It was also apparent, that the lower-copy alternatives available for the *accABCD* gene complex (pMSD8, pETM6-*MaccABCD*, and pETM6-*PaccABCD*) outperformed all or most members of the RSF-collection. Therefore, a partial optimization of expression levels was carried out by using smaller copy-number vectors to obtain viable cells with measurable malonyl-CoA production upon induction, leading to a successful circumvention of the possible toxicity caused by the high-copy expression vectors. Due to the relatively high variation of the resulting data, comparison of these constructs was not conclusive enough to validate

all our predicted rankings. However, the obtained measurements were quite useful for the practical purpose of selecting the most efficient implementation of our malonyl-CoA-producing constructs. From this aspect, the best performer of our construct set (in terms of fluorescence at OD=0.6) turned out to be pACYC $_{matCatoDA}$ carrying the malonate-CoA transferase on a P15a-derived low-copy-number plasmid (Figure 13). Its performance could be nonetheless matched by the expression of the acetyl-CoA carboxylase complex, which, in turn was indiscernible from the *matB*. The two genes encoding malonyl-CoA reductase (*mmsA* and *cagg1256*) performed significantly weaker. Importantly, the activity of these constructs uncovered two pathways, malonyl-CoA reductase and malonate-CoA transferase, which have not been implemented before in boosting malonyl-CoA biosynthesis.

When comparing the variation of measured fluorescence values, we observed a consistent decrease in the coefficient of variation for all constructs when recloned into the low-copy pACYC vector, possibly marking an increased genetic stability (Table S2 in Supplementary Material). In theory, this change could



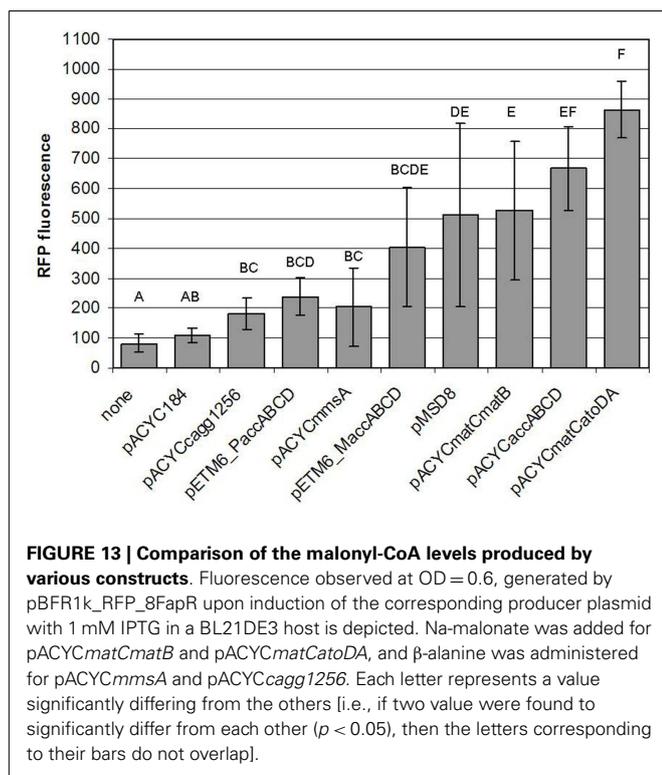
also be caused by the fact of using different sensor plasmids in the comparison. However, no such difference was seen in the case of the pET-based plasmids when switching from pCFR to pBFR1k_RFP_8FapR used for sensing (Table S2 in Supplementary Material), indicating that it is the change in copy-number of the malonyl-CoA producer plasmid, and not the change in the sensor plasmid that caused the decrease of variation. This variation can be, at least in part, explained by burden-relieving mutations that sweep through the population early after induction. In fact, such an expansion of a non-fluorescent subpopulation was seen real time during one of our tests concerning pRSF $matCmatB$ (Figure 4C), which was the least reproducible among all of our constructs. In that specific experiment, the fluorescence of the population ceased to grow after a certain point of time, despite the accelerating increase of the OD, leading to a decrease of the fluorescence/OD ratio (Figure S2 in Supplementary Material). The fact that no such effect is seen in Figure 4D indicates the random and incidental nature of this phenomenon, supporting the assumption that it is caused by a mutation. The specific mutations inactivating the production were not sought for.

In their original publication, the designers of this sensor detected a saturation of malonyl-CoA levels at 25 μ M, when inducing the expression of the *acc* operon at different strengths (Liu et al., 2015). This phenomenon was not investigated further, but could have been caused by the toxicity of the produced compound. Although we did not observe an unambiguous plateau of fluorescence values in our experiments, the possibility of such a scenario

should be taken into account when using similar systems in the future. One possible solution for such cases could be to transform the molecule of interest into a compound *in vivo* that is better tolerated by the cell. The prerequisites of this strategy are to have a high-capacity, unsaturated pathway for transformation as well as a sensory device for the downstream product.

To characterize further the dynamic response of the sensor, we proposed and validated a second-order model linking biomass to fluorescence. We showed that dynamic parameters were specific to each construct and that they were not significantly altered by changes in the production rate of malonyl-CoA. This result shows the robustness of the sensor for its use in pathway regulation, since a dynamic response uncoupled from the measured signal is often necessary in order to assure the stability of the feedback loop. Moreover, in such regulation strategies, knowledge of the sensor's parameters is of critical importance in the choice of one particular control architecture over another (Stevens and Carothers, 2014). Therefore, we hope our effort of modeling experimental data provided by this malonyl-CoA sensor will facilitate future developments in dynamic regulation of pathways such as the ones involved in fatty acid or flavonoid production.

Our most unexpected findings arose when we investigated the effect of providing the substrates of malonyl-CoA production on the fluorescence levels produced by the sensor. On one hand, providing Na-malonate to *matB*, resulted in an increased fluorescence, possibly indicating more product formation. On the other hand however, administering β -alanine or Na-malonate to *mmsA* and



atoDA, respectively resulted in an opposite effect. This phenomenon was especially pronounced for *atoDA*, which was notably the most effective malonyl-CoA producer among the constructs tested in this study. We can only speculate on the mechanism, which may even be a direct effect on the sensor plasmid, and may not represent an actually reduced malonyl-CoA synthesis. According to our hypothesis, it could be due to the fact that *atoDA* (as well as *matB*) are consumers of cellular acetyl-CoA, thereby causing an energy depletion. This is probably more severe if we provide the missing substrate (malonate), with very high enzyme concentrations, and thereby push the reaction toward the product. Perhaps, this energy depletion inhibits RFP production more than the newly produced malonyl-CoA would elevate it. Clearing this issue nevertheless requires further experiments.

When summarizing the observed results, several conclusions important to the metabolic engineer in general can be drawn. First, it is useful to test the performance of several parallel clones after transformation of the constructs, and choose the best for further experiments. Second, the reproducibility of cultures re-grown from glycerol stocks is acceptable, but repeated measurements are advisable to uncover the repeated emergence of deleterious mutants, possibly indicating genetic instability. Third, this instability was minimal when using the lowest copy-number vector for expressing the producer construct. This provides a further argument to start testing various alternative enzymes by expressing them from low-copy vectors with gradually controllable promoters, and varying ribosome binding sites to increase expression and find a construct that is optimal both in efficiency and stability. And finally, the variation among measurements obtained in the same run were small enough to test the effect of adding the substrates

of the tested pathways to the cell culture, and obtain the optimal IPTG levels for most effective utilization. This approach could be useful in the future to optimize the substrate concentration itself, as well as any other component of the culture medium or the induction process in a high throughput, combinatorial fashion.

ACKNOWLEDGMENTS

We thank Prof. John Cronan, Prof. Mattheos Koffas, Prof. Fuzhong Zhang, and Dr. Jingwen Zhou for providing their plasmids and Dr. Brian Jester for helpful discussions. PC is supported by Genopole through an ATIGE Grant by PRES UniverSud Paris, by Agence Nationale de la Recherche, and by UPFellows program with the support of the Marie Curie COFUND program. VL is supported by a DGA (French Ministry of Defense) graduate fellowship. The authors acknowledge funding provided by Toulouse White Biotech (TWB) and the GIP GENOPOLE.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00046>

REFERENCES

- Anesiadis, N., Cluett, W. R., and Mahadevan, R. (2008). Dynamic metabolic engineering for increasing bioprocess productivity. *Metab. Eng.* 10, 255–266. doi:10.1016/j.ymben.2008.06.004
- Binder, S., Schendzielorz, G., Stähler, N., Krumbach, K., Hoffmann, K., Bott, M., et al. (2012). A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.* 13, R40. doi:10.1186/gb-2012-13-5-r40
- Carbonell, P., Planson, A.-G., Fichera, D., and Faulon, J.-L. (2011). A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst. Biol.* 5:122. doi:10.1186/1752-0509-5-122
- Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nat. Rev. Microbiol.* 12, 327–340. doi:10.1038/nrmicro3238
- D'Agnolo, G., Rosenfeld, I. S., Awaya, J., Omura, S., and Vagelos, P. R. (1973). Inhibition of fatty acid synthesis by the antibiotic cerulenin. Specific inactivation of beta-ketoacyl-acyl carrier protein synthetase. *Biochim. Biophys. Acta* 326, 155–156. doi:10.1016/0005-2760(73)90241-5
- Davis, M. S., Solbiati, J., and Cronan, J. E. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 275, 28593–28598. doi:10.1074/jbc.M004756200
- Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986). Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* 5, 2987–2994.
- Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010). High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* 79, 563–590. doi:10.1146/annurev-biochem-062608-095938
- Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013). Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol.* 2, 47–58. doi:10.1021/sb300091d
- Farmer, W. R., and Liao, J. C. (2000). Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* 18, 533–537. doi:10.1038/75398
- Fehér, T., Planson, A.-G., Carbonell, P., Fernández-Castané, A., Grigoras, I., Dariy, E., et al. (2014). Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol. J.* 9, 1446–1457. doi:10.1002/biot.201400055
- Flores, S., de Anda-Herrera, R., Gosset, G., and Bolívar, F. G. (2004). Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol. Bioeng.* 87, 485–494. doi:10.1002/bit.20137
- Fowler, Z. L., Gikandi, W. W., and Koffas, M. A. G. (2009). Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl. Environ. Microbiol.* 75, 5831–5839. doi:10.1128/AEM.00270-09

- Hall, B. G. (1998). Activation of the *bgl* operon by adaptive mutation. *Mol. Biol. Evol.* 15, 1–5. doi:10.1093/oxfordjournals.molbev.a025842
- Holtz, W. J., and Keasling, J. D. (2010). Engineering static and dynamic control of synthetic pathways. *Cell* 140, 19–23. doi:10.1016/j.cell.2009.12.029
- Liu, D., Xiao, Y., Evans, B., and Zhang, F. (2015). Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. *ACS Synth. Biol.* 4, 132–140. doi:10.1021/sb400158w
- Michener, J. K., and Smolke, C. D. (2014). Synthetic RNA switches for yeast metabolic engineering: screening recombinant enzyme libraries. *Methods Mol. Biol.* 1152, 125–136. doi:10.1007/978-1-4939-0563-8_7
- Oppenheim, A. V., and Schafer, R. W. (2010). *Discrete-Time Signal Processing*, 3rd Edn. Upper Saddle River, NJ: Pearson Higher Education, Inc., 504.
- Oyarzún, D. A., and Stan, G.-B. V. (2013). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J. R. Soc. Interface* 10:20120671. doi:10.1098/rsif.2012.0671
- Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014). Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17803–17808. doi:10.1073/pnas.1409523111
- Schallmeyer, M., Frunzke, J., Eggeling, L., and Marienhagen, J. (2014). Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr. Opin. Biotechnol.* 26, 148–154. doi:10.1016/j.copbio.2014.01.005
- Schujman, G. E., Guerin, M., Buschiazio, A., Schaeffer, F., Llarrull, L. I., Reh, G., et al. (2006). Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *EMBO J.* 25, 4074–4083. doi:10.1038/sj.emboj.7601284
- Schujman, G. E., Paoletti, L., Grossman, A. D., and de Mendoza, D. (2003). FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev. Cell* 4, 663–672. doi:10.1016/S1534-5807(03)00123-0
- Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014). Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metab. Eng.* 21, 2–8. doi:10.1016/j.ymben.2013.10.011
- Smolke, C. D., and Michener, J. K. (2012). High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng.* 14, 306–316. doi:10.1016/j.ymben.2012.04.004
- Stevens, J. T., and Carothers, J. M. (2014). Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synth. Biol.* 4, 107–115. doi:10.1021/sb400201u
- Tang, S.-Y., and Cirino, P. C. (2011). Design and application of a mevalonate-responsive regulatory protein. *Angew. Chem. Int. Ed. Engl.* 50, 1084–1086. doi:10.1002/anie.201006083
- Wu, J., Du, G., Zhou, J., and Chen, J. (2013). Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab. Eng.* 16, 48–55. doi:10.1016/j.ymben.2012.11.009
- Xu, P., Gu, Q., Wang, W., Wong, L., Bower, A. G. W., Collins, C. H., et al. (2013). Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. *Nat. Commun.* 4, 1409. doi:10.1038/ncomms2425
- Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11299–11304. doi:10.1073/pnas.1406401111
- Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., and Koffas, M. A. G. (2011). Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.* 13, 578–587. doi:10.1016/j.ymben.2011.06.008
- Xu, P., Vansiri, A., Bhan, N., and Koffas, M. A. G. (2012). ePathBrick: a synthetic biology platform for engineering metabolic pathways in *E. coli*. *ACS Synth. Biol.* 1, 256–266. doi:10.1021/sb300016b
- Xu, Y., Piston, D. W., and Johnson, C. H. (1999). A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 151–156. doi:10.1073/pnas.96.1.151
- Zeileis, A. (2014). dynlm: Dynamic Linear Regression. R package version 0.3-3. Available from: <http://CRAN.R-project.org/package=dynlm>
- Zhang, F., Carothers, J. M., and Keasling, J. D. (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

