



ENGINEERING SYNTHETIC METABOLONS: FROM METABOLIC MODELLING TO RATIONAL DESIGN OF BIOSYNTHETIC DEVICES

EDITED BY : Lars M. Voll and Zoran Nikoloski

PUBLISHED IN: Frontiers in Bioengineering and Biotechnology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 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 978-2-88919-921-1

DOI 10.3389/978-2-88919-921-1

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?

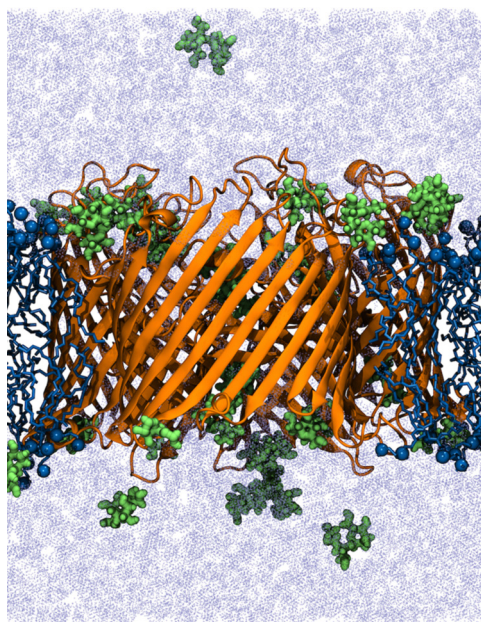
Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most 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

ENGINEERING SYNTHETIC METABOLONS: FROM METABOLIC MODELLING TO RATIONAL DESIGN OF BIOSYNTHETIC DEVICES

Topic Editors:

Lars M. Voll, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Zoran Nikoloski, Max-Planck-Institute for Molecular Plant Physiology, Germany



Side view of an ScrY trimer (orange ribbons) embedded in a phospholipid membrane (blue spheres and sticks) with the periplasmic side facing downward. The substrate of this porin, sucrose, is depicted as green stick-and-balls, water molecules are represented by blue dots.

Image adapted from Sun et al. (2016), DOI: 10.3389/fbioe.2016.00009.

The discipline of Synthetic Biology has recently emerged at the interface of biology and engineering. The definition of Synthetic Biology has been dynamic over time ever since, which exemplifies that the field is rapidly moving and comprises a broad range of research areas.

In the frame of this Research Topic, we focus on Synthetic Biology approaches that aim at rearranging biological parts/ entities in order to generate novel biochemical functions with inherent metabolic activity. This Research Topic encompasses Pathway Engineering in living systems as well as the *in vitro* assembly of biomolecules into nano- and microscale bioreactors.

Both, the engineering of metabolic pathways *in vivo*, as well as the conceptualization of bioreactors *in vitro*, require rational design of assembled synthetic pathways and depend on careful selection of individual biological functions and their optimization. Mathematical modelling has proven to be a powerful tool in predicting metabolic flux in living and artificial systems, although modelling approaches have to cope with a limitation in experimentally verified, reliable

input variables. This Research Topic puts special emphasis on the vital role of modelling approaches for Synthetic Biology, i.e. the predictive power of mathematical simulations for (i) the manipulation of existing pathways and (ii) the establishment of novel pathways in vivo as well as (iii) the translation of model predictions into the design of synthetic assemblies.

Citation: Voll, L. M., Nikoloski, Z., eds. (2016). Engineering Synthetic Metabolons: From Metabolic Modelling to Rational Design of Biosynthetic Devices. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-921-1

Table of Contents

06	<i>Editorial: Engineering Synthetic Metabolons: From Metabolic Modeling to Rational Design of Biosynthetic Devices</i>
	Lars M. Voll and Zoran Nikoloski
	Chapter 1: Molecular Dynamics Simulations at Biomembranes
08	<i>Building synthetic sterols computationally – unlocking the secrets of evolution?</i>
	Tomasz Róg, Sanja Pöyry and Ilpo Vattulainen
14	<i>Characteristics of Sucrose Transport through the Sucrose-Specific Porin ScrY Studied by Molecular Dynamics Simulations</i>
	Liping Sun, Franziska Bertelshofer, Günther Greiner and Rainer A. Böckmann
24	<i>GroPBS: Fast Solver for Implicit Electrostatics of Biomolecules</i>
	Franziska Bertelshofer, Liping Sun, Günther Greiner and Rainer A. Böckmann
	Chapter 2: From Mathematical Modelling to Biochemical Microreactors
32	<i>Model-Based Design of Biochemical Microreactors</i>
	Tobias Elbinger, Markus Gahn, Maria Neuss-Radu, Falk M. Hante, Lars M. Voll, Günter Leugering and Peter Knabner
45	<i>Underpinning starch biology with in vitro studies on carbohydrate-active enzymes and biosynthetic glycomaterials</i>
	Ellis C. O'Neill and Robert A. Field
51	<i>Compartmentalization and Transport in Synthetic Vesicles</i>
	Christine Schmitt, Anna H. Lippert, Navid Bonakdar, Vahid Sandoghdar and Lars M. Voll
	Chapter 3: Metabolic Modelling and Metabolic Engineering in Plants
63	<i>Metabolomics, standards, and metabolic modeling for synthetic biology in plants</i>
	Camilla Beate Hill, Tobias Czauderna, Matthias Klapperstück, Ute Roessner and Falk Schreiber
75	<i>Photorespiratory Bypasses Lead to Increased Growth in Arabidopsis thaliana: Are Predictions Consistent with Experimental Evidence?</i>
	Georg Basler, Anika Küken, Alisdair R. Fernie and Zoran Nikoloski
86	<i>Optimization of Engineered Production of the Glucoraphanin Precursor Dihomomethionine in Nicotiana benthamiana</i>
	Christoph Crocoll, Nadia Mirza, Michael Reichelt, Jonathan Gershenzon and Barbara Ann Halkier

Chapter 4: Protein Engineering

95 *Synthetic Peptides as Protein Mimics*

Andrea Groß, Chie Hashimoto, Heinrich Sticht and Jutta Eichler

111 *Synthetic Protein Scaffolds Based on Peptide Motifs and Cognate Adaptor Domains for Improving Metabolic Productivity*

Anselm H. C. Horn and Heinrich Sticht

118 *Engineering of metabolic pathways by artificial enzyme channels*

Marlene Pröschel, Rainer Detsch, Aldo R. Boccaccini and Uwe Sonnewald



Editorial: Engineering Synthetic Metabolons: From Metabolic Modeling to Rational Design of Biosynthetic Devices

Lars M. Voll^{1*} and Zoran Nikoloski²

¹Division of Biochemistry, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ²Max-Planck-Institute for Molecular Plant Physiology, Potsdam-Golm, Germany

Keywords: synthetic biology, pathway engineering, bioreactors, model-based design, computational models, molecular dynamics modeling

The Editorial on the Research Topic

Engineering synthetic metabolons: from metabolic modelling to rational design of biosynthetic devices

OPEN ACCESS

Edited by:

Zhanglin Lin,
Tsinghua University, China

Reviewed by:

Yinjie Tang,
Washington University, USA

*Correspondence:

Lars M. Voll
lars.voll@fau.de

Specialty section:

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

Received: 17 March 2016

Accepted: 21 April 2016

Published: 06 May 2016

Citation:

Voll LM and Nikoloski Z (2016)
Editorial: Engineering Synthetic
Metabolons: From Metabolic
Modeling to Rational Design of
Biosynthetic Devices.
Front. Bioeng. Biotechnol. 4:39.
doi: 10.3389/fbioe.2016.00039

Synthetic Biology is an emerging discipline that enjoys rapidly increasing popularity. However, there is no widely accepted definition for the term *Synthetic Biology*, because the field of Synthetic Biology is multifaceted and rapidly moving. As viewed from its underlying principle, Synthetic Biology can be defined as the “application of engineering principles to biological systems.” It reaches from pathway engineering in living systems, i.e., the introduction of functional biochemical pathways into existing organisms (see featured research topic review by Pröschel et al. and references cited therein), and the design of vesicle-based multicompartimented biochemical reactors and protocells (see featured research topic review by Schmitt et al. and references cited therein) to the creation of entirely synthetic cells with reproductive potential that is encoded by synthetic genes [with the work of Gibson et al. (2008, 2010) that attracted the greatest publicity in the recent years]. Likewise, the definition provided at <http://syntheticbiology.org> covers these two extremes in that Synthetic Biology is defined as “(A) the design and construction of new biological parts, devices, and systems and (B) the re-design of existing, natural biological systems for useful purposes.” This also means that Synthetic Biology is not limited to living systems but also comprises the combination and re-design of biological parts into artificial bioreactors. In this respect, Synthetic Biology is highly interdisciplinary, bringing together molecular biology with biophysics, material sciences, bioengineering, and computational approaches.

Besides the invaluable role in helping biochemists understand the metabolic complexity of living systems and the extent to which observations from existing metabolic engineering strategies match predictions from large-scale modeling (see, e.g., the featured research topic article by Basler et al.), models are crucial for the rational design of synthetic systems (see, e.g., the featured research topic article by Elbinger et al.). Metabolic modeling has the capacity to guide the biochemical layout of engineered pathways *in vivo* and the design of artificial bioreactors *in vitro*. For instance, computational modeling can play an important role for the selection of individual enzyme isoforms for engineered metabolic pathways. *Vice versa*, *in vitro* studies in artificial bioreactors has advanced our understanding of biochemically complex processes, such as starch biosynthesis, as summarized in the featured research topic review by O'Neill and Field. On the other hand, molecular dynamics simulations help to understand complex biochemical and biophysical processes such as enzymatic catalysis, molecular interactions of biomolecules (see featured research topic reviews by Horn and Sticht

and Groß et al.), or membrane transport (see featured research topic articles by Róg et al., Bertelshofer et al. and Sun et al.). The knowledge gained by these simulations can be exploited for the structural and molecular design of synthetic systems, i.e., for the choice of artificial protein or nucleic acid scaffolds for the spatial modular organization of artificial systems (see featured research topic review by Pröschel et al. and references cited therein).

This research topic focuses on the importance of computational modeling for the rational design of synthetic systems on all these levels. Consequently, a wide variety of approaches is being covered in this issue, from pathway engineering in eukaryotic cells to molecular dynamics simulations of transport processes. The modeling approaches range from partial differential equations, allowing predictions of spatiotemporal concentration of metabolites in a biochemical microreactor (see featured research topic original research article by Elbinger et al.), to steady-state equations, testing the effect of metabolic engineering strategy in large-scale metabolic networks (see, e.g., the featured research topic article by Basler et al.). A new metabolic engineering strategy for the glucosinolate pathway that increases dihomomethionine levels without increasing the levels of leucine-derived side products was experimentally validated in *Nicotiana benthamiana*

(see featured research topic original research article by Crocoll et al. and references to previous attempts therein). Sharing of modeling results and ensuring reproducibility of model-data integration requires setting and following a set of standards, which are currently established in the context of metabolomics research (see featured research topic review by Hill et al.).

The majority of the articles covered in this research topic article is contributed by members of the interdisciplinary project SynBio that is funded by the emerging fields initiative of the Friedrich-Alexander-Universität Erlangen-Nürnberg (Germany).

AUTHOR CONTRIBUTIONS

LV and ZN have conceived and written the article.

FUNDING

The topic editor would like to gratefully acknowledge generous financial support by the Emerging Fields Initiative of the FAU Erlangen-Nürnberg in the framework of the EFI SynBio program and funding by the FAU open access fund.

REFERENCES

- 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. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi:10.1126/science.1190719

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 © 2016 Voll and Nikoloski. 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.

Building synthetic sterols computationally – unlocking the secrets of evolution?

Tomasz Róg¹, Sanja Pöyry¹ and Ilpo Vattulainen^{1,2*}

¹ Department of Physics, Tampere University of Technology, Tampere, Finland, ² MEMPHYS-Center for Biomembrane Physics, University of Southern Denmark, Odense, Denmark

OPEN ACCESS

Edited by:

Lars Matthias Voll,
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany

Reviewed by:

Rainer A. Böckmann,
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany
Olle Edholm,
KTH Royal Institute of Technology,
Sweden

*Correspondence:

Ilpo Vattulainen,
Department of Physics, Tampere
University of Technology, POB 692,
Tampere FI-33101, Finland
ilpo.vattulainen@tut.fi

Specialty section:

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

Received: 27 May 2015

Accepted: 07 August 2015

Published: 21 August 2015

Citation:

Róg T, Pöyry S and Vattulainen I
(2015) Building synthetic sterols
computationally – unlocking the
secrets of evolution?
Front. Bioeng. Biotechnol. 3:121.
doi: 10.3389/fbioe.2015.00121

Cholesterol is vital in regulating the physical properties of animal cell membranes. While it remains unclear what renders cholesterol so unique, it is known that other sterols are less capable in modulating membrane properties, and there are membrane proteins whose function is dependent on cholesterol. Practical applications of cholesterol include its use in liposomes in drug delivery and cosmetics, cholesterol-based detergents in membrane protein crystallography, its fluorescent analogs in studies of cholesterol transport in cells and tissues, etc. Clearly, in spite of their difficult synthesis, producing the synthetic analogs of cholesterol is of great commercial and scientific interest. In this article, we discuss how synthetic sterols non-existent in nature can be used to elucidate the roles of cholesterol's structural elements. To this end, we discuss recent atomistic molecular dynamics simulation studies that have predicted new synthetic sterols with properties comparable to those of cholesterol. We also discuss more recent experimental studies that have vindicated these predictions. The paper highlights the strength of computational simulations in making predictions for synthetic biology, thereby guiding experiments.