Conflict of Interest Statement: The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript. The Associate Editor, Jean Marie François, declares that, despite hosting a Frontiers Research Topic alongside the author Pablo Carbonell, the review process was handled objectively and no conflict of interest exists.

Received: 02 September 2014; accepted: 23 March 2015; published online: 08 April 2015.

Citation: Fehér T, Libis V, Carbonell P and Faulon J-L (2015) A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*. *Front. Bioeng. Biotechnol.* 3:46. doi: 10.3389/fbioe.2015.00046

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Fehér, Libis, Carbonell and Faulon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



New transposon tools tailored for metabolic engineering of Gram-negative microbial cell factories

Esteban Martínez-García, Tomás Aparicio, Víctor de Lorenzo and Pablo I. Nikel *

Systems and Synthetic Biology Program, Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain

Edited by:

Jean Marie François, CNRS, France

Reviewed by:

M. Kalim Akhtar, University College London, UK

Daehee Lee, Korea Research Institute of Bioscience and Biotechnology, South Korea

*Correspondence:

Pablo I. Nikel, Systems and Synthetic Biology Program (CNB-CSIC), C/Darwin, 3, Madrid 28049, Spain
e-mail: pablo.nikel@cnb.csic.es

Re-programming microorganisms to modify their existing functions and/or to bestow bacteria with entirely new-to-Nature tasks have largely relied so far on specialized molecular biology tools. Such endeavors are not only relevant in the burgeoning metabolic engineering arena but also instrumental to explore the functioning of complex regulatory networks from a fundamental point of view. *À la carte* modification of bacterial genomes thus calls for novel tools to make genetic manipulations easier. We propose the use of a series of new broad-host-range mini-Tn5-vectors, termed pBAMDs, for the delivery of gene(s) into the chromosome of Gram-negative bacteria and for generating saturated mutagenesis libraries in gene function studies. These delivery vectors endow the user with the possibility of easy cloning and subsequent insertion of functional cargoes with three different antibiotic-resistance markers (kanamycin, streptomycin, and gentamicin). After validating the pBAMD vectors in the environmental bacterium *Pseudomonas putida* KT2440, their use was also illustrated by inserting the entire poly(3-hydroxybutyrate) (PHB) synthesis pathway from *Cupriavidus necator* in the chromosome of a phosphotransacetylase mutant of *Escherichia coli*. PHB is a completely biodegradable polyester with a number of industrial applications that make it attractive as a potential replacement of oil-based plastics. The non-selective nature of chromosomal insertions of the biosynthetic genes was evidenced by a large landscape of PHB synthesis levels in independent clones. One clone was selected and further characterized as a microbial cell factory for PHB accumulation, and it achieved polymer accumulation levels comparable to those of a plasmid-bearing recombinant. Taken together, our results demonstrate that the new mini-Tn5-vectors can be used to confer interesting phenotypes in Gram-negative bacteria that would be very difficult to engineer through direct manipulation of the structural genes.

Keywords: metabolic engineering, *Pseudomonas putida*, *Escherichia coli*, transposon mini-Tn5, central metabolism, chromosomal integration, polyhydroxyalkanoates

INTRODUCTION

Over the last few years, a number of Gram-negative bacteria have become increasingly attractive *chassis* for a number of synthetic biology and metabolic engineering purposes. One conspicuous case involves the environmental bacterium *Pseudomonas putida* as a robust host for strong oxidative bioreactions, together with its GRAS (generally recognized as safe) status and its inherent ability to grow on a wide range of substrates (Nikel et al., 2014a). This situation calls for the expansion of the available tools for re-wiring its extant genetic features to further extend its metabolic potential – or even introducing new-to-Nature functions.

One frequently used molecular biology resource for analyses and manipulations of bacterial genomes is the Tn5 transposon. Historically, a number of plasmid vectors based on both wild-type and minimized versions of Tn5 (i.e. mini-transposons) had allowed the user to introduce stable insertions of foreign DNA into the chromosome of virtually any Gram-negative bacteria (de Lorenzo et al., 1990, 1998; Herrero et al., 1990; de Lorenzo and Timmis, 1994; Martínez-García et al., 2011; Martínez-García and de Lorenzo, 2012; Nikel and de Lorenzo, 2013a). Such Tn5-derived

elements present clear advantages over the use of their plasmid-based counterparts for the introduction and expression of heterologous genes into several bacterial species. These features include (but are not limited to) (i) the maintenance of the corresponding transgenes without antibiotic selective pressure, (ii) the long-term stability of the constructs and the re-usability of the functional parts, and, furthermore, (iii) Tn5 vectors admit cloning and chromosomal delivery of considerably long DNA fragments. Finally, as the transposase gene (*tnpA*) is lost following each transposition event (Berg, 1989; Reznikoff, 2006, 2008), one added value of mini-Tn5-vectors is the possibility to use them recursively in the same host, provided that they bear different selection markers. Moreover, as the TnpA transposase tends to act in *cis* (Phadnis et al., 1986), it promotes the insertion of DNA sequences borne by the plasmid, irrespective of previous insertions in a given target chromosome. These features allow for the integration of more than one DNA cargo into the same genome.

In this study, we report a series of synthetic, modular broad-host-range mini-Tn5-vectors for the delivery of gene(s) into the chromosome of a diversity of Gram-negative bacteria and to

construct saturated mutagenesis libraries for gene function studies. These vectors were termed pBAMDs, and they enable the possibility of easy cloning and subsequent chromosomal insertion of functional cargoes with three different and interchangeable antibiotic-resistance markers (kanamycin, streptomycin, and gentamicin). Moreover, the functional parts of the new vectors can be easily swapped by digestion with the appropriate restriction enzymes, allowing for the shuffling of each element as needed. Potential applications of the new tools are illustrated in two different genetic contexts. In one case, a systematic validation of the Tn5 vectors was carried out in *P. putida* KT2440, demonstrating the potential of the new insertional plasmids to be used in a sequential fashion for constructing (and deconstructing) complex phenotypes. In a second case, one of the pBAMD plasmids was used to insert a gene cluster from *Cupriavidus necator*, encoding all the biochemical functions needed for the formation of poly(3-hydroxybutyrate) (PHB), into the chromosome of *Escherichia coli*, thereby resulting in a new microbial cell factory tailored for biopolymer synthesis.

MATERIALS AND METHODS

BACTERIAL STRAINS, PLASMIDS, AND GROWTH CONDITIONS

The bacterial strains and plasmids used in this study are described in **Table 1**. Bacteria were routinely grown batchwise in LB medium (10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, and 5 g l⁻¹ NaCl) with rotary agitation (170 rpm). *P. putida* was grown at 30°C while *E. coli* cells were grown at 37°C. Selection of *P. putida* transconjugants was performed by spotting the cells onto M9 minimal medium agar plates (Sambrook et al., 1989) added with 0.2% (w/v) sodium citrate as the sole carbon source (to counterselect *E. coli* cells). PHB accumulation was assessed in selected *E. coli* transconjugants cultured in M9 minimal medium containing 30 g l⁻¹ glucose as the sole carbon source. Aerobic culture conditions in experiments aimed at polymer synthesis were achieved essentially as described by Nikel et al. (2010a), by using a 1:10 culture medium-to-flask volume ratio. Antibiotics were added at the following final concentrations whenever needed: ampicillin, 150 μg ml⁻¹ for *E. coli* or 500 μg ml⁻¹ for *P. putida*, chloramphenicol, 30 μg ml⁻¹; kanamycin, 50 μg ml⁻¹, streptomycin, 80 μg ml⁻¹; and gentamicin, 10 μg ml⁻¹. All solid media also contained 15 g l⁻¹ agar. Growth was estimated spectrophotometrically by measuring the optical density at 600 nm (OD₆₀₀) of the cultures (appropriately diluted in 9 g l⁻¹ NaCl whenever needed) in a Ultrospec 3000 *pro* UV/Visible spectrophotometer (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA). When culturing *E. coli* strains that accumulate PHB, for which OD₆₀₀ readings are no longer useful to estimate the cell dry weight (CDW), cells from 15 ml aliquots were washed, concentrated, and the CDW determined after drying the samples at 80°C to constant weight as previously indicated by Nikel et al. (2008a,b).

NUCLEIC ACID MANIPULATIONS AND GENERAL CLONING TECHNIQUES

DNA manipulations followed routine laboratory techniques as described by Sambrook et al. (1989) and Martínez-García and de Lorenzo (2012). Plasmid DNA was obtained using the QIAprep Spin™ Miniprep kit (Qiagen, Inc., Valencia, CA, USA). Restriction enzymes were obtained from New England Biolabs Inc. (Ipswich,

MA, USA), and T4 DNA ligase was purchased from Roche Applied Science Co. (Indianapolis, IN, USA). Plasmid p-R-SETA111 was constructed using isothermal assembly essentially as detailed by Gibson et al. (2009) but using a home-made mixture of enzymes. Colony PCR was performed using a single colony from a fresh LB agar plate and transferred directly into the reaction tube. PCR reactions were purified either with the NucleoSpin™ Gel and PCR clean-up kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) or the ExoSAP-IT™ PCR product clean-up kit (USB, Affymetrix Ltd., Santa Clara, CA, USA). Oligonucleotides were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA). The oligonucleotides used in this work for specific DNA constructions are indicated in Table S1 in the Supplementary Material; the oligonucleotides used to identify the location of the chromosomal insertions are indicated in **Table 2** (see also next section). The three different mini-Tn5 modules were chemically synthesized *de novo* by GeneCust Europe S.A. (Dudelange, Luxembourg). DNA sequencing was carried out by Secugen SL (Madrid, Spain).

Three separate DNA segments, carrying the transposon module along with the corresponding antibiotic-resistance determinant, were obtained as follows. In the first step, we used pBAM1 (Martínez-García et al., 2011) as the template to amplify the *bla* gene using primers Ap-AsiSI-F and Ap-MluI-R (Table S1 in the Supplementary Material), thereby substituting the *Swa*I and *Psh*AI restriction sites by target recognition sites for *Asi*SI and *Mlu*I, respectively, while maintaining the transcriptional control of *bla* through the native P3 promoter (Brosius et al., 1982). The backbone of plasmid pSEVA111 (Silva-Rocha et al., 2013) was amplified with the SEVA111-F and SEVA111-R oligonucleotide pair to obtain the second DNA fragment. Finally, the *tnpA* gene from pBAM1 was obtained using the oligonucleotides *tnpA*-*San*DI-F and *tnpA*-*Asi*SI-R that add the corresponding *San*DI and *Asi*SI restriction sites to the amplified fragment. These fragments were joined together by isothermal assembly, giving rise to plasmid p-R-SETA111 (Figure S1 in Supplementary Material).

Tripartite conjugative matings were set using *E. coli* CC118 λ*pir* (carrying pBAMD1-*x*, where *x* stands for any of the three antibiotic markers; see below) as the donor strain, the mating-helper strain *E. coli* HB101 (carrying pRK600), and *P. putida* KT2440 as the recipient strain. Conjugative matings were performed as described elsewhere (Martínez-García et al., 2011; Martínez-García and de Lorenzo, 2012). Briefly, the OD₆₀₀ from overnight cultures grown in LB medium with the appropriate antibiotics was adjusted to 1, then the cells were washed twice with 10 mM MgSO₄ to remove antibiotics from the culture medium, and each bacterial suspension was added to a 10 ml tube containing 5 ml of 10 mM MgSO₄ to obtain a final OD₆₀₀ of ca. 0.03. Biparental matings were done by following a similar procedure, but using *E. coli* S17-1 λ*pir* as the donor strain.

The mixture was concentrated by filtration and the cells were laid onto a filter disk (0.45 μm pore-size, 23 mm diameter, EMD Millipore Corp., Billerica, MA, USA). The filter was placed onto the surface of an LB agar plate and incubated at 30°C for 6 h. Finally, the biomass from the filter was suspended in 5 ml of 10 mM MgSO₄ and different dilutions were plated onto a suitable selective media.

Table 1 | Bacterial strains and plasmids used in this work.

Bacterial strain or plasmid	Relevant characteristics ^a	Reference or source
<i>Escherichia coli</i>		
DH5α λ _{pir}	Cloning host; F ⁻ λ ⁻ <i>endA1 glnX44(AS) thiE1 recA1 relA1 spoT1 gyrA96(Nal^R) rfbC1 deoR nupG Φ80(lacZΔM15) Δ(argF-lac)U169 hsdR17(r_K⁻ m_K⁺), λ_{pir} lysogen</i>	Hanahan and Meselson (1983)
CC118 λ _{pir}	Cloning host; Δ(<i>ara-leu</i>) <i>araD</i> Δ <i>lacX174 galE galK phoA thiE1 rpsE rpoB</i> (Rif ^R) <i>argE</i> (Am) <i>recA1</i> , λ _{pir} lysogen	Herrero et al. (1990)
S17-1 λ _{pir}	Cloning host; F ⁻ <i>recA1 endA1 thiE1 pro-82 creC510 hsdR17</i> RP4-2(Km:Tn7 Tc:Mu-1), λ _{pir} lysogen, Sm ^R	de Lorenzo et al. (1993)
S17P	Same as S17-1 λ _{pir} but carrying plasmid pBAMD1-6- <i>pha</i> (<i>phaC1AB1</i> ⁺), Sm ^R Ap ^R Gm ^R	This work
HB101	Helper strain; F ⁻ λ ⁻ <i>hsdS20</i> (r _B ⁻ m _B ⁻) <i>recA13 leuB6</i> (Am) <i>araC14</i> Δ(<i>gpt-proA</i>)62 <i>lacY1 galK2</i> (Oc) <i>xyl-5 mtl-1 thiE1 rpsL20</i> (Sm ^R) <i>glnX44</i> (AS)	Boyer and Roulland-Dussoix (1969)
BW25113	Wild-type strain; F ⁻ λ ⁻ Δ(<i>araD-araB</i>)567 Δ <i>lacZ4787</i> (: <i>rrnB-3</i>) <i>rph-1</i> Δ(<i>rhaD-rhaB</i>)568 <i>hsdR514</i>	Datsenko and Wanner (2000)
JW2293-1 ^b	Same as BW25113 but Δ <i>pta779:aphA</i> , Km ^R	Baba et al. (2006)
JW2293P	Same as JW2293-1 but carrying plasmid pAeT41 (<i>phaC1AB1</i> ⁺), Ap ^R	This work
TA2293P	Same as JW2293-1 but <i>ykgH::mini-Tn5(phaC1AB1)</i> , expresses the <i>pha</i> gene cluster from <i>Cupriavidus necator</i> as a chromosomal integration, Km ^R Gm ^R	This work
<i>Pseudomonas putida</i>		
KT2440	Wild-type strain; mt-2 derivative cured of the TOL plasmid pWW0	Bagdasarian et al. (1981)
Plasmids		
pRK600	Helper plasmid used for conjugation; <i>ori</i> (ColE1), RK2(<i>mob</i> ⁺ <i>tra</i> ⁺); Cm ^R	Kessler et al. (1992)
pBAM1	Mini-Tn5 delivery plasmid; <i>ori</i> (R6K), standard multiple cloning site; Ap ^R Km ^R	Martínez-García et al. (2011)
pSEVA111	Cloning vector; <i>ori</i> (R6K), standard multiple cloning site; Ap ^R	Silva-Rocha et al. (2013)
p-R-SETA111	Cloning vector; <i>ori</i> (R6K), standard multiple cloning site, <i>tnpA</i> ; Ap ^R	This work
pBAMD1-2	Mini-Tn5 delivery plasmid; <i>ori</i> (R6K); Ap ^R Km ^R	This work
pBAMD1-4	Mini-Tn5 delivery plasmid; <i>ori</i> (R6K); Ap ^R Sm ^R /Sp ^R	This work
pBAMD1-6	Mini-Tn5 delivery plasmid; <i>ori</i> (R6K); Ap ^R Gm ^R	This work
pAeT41	Derivative of cloning vector pUC18 (Thermo Fisher Scientific Inc.; Pittsburgh, PA, USA) bearing a ca. 5-kb <i>Sma</i> I/ <i>Eco</i> RI DNA fragment spanning the <i>phaC1AB1</i> gene cluster from <i>C. necator</i> ; Ap ^R	Peoples and Sinskey (1989)
pBAMD1-6- <i>pha</i>	Derivative of pBAMD1-6; delivery plasmid used to insert the <i>phaC1AB1</i> gene cluster from <i>C. necator</i> into the chromosome of recipient <i>E. coli</i> strains; Ap ^R Gm ^R	This work

^aAntibiotic markers: Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Nal, nalidixic acid; Rif, rifampicin; Sm, streptomycin; Sp, spectinomycin; and Tc, tetracycline.

^bStrain obtained from the *E. coli* Genetic Stock Center (Yale University, New Haven, CT, USA).

LOCALIZATION OF THE mini-Tn5 TRANSPOSON INSERTION SITES BY ARBITRARY PCR

In order to specifically map the landing sites of the mini-transposon within the chromosome, a set of oligonucleotides was designed that specifically hybridizes in each of the mini-Tn5 elements (Table 2). Transconjugants were streaked onto M9 minimal medium agar plates with 0.2% (w/v) sodium citrate as the sole carbon source and supplemented with the appropriate antibiotics to obtain isolated colonies, which were re-streaked again in the

same culture medium to obtain isolated clones. These colonies were used as the template for arbitrary PCR. The identification of the transposon insertion site could be independently obtained using either the oligonucleotides that are close to the ME-O end or with the ones designed for the ME-I end. However, when these plasmids carry any heterologous DNA in the multiple cloning site (MCS), the use of the oligonucleotides that are close to the ME-O end is the preferable option, since the ones for the ME-I end are located downstream the MCS and therefore would amplify

Table 2 | Oligonucleotides used in this work to identify and sequence the chromosomal locus of mini-Tn5 insertions.

Oligonucleotide	Sequence (5' → 3')	Use and reference
ARB6	GGCACGCGTCGACTAGTACNNNNNNNNNACGCC	Arbitrary PCR, round 1 (Pratt and Kolter, 1998)
ARB2	GGCACGCGTCGACTAGTAC	Arbitrary PCR, round 2 (Pratt and Kolter, 1998)
pBAM-ME-I-Ext-R	CTCGTTTCACGCTGAATATGGCTC	Arbitrary PCR for pBAMD1-2, round 1 (Martínez-García et al., 2011)
pBAM-ME-I-Int-R	CAGTTTTATTGTTTCATGATGATATA	Arbitrary PCR for pBAMD1-2, round 2, and sequencing (Martínez-García et al., 2011)
ME-O-Km-Ext-F	CGTCTGTTTCAGAAATATGGCAT	Arbitrary PCR for pBAMD1-2, round 1
ME-O-Km-Int-F	ATCTGATGCTGGATGAATTTTTC	Arbitrary PCR for pBAMD1-2, round 2, and sequencing
ME-I-Sm-Ext-R	ATGACGCCAACTACCTCTGATA	Arbitrary PCR for pBAMD1-4, round 1
ME-I-Sm-Int-R	TCACCGCTCCCTCATGATGTT	Arbitrary PCR for pBAMD1-4, round 2, and sequencing
ME-O-Sm-Ext-F	CTTGGCCTCGCGCGAGATCAG	Arbitrary PCR for pBAMD1-4, round 1
ME-O-Sm-Int-F	CACCAAGGTAGTCGGCAAAT	Arbitrary PCR for pBAMD1-4, round 2, and sequencing
ME-O-Gm-Ext-F	GCACTTTGATATCGACCCAAGT	Arbitrary PCR for pBAMD1-6, round 1
ME-O-Gm-Int-F	TCCCGGCCGCGGAGTTGTTCCGG	Arbitrary PCR for pBAMD1-6, round 2, and sequencing
ME-I-Gm-Ext-R	GTTCTGGACCAGTTGCGTGAG	Arbitrary PCR for pBAMD1-6, round 1
ME-I-Gm-Int-R	GAACCGAACAGGCTTATGTCA	Arbitrary PCR for pBAMD1-6, round 2, and sequencing

through it. In the later case, and depending on the length of the DNA cloned in the MCS, the corresponding amplicon may not provide long enough a sequence to ascertain the insertion site of the mini-transposon module.

The conditions of the first round of arbitrary PCR were as follows: 5 min at 95°C (initial denaturation); six cycles of 30 s at 95°C, 30 s at 30°C, and 90 s at 72°C; and 30 cycles of 30 s at 95°C, 30 s at 45°C, and 90 s at 72°C (Das et al., 2005). The ARB6 oligonucleotide was used together with the external oligonucleotides within the mini-transposon (indicated with the “Ext” acronym in **Table 2**). Then, we used 1 µl of the first PCR round as the template for the second round of arbitrary PCR by applying the following conditions: 1 min at 95°C (initial denaturation); 30 cycles of 30 s at 95°C, 30 s at 52°C, and 90 s at 72°C; followed by an extra extension of 4 min at 72°C (Das et al., 2005). For the second round of arbitrary PCR, the ARB2 oligonucleotide was used together with the internal primers within the mini-transposon (indicated with the “Int” acronym in **Table 2**). Finally, the PCR amplification product obtained in the second round was directly purified and sent for sequencing with the corresponding internal oligonucleotide.

DNA sequences were thoroughly inspected visually for any error and analyzed using the *Pseudomonas* Genome Database (Winsor et al., 2011), and BlastN (Altschul et al., 1990, 1997) was subsequently employed to map the precise transposon insertion point. To ascertain the conservation level of the 9-bp target sequence of Tn5 transposase, we resorted to the web-based application WebLogo 3.4 (Crooks et al., 2004).

ANALYTICAL PROCEDURES

For a coarse estimation of the PHB content in *E. coli* transconjugants, we resorted to a fluorimetric assay based on Nile red staining (Spiekermann et al., 1999). Cells were grown overnight in LB medium with the proper antibiotic. The cultures were diluted

to an OD₆₀₀ of 0.1 in fresh LB medium containing 30 g l⁻¹ glucose and 200 µl aliquots were placed in a 96 well microtiter plate (Costar™ black plates with clear bottom; Thermo Fisher Scientific Inc.). After growing the cells for 24 h at 37°C, 0.002 volume of a Nile red stock solution, freshly prepared by dissolving the dye (purchased from Sigma-Aldrich Co.) to 1 mg ml⁻¹ in dimethyl sulfoxide, were added to each well. The microtiter plates were incubated at 37°C in the dark for 30 min, and the fluorescence at 585 nm was measured in a SpectraMax M2e plate reader (Molecular Devices, LLC., Sunnyvale, CA, USA) in cells before and after staining with Nile red. The raw fluorescence readings were normalized to the biomass in each well by dividing the values by the OD₆₀₀ of the corresponding culture.