Keywords: cholesterol, synthetic sterol, computer simulation, molecular dynamics simulation

Why Synthetic Lipids and Sterols are Important?

As nature has designed thousands of lipid species, why then would we need synthetic lipids in addition? Clearly, however, the use of synthetic lipids is commonplace in both applied and basic sciences. The largest applications of synthetic lipids are in pharmacology, where synthetic lipids are used, e.g., in drug delivery and gene transfection. In drug delivery, the most commonly used carriers are liposomes, however, simple micelles or nanodiscs can be used as well. Technical requirements for the carriers include optimal lifetime, just-in-time triggered release of their contents, feasible targeting agents, etc. Numerous synthetic lipids have been synthesized and tested for this purpose [for a recent review, see Kohli et al. (2014)]. As in several other cases, here also atomistic molecular dynamics simulations have been used to unravel the physicochemical properties of these lipids [e.g., Bunker (2012)]. In gene transfection, one possible form of DNA packaging is the so-called *genosome*, commonly also called the *lipoplex*. Lipoplex is an aggregate of DNA and lipids; however, the cationic lipids needed to form this aggregate do not exist in nature. Consequently, only synthetic lipids can be used for this purpose.

Synthetic lipids have also numerous applications in basic research. Possibly, the most apparent example is labeling lipids with fluorescent or spin labels. For instance, cholesterol labeled with BOD-IPY or NBD has been used to study cholesterol trafficking in cells. Here also, molecular dynamics (MD) simulations have been used to examine the different behaviors of native and modified

molecules, thus complementing and explaining experiments (Hölttävuori et al., 2008; Robalo et al., 2013). Synthetic detergents like cholesteryl hemisuccinate are commonly used in G-protein coupled receptor crystallography, and again MD simulations have elucidated the differences between native and modified molecules (Kulig et al., 2014, 2015). More sophisticated applications of synthetic lipids include modifying the molecule's native structure by removing functional groups, in order to understand their individual function. Particularly, sphingolipids have been extensively studied in this manner (Slotte, 2013).

In this perspective article, we show an example of this last approach. The studies discussed in this article aimed at understanding the detailed structure–function relationships of cholesterol, in particular, the role of methyl groups attached to the steroid ring system. As we next explain in detail, these groups might with good reasons be thought of as unnecessary molecular fossils. However, as the below discussion highlights, extensive atomistic MD simulations showed that the methyl groups are indeed important parts of the cholesterol molecule, and the simulation results were later confirmed by experiments.

What is So Special About Cholesterol?

Cholesterol is a truly special molecule and absolutely vital for animals' wellbeing. This is probably best proved by the complete lack of mutations that would totally block the synthesis of cholesterol. Furthermore, some rare genetic syndromes caused by impaired cholesterol synthesis lead to serious conditions or death (Kelley and Herman, 2001). To ensure proper function, cholesterol needs a high degree of structural specificity. Indeed, cholesterol's precursors that have one additional double bond compared to cholesterol cannot substitute it independently, irrespective of whether the bond is located in the ring structure (7-dehydrocholesterol) or in the hydrocarbon tail (desmosterol) (Kelley and Herman, 2001). Highlighting its pivotal role, cholesterol is the single most common lipid species in our body. Its concentration in cell membranes varies from 30 to 50 mol% (van Meer et al., 2008), whereas in specialized membranes, such as the ocular lens (Mason et al., 2003), its concentration may reach 75 mol%. Ten percent of brain dry mass is cholesterol (Snipes and Suter, 1997). In the intracellular membranes, the concentration of cholesterol is lower but still typically 10–20 mol%. Deservedly, cholesterol is one of the most studied lipid molecules of all time.

Many of the various functions of cholesterol are related to modifying the structural properties of membranes. For example, cholesterol increases the mechanical strength of membranes, decreases their permeability, and affects membrane thickness and condensation [for reviews, see Ohvo-Rekila et al. (2002), Almeida (2009), and Róg and Vattulainen (2014)]. Presence of cholesterol alters the pressure profile across membranes (Ollila et al., 2007); this effect is sensitive to even small modifications in sterol structure. Cholesterol also modulates the phase behavior of lipid bilayers in a complex way (Ipsen et al., 1987; Vist and Davis, 1990). At larger cholesterol concentrations, a new phase called the *liquid ordered (Lo) phase* occurs, while at lower concentrations a *liquid disordered phase* is observed. Cholesterol is able to promote the formation of so-called *lipid rafts*, functional nanoscale domains

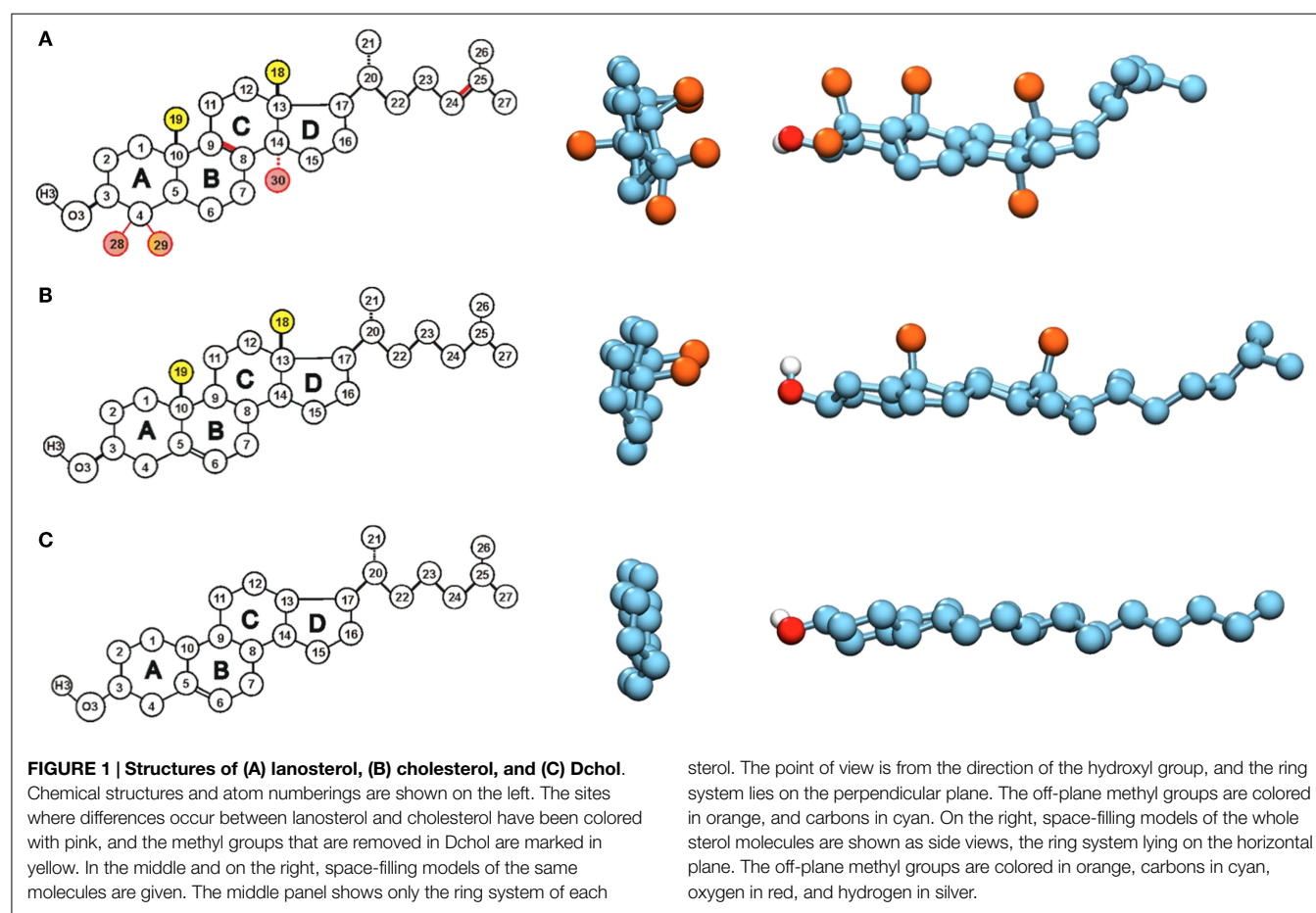
that are rich in cholesterol, sphingolipids, and saturated phospholipids (Lingwood and Simons, 2010), and numerous cellular functions, such as signaling and intracellular trafficking, actually depend on cholesterol (Coskun and Simons, 2011). Other cellular functions of cholesterol include its role as a metabolite and precursor of bile salts, some vitamins, and adrenal, pituitary, and sex (steroid) hormones.

All of the discussed points give rise to a picture of cholesterol having a very special and specific structure. Already during the seventies, cholesterol was established to be composed of three structural elements: a small hydroxyl head group, a rigid steroid ring system, and a short iso-octyl tail (Demel et al., 1972; Wenz, 2012). Modifications of these elements typically decrease the strength of cholesterol's effects on the physical properties of lipid bilayers and, as mentioned above, other sterols cannot substitute cholesterol in its biological function.

Does Cholesterol's Biosynthetic Pathway Reflect Molecular Evolution?

The biosynthesis of cholesterol is a complex process. The first sterol on the path is lanosterol (Figure 1), which is synthesized from squalene in a reaction that requires molecular oxygen. Consequently, the occurrence of this sterol can be located in the history of earth to a time after prokaryotic life had developed. Thus, perhaps not surprisingly, sterols are not typical bacterial lipids with the exception of *Mycoplasmata*, one of the most simple parasitic bacteria that utilizes lipids produced by their hosts and is actually often thought of as an intermediate form of life between viruses and bacteria. Next, lanosterol is converted into cholesterol through two alternative pathways: one ending in desmosterol and another with 7-dehydrocholesterol – the direct precursors of cholesterol. Although textbooks show these as separate pathways, it should be kept in mind that at each of the individual steps, it is possible to swap to the other pathway, as appropriate enzymes for this do exist. The conversion of lanosterol to cholesterol needs a minimum of only 7 steps; however, 18 steps are possible and thus also 18 enzymes exist! This has to be energetically very expensive for cells, once more stressing the great importance of cholesterol.

This amazing redundancy has been noticed a long time ago and it has given rise to a question as to what is so special about the structure of cholesterol that sets it apart from lanosterol and other precursors. When looking at the structures of lanosterol and cholesterol in Figure 1, one notices that the differences are limited to the number and position of double bonds (one more in lanosterol) and the number of methyl groups attached to the steroid ring system (three more in lanosterol). While these do not seem such large differences, they have substantial consequences. First, it has been shown that lanosterol does not induce the existence of the *Lo* phase and thus lipid rafts cannot be formed by this sterol (Miao et al., 2002). Even more intriguingly, it has been shown, already in the sixties, that cholesterol's precursors affect the properties of lipid bilayers step by step more, ending in cholesterol whose effect is the strongest of all. Thus, it has been proposed that the biosynthetic pathway of cholesterol reflects the evolutionary optimization of its structure (Bloch, 1979; Nielsen et al., 2000).



This idea was the starting point for our first investigation into the matter using atomistic MD simulations. Intriguingly, the methyl groups stick out from one side of the cholesterol molecule, called the β -side, while the other side, called the α -side, is flat (**Figure 1**). Lanosterol has three additional methyl groups as compared to cholesterol. Two of these additional methyl groups stick out from the α -side, while the third is directed along the ring plane. Our first results showed greater ordering of saturated lipids neighboring the α -side of cholesterol as compared to lipids next to the β -side (Róg and Pasenkiewicz-Gierula, 2001). Subsequent studies showed that the packing of lipid carbon atoms near the α -side is tight; while near the β -side it is much looser (Róg and Pasenkiewicz-Gierula, 2004). In other words, we showed that the flatness of the ring is associated with higher ordering of lipids. These results fit perfectly with the idea of considering the removal of methyl groups as optimization of cholesterol's structure.

At this point, another open question remains about the role of double bonds in the sterols' structure. In case of desmosterol, atomistic MD studies showed it to be inferior to cholesterol in its ordering capability of saturated lipids; while in the case of unsaturated lipids, there is no significant difference between the two sterols (Vainio et al., 2006; Róg et al., 2008). These results agree with experimental data (Huster et al., 2005; Scheidt et al., 2005). Subsequently, studies of 7-dehydrocholesterol showed very small or non-existent differences as compared to cholesterol. This was observed both in MD simulations (Róg et al., 2008; Liu et al.,

2011) and experimental studies (Chen and Tripp, 2012). However, there are two conjugated double bonds in the ring structure of 7-dehydrocholesterol, which may render the molecule prone to oxidation. This might be the reason why 7-dehydrocholesterol is not the sterol of choice for biological membranes.

Eukaryotic cells require the ordering properties of sterols. At the same time, all of the above considerations lead us to the conclusion that these ordering properties are decreased in the presence of methyl groups. Then, why would any methyl groups remain on the β -side of the ring system? Are they molecular fossils? Could we further optimize the structure of sterols by removing these last remaining groups?