The quantification of PHB content in *E. coli* via fluorescence-activated cell sorting (FACS) was conducted by following a slight modification of the protocol described by Tyo et al. (2006). In brief, cultures to be analyzed were promptly cooled to 4°C by placing them in an ice bath for 15 min. Cells were harvested by centrifugation (5 min, 5,000 × g, 4°C), resuspended to an OD₆₀₀ of 0.4 in cold TES buffer [10 mM Tris·HCl (pH = 7.5), 2.5 mM EDTA, and 10% (w/v) sucrose], and further incubated on ice for 15 min. Bacteria were recovered by centrifugation as explained above, and finally resuspended in the same volume of cold 1 mM MgCl₂. A 1 ml aliquot of this suspension was added with 3 µl of an 1 mg ml⁻¹ Nile red solution and incubated in the dark at 4°C for 30 min. Cells were analyzed by FACS immediately after the staining procedure. FACS was carried out in a MACSQuant™ VYB cytometer (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany). Cells were excited with an Ar laser (488 nm, diode-pumped solid state), and the Nile red fluorescence at 585 nm was detected with a 614/50 nm band-pass filter. FACS analysis was done on at least 50,000 cells and the results were analyzed with the built-in MACSQuantify™ software 2.5 (Miltenyi Biotec). The geometric

mean of fluorescence in each sample was correlated to the PHB content (expressed as a percentage) through a calibration curve as described previously (Tyo et al., 2006).

Cell-free extracts were obtained from cells harvested by centrifugation from an appropriate culture volume at $4,000 \times g$ at 4°C for 10 min and processed as described previously (Nikel and de Lorenzo, 2013b; Nikel et al., 2014b). The total protein concentration in cell extracts was assessed by means of the Bradford method (Bradford, 1976) using a commercially available kit from BioRad Laboratories, Inc. (Hercules, CA, USA), with crystalline bovine serum albumin as the standard for determinations. *In vitro* quantification of the specific 3-ketoacyl-coenzyme A (CoA) thiolase activity in the thiolysis direction was conducted according to the protocols developed by Palmer et al. (1991) and Slater et al. (1998), with some modifications. The assay mixture (1 ml) contained 65 mM Tris-HCl (pH = 7.5), 50 mM MgCl_2 , 62.5 μM CoA, and 65 μM 3-acetoacetyl-CoA (Sigma-Aldrich Co.). Solutions of both CoA and 3-acetoacetyl-CoA were freshly prepared just prior to the assay. The assay was initiated upon the prompt addition of the cell-free extract, and the disappearance of 3-acetoacetyl-CoA was measured with time at 304 nm (using an extinction coefficient for 3-acetoacetyl-CoA $\epsilon_{304} = 16.9 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$). One enzyme unit was defined as the amount of enzyme catalyzing the conversion of 1 μmol of substrate to product per min at 25°C .

Residual glucose and acetate concentrations in culture supernatants were determined in selected samples using adequate enzymatic kits (R-Biopharm AG, Darmstadt, Germany), essentially as per the manufacturer's instructions. In either case, control mock assays were made by spiking M9 minimal medium with different amounts of the metabolite under examination. Metabolite yields and kinetic culture parameters were analytically calculated from the raw growth data as described elsewhere (Nikel et al., 2008a, 2010a, 2014b; Nikel and de Lorenzo, 2013a,b).

STATISTICAL ANALYSIS

The reported experiments were independently repeated at least twice (as indicated in the corresponding figure legend), and the mean value of the corresponding parameter \pm SD is presented. When appropriate, data were statistically treated with an unpaired Student's *t* test, and 95% confidence intervals for each parameter were calculated to demonstrate a statistically significant difference in means among the experimental samples. For the flow cytometry experiments, the geometric mean values (from which the PHB content is derived) were analyzed *via* the Mann-Whitney *U* test.

NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

The sequences of the pBAMD vectors were deposited in the GenBank database with the following GenBank accession numbers: KM403113 (pBAMD1-2), KM403114 (pBAMD1-4), and KM403115 (pBAMD1-6).

RESULTS AND DISCUSSION

RATIONALE, DESIGN, AND GENERAL CHARACTERISTICS OF THE mini-Tn5-BASED VECTORS pBAMDS

Vector pBAM1 (*born again mini-transposon*) is a synthetic and modular plasmid with a number of features to facilitate the genome editing of Gram-negative bacteria (Martínez-García et al.,

2011). We decided to further extend the range of such applications by constructing a new set of pBAM1-derivative plasmids, which are compatible with the rules set in the *Standard European Vector Architecture* (SEVA) format (Silva-Rocha et al., 2013). We decided to accomplish this challenge by constructing a standardized version of the mini-transposon delivery plasmid that take full advantage of all the benefits of the pBAM1 plasmid together with the functional elements available within the SEVA collection. The starting idea was to design a plasmid series in which the mini-transposon module, the antibiotic-resistance marker, and the *tnpA* could be easily interchanged at the user's will. In doing so, several changes were needed to re-structure the constituents of pBAM1, giving rise to three plasmids that were collectively termed pBAMD (i.e., *pBAM derivative*) vectors. These insertion plasmids share all the advantages and several structural features of their pBAM1 predecessor. In brief, these features include (i) the narrow host-range origin of replication of plasmid R6K [*ori*(R6K)], dependent on the Π protein (encoded by the *pir* gene of plasmid R6K); (ii) an origin of transfer, *oriT*, that allows for the conjugative transfer of the plasmid from a host strain to a new bacterial recipient through RK2-mediated mobilization; (iii) the *bla*-encoded β -lactamase marker that confers resistance to ampicillin as a selective marker of the backbone vector; and [iv] a modified, hyper-active transposase encoded by *tnpA* just outside (but adjacent to) a DNA segment that is flanked by the terminal sequences of Tn5 (i.e., the mini-transposon module itself). The Tn5-transposition system is an optimal source of biological parts because of its genetic promiscuity, and it can be considered to operate as a virtually orthogonal part, since it displays an autonomous behavior with respect to the host metabolic and regulatory traits.

We first constructed an intermediary plasmid, named p-R-SETA111, that was later used as the backbone in which the different mini-Tn5 antibiotic modules were implanted to obtain the pBAMD1-*x* vectors. A mixture of three separate DNA fragments were isothermally assembled (Gibson et al., 2009) to construct the p-R-SETA111 intermediary plasmid (Figure S1 in Supplementary Material). The sequence of p-R-SETA111 was thoroughly checked after assembling with the set of oligonucleotides described in Table S1 in the Supplementary Material. This intermediate vector bears an R6K origin of replication that depends on the Π protein supplied in *trans* (Kolter et al., 1978) for replication. This situation calls for the use of *E. coli pir*⁺ strains to propagate these plasmids (Miller and Mekalanos, 1988; Herrero et al., 1990), such as *E. coli* DH5 α λ *pir*, CC118 λ *pir*, or S17-1 λ *pir* (Table 1). Another feature of plasmid p-R-SETA111 is the presence of a minimized origin of transfer (*oriT*) from the promiscuous conjugative plasmid RP4 (Lyras and Rood, 1998; Silva-Rocha et al., 2013). Another trait of this intermediary vector shared with the SEVA plasmids is that the cargo module is flanked by the strong T1 and T0 transcriptional terminators (Silva-Rocha et al., 2013), which isolate transcriptionally any DNA sequence cloned in the MCS of p-R-SETA111. Importantly, the modular design of this plasmid allows for the convenient exchange of the *tnpA* by restriction of p-R-SETA111 with the rare cutters *SanDI* (5'-GG/GWCCC-3', W = A or T; Simcox et al., 1995) and *AsiSI* (5'-GCGAT/CGC-3'). Likewise, the antibiotic marker of the plasmid backbone (*bla* in this

particular case), can be easily exchanged by enzymatic restriction with *AsiSI* and *MluI* (5'-A/CGCGT-3').

Three different mini-Tn5 modules were designed as the cargo segments to be implanted into the p-R-SETA111 plasmid. These elements were devised as cargoes for the SEVA plasmid collection (Silva-Rocha et al., 2013; Durante-Rodríguez et al., 2014), and so they were bracketed by *PacI* and *SpeI* restriction sites. The cargo modules have a similar structural design, only differing in the antibiotic marker placed within the mini-transposon (see below). The mini-Tn5 modules are flanked by the two mosaic end (ME) sequences. ME elements are optimized 19-bp DNA sequences recognized by the Tn5 TnpA transposase to promote the specific transposition of any DNA segment bracketed by these elements (Zhou et al., 1998). Even though they are identical in sequence, and with the aim to facilitate the orientation of each functional element within the plasmid, we termed them as either ME-I (the one just after the *PacI* recognition site) or ME-O (the one close to the *SpeI* recognition site).

The next relevant feature, placed right after the ME-I element, is a 10-bp buffer DNA sequence that is immediately followed by a MCS (spanning recognition sites for *AvrII*, *SfiI*, *NotI*, *EcoRI*, *SacI*, *SmaI*, *BamHI*, *XbaI*, *SalI*, *PstI*, *SphI*, *HindIII*, and *NotI*, in the 5' → 3' direction). This DNA stretch has the same restriction sites as those present in cargo 1 in the SEVA database. The synthetic T500 terminator (Yarnell and Roberts, 1999) was placed after the MCS in order to avoid any transcriptional read-through that any heterologous DNA cloned within the cargo can leak into the following component of the plasmid. The next functional

element of the mini-transposon is the antibiotic selection marker that allows for the proper selection of transconjugants. For this set of plasmids, we have used resistances to kanamycin (*aphA*, encoding an aminoglycoside 3'-phosphotransferase; Martínez-García et al., 2011), streptomycin/spectinomycin (*aadA*, encoding a streptomycin 3''(9)-O-nucleotidyl transferase; Fling et al., 1985), and gentamicin (*aacC1*, encoding a gentamicin 3'-N-acetyltransferase; Kovach et al., 1995). These antibiotic-resistance cassettes are coded as 2, 4, and 6 in the SEVA database. Since the antibiotic selection cassettes are flanked by *SwaI* and *PshAI* recognition sites, the user has the ability to further expand the plasmid collection by exchanging the cognate resistance genes with any of the two other markers present in the SEVA collection, i.e., *cat* (encoding a chloramphenicol O-acetyltransferase) or *tetA* (encoding a tetracycline efflux protein). The *rho*-independent transcriptional terminator from the gene 32 encoded in the phage T4 genome (Gorski et al., 1985; Miller et al., 2003; Martínez-García et al., 2011) was placed immediately downstream to the antibiotic-resistance gene to prevent any possible read-through from the corresponding promoter elements. Furthermore, the motif 5'-GGGACCC-3' was changed to 5'-CGGACCC-3' to eliminate a *SanDI* restriction target within the existing terminator sequence.

The last step in the construction of the pBAMD1-*x* vectors was the insertion of the mini-transposon modules themselves into the p-R-SETA111 backbone. The sequences of the three mini-Tn5 modules were edited *in silico* to follow the SEVA rules and synthesized *de novo*. The mini-transposon modules were restricted

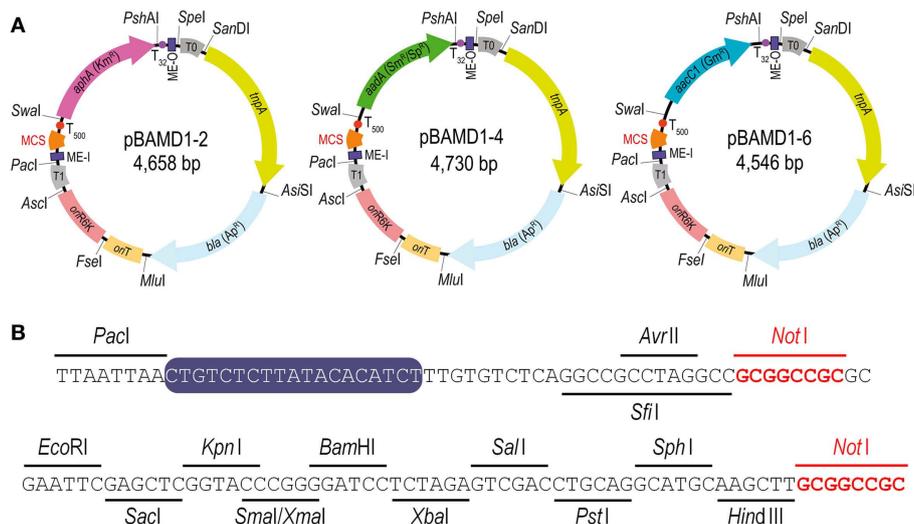


FIGURE 1 | Functional features of the pBAMD1-*x* delivery vectors.