Are Cholesterol's Methyl Groups Molecular Fossils? – Simulations said No!

Molecular dynamics is a very flexible method and provides an inexpensive way to start investigating a new molecule. Surely, if the new molecule does not exist yet, validating the model may be problematic. Nevertheless, taking into account the current development of organic synthesis methods, one may expect the results from MD to be eventually validated.

In the second phase of our investigations, we designed our first sterol, which lacks the methyl groups C19 and C18: 18-19-di-nor-cholesterol, which we called Dchol (see **Figure 1**) (Róg et al., 2007). To our surprise, this sterol does not induce more

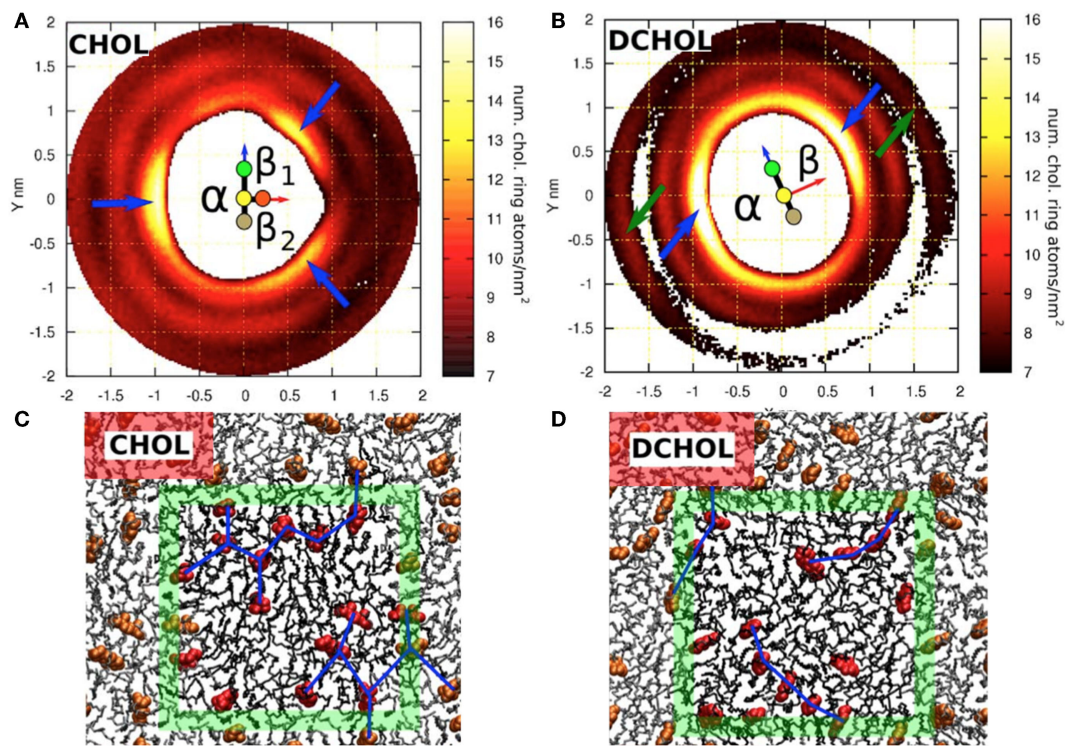


FIGURE 2 | Sterol–sterol in-plane distribution and configurations of sterol molecules in a DSPC bilayer with 20 mol% sterol.

Two-dimensional density distribution for the ring atoms of (A) cholesterol around a tagged cholesterol and (B) Dchol around a tagged Dchol. Both (A,B) show a schematic representation of the tagged sterol (see also Figure 1). The β -face of cholesterol is divided into two sub-faces: β_1 and β_2 . (A) shows that cholesterol avoids the first coordination shell, instead forming a clear second coordination shell. The three emerging peaks, each on a different face, are marked with blue arrows. (B) shows that the two sides of Dchol behave in a similar manner as the smooth α -face of cholesterol. No Dchol is seen in the first coordination shell, and peaks (marked with blue arrows) are observed on both faces. Some structure is still

visible in the outer coordination shell around 1.8 nm. Two peaks, which are collinear with the previous ones, are marked with green arrows. This reflects a strong preference to form linear Dchol–Dchol structures. (C,D) show a top view of an equilibrated configuration of (C) a DSPC/cholesterol bilayer and (D) a DSPC/Dchol bilayer. Only one leaflet is drawn for clarity. PC molecules are shown as black sticks and sterols with a red space-filling model. The boundary of the simulation box is marked with the green square and color brightness. (C) shows the connections between neighboring cholesterol molecules forming triangular patterns, whereas in (D), the connection patterns formed by Dchol molecules are clearly linear. This fundamental difference is due to the missing out-of-plane methyl groups in the Dchol molecule. Figure adapted from Martinez-Seara et al. (2010).

order in saturated bilayers than cholesterol does, even though packing of lipid tails' atoms is almost identical at both sides of Dchol and even slightly higher than in the case of cholesterol. On the contrary, Dchol's ordering capability is clearly worse. In unsaturated bilayers, the differences were smaller; however, cholesterol was still superior to our artificial Dchol. The molecular level mechanism behind the weaker ordering and condensing effects was related to the larger tilt of Dchol in the bilayer (Aittoniemi et al., 2006). Studies of several sterols have shown that the sterol's tilt correlates with its ordering capability (Aittoniemi et al., 2006; Khelashvili and Harries, 2013). Thus, our conclusion was that the methyl groups at the β -side are needed to ensure the proper orientation of cholesterol. Following the initial idea, we then designed alternative sterols with the methyl groups removed one by one – we expected that maybe not all of the methyl groups are needed for maintaining the optimal tilt (Pöyry et al., 2008). Contrary to expectations, however, all the designed sterols turned out again to be inferior in their ordering capabilities to cholesterol, although in some cases, the differences were very small. These studies also showed the C18

methyl group to be the most important one, as its removal had the largest effect. Still, other methyl groups also enhanced the sterols' ordering abilities. All this was very surprising and was in contrast to our expectations, so we continued our investigations even further.

The observation of the most important methyl group being C18 has interesting connotations. The most common lipid chain is an 18-carbon, monounsaturated chain, with the double bond located at position 9–10, and attached at the *sn*-2 position of a glycerol moiety. Cholesterol's effect on unsaturated lipids is weaker than on saturated ones. However, as our studies have shown (Martinez-Seara et al., 2008), the position of the double bond is significant. The largest differences between saturated and unsaturated lipids were observed when the double bond was located at position 9–10. Shifting the double bond up or down leads to stronger effects of cholesterol, and gradually the interactions of the unsaturated and saturated tails with cholesterol converged. Even shifting an unsaturated tail from the *sn*-2 position to *sn*-1 slightly increased cholesterol's effects (Martinez-Seara et al., 2009). Plausibly, the reason for this may be the difference in equivalent

atom positions in the two tails. Consequently, we proposed an additional function for the C18 methyl group: discrimination between saturated and unsaturated chain. We also hypothesized that lipids and sterols coevolved, leading to the known cholesterol structure, and selection of hydrocarbon chain, which together optimize the desired membrane properties. Moreover, the differences in cholesterol effects on saturated and unsaturated lipids affect phase separation and properties of the formed domains.

Another difference between cholesterol and Dchol can be easily visualized. If we look at the cholesterol molecule perpendicularly from its side (**Figure 1**), we see a clear pattern – a flat and a rough face. Now, if we instead look at the cholesterol molecule from top down, we see a kind of threefold symmetry, shown in **Figures 1** and **2**. This is caused by the β -face being subdivided into two further faces (Martinez-Seara et al., 2010). Dchol, due to its lack of methyl groups on the β -face, does not display this kind of threefold symmetry. The difference can be visualized well by looking at the two-dimensional radial distribution of cholesterols around a tagged cholesterol shown in **Figure 2**. This difference may affect the phase behavior of lipid bilayers. As we mentioned above, lanosterol does not promote the *Lo* phase formation, and due to the additional methyl group does not possess the threefold symmetry. As depicted in **Figure 2**, our preliminary data suggest that the symmetry of cholesterol's ring affects the sterol–sterol arrangement. Sterols tend to locate in the second coordination shell of each other, with a lipid molecule in between (Martinez-Seara et al., 2010). Due to the threefold symmetry, cholesterol molecules are able to form a fork net (**Figure 2**) that is likely capable of covering large areas. By contrast, Dchol has only twofold symmetry and thus forms linear structures. It seems plausible that this different form of molecular packing will affect also the phase behavior of Dchol. At this point, we need more extensive studies to further clarify the matter.

Experiments Confirmed the Results from Atomistic MD Simulations

To validate the results from these MD simulation studies, one first has to synthesize the de-methylated form of cholesterol. This task is not to be taken lightly, as cholesterol has seven chiral centers, which make its synthesis particularly complicated. Nevertheless, Dchol was recently synthesized, 7 years after our first simulation studies of de-methylated sterols (Mydock-McGrane et al., 2014). Synthesis was started from a compound whose synthesis was known before: perhydrochrysenone from which 18-19-dinor-cholesterol was obtained in eighteen steps. The yield from the whole synthesis was 3.5%, which taking into account the complexity of the process is a very good result.

The properties of Dchol were carefully examined via an extensive set of biophysical methods (Krause et al., 2014). Langmuir monolayers and fluorescence anisotropy measurements showed that Dchol has slightly weaker condensing and ordering ability than cholesterol, in agreement with our simulation data. Calorimetric study showed that the temperature of the main phase

transition is within error range for lipid bilayers with both sterol types. Nevertheless, excess heat capacity endotherms showed that cholesterol affects the phase transition more strongly than Dchol, indicating differences in interactions of both sterols with phospholipids. Most interestingly, the results of this study showed decreased nearest neighbor interactions in bilayers with Dchol, compared to those with cholesterol. This result cannot be directly compared to results from MD simulations; however, it has interesting consequences. The difference of nearest neighbor interactions of tens of calories per mole, as observed in this experimental study, might lead to substantial changes in domain size distribution as documented by Monte Carlo simulations (Almeida, 2009).

Conclusion

Both atomistic MD simulations and experimental studies have shown that cholesterol's methyl groups are important structural elements of cholesterol and definitely are not molecular fossils. On the contrary, they are important structural elements. Removal of these groups clearly decreases the sterol's ordering and condensing effects. MD simulation studies have indicated that the decreased ordering is related to a larger tilt of the de-methylated sterols, suggesting that the methyl groups are involved in maintaining the proper orientation of cholesterol in lipid bilayers. Both experimental and MD studies imply that the presence of methyl groups might affect the sterol's ability to induce phase separation by affecting domain sizes or changing the structure of the formed sterol–lipid–sterol patches. This problem clearly requires more studies, as it might potentially be the most important reason for nature to select cholesterol.

This perspective article has provided a clear example of how MD simulations can independently provide powerful predictions and thus guide experiments. We have shown how constructing molecules that do not exist in nature can increase our understanding of molecular design of lipids and how simulations of these systems are capable of providing correct, valuable predictions later confirmed by experiments. Despite these kinds of successes, the current editorial practice in high-impact journals clearly favors papers that include both MD simulations and experiments. This happens at the expense of pure simulation papers, which are considerably harder to publish. Surely, every model has to be validated with experimental data, yet lifting the requirement that every single simulation study has to be coupled to experiments *in the same paper* might result in the publication of a greater number of progressive, high-quality simulation articles, which would likely provide fresh, valuable ideas, and predictions for experimental scientists.

Acknowledgments

This work was supported by the Academy of Finland (Center of Excellence program, project no. 272130), the European Research Council (Advanced Grant CROWDED-PRO-LIPIDS), and the Sigrid Juselius Foundation. CSC-IT Center for Science (Espoo, Finland) is acknowledged for computational resources.