(A) Schematic representation of the pBAMD1-*x* plasmid series. The functional elements of each plasmid include the relevant restriction sites used for assembling the vectors, the antibiotic-resistance markers (Ap, ampicillin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin; and Gm, gentamicin), the hyper-active TnpA transposase encoded by *tnpA*, a conditional origin of replication [*ori*(R6K)] dependent of the π protein, an origin of transfer (*oriT*), two mosaic elements (termed ME-I and ME-O), the transcriptional terminators T0 and T1 located just outside the transposon module, and a multiple cloning site (MCS)

compatible with any plasmid belonging to the *Standard European Vector Architecture* (SEVA) initiative (Silva-Rocha et al., 2013; Durante-Rodríguez et al., 2014). Note that the antibiotic-resistance gene determines the full name of each plasmid. **(B)** Restriction enzymes targets within the multiple cloning site of the pBAMD1-*x* delivery vectors ($x = 2$, Km^r; $x = 4$, Sm^r/Sp^r; and $x = 6$, Gm^r). The MCS starts with the unique *AvrII*/*SfiI* recognition sites, and two *NotI* recognition sites (highlighted in red) were included in the MCS sequence to enable the consecutive assembly of different cargoes from the SEVA collection. The sequence of ME-I is indicated by a purple box.

using *PacI* and *SpeI* and inserted into p-R-SETA111 previously digested with the same enzymes. The correctness of this last cloning step was verified by DNA sequencing with the SEVA oligonucleotides PS1 and PS2 (Table S1 in the Supplementary Material), that flank the cargo module. The resulting delivery plasmids were termed pBAMD1-2 (Km^R), pBAMD1-4 (Sm^R/Sp^R), and pBAMD1-6 (Gm^R) (Figure 1A). Note that the first digit numeric nomenclature stems for the fact that all the pBAMD1-x vectors are derivatives of the pSEVA111 vector, while the second number identify the antibiotic-resistance marker. All the vectors share the same MCS (Figure 1B), also compatible with the rest of SEVA vectors already available.

FUNCTIONAL VALIDATION OF THE pBAMD1-x VECTORS IN *P. PUTIDA* KT2440

To evaluate the functionality of the new set of Tn5 plasmids, the frequency of transposition into the environmental bacterium *P. putida* KT2440 was firstly tested in 6 h triparental mating assays. In order to estimate the frequency of transposition events, the number of antibiotic-resistant colonies was assessed after 24 h of incubation at 30°C, and normalized to the total of 1.5×10^8 recipient cells used in each experiment. The average frequency of transposition obtained at 6 h was $(3.8 \pm 2.9) \times 10^{-4}$ transconjugants (ranging from a minimum of 2.5×10^{-5} to a maximum of 1.0×10^{-3} transconjugant cells). These figures were independent of the antibiotic marker used (either pBAMD1-2, pBAMD1-4, or pBAMD1-6), and no spontaneous antibiotic-resistant clones (i.e., Km^R, Sm^R, or Gm^R, respectively) were detected under such growth conditions (data not shown).

The next step was to differentiate between *bona fide* transposition events and non-specific plasmid integration events in the *P. putida* genome. To do so, transconjugants cells were re-streaked onto LB medium plates containing 500 µg ml⁻¹ ampicillin and incubated for 24 h to check for possible growth – which would indicate that the corresponding pBAMD vector had integrated into the target chromosome instead of transposing the MEs-flanked DNA cargo. Besides this simple test, the user could also perform colony PCR amplifications to confirm that the transconjugants do not have the pBAMD plasmid backbone by using any of the following two SEVA oligonucleotides combination. If the plasmid is present, the PS5-PS4 oligonucleotide pair will produce a 225-bp amplicon within the *oriT* region, and the PS5-PS6 oligonucleotide pair will generate a 665-bp amplicon including the *oriT* segment and the R6K origin of replication (Silva-Rocha et al., 2013). By conducting these two assays, we noticed that $4.2 \pm 2.4\%$ of the potential transconjugants obtained with the pBAMD vectors resulted from plasmid co-integration events (a percentage very similar to that reported for other Tn5-based plasmid systems; de Lorenzo et al., 1990). Therefore, it is highly recommended to confirm the nature of the antibiotic-resistant clones obtained after each round of insertions.

We also studied whether the pBAMD mini-transposon delivery plasmids can be used serially to generate double and even triple transconjugant mutants by taking advantage of the three different antibiotic resistance markers (Figure 2A). A first round of transposition was performed with the three individual pBAMD plasmids using *P. putida* KT2440 as the recipient strain. Transconjugants

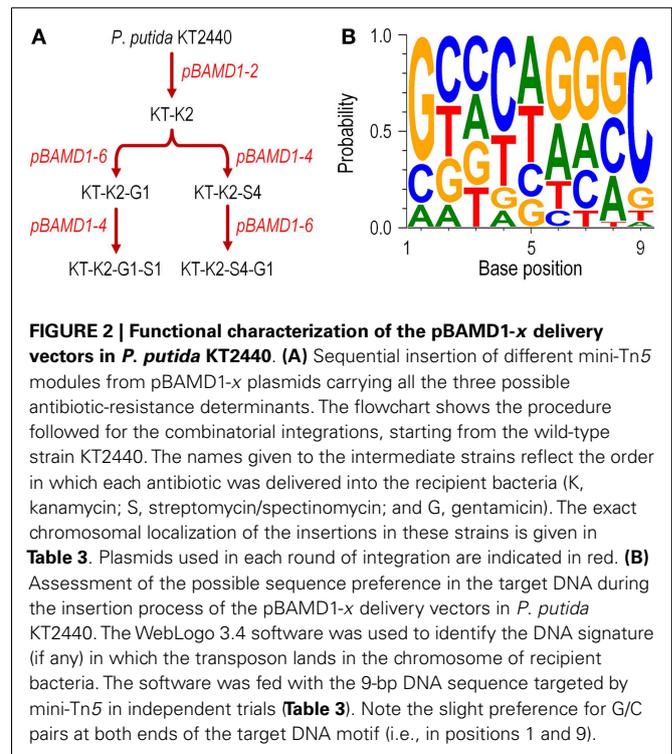


FIGURE 2 | Functional characterization of the pBAMD1-x delivery vectors in *P. putida* KT2440. (A) Sequential insertion of different mini-Tn5 modules from pBAMD1-x plasmids carrying all the three possible antibiotic-resistance determinants. The flowchart shows the procedure followed for the combinatorial integrations, starting from the wild-type strain KT2440. The names given to the intermediate strains reflect the order in which each antibiotic was delivered into the recipient bacteria (K, kanamycin; S, streptomycin/spectinomycin; and G, gentamicin). The exact chromosomal localization of the insertions in these strains is given in **Table 3**. Plasmids used in each round of integration are indicated in red. **(B)** Assessment of the possible sequence preference in the target DNA during the insertion process of the pBAMD1-x delivery vectors in *P. putida* KT2440. The WebLogo 3.4 software was used to identify the DNA signature (if any) in which the transposon lands in the chromosome of recipient bacteria. The software was fed with the 9-bp DNA sequence targeted by mini-Tn5 in independent trials (**Table 3**). Note the slight preference for G/C pairs at both ends of the target DNA motif (i.e., in positions 1 and 9).

were re-streaked onto M9 minimal medium plates containing 0.2% (w/v) citrate and supplemented with the appropriate antibiotics to obtain isolated colonies. These colonies were re-streaked again in the same culture medium to obtain pure clones. We randomly picked 22 clones obtained with each pBAMD1-x plasmid and characterized the corresponding transposon insertion site. In 12 out of 22 transconjugant clones, the ME-O related oligonucleotides (**Table 2**) were used, while in the remaining 10 transconjugant clones, the sequence was obtained by using the ME-I related oligonucleotides for the arbitrary PCR amplification. One Km^R clone after the first round of transposition with pBAMD1-2 was selected and used as the recipient strain for the pBAMD1-4 and pBAMD1-6 mini-transposon plasmids. After this second insertion round, we selected four transconjugants obtained with each system, and mapped the landing point of the mini-transposon using only the ME-O related oligonucleotides. Finally, we selected one *P. putida* KT2440 Km^R and Gm^R and one *P. putida* KT2440 Km^R and Sm^R to perform the third round of transposition with either pBAMD1-4 or pBAMD1-6, respectively. In the last step, we have chosen one mutant per plasmid system and characterized again the insertion site of the three mini-transposons using the ME-O oligonucleotides. The precise site of transposon insertion could be ascertained in 30 transconjugants out of 32 independent clones (**Table 3**). Specifically, in one of the cases in which it was not possible to identify the localization of the mini-Tn5 element, the transposon insertion could have happened in either PP2612 or PP3616, since both genes share a 97% sequence identity. In the other case, the transposon insertion site could not be precisely mapped due to the presence of a large number of internal repeats within the *lapA* gene (PP0168).

Table 3 | Insertion sites of the mini-transposon born by the different pBAMD1-x vectors in the genome of *P. putida* KT2440^a.

Clone	End	Genome coordinate (bp)	PP number	Strand	Gene name and putative function
KTK1	ME-I	5,825,403	PP5103	–	<i>trmB</i> , tRNA (guanine- <i>N</i> ₇ -) methyltransferase
KTK2	ME-I	425,161	PP0349	–	PepSY-associated transmembrane helix domain-containing protein
KTK3	ME-I	5,272,887	PP4648	+	rRNA (guanine- <i>N</i> ₂ -) methyltransferase
KTK5	ME-O	864,889	PP0745	–	<i>uraA</i> , uracil-xanthine permease
KTK6	ME-O	4,661,528	PP4124	+	<i>nuoG</i> , NADH dehydrogenase subunit G
KTK7	ME-O	1,261,995	PP1104	–	Succinyl-glutamate desuccinylase/aspartoacylase
KTK8	ME-O	727,311	PP0621	–	Hypothetical protein
KT-G1 ^b	ME-O	2,988,216	PP2612	–	Hypothetical protein
		4,108,354	PP3616	+	Hypothetical protein
KT-G2	ME-O	1,313,801	PP1145	–	<i>hepA</i> , ATP-dependent helicase
KT-G3	ME-O	3,002,652	PP2626	+	Hypothetical protein
KT-G4	ME-O	3,251,143	PP2847	+	<i>ureJ</i> , hydrogenase/urease accessory protein
KT-G6	ME-I	4,444,716	PP3941	–	Isochorismatase superfamily hydrolase
KT-G7	ME-I	4,792,981	PP4221	–	Non-ribosomal peptide synthetase
KT-G8	ME-I	1,959,862	PP1757	+	<i>bolA</i> , BolA family protein
KTS1	ME-I	4,463,594	PP3956	+	Hypothetical protein
KTS2	ME-I	3,552,337	PP3136	NA	Intergenic region (PP3136 and PP3137)
KTS3	ME-I	4,671,241	PP4132	–	Hypothetical protein
KTS4	ME-I	35,484	PP0031	–	Hypothetical protein
KTS5	ME-O	2,422,455	PP2123	+	<i>moeA</i> , molybdopterin biosynthesis enzyme
KTS6	ME-O	6,021,399	PP5272	–	Hypothetical protein
KTS7	ME-O	381,423	PP0317	–	Methyl-accepting chemotaxis sensory transducer
KTS8	ME-O	328,477	PP0270	–	Integral membrane sensor signal transduction histidine kinase
KTK2-G1	ME-O	4,892,623	PP4301	+	<i>pykF</i> , pyruvate kinase
KTK2-G2	ME-O	3,749,626	PP3313	–	Heat shock protein
KTK2-G3	ME-O	5,842,545	PP5120	+	Conifer aldehyde dehydrogenase
KTK2-G4	ME-O	1,373,720	PP1199	–	Hypothetical protein
KTK2-S1	ME-O	2,904,080	PP2556	–	Chromate transporter
KTK2-S2	ME-O	5,347,381	PP4703	+	Hypothetical protein
KTK2-S3	ME-O	1,542,387	PP1353	+	Mechanosensitive ion channel protein MscS
KTK2-S4	ME-O	5,892,275	PP5166	–	Fis family transcriptional regulator
KTK2-G1-S1	ME-O	3,635,742	PP3202	+	Predicted exporter of the RND superfamily
KTK2-S4-G1 ^b	ME-O	NA	PP0168	+	<i>lapA</i> , surface adhesion protein

^aThe insertion sites were ascertained by means of arbitrary PCR with the oligonucleotides indicated in **Table 2**, and the assigned function of the corresponding ORF is given according to the information available in the *Pseudomonas* Genome Database (Winsor et al., 2011). NA, not available.

^bIn the two transconjugant clones indicated, the insertion site of the mini-Tn5 element could not be unambiguously identified.

We then used the insertion sequence data of each mapped clone in **Table 3** to detect any sequence preference for the integration of the mini-Tn5 cassettes within the genome of *P. putida* KT2440. The web-based application WebLogo 3.4 (Crooks et al.,

2004) was fed with the 9-bp landing sequence targeted by the mini-transposon. The program was set to show the probability of having a defined base at any specific position within the 9-bp motif, and the G + C percentage of the genome was adjusted to the value of *P.*

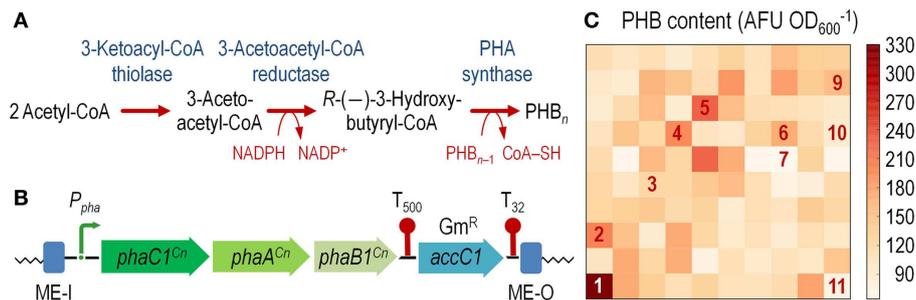


FIGURE 3 | Construction of an *E. coli* cell factory expressing the *phaC1AB1* gene cluster from *C. necator* as a chromosomal insertion. (A) Poly(3-hydroxybutyrate) (PHB) biosynthesis pathway. Three enzymes are necessary for *de novo* synthesis of PHB in *C. necator*: a 3-ketoacyl-coenzyme A (CoA) thiolase (PhaA), a NADPH-dependent 3-acetoacetyl-CoA reductase (PhaB1), and a PHB synthase (PhaC1). PhaA and PhaB1 catalyze the condensation of two molecules of acetyl-CoA to 3-acetoacetyl-CoA and the reduction of acetoacetyl-CoA to *R*-(-)-3-hydroxybutyryl-CoA, respectively. PhaC1 polymerizes these monomers to PHB, whereas one CoA-SH molecule per monomer is released. The resulting PHB polymer is stored as water-insoluble granules in the cytoplasm of the cells. **(B)** Organization of the functional elements borne by plasmid pBAM1-6-*pha* and transferred into the

chromosome of the recipient *E. coli* strain. The transcriptional terminators included in the plasmid backbone, which flank the gentamicin resistance (Gm^R) determinant (*accC1*), are depicted as T₅₀₀ and T₃₂. Note that the elements in this outline are not drawn to scale. **(C)** Exploring the landscape of PHB synthesis in *E. coli* transconjugants. The *phaC1AB1* gene cluster from *C. necator* was randomly integrated into the chromosome of *E. coli* JW2293-1 (Δ *pta*), and 24-h cultures of individual colonies were analyzed for PHB accumulation by fluorimetry after staining the cells with Nile red (see Materials and Methods for details). Several colonies, identified by numbers in the heat-map, were kept and further analyzed to establish the precise site of mini-Tn5(*phaC1AB1*) insertion (Table S2 in the Supplementary Material). AFU, arbitrary fluorescence units.

putida KT2440 (G + C = 61.5%; Nelson et al., 2002). The results shown in **Figure 2B** reveal that there is no DNA sequence bias for the integration site of the mini-transposon, in a similar fashion as observed for plasmid pBAM1 (Martínez-García et al., 2011). However, a relatively minor preference for G/C pairs at both ends of the target DNA motif could be observed, as detected for other systems based on Tn5 (Lodge et al., 1988). These experiments confirm that the three mini-transposons borne by the pBAMD1-*x* vectors could be used serially to generate a second or even a third round of insertion mutagenesis procedure, or also to stably integrate multiple genetic devices into the genome of a single microbial strain.

DESIGN AND CONSTRUCTION OF A MICROBIAL CELL FACTORY FOR POLY(3-HYDROXYBUTYRATE) SYNTHESIS

Engineering an stable PHB⁺ phenotype in *E. coli*

Poly(3-hydroxybutyrate) is an isotactic polyester composed by 3-hydroxybutyrate units (Anderson and Dawes, 1990). The PHB synthesis pathway in *C. necator* (formerly known as *Ralstonia eutropha*) comprises three enzymes (**Figure 3A**) (Steinbüchel and Hein, 2001). PhaA, a 3-ketoacyl-CoA thiolase, condenses two acetyl-CoA moieties, yielding 3-acetoacetyl-CoA. This intermediate is the substrate for PhaB, a NADPH-dependent 3-acetoacetyl-CoA reductase (encoded by *phaB1*). In the final step of this biosynthetic pathway, (*R*)-(-)-3-hydroxybutyryl-CoA is polymerized to PHB by PhaC, a poly(3-hydroxyalkanoate) synthase (encoded by *phaC1*). The very idea of a thermoplastic and biocompatible material, which is also readily biodegraded by a number of bacteria has become very attractive in an era of increasing environmental concern (Keshavarz and Roy, 2010). A number of different recombinant *E. coli* strains designed for polymer accumulation have been constructed thus far (Li et al., 2007; Chen et al., 2013; Ruiz et al., 2013), outsourcing the *pha* genes from several bacteria (Verlinden et al., 2007). However, most of the PHB production systems

available thus far suffer from a number of drawbacks (Wang et al., 2014). Among them, the transcriptional regulation of the *pha* genes is of particular importance. In natural producer bacteria, PHB accumulation is triggered by an imbalance in the availability of critical nutrients (e.g., the N or S source; Anderson and Dawes, 1990). In recombinant *E. coli*, however, the constitutive expression of the *pha* genes leads to a growth-dependent accumulation of PHB, which normally results in metabolic burden in the producing cells (Wang and Lee, 1997). Controlling the rate of polymer accumulation in recombinant *E. coli* is thus of paramount importance for the design of efficient microbial cell factories. Besides this feature, the segregational stability of plasmids in *E. coli* recombinants could also be an issue in prolonged fermentation processes aimed at biopolymer production (Nikel et al., 2010b; Ruiz et al., 2013).

To overcome this state of affairs, we decided to explore the landscape of potentially useful transcription levels in *E. coli* by randomly integrating the *pha* genes from *C. necator* into the chromosome. To this end, we used vector pBAMD1-6 as the backbone to clone a 5.3-kb DNA fragment spanning the *phaC1AB1* genes. Plasmid pAeT41 (Peoples and Sinskey, 1989) was digested with *EcoRI* and *SmaI* to liberate the aforementioned DNA segment, and inserted into the corresponding restriction sites of pBAMD1-6 to generate plasmid pBAMD1-6-*pha* (**Table 1**). The functional parts of the DNA element to be transferred into *E. coli* are shown in **Figure 3B**.

As acetyl-CoA is the precursor metabolite for PHB formation, competing pathways that use this intermediate are expected to drain building blocks of the biopolymer synthesis. In *E. coli*, the acetate formation pathway, comprising Pta and AckA (phosphotransacetylase and acetate kinase), uses acetyl-CoA as the starting metabolite (Neidhardt et al., 1990). This pathway, which diverts a considerable amount of carbon from the central metabolism (Wolfe, 2005), is active under both oxic and anoxic conditions (Clark, 1989). For this reason, the *phaBAC* gene cluster was

delivered into a Δpta recipient strain (Table 1), in which the acetate formation is expected to be low. Plasmid pBAMD1-6-*pha* was first transferred to *E. coli* S17-1 λpir to perform a biparental mating. *E. coli* S17-1 λpir or SM10 λpir are the preferred donor strains when mobilizing RP4-based plasmid to *E. coli* recipient strains since they bear the functional *tra* and *mob* elements integrated in the genome, thus avoiding the inadvertent transfer of the mating-helper plasmid (pRK600) alongside the mini-transposon delivery system. Alternatively, *E. coli* strain MFD *pir*⁺ (Ferrières et al., 2010), devoid of the Mu element present in the strains detailed above, could be used for transposon insertions.

After integration of the mini-Tn5:*phaC1AB1* device into *E. coli* JW2293-1, individual Gm^R colonies were purified and separately grown in microtiter plates as explained in the Section “Materials and Methods” to explore PHB accumulation in 100 independent transconjugants after 24 h of incubation. Figure 3C shows the level of PHB accumulation in the transconjugants as compared to that of *E. coli* JW2293-1 carrying the pAeT41 plasmid, in which the expression of the *phaC1AB1* gene cluster is driven by the native promoter. All the transconjugants tested accumulated the polymer from mono-copy chromosomal insertions of the PHB biosynthesis pathway to some extent, ranging from 3% up to 78% of the accumulation levels observed in *E. coli* JW2293P (which carries plasmid pAeT41). Table S2 in the Supplementary Material shows the insertion locus for some selected *phaC1AB1*⁺ transconjugants, clearly illustrating the non-selective nature of the chromosomal incorporation of Tn5 (and consequently, the wide range of PHB accumulation levels), as it was already observed in *P. putida* transconjugants. We selected one of the transconjugants, termed *E. coli* TA2293P, that accumulated high polymer levels (clone 1 in Figure 3C and Table S2 in Supplementary Material), and the insertion site of the transposon was determined to be *ykgH*, an open reading frame encoding a predicted inner membrane protein (Table S2 in the Supplementary Material). The physiology of PHB accumulation of this strain was studied as detailed below.

Physiological and biochemical characterization of *E. coli* TA2293P as a microbial cell factory for PHB synthesis

The growth parameters of several *E. coli* strains were determined to explore their potential as biopolymer cell factories (Table 4).

We decided to compare the performance of strains in which the *pha* genes are expressed either in a multi-copy plasmid or as a mono-copy insertion in the bacterial chromosome side-by-side. Interestingly, the elimination of Pta resulted in a reduction of the specific growth rate, probably by an imbalance in the acetyl-CoA pool that was partially restored by the heterologous expression of the *phaC1AB1* gene cluster. This positive effect was more evident in the strain in which the genes were inserted into the chromosome (the specific growth rate attained ca. 85% of that in the wild-type strain), thus suggesting that the adequate expression level of the PHB biosynthesis genes is important to recover a homeostatic acetyl-CoA balance. This kinetic pattern was also mirrored in the final biomass density of the cultures. In fact, among the strains tested, *E. coli* TA2293P attained the highest cell density. This result highlights the advantage of integrating the *pha* genes in the chromosome, as such approach not only avoids the metabolic burden usually associated with heterologous gene expression from plasmids and other extra-chromosomal elements but it also allows for the selection of an integrant strain exhibiting the appropriate level of transcription of the corresponding genes. Moreover, the selection of *ykgH* as a target was not at all obvious, indicating, again, the value of random insertion of the genes of interest in the bacterial chromosome.