References

- Aittoniemi, J., Róg, T., Niemelä, P. S., Pasenkiewicz-Gierula, M., Karttunen, M., and Vattulainen, I. (2006). Tilt: major factor in sterols' ordering capability in membranes. *J. Phys. Chem. B* 110, 25562–25564. doi:10.1021/jp064931u
- Almeida, P. F. (2009). Thermodynamics of lipid interactions in complex bilayers. *Biochim. Biophys. Acta* 1788, 72–85. doi:10.1016/j.bbame.2008.08.007
- Bloch, K. (1979). Speculation on the evolution of sterol structure and function. *CRC Crit. Rev. Biochem.* 7, 1–5. doi:10.3109/10409238309102790
- Bunker, A. (2012). Poly(ethylene glycol) in drug delivery, why does it work, and can we do better? All atom molecular dynamics simulation provides some answers. *Phys. Procedia* 34, 24–33. doi:10.1016/j.phpro.2012.05.004
- Chen, C., and Tripp, C. P. (2012). A comparison of the behavior of cholesterol, 7-dehydrocholesterol and ergosterol in phospholipid membranes. *Biochim. Biophys. Acta* 1818, 1673–1681. doi:10.1016/j.bbame.2012.03.009
- Coskun, U., and Simons, K. (2011). Cell membranes: the lipid perspective. *Structure* 19, 1543–1548. doi:10.1016/j.str.2011.10.010
- Demel, R., Bruckdorfer, K., and Van Deenen, L. (1972). The effect of sterol structure on the permeability of liposomes to glucose, glycerol and Rb⁺. *Biochim. Biophys. Acta* 255, 321–330. doi:10.1016/0005-2736(72)90031-4
- Hölttävuori, M., Uronen, R. L., Repakova, J., Salonen, E., Vattulainen, I., Panula, P., et al. (2008). BODIPY-cholesterol: a new tool to visualize sterol trafficking in living cells and organisms. *Traffic* 9, 1839–1849. doi:10.1111/j.1600-0854.2008.00801.x
- Huster, D., Scheidt, H. A., Arnold, K., Herrmann, A., and Muller, P. (2005). Desmosterol may replace cholesterol in lipid membranes. *Biophys. J.* 88, 1838–1844. doi:10.1529/biophysj.104.048926
- Ipsen, J. H., Karlstrom, G., Mouritsen, O. G., Wennerstrom, H., and Zuckermann, M. J. (1987). Phase equilibria in the phosphatidylcholine-cholesterol system. *Biochim. Biophys. Acta* 905, 162–172. doi:10.1016/0005-2736(87)90020-4
- Kelley, R. I., and Herman, G. E. (2001). Inborn errors of sterol biosynthesis. *Annu. Rev. Gen. Hum. Genet.* 2, 299–341. doi:10.1146/annurev.genom.2.1.299
- Khelashvili, G., and Harries, D. (2013). How sterol tilt regulates properties and organization of lipid membranes and membrane insertions. *Chem. Phys. Lipids* 169, 113–123. doi:10.1016/j.chemphyslip.2012.12.006
- Kohli, A. G., Kierstead, P. H., Venditto, V. J., Walsh, C. L., and Szoka, F. C. (2014). Designer lipids for drug delivery: from heads to tails. *J. Control Release* 190, 274–287. doi:10.1016/j.jconrel.2014.04.047
- Krause, M. R., Wang, M., Mydock-McGrane, L., Covey, D. F., Tejada, E., Almeida, P. F., et al. (2014). Eliminating the roughness in cholesterol's β -face: does it matter? *Langmuir* 30, 12114–12118. doi:10.1021/la503075e
- Kulig, W., Jurkiewicz, P., Olżyńska, A., Tynkynen, J., Javanainen, M., Manna, M., et al. (2015). Experimental determination and computational interpretation of biophysical properties of lipid bilayers enriched by cholesterol hemisuccinate. *Biochim. Biophys. Acta* 1848, 422–432. doi:10.1016/j.bbame.2014.10.032
- Kulig, W., Tynkynen, J., Javanainen, M., Manna, M., Róg, T., Vattulainen, I., et al. (2014). How well does cholesterol hemisuccinate mimic cholesterol in saturated phospholipid bilayers? *J. Mol. Model.* 20, 2121. doi:10.1007/s00894-014-2121-z
- Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. *Science* 327, 46–50. doi:10.1126/science.1174621
- Liu, Y. Z., Chipot, C., Shao, X. G., and Cai, W. S. (2011). The effects of 7-dehydrocholesterol on the structural properties of membranes. *Phys. Biol.* 8, 056005. doi:10.1088/1478-3975/8/5/056005
- Martinez-Seara, H., Róg, T., Karttunen, M., Vattulainen, I., and Reigada, R. (2009). Why is the sn-2 chain of monounsaturated glycerophospholipids usually unsaturated whereas the sn-1 chain is saturated? Studies of 1-stearoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (SOPC) and 1-oleoyl-2-stearoyl-sn-glycero-3-phosphatidylcholine (OSPC) membranes with and without cholesterol. *J. Phys. Chem. B* 113, 8347–8356. doi:10.1021/jp902131b
- Martinez-Seara, H., Róg, T., Karttunen, M., Vattulainen, I., and Reigada, R. (2010). Cholesterol induces specific spatial and orientational order in cholesterol/phospholipid membranes. *PLoS ONE* 5:e11162. doi:10.1371/journal.pone.0011162
- Martinez-Seara, H., Róg, T., Pasenkiewicz-Gierula, M., Vattulainen, I., Karttunen, M., and Reigada, R. (2008). Interplay of unsaturated phospholipids and cholesterol in membranes: effect of double bond position. *Biophys. J.* 95, 3295–3305. doi:10.1529/biophysj.108.138123
- Mason, R. P., Tulenko, T. N., and Jacob, R. F. (2003). Direct evidence for cholesterol crystalline domains in biological membranes: role in human pathobiology. *Biochim. Biophys. Acta* 1610, 198–207. doi:10.1016/S0005-2736(03)00018-X
- Miao, L., Nielsen, M., Thewalt, J., Ipsen, J. H., Bloom, M., Zuckerman, M. J., et al. (2002). From lanosterol to cholesterol: structural evolution and differential effects on lipid bilayers. *Biophys. J.* 82, 1429–1444. doi:10.1016/S0006-3495(02)75497-0
- Mydock-McGrane, L. N., Rath, P., and Covey, D. F. (2014). Synthesis of a smoothened cholesterol: 18,19-Di-nor-cholesterol. *J. Org. Chem.* 79, 5636–5643. doi:10.1021/la105039q
- Nielsen, M., Thewalt, J., Miao, L., Ipsen, J. H., Bloom, M., Zuckermann, M. J., et al. (2000). Sterol evolution and the physics of membranes. *Europhys. Lett.* 52, 368–374. doi:10.1209/epl/i2000-00448-8
- Ohvo-Rekila, H., Ramstedt, B., Leppimäki, P., and Slotte, J. P. (2002). Cholesterol interactions with phospholipids in membranes. *Prog. Lip. Res.* 41, 66–97. doi:10.1016/S0163-7827(01)00020-0
- Ollila, O. H. S., Róg, T., Karttunen, M., and Vattulainen, I. (2007). Role of sterol type on lateral pressure profiles of lipid membranes affecting membrane protein functionality: comparison between cholesterol, desmosterol, 7-dehydrocholesterol and ketosterol. *J. Struct. Biol.* 159, 311–323. doi:10.1016/j.jsb.2007.01.012
- Pöyry, S., Róg, T., Karttunen, M., and Vattulainen, I. (2008). Significance of cholesterol methyl groups. *J. Phys. Chem. B* 112, 2922–2929. doi:10.1021/jp7100495
- Robalo, J. R., Ramalho, J. P. P., and Loura, L. M. S. (2013). NBD-labeled cholesterol analogues in phospholipid bilayers: insights from molecular dynamics. *J. Phys. Chem. B* 117, 13731–13742. doi:10.1021/jp406135a
- Róg, T., and Pasenkiewicz-Gierula, M. (2001). Cholesterol effects on the phosphatidylcholine bilayer nonpolar region: a molecular simulation study. *Biophys. J.* 81, 2190–2202. doi:10.1016/S0006-3495(01)75867-5
- Róg, T., and Pasenkiewicz-Gierula, M. (2004). Cholesterol-phospholipid hydrophobic interactions: a molecular simulation study. *Biophys. Chem.* 107, 151–164. doi:10.1016/S0301-4622(03)00255-2
- Róg, T., and Vattulainen, I. (2014). Cholesterol, sphingolipids, and glycolipids: what do we know about their role in raft-like membranes. *Chem. Phys. Lipids* 184, 82–104. doi:10.1016/j.chemphyslip.2014.10.004
- Róg, T., Vattulainen, I., Jansen, M., Ikonen, E., and Karttunen, M. (2008). Comparison of cholesterol and its direct precursors along the biosynthetic pathway: effects of cholesterol, desmosterol and 7-dehydrocholesterol on saturated and unsaturated lipid bilayers. *J. Chem. Phys.* 129, 154508. doi:10.1063/1.2996296
- Róg, T., Vattulainen, I., Pasenkiewicz-Gierula, M., and Karttunen, M. (2007). What happens if cholesterol is made smoother: importance of methyl substituents in cholesterol ring structure on phosphatidylcholine-sterol interaction. *Biophys. J.* 92, 3346–3357. doi:10.1529/biophysj.106.095497
- Scheidt, H. A., Huster, D., and Gawrisch, K. (2005). Diffusion of cholesterol and its precursors in lipid membranes studied by 1H pulsed field gradient magic angle spinning NMR. *Biophys. J.* 89, 2504–2512. doi:10.1529/biophysj.105.062018
- Slotte, J. P. (2013). Molecular properties of various structurally defined sphingomyelins – correlation of structure with function. *Prog. Lip. Res.* 52, 206–219. doi:10.1016/j.plipres.2013.05.001
- Snipes, G., and Suter, U. (1997). Cholesterol and myelin. *Sub-cellular Biochem.* 28, 173–204. doi:10.1007/978-1-4615-5901-6_7
- Vainio, S., Jansen, M., Koivusalo, M., Róg, T., Karttunen, M., Vattulainen, I., et al. (2006). Significance of sterol structural specificity. Desmosterol cannot replace cholesterol in lipid rafts. *J. Biol. Chem.* 281, 348–355. doi:10.1074/jbc.M509530200
- van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* 9, 112–124. doi:10.1038/nrm2330
- Vist, M. R., and Davis, J. H. (1990). Phase equilibria of cholesterol/dipalmitoylphosphatidylcholine mixtures: 2H nuclear magnetic resonance and differential scanning calorimetry. *Biochemistry* 29, 451–464. doi:10.1021/bi00454a021
- Wenz, J. J. (2012). Predicting the effect of steroids on membrane biophysical properties based on the molecular structure. *Biochim. Biophys. Acta* 1818, 896–906. doi:10.1016/j.bbame.2011.12.021

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 Róg, Pöyry and Vattulainen. 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.



Characteristics of Sucrose Transport through the Sucrose-Specific Porin ScrY Studied by Molecular Dynamics Simulations

Liping Sun¹, Franziska Bertelshofer^{1,2}, Günther Greiner² and Rainer A. Böckmann^{1*}

¹ Computational Biology, Department of Biology, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany,

² Computer Graphics Group, Department of Computer Science, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

Alexander Schulz,
University of Copenhagen, Denmark
Mario Andrea Marchisio,
Harbin Institute of Technology, China
Ulrich Kleinekathöfer,
Jacobs University Bremen, Germany

*Correspondence:

Rainer A. Böckmann
rainer.boeckmann@fau.de

Specialty section:

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

Received: 10 September 2015

Accepted: 25 January 2016

Published: 15 February 2016

Citation:

Sun L, Bertelshofer F, Greiner G and
Böckmann RA (2016) Characteristics
of Sucrose Transport through the
Sucrose-Specific Porin ScrY Studied
by Molecular Dynamics Simulations.
Front. Bioeng. Biotechnol. 4:9.
doi: 10.3389/fbioe.2016.00009

Sucrose-specific porin (ScrY) is a transmembrane protein that allows for the uptake of sucrose under growth-limiting conditions. The crystal structure of ScrY was resolved before by X-ray crystallography, both in its uncomplexed form and with bound sucrose. However, little is known about the molecular characteristics of the transport mechanism of ScrY. To date, there has not yet been any clear demonstration for sucrose transport through the ScrY. Here, the dynamics of the ScrY trimer embedded in a phospholipid bilayer as well as the characteristics of sucrose translocation were investigated by means of atomistic molecular dynamics (MD) simulations. The potential of mean force (PMF) for sucrose translocation through the pore showed two main energy barriers within the constriction region of ScrY. Energy decomposition allowed to pinpoint three aspartic acids as key residues opposing the passage of sucrose, all located within the L3 loop. Mutation of two aspartic acids to uncharged residues resulted in an accordingly modified electrostatics and decreased PMF barrier. The chosen methodology and results will aid in the design of porins with modified transport specificities.

Keywords: molecular dynamics, ScrY, porin, sucrose binding, transport mechanism, potential of mean force

1. INTRODUCTION

Most bacteria produce cell walls surrounding the cytoplasmic membrane to protect their vulnerable cell structure and to maintain their mechanical rigidity. Gram-negative bacteria possess an outer membrane surrounding the inner cell wall with a peptidoglycan layer in between (Glauert and Thornley, 1968). The outer membrane acts as a selective permeability barrier to exclude noxious compounds and exchange nutrients and waste products with the external medium (Nikaido and Nakae, 1980). For this purpose, the outer membrane contains porins, a special class of proteins, which function as channels across the membrane. Through porins a variety of substrates can pass in a diffusion-like process (Nakae, 1976). Such porins can be classified into (i) general diffusion porins that are responsible for the non-specific and spontaneous transport of ions and small hydrophilic molecules, and (ii) specific diffusion channels that contain stereospecific binding sites within the pore, facilitating the uptake of solutes of certain types (Nikaido, 1992). The expression of these specific porins is usually induced under special environmental conditions (Nikaido and Vaara, 1985).

General diffusion porins usually form tightly assembled homotrimers. Each subunit is a water-filled β -barrel consisting of typically 16 or 18 antiparallel β -strands oriented perpendicular to the membrane plane and tilted by 30–60° with respect to the molecular symmetry axis (Nabedryk et al., 1988; Jap, 1989). The β -strands are connected by short turns on the periplasmic side and by long loops on the external side (Nikaido, 1994). Interestingly, porin channels are constricted by the so-called eyelet loop that folds inwardly and is attached to the inner side of the barrel wall, resulting in a cross-section of a minimal size of $\sim 7 \times 11 \text{ \AA}$, which allows the passage of hydrophilic solutes up to an exclusion size of $\approx 600 \text{ Da}$ (Weiss et al., 1991; Schirmer, 1998). Since the pore size is similar to the diameter of most nutrient molecules, the diffusion rates are strongly affected by the physical properties of the substrates. As typical general porins in *E. coli*, OmpF and OmpC were reported to favor both neutral molecules and cations, while PhoE favors anions (Nikaido and Vaara, 1985; Bauer et al., 1989).