Since we selected a *pta* mutant of *E. coli* as the recipient strain, in which acetate formation is expected to be impaired, by-product formation was also explored in these cultures as a measure of the carbon flow from glucose to PHB (Table 4). In cultures of the wild-type strain, up to 60% of the total carbon source was converted into acetate, pinpointing this metabolite as the key by-product of hexose catabolism in *E. coli* (and therefore, as the main side pathway competing for acetyl-CoA). The *pta* mutant still produced some acetate (probably through the action of pyruvate oxidase, PoxB); however, the molar conversion of glucose into acetate reached only ca. 20% of that observed in the wild-type strain. Acetate formation in both *E. coli* strains carrying *PhaC1AB1* was comparable, and much lower than the other two strains. In all, these results bear witness of (i) the suitability of a *pta* mutant, deficient in acetate formation, as the starting point to construct a PHB cell factory, and (ii) the effect of the PHB biosynthetic pathway in using acetyl-CoA as the precursor metabolite. Once the coarse physiological

Table 4 | Physiological characterization of wild-type and mutant *E. coli* strains as microbial cell factories for PHB biosynthesis in shaken-flask cultures.

<i>E. coli</i> strain	Relevant characteristics	Physiological parameter ^a		
		μ (h ⁻¹)	CDW (g l ⁻¹)	$Y_{A/S}$ (mol mol ⁻¹)
BW25113	Wild-type strain	0.74 ± 0.04	3.8 ± 0.2	0.64 ± 0.05
JW2293-1	Δpta	0.44 ± 0.03	2.9 ± 0.3	0.13 ± 0.02
JW2293P	$\Delta pta phaC1AB1^+$	0.51 ± 0.07	3.5 ± 0.1	0.05 ± 0.02
TA2293P	$\Delta pta ykgH$:mini-Tn5(<i>phaC1AB1</i>)	0.63 ± 0.05	4.2 ± 0.4	0.09 ± 0.03

^a Cells were grown aerobically in M9 minimal medium containing 30 g l⁻¹ glucose as the sole carbon source. The specific growth rate (microns) was determined during logarithmic growth, whereas the final cell density (expressed as the cell dry weight, CDW) and the molar yield of acetate on glucose ($Y_{A/S}$) were calculated after 24 h of incubation. Reported results represent the mean value ± SD of triplicate measurements from at least two independent cultures.

characterization of these strains was completed, the next relevant question was how they perform as PHB producers.

Since PhaA is the first committed enzymatic step of PHB formation from acetyl-CoA, the *in vitro* activity of this enzyme was assayed as a proxy of the activity of the whole biosynthetic pathway in shaken-flask cultures using glucose as the sole carbon source (Figure 4A). Note that there was some degree of thiolase activity in both *E. coli* BW25113 and JW2293-1, probably represented by FadA (an enzyme normally involved in the degradation of fatty acids via the β -oxidation cycle). However, the specific PhaA activity was six and fourfold higher in the strains carrying the *phaC1AB1* gene cluster (*E. coli* JW2293P and TA2293P, respectively) as compared to that of *E. coli* JW2293-1. As expected, the highest enzymatic activity corresponded to the strain carrying the *pha* genes in a multi-copy plasmid. Nevertheless, *E. coli* TA2293P had an activity level *ca.* 60% of the strain bearing pAeT41, indicating that the appropriate insertion of the gene cluster could result in thiolase activities similar to those of a typical recombinant PHB producer. Yet, how do these activities translate into PHB accumulation?

Figure 4B shows that *E. coli* TA2293P accumulated PHB up to 62% of the level observed in the same strain but expressing the *pha* genes in a plasmid (i.e., *E. coli* JW2293P). As previously observed in other traits during the physiological characterization, the later strain had the highest PHB accumulation level among all the strains tested. That *E. coli* TA2293P accumulates such a high amount of PHB is a somewhat surprising (and welcome) result considering the difference in copy number between the two *phaC1AB1*⁺ strains under comparison. Note that a possible effect of the absence (or an altered expression level) of YkgH on the properties of *E. coli* TA2293P cannot be completely ruled out. Interestingly, when this strain was persistently cultured in LB medium without any selective pressure, its phenotypic traits, particularly regarding polymer accumulation, remained unchanged. By contrast, the segregational stability of pAeT41 was assessed in cultures of *E. coli* JW2293P, and, after growing the cells in LB medium and sub-culturing them daily seven times without any antibiotic, <25% of the cells were resistant to ampicillin.

CONCLUSION

In our current study, we presented a set of new mini-Tn5-derived vectors that can be used to engineer the genome of Gram-negative bacteria. While the worth of these tools has been exposed in two model bacteria, *P. putida* and *E. coli*, the inherent promiscuity of Tn5 ensures its functioning in a number of different microbial hosts. Of particular importance is the possibility of sequentially using the three pBAMD1-*x* vectors, thereby enabling the user to accumulate insertions in the same genetic background in a combinatorial fashion. This is a particularly interesting feature for the construction of complex phenotypes, such as biopolymer formation, which depend on more than one enzyme. Although the expression of the biosynthetic genes in a multi-copy plasmid is in principle enough to bestow the desired phenotype on the recipient bacterium, fine-tuned expression levels (together with the appearance of emergent phenotypic properties in the host, brought about by the insertion process itself) can be easily achieved by randomly

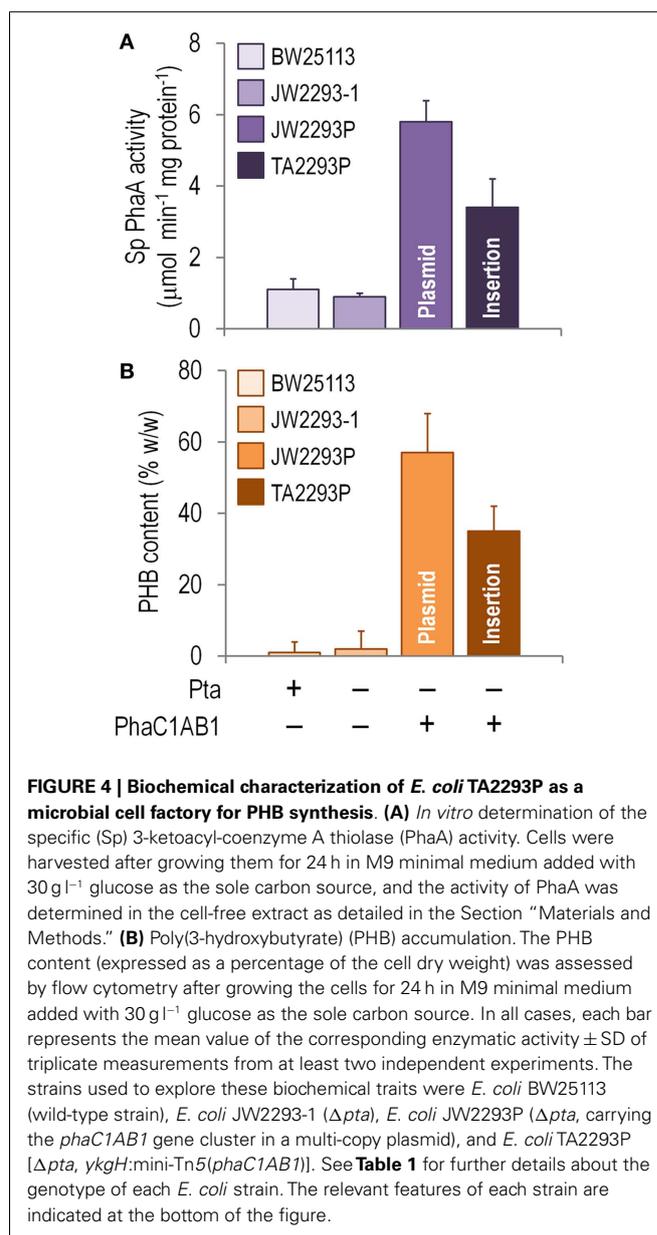


FIGURE 4 | Biochemical characterization of *E. coli* TA2293P as a microbial cell factory for PHB synthesis. (A) *In vitro* determination of the specific (Sp) 3-ketoacyl-coenzyme A thiolase (PhaA) activity. Cells were harvested after growing them for 24 h in M9 minimal medium added with 30 g l⁻¹ glucose as the sole carbon source, and the activity of PhaA was determined in the cell-free extract as detailed in the Section “Materials and Methods.” **(B)** Poly(3-hydroxybutyrate) (PHB) accumulation. The PHB content (expressed as a percentage of the cell dry weight) was assessed by flow cytometry after growing the cells for 24 h in M9 minimal medium added with 30 g l⁻¹ glucose as the sole carbon source. In all cases, each bar represents the mean value of the corresponding enzymatic activity \pm SD of triplicate measurements from at least two independent experiments. The strains used to explore these biochemical traits were *E. coli* BW25113 (wild-type strain), *E. coli* JW2293-1 (Δpta), *E. coli* JW2293P (Δpta , carrying the *phaC1AB1* gene cluster in a multi-copy plasmid), and *E. coli* TA2293P [Δpta , *ykgH*:mini-Tn5(*phaC1AB1*)]. See Table 1 for further details about the genotype of each *E. coli* strain. The relevant features of each strain are indicated at the bottom of the figure.

integrating the structural genes into the chromosome. In this way, and since the genetic context of each integration will surely result in different regulatory patterns at the transcriptional level, the user could choose among a library of insertions those clones that meet any desired criterion (in our case, PHB accumulation). Moreover, as the delivery plasmids described in this study can be used in a sequential manner, other polymer-associated enzymes, such as phasins, can also be incorporated in the same strain to enhance further polymer production.

ACKNOWLEDGMENTS

We are indebted to Prof. A. Sinskey (Massachusetts Institute of Technology) for sharing research materials and to A. Goñi (CNB-CSIC) for his help in image processing. I. Benedetti (CNB-CSIC)