One example of a specific porin is the sucrose-specific porin (ScrY) of enteric bacteria, expressed when *E. coli* is starved for sucrose (Schmid et al., 1982). It permits the rapid influx of sucrose across the outer membrane, allowing cells to grow on sucrose as a solo carbon source (García, 1985; Schmid et al., 1991). ScrY was found during the investigation of the plasmid-encoded metabolic pathway of sucrose in *Salmonella typhimurium*, where sucrose has only a small rate of translocation through the outer membrane in the absence of the *scr* genes pUR400 (Schmid et al., 1988). The crystal structure of sucrose-specific porin was determined at a resolution of 2.4 Å, both in its uncomplexed form and with bound sucrose (Forst et al., 1998). Apart from the shared architectural properties with general porins, each polypeptide chain of the ScrY channel, containing 413 structurally well-defined amino acids, traverses the membrane 18 times as antiparallel β -strands surrounding a hydrophilic pore. Importantly, the eyelet loop (L3) folds inwardly into the lumen of the β -barrel forming a selective gate, comparable to general porins as mentioned above. This is in excellent agreement with the physiological function of the porin to exclude toxic compounds and to maximize the uptake of nutrients using wide openings and a greasy pathway for sugars (see below), minimizing frictional interactions to the constriction site in the L3 region (Welte et al., 1995). Sucrose-specific porin binds two sucrose molecules at the same time in a certain configuration (Forst et al., 1998) [similar to maltoporin (Dutzler et al., 1996)]. Besides, it also has the features of a general diffusion pore with a comparable single-channel conductance, which is much smaller for other specific porins (Schmid et al., 1991; Schüle et al., 1995).

The essential role of the inner loop L3 for modulating the translocation of molecules was also suggested for the outer membrane porin OprD, specific for the uptake of small natural substrates like cationic amino acids, by means of *in silico* electrophysiology and metadynamics simulation techniques (Samanta et al., 2015). A combination of *in silico* and *in vitro* studies was used in the study of the outer membrane channels OprP (specific for phosphate transportation) and OprO (specific for diphosphate transportation). Two amino acids in the central constriction region were suggested to generate the substrate specificity. Reciprocal exchange of these amino acids resulted in an interchange of substrate specificities for these channels (Modi et al., 2015). For

OmpF, the cation selectivity was reported to be highly influenced by the electrostatic environment of the constriction region. It was found that removing the cationic residues in the cross-sectional area enhanced the cation selectivity, whereas removal of the anionic residues reversed the selectivity (Pezeshki et al., 2009).

Another specific porin, Maltoporin (LamB), was first identified as the receptor for λ -phage in *E. coli* (Randall-Hazelbauer and Schwartz, 1973). It forms trimeric channels, which are specific for the transport of maltose and malto-oligosaccharides and is synthesized for maltodextrin concentrations below 10 μM (Nikaido and Vaara, 1985; Saurin et al., 1989; Death et al., 1993). Noteworthy, ScrY can also function as a malto-oligosaccharide porin, and LamB is also able to activate the influx of sucrose, but only at high substrate concentrations (Szmecman and Hofnung, 1975; Schüle et al., 1990; Schmid et al., 1991). Therefore, it was suggested that ScrY is a sugar-specific porin with a similar function as LamB. This is further stressed by their structural similarity (Dali Server yields a Z score of 36.8) and their sequence similarity (Blast E value = $4e - 173$) (Altschul et al., 1990; Holm and Rosenström, 2010). Although the similarity in topology is remarkable, the depicted porins share only 23% amino acid sequence homology (Protein Data Bank). The sequence identity is increased at the prospective glucose-binding regions, located in the first half of the primary sequence from the N-terminal end (Hardesty et al., 1991; Schüle et al., 1995). Interestingly, the binding constant of ScrY and LamB with malto-oligosaccharides increases with the number of glucose residues but becomes saturated after five residues, as inferred from current noise and ion flow inhibition studies respectively (Benz et al., 1987; Schüle et al., 1991). Based on these observations, Benz and colleagues proposed that the binding site has a length of about five glucose residues along the wall of the channel and the dextrine penetrates in a single file because of the hourglass-shaped constriction of the pore. However, evidence for this mechanism is still scarce (Benz et al., 1987; Benz and Bauer, 1988).

Very few experiments have been carried out so far that provide insights into the structure–function relationship of ScrY. The first clear demonstration of sucrose uptake was demonstrated by Schmid et al. (1982) in their *in vivo* research of *E. coli* using chromatography techniques, where the apparent K_m value (i.e., the Michaelis constant) was determined to 10 μM . Simultaneously, the binding of sucrose to ScrY has been investigated by Schüle et al. who reported a stability constant K defined as the ratio between on and off rate constants for sugar binding of 20 l mol⁻¹ based on the relative rates of permeation of ScrY and LamB and the stability constant of sucrose binding to LamB (Schüle et al., 1995). The penetration rates for sugar through porin channels is likely to be strongly affected by its concentration on each side of the membrane, as revealed by current noise studies of Jordy on LamB due to the blockage of the ionic flow during the passage (Jordy et al., 1996). Besides, the conception of a “greasy slide” was described both for ScrY and LamB containing five or six contiguous aromatic side chains lining up on one side of the channel and forming a smooth hydrophobic path (Schirmer et al., 1995; Wang et al., 1997; Forst et al., 1998). The greasy slide extends from the channel vestibule through the constriction zone to the periplasmic exit and functions as a guidance for sugar molecules sliding

through the channel by engagement in non-specific hydrophobic interactions with the pyranose rings (Schirmer et al., 1995). Based on this concept, the passage of sucrose has been hypothesized to consist of a series of steps: first, sucrose diffuses from the external solution to a trapping zone in sucrose-specific porin. Second, a sucrose molecule slides along the greasy slide and enters the binding region. Finally, sucrose passes through the binding region and enters the cell (Forst et al., 1998). However, to date there is no convincing evidence supporting this hypothesis.

In this study, we have investigated the dynamics of the ScrY trimer embedded in a phospholipid bilayer as well as the characteristics of sucrose translocation by means of atomistic molecular dynamics (MD) simulations. Various structural properties of ScrY and its implications for sucrose transport are discussed. Based on results of potential of mean force (PMF) calculations for sucrose transport through ScrY, key residues were pinpointed and a mutant suggested that showed a significantly decreased free energy barrier for the passage of sucrose.

2. COMPUTATIONAL METHODS

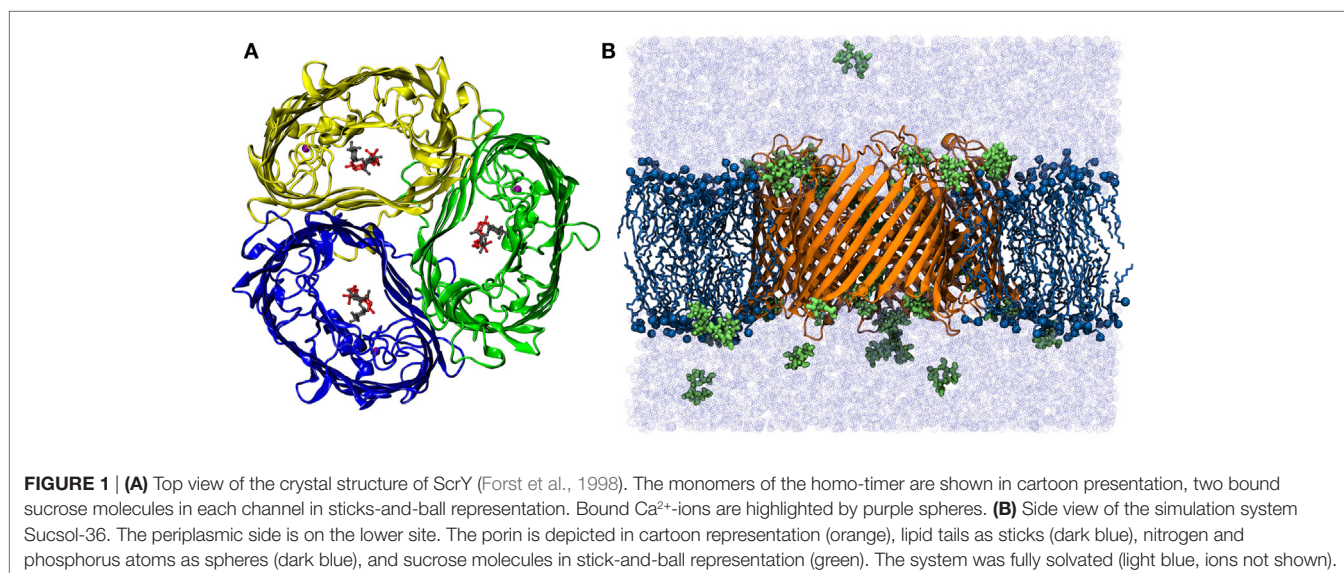
2.1. System Setup

The simulation system was prepared by embedding the refined protein model in the lipid bilayer using the INSANE approach (Pluhackova et al., 2013; Wassenaar et al., 2015). The crystal structure of sucrose-specific porin was taken from the Protein Data Bank (PDB entry 1A0T, see **Figure 1A**) (Forst et al., 1998). The parameters for sucrose were generated using the GLYCAM06 Carbohydrate Builder. The lipid type and the bilayer composition have been proven to be non-essential for both the porin functions and for the pore properties (Parr et al., 1986; Wiese et al., 1994, 1996). Here, a bilayer patch composed of 331 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids was constructed in the fluid phase with equal cell lengths in the membrane plane ($L_x = L_y$, thickness of approx. 3.76 nm). The protein membrane system was dissolved with TIP3P water (Jorgensen et al., 1983) in

a rhombic box ($>36,000$ water molecules). Na^+ and Cl^- ions were added at a physiological concentration of 50 mM to the solution, the protein net charge was compensated by ions. All starting structures were initially energy minimized using the steepest descent algorithm. Two setups constructed from the crystal structure without sucrose (system “ScrY-nosuc”) and with 6 bound sucrose molecules (“SrcY-suc”) were simulated for 100 ns each. Subsequently, the system without sucrose after 100 ns was taken as the start configuration for two additional simulations, with 36 (system “Sucsol-36”) and 72 sucrose molecules (“Sucsol-72”) added to the solvent, respectively. These systems were simulated for 0.5 μs each. A snapshot of one simulation system is shown in **Figure 1B**, it contains in total more than 173,000 atoms.

2.2. Simulation Details

Atomistic molecular dynamics simulations were carried out using the open-source software package GROMACS version 4.6.3 with a time step of 2 fs (Berendsen et al., 1995; Hess et al., 2008). The AMBER99SB-ILDN force field (Hornak et al., 2006; Lindorff-Larsen et al., 2010) for proteins was combined with GLYCAM06 parameters (Kirschner et al., 2008; Tessier et al., 2008) for sucrose and the SLIPIDS force field (Jämbeck and Lyubartsev, 2012a,b,c) for lipids. The cutoff for the van der Waals and the (short-range) Coulomb potential was chosen to 1.2 nm. Short-range electrostatic interactions were calculated explicitly, whereas long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method (Darden et al., 1993). Temperature coupling was achieved with the Nose-Hoover scheme with a reference temperature of 310 K and a time constant of 1.0 ps (Nosé, 1984; Hoover, 1985). The semiisotropic pressure coupling was controlled using the Berendsen-thermostat algorithm with a time constant of 5.0 ps (Berendsen et al., 1984). The Lincs algorithm was applied to constrain the bond lengths of the hydrogen atoms to a constant value (Hess et al., 1997). Periodic boundary conditions were applied in all three dimensions to avoid boundary effects caused by a finite simulation system.



Analysis was conducted using in-house routines, GROMACS analysis utilities, and the HOLE program (Smart et al., 1997). Molecular visualization was performed using VMD (Humphrey et al., 1996) and PyMOL (DeLano and Bromberg, 2004).

2.3. Potential of Mean Force

The Potential of Mean Force (PMF) was computed using umbrella sampling in order to gain insight into the energetic determinants of sucrose transport through ScrY. In umbrella sampling simulations, the sampling of high-energy regions is improved by adding a biasing “umbrella” potential [e.g., Christ et al. (2010)]. For this purpose, Steered Molecular Dynamics (SMD) simulations were initially carried out by pulling a sucrose molecule through one of the three porins in the trimer (Isralewitz et al., 2001). To avoid ambiguities at the channel openings, a structure with a sucrose bound inside the pore was taken as a start configuration. This sucrose molecule was pulled in both directions (along z -axis, 50 ns for each direction). From these pulling simulations, 105 starting structures for subsequent umbrella sampling simulations were extracted with a spacing of the sucrose positions along the channel axis of ≤ 0.08 nm (in total 103 simulations for the mutant). The umbrella potential was introduced between the center of mass of the channel and the sucrose. The harmonic force constant was chosen to 1,000 kJ/mol/nm². Other molecular dynamics parameters were identical to the equilibrium simulations as described above.

Each umbrella was simulated for at least 10 ns. Several samples where the sucrose was close to the pore constriction were equilibrated for longer times in order to obtain an improved equilibration of the system. Histograms distal to the mutation site and the constriction region were used both for the wild type and the mutant. **Figure 2** shows the umbrella histograms exhibiting sufficient overlap between the adjacent windows. The potential of mean force as a function of the channel coordinate (z -coordinate along the membrane normal) was calculated using the Weighted Histogram Analysis Method (WHAM) based on these umbrella

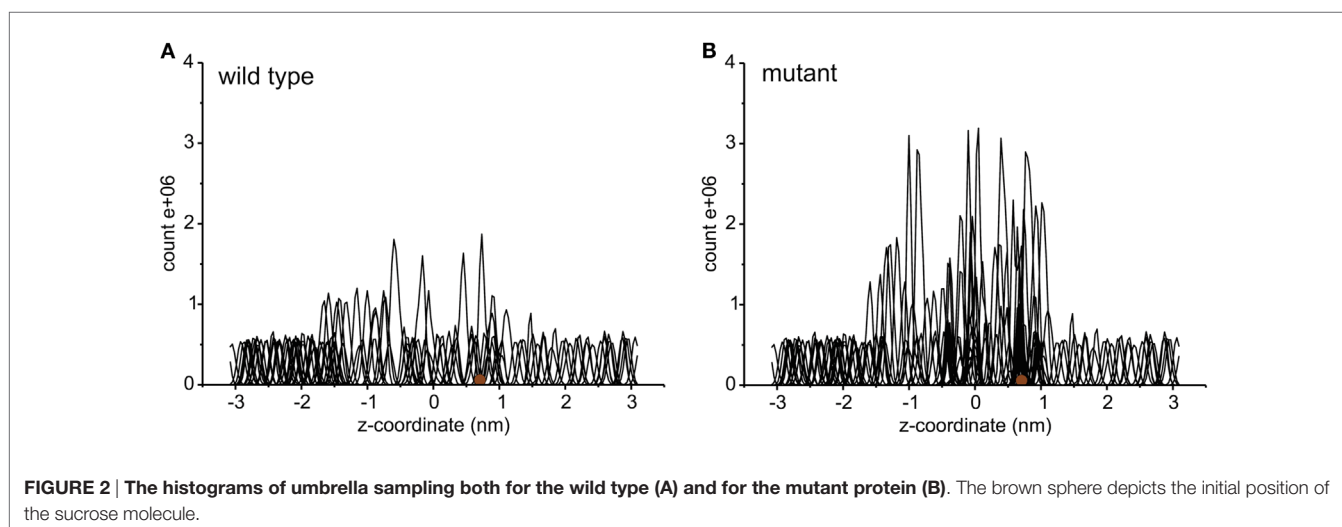
sampling simulations (Rosenberg, 1992) as implemented in the Gromacs tool *g_wham* (Hub et al., 2010). The error was obtained from bootstrapping of the histograms using *Bayesian bootstrap* with 200 bootstraps. The PMF was analyzed for a periodic reaction coordinate, and the integrated autocorrelation times of the bootstrapped trajectories were smoothed using a Gaussian filter (Hub et al., 2010).

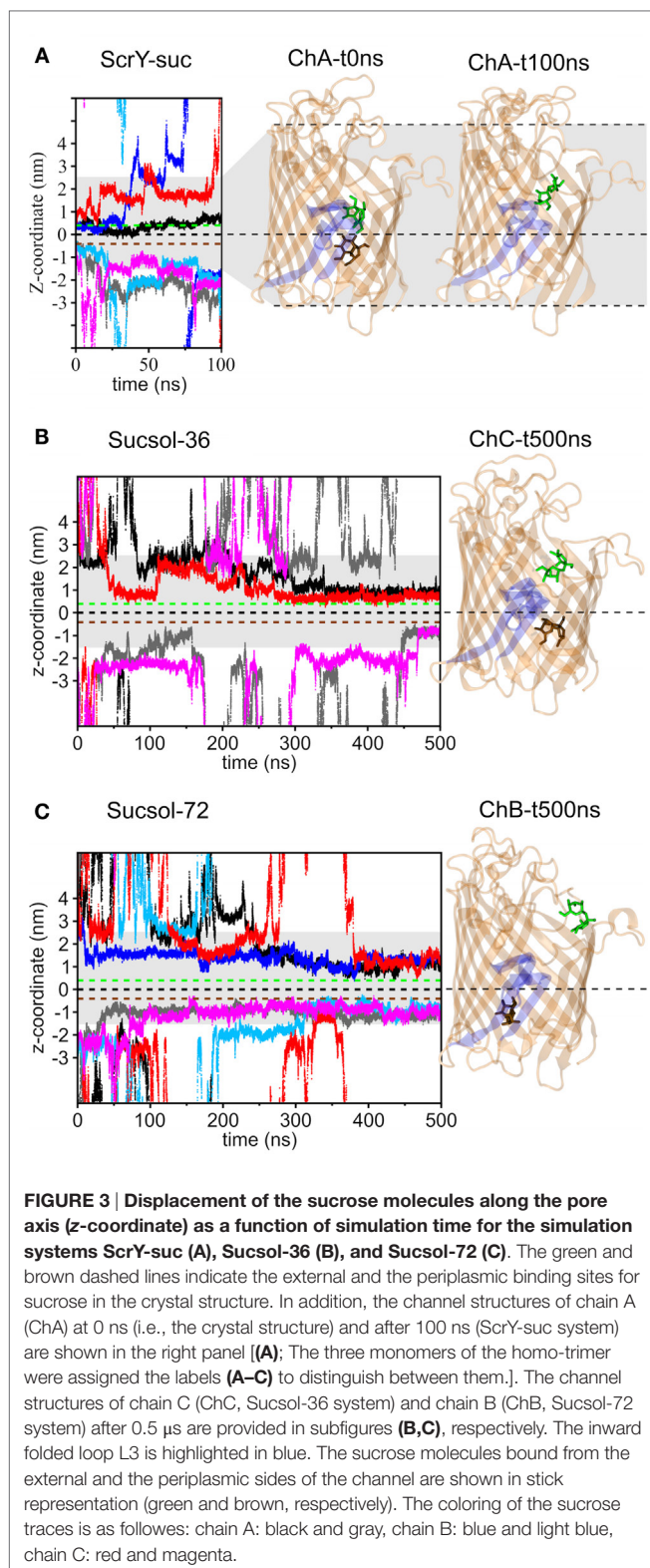
3. RESULTS

The ScrY-trimer with and without bound sucrose were stable in 100 ns simulations with comparably small root mean square deviations (rmsd) of 1–1.5 Å for the backbone atoms of each monomer (not shown, similar rmsd for backbone atoms of trimer structure empty/with bound sucrose). The spontaneous binding of sucrose to the channel was addressed in two 0.5- μ s simulations for different sucrose concentrations in solution (36 and 72 sucrose molecules, corresponding to a concentration of ~ 50 and 100 mM in solution, respectively).

3.1. Sucrose Binding to ScrY

Two bound sucrose molecules per monomer were resolved in the crystal structure of ScrY (Forst et al., 1998). In the simulations, the sucrose in the periplasmic binding sites spontaneously left ScrY within the first 20 ns of simulation (see **Figure 3A**, traces colored light blue, gray, and magenta; sucrose colored brown in ScrY structure). For the external binding sites, only one sucrose molecule stayed bound for the full 100 ns (colored green). It is located above the inner loop L3 and below the more flexible loops connecting the β -strands of the barrel at the external side. No sucrose molecule passed the center of the channel defined by the midpoint between the bound sucrose molecules of the crystal structure ($z = 0$ nm). These simulation results together with the reported sucrose occupancies of only 0.71 (periplasmic binding site) and 0.80 (external binding site) in the crystal structure (Forst et al., 1998) suggest that not all ScrY sucrose binding sites are occupied under physiological conditions. In addition, sucrose





may be trapped in the region between the inner loop L3 and the free loops at the external side of the channel.

Figures 3B,C show the binding of sucrose molecules from the solution to the apo-ScrY as observed in 0.5 μ s simulations

TABLE 1 | Number of ions passing through ScrY in equilibrium simulations.

Channel	ScrY-nosuc	ScrY-suc	Sucsol-36	Sucsol-72
Chain A	4 Na ⁺	0	9 Na ⁺	7 Na ⁺
Chain B	0	2 Na ⁺	7 Na ⁺	4 Na ⁺
Chain C	0	1 Na ⁺ , 1 Cl ⁻	4 Na ⁺	2 Na ⁺ , 1 Cl ⁻
Sum	4 Na ⁺	3 Na ⁺ , 1 Cl ⁻	20 Na ⁺	13 Na ⁺ , 1 Cl ⁻

The three monomers of the homo-trimer were assigned the labels A, B, and C to distinguish between them.

(Sucsol-36 and Sucsol-72). No sucrose molecule passed the channel in the simulations. However, a number of events is observed for sucrose binding to both periplasmic and external openings of the channel. Each two sucrose molecules got bound to two of the three chains at intermediate concentrations (Sucsol-36), at high concentration two sucrose molecules were bound to each chain. The binding positions were, however, shifted with respect to the positions reported in the crystal structure. Sample snapshots after 0.5 μ s of simulation with marked sucrose positions are provided in Figure 3. In conclusion, sucrose is able to spontaneously move into and bind within the channel close to the periplasmic binding site and within the external trapping area, but is not transported on the submicrosecond timescale.

In addition to sucrose binding, transport of Na⁺ and Cl⁻ ions could be observed. As detailed in Table 1, in total seven Na⁺ ions and one Cl⁻ ion moved through ScrY during the 100 ns equilibrium simulations (ScrY-nosuc and ScrY-suc), and 33 (1 Na⁺ (Cl⁻) ions during the 0.5 μ s simulations (Sucsol-36, Sucsol-72). Thus the passage of positively charged ions through ScrY (both directions) is strongly preferred over the passage of negatively charged ions. This may be explained by the negative electrostatic potential at the external opening and through the pore as shown in Figure 4 that is probably caused by several (negatively charged) aspartic acids in the inner L3 loop of ScrY. Exchange of two aspartic acids of L3 (Asp194 and Asp201) by alanines significantly shifted the electrostatic potential within the pore (see below for the selection of mutants). Interestingly, ion passage was blocked for the wild type ScrY if a sucrose molecule bound close to the inner constriction zone. This is in agreement with earlier experimental studies on porins reporting ion blockage by sugar binding to the porin (Andersen et al., 1998; Kullman et al., 2002).

3.2. ScrY Pore Size

The shape and size of the ScrY channel was addressed using the HOLE program (Smart et al., 1996). Figures 5A,B provide both the average pore radius along the channel axis as well the pore flexibility or pore fluctuations (gray shaded area). The radius profile as well as the channel structure (Figure 5C) reflects an hourglass shape of the pore with a minimum radius of only 0.3 nm in the constriction area (0–1 nm). In this region, the pore is narrowed down by the inwardly folded loop L3; it forms the external binding site for sucrose. The loop L3 is stabilized by a hydrogen bond network in particular with loop L1 (Figure 5D). Arg110 of L1 forms hydrogen bonds with His196 and Trp197

(both L3), and Tyr97 (L1) forms hydrogen bonds with Asp199 and Ser200 (both L3).

The external opening of the porin shows a comparable high flexibility. The shape of the pore and also its flexibility appear unchanged for high sucrose concentrations in the medium.

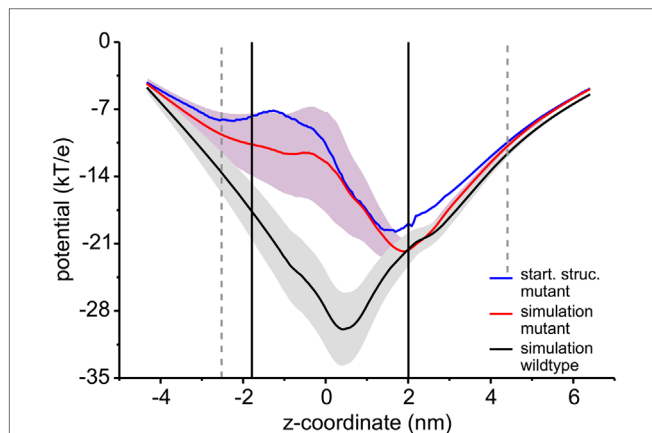


FIGURE 4 | Electrostatic potential along the channel axis. The electrostatic potential was calculated using a recently developed Poisson–Boltzmann solver (Bertelshofer et al., 2015). The membrane was implicitly modeled using a dielectric constant of 2 (Böckmann et al., 2008), the region of the membrane boundaries are shown by black lines, the extension of the ScrY protein into the solvent phase by dashed gray lines. The potential for the wild type (black line) was averaged over 100 ns of simulation, the potential for the Asp194Ala:Aps201Ala mutant averaged over 50 ns (red line). The path through the individual channels was analyzed using the HOLE program (Smart et al., 1997). The blue line shows the electrostatic potential for the crystal structure.

3.3. Potential of Mean Force for Passage of Sucrose through ScrY

Umbrella sampling was applied to study the energetics of sucrose passage through ScrY. Starting structures for the individual umbrellas along the channel were extracted from pulling simulations; in these, starting from an initial configuration after 100 ns of simulation (ScrY-suc system, see Figure 3A) a bound sucrose was pulled to both channel openings.