is gratefully acknowledged for her help in FACS measurements. This study was supported by the BIO Program of the Spanish Ministry of Economy and Competitiveness, the ST-FLOW and ARISYS Contracts of the EU, the ERANET-IB Program, and the PROMT Project of the CAM to VDL. PIN is a researcher from the Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina) and holds a Marie Curie Actions Program grant from the EC (ALLEGRO, UE-FP7-PEOPLE-2011-IIF-300508). The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fbioe.2014.00046/abstract>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389
- Anderson, A. J., and Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol. Rev.* 54, 450–472.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008. doi:10.1038/msb4100050
- Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F. C., Bagdasarian, M. M., Frey, J., et al. (1981). Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* 16, 237–247. doi:10.1016/0378-1119(81)90080-9
- Berg, D. E. (1989). “Transposon Tn5,” in *Mobile DNA*, eds D. E. Berg and M. M. Howe (Washington, DC: American Society for Microbiology Press), 185–210.
- Boyer, H. W., and Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41, 459–472. doi:10.1016/0022-2836(69)90288-5
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi:10.1016/0003-2697(76)90527-3
- Brosius, J., Cate, R. L., and Perlmutter, A. P. (1982). Precise location of two promoters for the β -lactamase gene of pBR322. S1 mapping of ribonucleic acid isolated from *Escherichia coli* or synthesized *in vitro*. *J. Biol. Chem.* 257, 9205–9210.
- Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A., et al. (2013). Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* 31, 1200–1223. doi:10.1016/j.biotechadv.2013.02.009
- Clark, D. P. (1989). The fermentation pathways of *Escherichia coli*. *FEMS Microbiol. Rev.* 5, 223–234. doi:10.1111/j.1574-6968.1989.tb03398.x
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi:10.1101/gr.849004
- Das, S., Noe, J. C., Paik, S., and Kitten, T. (2005). An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *J. Microbiol. Methods* 63, 89–94. doi:10.1016/j.mimet.2005.02.011
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi:10.1073/pnas.120163297
- de Lorenzo, V., Cases, I., Herrero, M., and Timmis, K. N. (1993). Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in *Pseudomonas putida* with *lacZ-tet* bicistronic reporters. *J. Bacteriol.* 175, 6902–6907.
- de Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K. N. (1990). Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in Gram-negative Eubacteria. *J. Bacteriol.* 172, 6568–6572.
- de Lorenzo, V., Herrero, M., Sánchez, J. M., and Timmis, K. N. (1998). Mini-transposons in microbial ecology and environmental biotechnology. *FEMS Microbiol. Ecol.* 27, 211–224. doi:10.1111/j.1574-6941.1998.tb00538.x
- de Lorenzo, V., and Timmis, K. N. (1994). Analysis and construction of stable phenotypes in Gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Meth. Enzymol.* 235, 386–405. doi:10.1016/0076-6879(94)35157-0
- Durante-Rodríguez, G., de Lorenzo, V., and Martínez-García, E. (2014). The Standard European Vector Architecture (SEVA) plasmid toolkit. *Methods Mol. Biol.* 1149, 469–478. doi:10.1007/978-1-4939-0473-0_36
- Ferrières, L., Hémerly, G., Nham, T., Guérout, A. M., Mazel, D., Beloin, C., et al. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* 192, 6418–6427. doi:10.1128/JB.00621-10
- Fling, M. E., Kopf, J., and Richards, C. (1985). Nucleotide sequence of the transposon Tn7 gene encoding an aminoglycoside-modifying enzyme, 3⁹(9)-O-nucleotidyltransferase. *Nucleic Acids Res.* 13, 7095–7106. doi:10.1093/nar/13.19.7095
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318
- Gorski, K., Roch, J. M., Prentki, P., and Krisch, H. M. (1985). The stability of bacteriophage T4 gene 32 mRNA: a 5' leader sequence that can stabilize mRNA transcripts. *Cell* 43, 461–469. doi:10.1016/0092-8674(85)90176-X
- Hanahan, D., and Meselson, M. (1983). Plasmid screening at high colony density. *Meth. Enzymol.* 100, 333–342. doi:10.1016/0076-6879(83)00066-X
- Herrero, M., de Lorenzo, V., and Timmis, K. N. (1990). Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in Gram-negative bacteria. *J. Bacteriol.* 172, 6557–6567.
- Keshavarz, T., and Roy, I. (2010). Polyhydroxyalkanoates: bioplastics with a green agenda. *Curr. Opin. Microbiol.* 13, 321–326. doi:10.1016/j.mib.2010.02.006
- Kessler, B., de Lorenzo, V., and Timmis, K. N. (1992). A general system to integrate *lacZ* fusions into the chromosomes of Gram-negative Eubacteria: regulation of the *Pm* promoter of the TOL plasmid studied with all controlling elements in monocopy. *Mol. Gen. Genet.* 233, 293–301. doi:10.1007/BF00587591
- Kolter, R., Inuzuka, M., and Helinski, D. R. (1978). *trans*-Complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* 15, 1199–1208. doi:10.1016/0092-8674(78)90046-6
- Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., et al. (1995). Four new derivatives of the broad-host-range cloning vector pBRR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176. doi:10.1016/0378-1119(95)00584-1
- Li, R., Zhang, H., and Qi, Q. (2007). The production of polyhydroxyalkanoates in recombinant *Escherichia coli*. *Bioresour. Technol.* 98, 2313–2320. doi:10.1016/j.biortech.2006.09.014
- Lodge, J. K., Weston-Hafer, K., and Berg, D. E. (1988). Transposon Tn5 target specificity: preference for insertion at G/C pairs. *Genetics* 120, 645–650.
- Lyras, D., and Rood, J. I. (1998). Conjugative transfer of RP4-*oriT* shuttle vectors from *Escherichia coli* to *Clostridium perfringens*. *Plasmid* 39, 160–164. doi:10.1006/plas.1997.1325
- Martínez-García, E., Calles, B., Arévalo-Rodríguez, M., and de Lorenzo, V. (2011). pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* 11:38. doi:10.1186/1471-2180-11-38
- Martínez-García, E., and de Lorenzo, V. (2012). Transposon-based and plasmid-based genetic tools for editing genomes of Gram-negative bacteria. *Methods Mol. Biol.* 813, 267–283. doi:10.1007/978-1-61779-412-4_16
- Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Rüger, W. (2003). Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86–156. doi:10.1128/MMBR.67.1.86-156.2003
- Miller, V. L., and Mekalanos, J. J. (1988). A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. *J. Bacteriol.* 170, 2575–2583.
- Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). *Physiology of the Bacterial Cell: A Molecular Approach*. Sunderland, MA: Sinauer Associates.
- Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A. P., et al. (2002). Complete genome sequence and comparative analysis

- of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* 4, 799–808. doi:10.1046/j.1462-2920.2002.00366.x
- Nikel, P. I., de Almeida, A., Pettinari, M. J., and Méndez, B. S. (2008a). The legacy of HfrH: mutations in the two-component system CreBC are responsible for the unusual phenotype of an *Escherichia coli* *arcA* mutant. *J. Bacteriol.* 190, 3404–3407. doi:10.1128/JB.00040-08
- Nikel, P. I., Pettinari, M. J., Galvagno, M. A., and Méndez, B. S. (2008b). Poly(3-hydroxybutyrate) synthesis from glycerol by a recombinant *Escherichia coli* *arcA* mutant in fed-batch microaerobic cultures. *Appl. Microbiol. Biotechnol.* 77, 1337–1343. doi:10.1007/s00253-007-1255-7
- Nikel, P. I., and de Lorenzo, V. (2013a). Implantation of unmarked regulatory and metabolic modules in Gram-negative bacteria with specialised mini-transposon delivery vectors. *J. Biotechnol.* 163, 143–154. doi:10.1016/j.jbiotec.2012.05.002
- Nikel, P. I., and de Lorenzo, V. (2013b). Engineering an anaerobic metabolic regime in *Pseudomonas putida* KT2440 for the anoxic biodegradation of 1,3-dichloroprop-1-ene. *Metab. Eng.* 15, 98–112. doi:10.1016/j.ymben.2012.09.006
- Nikel, P. I., Giordano, A. M., de Almeida, A., Godoy, M. S., and Pettinari, M. J. (2010a). Elimination of D-lactate synthesis increases poly(3-hydroxybutyrate) and ethanol synthesis from glycerol and affects cofactor distribution in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 76, 7400–7406. doi:10.1128/AEM.02067-10
- Nikel, P. I., Pettinari, M. J., Galvagno, M. A., and Méndez, B. S. (2010b). Metabolic selective pressure stabilizes plasmids carrying biosynthetic genes for reduced biochemicals in *Escherichia coli* redox mutants. *Appl. Microbiol. Biotechnol.* 88, 563–573. doi:10.1007/s00253-010-2774-1
- Nikel, P. I., Martínez-García, E., and de Lorenzo, V. (2014a). Biotechnological domestication of pseudomonads using synthetic biology. *Nat. Rev. Microbiol.* 12, 368–379. doi:10.1038/nrmicro3253
- Nikel, P. I., Kim, J., and de Lorenzo, V. (2014b). Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440. *Environ. Microbiol.* 16, 239–254. doi:10.1111/1462-2920.12224
- Palmer, M. A. J., Differding, E., Gamboni, R., Williams, S. F., Peoples, O. P., Walsh, C. T., et al. (1991). Biosynthetic thiolase from *Zoogloea ramigera*. Evidence for a mechanism involving Cys-378 as the active site base. *J. Biol. Chem.* 266, 8369–8375.
- Peoples, O. P., and Sinskey, A. J. (1989). Poly-β-hydroxybutyrate (PHB) biosynthesis in *Alcaligenes eutrophus* H16. Identification and characterization of the PHB polymerase gene (*phbC*). *J. Biol. Chem.* 264, 15298–15303.
- Phadnis, S. H., Sasakawa, C., and Berg, D. E. (1986). Localization of action of the IS50-encoded transposase protein. *Genetics* 112, 421–427.
- Pratt, L. A., and Kolter, R. (1998). Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* 30, 285–293. doi:10.1046/j.1365-2958.1998.01061.x
- Reznikoff, W. S. (2006). Tn5 transposition: a molecular tool for studying protein structure-function. *Biochem. Soc. Trans.* 34, 320–323. doi:10.1042/BST20060320
- Reznikoff, W. S. (2008). Transposon Tn5. *Annu. Rev. Genet.* 42, 269–286. doi:10.1146/annurev.genet.42.110807.091656
- Ruiz, J. A., de Almeida, A., Godoy, M. S., Mezzina, M. P., Bidart, G. N., Méndez, B. S., et al. (2013). *Escherichia coli* redox mutants as microbial cell factories for the synthesis of reduced biochemicals. *Comput. Struct. Biotechnol. J.* 3, e201210019. doi:10.5936/CSBJ.201210019
- Sambrook, J., Maniatis, T., and Fritsch, E. F. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de Las Heras, 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. doi:10.1093/nar/gks1119
- Simcox, T. G., Fabian, L., Kretz, K., Hedden, V., and Simcox, M. E. (1995). *SamDI*, a new type-II restriction endonuclease that recognizes 5'-GG/GWCCC-3'. *Gene* 155, 129–130. doi:10.1016/0378-1119(94)00711-Z
- Slater, S., Houmiel, K. L., Tran, M., Mitsky, T. A., Taylor, N. B., Padgett, S. R., et al. (1998). Multiple β-ketothiolases mediate poly(β-hydroxyalkanoate) copolymer synthesis in *Ralstonia eutropha*. *J. Bacteriol.* 180, 1979–1987.
- Spiekermann, P., Rehm, B. H., Kalscheuer, R., Baumeister, D., and Steinbüchel, A. (1999). A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. *Arch. Microbiol.* 171, 73–80. doi:10.1007/s002030050681
- Steinbüchel, A., and Hein, S. (2001). Biochemical and molecular basis of microbial synthesis of polyhydroxyalkanoates in microorganisms. *Adv. Biochem. Eng. Biotechnol.* 71, 81–123. doi:10.1007/3-540-40021-4_3
- Tyo, K. E., Zhou, H., and Stephanopoulos, G. N. (2006). High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. strain PCC6803. *Appl. Environ. Microbiol.* 72, 3412–3417. doi:10.1128/AEM.72.5.3412-3417.2006
- Verlinden, R. A., Hill, D. J., Kenward, M. A., Williams, C. D., and Radecka, I. (2007). Bacterial synthesis of biodegradable polyhydroxyalkanoates. *J. Appl. Microbiol.* 102, 1437–1449. doi:10.1111/j.1365-2672.2007.03335.x
- Wang, F., and Lee, S. Y. (1997). Production of poly(3-hydroxybutyrate) by fed-batch culture of filamentation-suppressed recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 63, 4765–4769.
- Wang, Y., Yin, J., and Chen, G. Q. (2014). Polyhydroxyalkanoates, challenges and opportunities. *Curr. Opin. Biotechnol.* 30, 59–65. doi:10.1016/j.copbio.2014.06.001
- Winsor, G. L., Lam, D. K. W., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., et al. (2011). *Pseudomonas* genome database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res.* 39, D596–D600. doi:10.1093/nar/gkq869
- Wolfe, A. J. (2005). The acetate switch. *Microbiol. Mol. Biol. Rev.* 69, 12–50. doi:10.1128/MMBR.69.1.12-50.2005
- Yarnell, W. S., and Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science* 284, 611–615. doi:10.1126/science.284.5414.611
- Zhou, M., Bhasin, A., and Reznikoff, W. S. (1998). Molecular genetic analysis of transposase-end DNA sequence recognition: cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase. *J. Mol. Biol.* 276, 913–925. doi:10.1006/jmbi.1997.1579

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 August 2014; accepted: 13 October 2014; published online: 28 October 2014.

Citation: Martínez-García E, Aparicio T, de Lorenzo V and Nikel PI (2014) New transposon tools tailored for metabolic engineering of Gram-negative microbial cell factories. *Front. Bioeng. Biotechnol.* 2:46. doi: 10.3389/fbio.2014.00046

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Martínez-García, Aparicio, de Lorenzo and Nikel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ADVANTAGES OF PUBLISHING IN FRONTIERS



FAST PUBLICATION

Average 90 days
from submission
to publication



COLLABORATIVE PEER-REVIEW

Designed to be rigorous –
yet also collaborative, fair and
constructive



RESEARCH NETWORK

Our network
increases readership
for your article



OPEN ACCESS

Articles are free to read,
for greatest visibility



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



GLOBAL SPREAD

Six million monthly
page views worldwide



COPYRIGHT TO AUTHORS

No limit to
article distribution
and re-use



IMPACT METRICS

Advanced metrics
track your
article's impact



SUPPORT

By our Swiss-based
editorial team