The PMF for sucrose transport through the ScrY using a periodic reaction coordinate shows a total barrier height of ≈ 22 kJ/mol (Figure 6, black line). Two equally steep regions were identified labeled B1 and B2. The first barrier B1 at the periplasmic part of the porin has a height of ≈ 19 kJ/mol, the second energy barrier B2 of ≈ 22 kJ/mol. The latter barrier is close to the external sucrose trapping region of ScrY. It is located within the constriction area. The two crystal binding sites for sucrose are found at two metastable positions, right above the B2 barrier, and close to the minimum following the intracellular B1 barrier (dashed lines in Figure 6). Smaller barriers preceding B1 ($z < -0.6$ nm) are caused by electrostatics interactions of sucrose with the mobile loops at the periplasmic pore entrance (not shown). Figure 7 shows sampled sucrose positions from the umbrella simulations. As suggested before, the passage of sucrose mainly follows the amino acids that form the greasy slide (highlighted).

An energy decomposition was applied to determine the influence of all residues aligning the pore through the porin on the passage of sucrose, and key residues were selected for subsequent mutation. The enthalpic contributions of the 115 amino acids forming the ScrY pore on the PMF for sucrose transport were estimated as the sum of the corresponding Lennard-Jones and (short-range) Coulomb interactions (short-range cutoff of 1.2 nm) between each amino acid and the sucrose molecule along

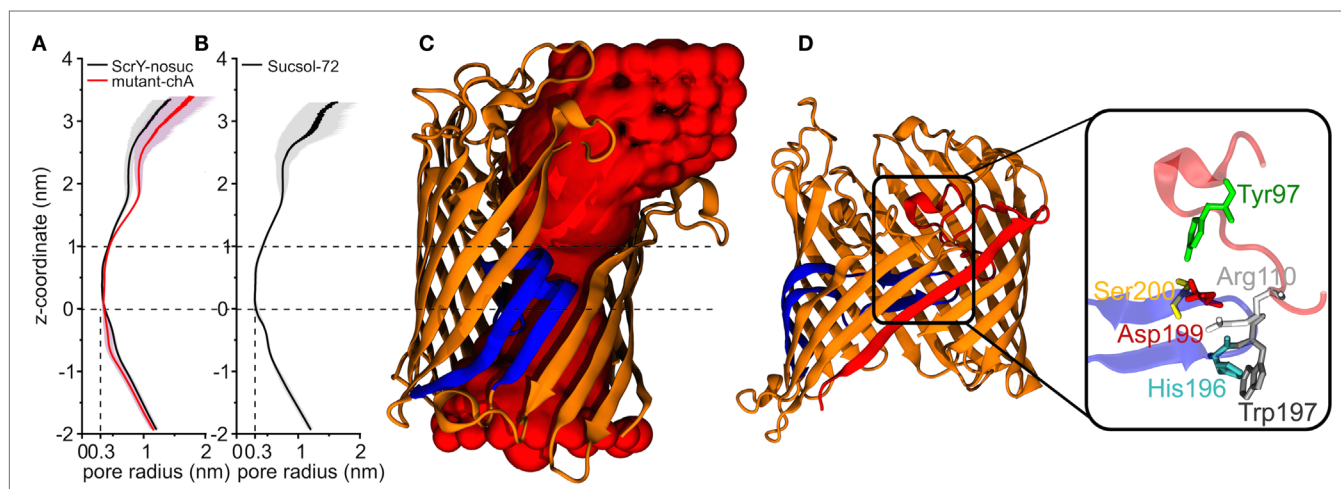


FIGURE 5 | Pore radius profiles along z-coordinate for equilibrium simulations of ScrY without sucrose [ScrY-nosuc, 100 ns, (A)] and mutant [mutant-chA, 50 ns, (A)], and for ScrY in a solution with high sucrose concentration [ScrY-72, 500 ns, (B)]. The gray shadow describes the flexibility caused by the channel fluctuation. The middle of two crystal sucrose binding sites is found at $z = 0$ nm. The pore for a monomer is visualized (C), where the porin is described by cartoon presentation in orange, and the inner loop L3 in blue. The red spherical probes describe the lumen along the pore axis. The hydrogen bond interactions between loop L1 (red) and loop L3 (blue) of the channel are shown in (D). The detailed description is shown in the enlarged square. The involved amino acids are described by sticks and spheres, where ARG110 (white) from L1 forms hydrogen bonds with HIS196 (blue) and TRP197 (gray) from L3, TYR97 (green) from L1 forms hydrogen bonds with ASP199 (red), and SER200 (yellow) from L3.

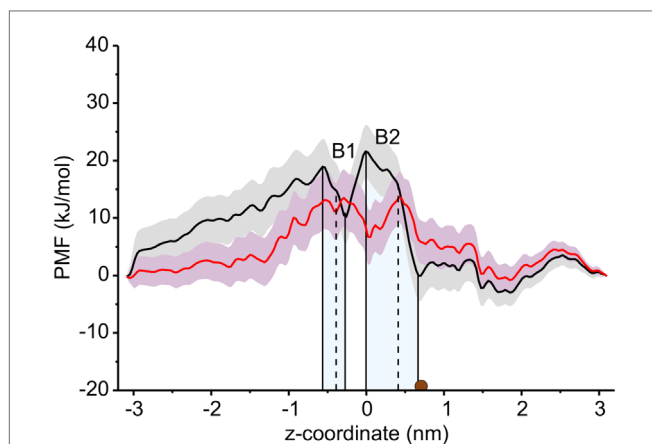


FIGURE 6 | Potential of mean force (PMF) profiles along the channel coordinate both for the wild type (black) and for the mutant protein (red). The shaded areas indicate the statistical uncertainty (67% confidence interval). The two main energy barriers for wt ScrY are highlighted by gray shaded areas (B1, B2). The brown dot depicts the initial position of the sucrose molecule for the initial pulling simulation. The two sucrose binding sites of the crystal structure are found at the dashed lines. The external side of ScrY is found on the right side. The PMF was analyzed using a periodic reaction coordinate in the Weighted Histogram Analysis Method (WHAM), a profile using non-periodic WHAM is provided as Supplementary Information.

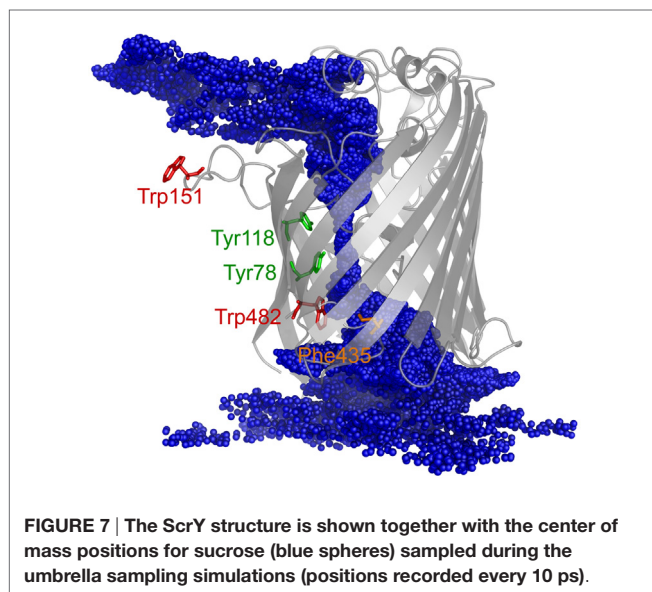


FIGURE 7 | The ScrY structure is shown together with the center of mass positions for sucrose (blue spheres) sampled during the umbrella sampling simulations (positions recorded every 10 ps).

the path. The average residue–sucrose interaction energy for each umbrella window was computed and analyzed.

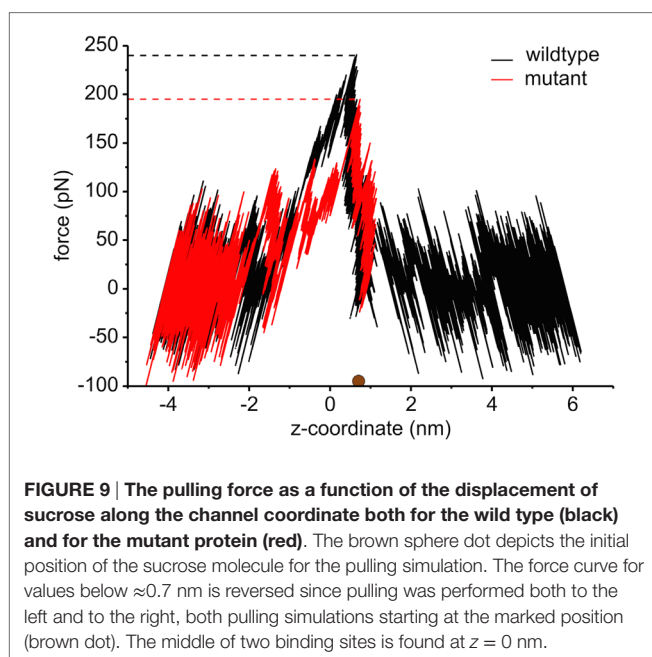
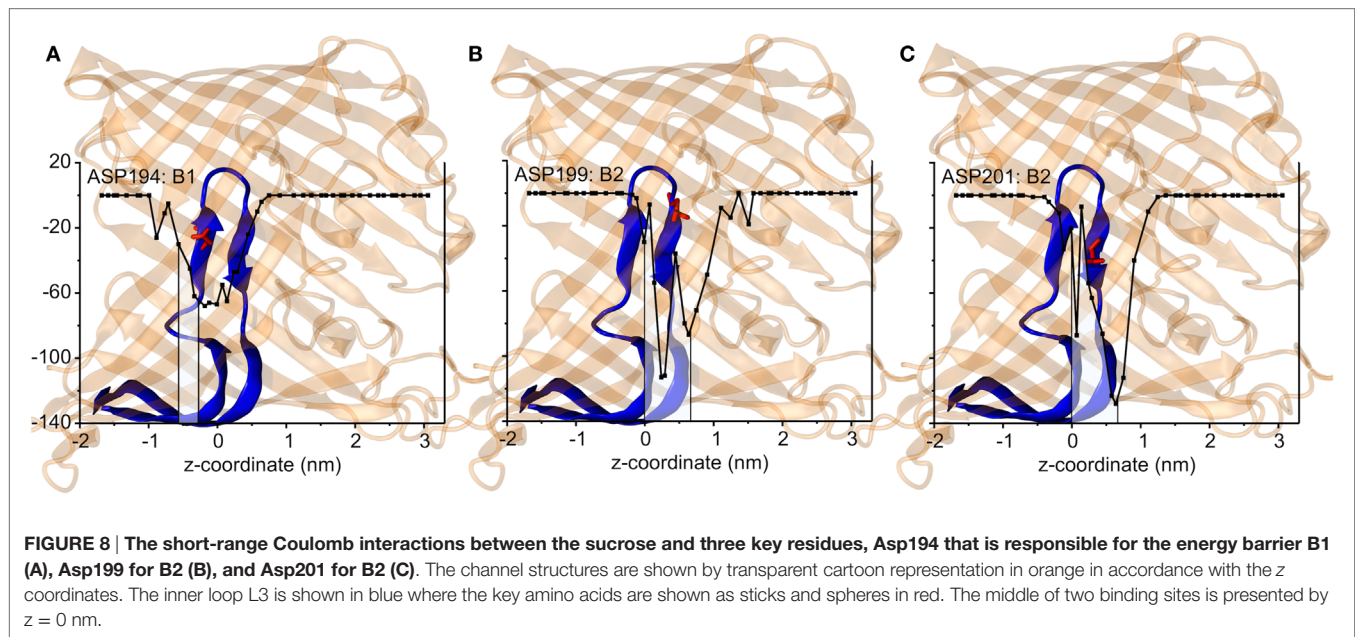
The interaction energies of the three most strongly interacting residues with sucrose are shown in **Figure 8**. Asp194, Asp199, and Asp201 show strong Coulomb interactions, which coincide with the largest barriers of the PMF profile (see **Figure 6**). Interestingly, these three amino acids are located in the central L3 loop, highlighting the importance of L3 for sucrose transport. For the design of a ScrY mutant with improved sucrose transport characteristics,

we accordingly chose Asp194 and Asp201 for mutation to alanine (termed ScrY^{mut}). Asp in position 199 was kept due to its role in fixing the position of the L3 loop within the pore (see **Figure 5D**), as shown above. The influence of the double mutant Asp194Ala:Asp201Ala on the PMF was subsequently tested in additional umbrella simulations, restricted to the region from the external sucrose binding site to the periplasmic ScrY opening. Starting structures for the umbrella simulations were obtained from a sucrose pulling simulation, applied pulling forces for the wt ScrY, and mutant ScrY are compared in **Figure 9**. The maximal pulling force is reduced from ≈ 240 pN for wt ScrY to <200 pN for the double mutant. However, considerable fluctuations for the pulling forces are to be expected due to activated processes. Still, the location of the main barrier is in agreement with the PMF result. As expected from the residue-resolved sucrose interaction energies and the pulling forces, the potential of mean force is significantly lowered for the designed ScrY double mutant (**Figure 6**, red line). Both the B1 and B2 barriers are diminished by the Asp-to-Ala mutations within the L3 loop. The overall barrier height was found decreased by ≈ 8 kJ/mol. Also the electrostatic potential along the pore axis increased significantly (see **Figure 4**, green line). Provided the stability of this mutant also *in vivo*, the change in the PMF profile suggests a significantly enhanced sucrose transport for this mutant.

4. DISCUSSION

The characteristics of sucrose passage through the porin ScrY was studied using atomistic molecular dynamics simulations. Equilibrium simulations showed that very high sucrose concentrations are required to see (meta-)stable binding of sucrose to both the periplasmic- and the external openings of the porin, i.e., the two binding sites reported for the crystal structure are most probably not simultaneously occupied at low or intermediate sucrose concentrations. The inside width of the hourglass shaped pore and thus the size of molecules allowed for passage is mainly restricted by the inwardly folded L3 loop. It forms a constriction zone with a radius of only 0.3 nm. Larger fluctuations of the pore width were observed only for the external pore mouth, providing a larger access volume for sucrose approach. It has been reported that the sucrose molecule has to form a specific configuration (left-handed helical conformation) in order to pass through this narrow pore (Dutzler et al., 1996; Forst et al., 1998).

The potential of mean force for sucrose passage through the porin was obtained from extensive umbrella sampling simulations. The PMF exhibits two main energy barriers with 19 and 22 kJ/mol. The enthalpic contribution to these barriers could be mainly ascribed to Coulombic interactions with the aspartic acids of the central L3 loop (Asp194, Asp199, Asp201), suggesting these acidic residues as mutation candidates for the design of ScrY variants with altered transport characteristics. How well will the porin structure be retained for such mutants? A study of OmpF with four point mutants and one deletion mutant showed no alteration of the barrel structure but only local effects on the structure of the pore constriction region (Lou et al., 1996). The authors mutated pore wall arginines at the constriction zone of OmpF by shorter uncharged residues. Similar to ScrY, OmpF contains acidic residues on the L3 loop and basic residues on the



facing barrel wall (Nikaido, 2003). Therefore, one may conclude from the OmpF study that also non-conservative mutations of the counterpart residues of the pore wall arginines, i.e., the mutations of acidic residues of the L3 loop to uncharged ones, will leave the overall porin structure unaltered.

Since one of the aspartic acids (Asp199) is involved in hydrogen bonding with loop L1, that appears essential for the configurational stabilization of loop L3, only two aspartic acids of L3 (Asp194 and Asp201) were mutated to ALA in order to generate a mutant showing enhanced sucrose passage capabilities. Comparisons of forces for pulling sucrose through the porin, of the electrostatic

potential along the pore, and of the PMF between wt ScrY and ScrY^{mut} showed a significantly altered energetics for sucrose passage. The PMF barrier is decreased by ≈ 8 kJ/mol. However, while sucrose passage through ScrY^{mut} will be enhanced, the mutation will possibly also affect the specificity of the porin. Additionally, despite using umbrellas being very close in space with long simulation times (total simulation time for wt was 1.29 μ s and for the mutant 1.72 μ s), the PMFs are not fully converged and show a substantial error. Still, the results indicate a substantially improved passage of sucrose through the mutated porin.

In summary, key residues for passage of sucrose through ScrY were identified and a double mutant with improved transport characteristics designed *in silico*. The suggested mutant is currently further characterized in experiments.

AUTHOR CONTRIBUTIONS

RB designed research, LS performed simulations, LS, FB, and RB performed analysis, and all wrote the manuscript.

ACKNOWLEDGMENTS

We acknowledge support by the Emerging Fields Initiative *Synthetic Biology* (EFI) and the Research Training Group 1962, *Dynamic Interactions at Biological Membranes: From Single Molecules to Tissue*. Liping Sun was supported by the China Scholarship Council (CSC). We acknowledge computational support from the Computer Center of the Friedrich-Alexander University of Erlangen-Nürnberg (RRZE).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2016.00009>

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
- Andersen, C., Cseh, R., Schülein, K., and Benz, R. (1998). Study of sugar binding to the sucrose-specific ScrY channel of enteric bacteria using current noise analysis. *J. Membr. Biol.* 164, 263–274. doi:10.1007/s002329900411
- Bauer, K., Struyve, M., Bosch, D., Benz, R., and Tommassen, J. (1989). One single lysine residue is responsible for the special interaction between polyphosphate and the outer membrane porin PhoE of *Escherichia coli*. *J. Biol. Chem.* 264, 16393–16398.
- Benz, R., and Bauer, K. (1988). Permeation of hydrophilic molecules through the outer membrane of gram-negative bacteria. *Eur. J. Biochem.* 176, 1–19.
- Benz, R., Schmid, A., and Vos-Scheperkeuter, G. H. (1987). Mechanism of sugar transport through the sugar-specific LamB channel of *Escherichia coli* outer membrane. *J. Membr. Biol.* 100, 21–29. doi:10.1007/BF02209137
- Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Berendsen, H. J., van der Spoel, D., and van Drunen, R. (1995). Gromacs: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56. doi:10.1016/0010-4655(95)00042-E
- Bertelshofer, F., Sun, L., Greiner, G., and Böckmann, R. A. (2015). Fast solver for implicit electrostatics of biomolecules. *Front. Bioeng. Biotechnol.* 3:186. doi:10.3389/fbioe.2015.00186
- Böckmann, R. A., de Groot, B. L., Kakorin, S., Neumann, E., and Grubmüller, H. (2008). Kinetics, statistics, and energetics of lipid membrane electroporation studied by molecular dynamics simulations. *Biophys. J.* 95, 1837–1850. doi:10.1529/biophysj.108.129437
- Christ, C. D., Mark, A. E., and van Gunsteren, W. F. (2010). Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* 31, 1569–1582. doi:10.1002/jcc.21450
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089. doi:10.1063/1.464397
- Death, A., Notley, L., and Ferenci, T. (1993). Derepression of LamB protein facilitates outer membrane permeation of carbohydrates into *Escherichia coli* under conditions of nutrient stress. *J. Bacteriol.* 175, 1475–1483.
- DeLano, W. L., and Bromberg, S. (2004). *Pymol User's Guide*. San Carlos, CA: DeLano Scientific LLC.
- Dutzler, R., Wang, Y., Rizkallah, P., Rosenbusch, J., and Schirmer, T. (1996). Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* 4, 127–134. doi:10.1016/S0969-2126(96)00016-0
- Forst, D., Welte, W., Wacker, T., and Diederichs, K. (1998). Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Mol. Biol.* 5, 37–46. doi:10.1038/nsb0198-37
- García, J. L. (1985). Cloning in *Escherichia coli* and molecular analysis of the sucrose system of the *Salmonella* plasmid SCR-53. *Mol. Gen. Genet.* 201, 575–577. doi:10.1007/BF00331358
- Glauert, A. M., and Thornley, M. J. (1968). The topography of the bacterial cell wall. *Annu. Rev. Microbiol.* 23, 159–198. doi:10.1146/annurev.mi.23.100169.001111
- Hardesty, C., Ferran, C., and DiRienzo, J. M. (1991). Plasmid-mediated sucrose metabolism in *Escherichia coli*: characterization of scrY, the structural gene for a phosphoenolpyruvate-dependent sucrose phosphotransferase system outer membrane porin. *J. Bacteriol.* 173, 449–456.
- Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997). Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-L
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi:10.1021/ct700301q
- Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38(Suppl. 2), W545–W549. doi:10.1093/nar/gkq366
- Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695. doi:10.1103/PhysRevA.31.1695
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi:10.1002/prot.21123
- Hub, J. S., de Groot, B. L., and van der Spoel, D. (2010). g-wham – free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* 6, 3713–3720. doi:10.1021/ct100494z
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K. (2001). Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* 19, 13–25. doi:10.1016/S1093-3263(00)00133-9
- Jämbeck, J. P., and Lyubartsev, A. P. (2012a). Another piece of the membrane puzzle: extending slipids further. *J. Chem. Theory Comput.* 9, 774–784. doi:10.1021/ct300777p
- Jämbeck, J. P., and Lyubartsev, A. P. (2012b). Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B* 116, 3164–3179. doi:10.1021/jp221503e
- Jämbeck, J. P., and Lyubartsev, A. P. (2012c). An extension and further validation of an all-atomistic force field for biological membranes. *J. Chem. Theory Comput.* 8, 2938–2948. doi:10.1021/ct300342n
- Jap, B. K. (1989). Molecular design of PhoE porin and its functional consequences. *J. Mol. Biol.* 205, 407–419. doi:10.1016/0022-2836(89)90351-3
- Jordy, M., Andersen, C., Schülein, K., Ferenci, T., and Benz, R. (1996). Rate constants of sugar transport through two LamB mutants of *Escherichia coli*: comparison with wild-type maltoporin and LamB of *Salmonella typhimurium*. *J. Mol. Biol.* 259, 666–678. doi:10.1006/jmbi.1996.0348
- Jorgensen, W. L., Chandrasekhar, J., and Madura, J. D. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., et al. (2008). Glycam06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29, 622–655. doi:10.1002/jcc.20820
- Kullman, L., Winterhalter, M., and Bezrukov, S. M. (2002). Transport of maltodextrins through maltoporin: a single-channel study. *Biophys. J.* 82, 803–812. doi:10.1016/S0006-3495(02)75442-8
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi:10.1002/prot.22711
- Lou, K.-L., Saint, N., Prilipov, A., Rummel, G., Benson, S. A., Rosenbusch, J. P., et al. (1996). Structural and functional characterization of OmpF porin mutants selected for larger pore size I. Crystallographic analysis. *J. Biol. Chem.* 271, 20669–20675. doi:10.1074/jbc.271.34.20669
- Modi, N., Ganguly, S., Bárcena-Uribarri, I., Benz, R., van den Berg, B., and Kleinekathöfer, U. (2015). Structure, dynamics, and substrate specificity of the OprO porin from *Pseudomonas aeruginosa*. *Biophys. J.* 109, 1429–1438. doi:10.1016/j.bpj.2015.07.035
- Nabedryk, E., Garavito, R., and Breton, J. (1988). The orientation of beta-sheets in porin. A polarized fourier transform infrared spectroscopic investigation. *Biophys. J.* 53, 671–676. doi:10.1016/S0006-3495(88)83148-5
- Nakae, T. (1976). Identification of the outer membrane protein of *E. coli* that produces transmembrane channels in reconstituted vesicle membranes. *Biochem. Biophys. Res. Commun.* 71, 877–884. doi:10.1016/0006-291X(76)90913-X
- Nikaido, H. (1992). Porins and specific channels of bacterial outer membranes. *Mol. Microbiol.* 6, 435–442. doi:10.1111/j.1365-2958.1992.tb01487.x
- Nikaido, H. (1994). Porins and specific diffusion channels in bacterial outer membranes. *J. Biol. Chem.* 269, 3905–3908.
- Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* 67, 593–656. doi:10.1128/MMBR.67.4.593-656.2003
- Nikaido, H., and Nakae, T. (1980). The outer membrane of Gram-negative bacteria. *Adv. Microb. Physiol.* 20, 163–250. doi:10.1016/S0065-2911(08)60208-8
- Nikaido, H., and Vaara, M. (1985). Molecular basis of bacterial outer membrane permeability. *Microbiol. Rev.* 49, 1.
- Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81, 511–519. doi:10.1063/1.447334
- Parr, T., Poole, K., Crockford, G., and Hancock, R. (1986). Lipopolysaccharide-free *Escherichia coli* OmpF and *Pseudomonas aeruginosa* protein P porins are functionally active in lipid bilayer membranes. *J. Bacteriol.* 165, 523–526.

- Pezeshki, S., Chimere, C., Bessonov, A. N., Winterhalter, M., and Kleinekathöfer, U. (2009). Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF. *Biophys. J.* 97, 1898–1906. doi:10.1016/j.bpj.2009.07.018
- Pluhackova, K., Wassenaar, T. A., and Böckmann, R. A. (2013). “Molecular dynamics simulations of membrane proteins,” in *Membrane Biogenesis, Volume 1033 of Methods in Molecular Biology*, eds D. Rapoport and J. M. Herrmann (New York: Humana Press), 85–101.
- Randall-Hazelbauer, L., and Schwartz, M. (1973). Isolation of the bacteriophage lambda receptor from *Escherichia coli*. *J. Bacteriol.* 116, 1436–1446.
- Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021. doi:10.1002/jcc.540130812
- Samanta, S., Scorciapino, M. A., and Ceccarelli, M. (2015). Molecular basis of substrate translocation through the outer membrane channel OprD of *Pseudomonas aeruginosa*. *Phys. Chem. Chem. Phys.* 17, 23867–23876. doi:10.1039/c5cp02844b
- Saurin, W., Francoz, E., Martineau, P., Charbit, A., Dassa, E., Duplay, P., et al. (1989). Periplasmic binding protein dependent transport system for maltose and maltodextrins: some recent studies. *FEMS Microbiol. Lett.* 63, 53–60. doi:10.1016/0168-6445(89)90008-9
- Schirmer, T. (1998). General and specific porins from bacterial outer membranes. *J. Struct. Biol.* 121, 101–109. doi:10.1006/jsbi.1997.3946
- Schirmer, T., Keller, T. A., Wang, Y.-F., and Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* 267, 512–514. doi:10.1126/science.7824948
- Schmid, K., Ebner, R., Altenbuchner, J., Schmitt, R., and Lengeler, J. (1988). Plasmid-mediated sucrose metabolism in *Escherichia coli* K12: mapping of the scr genes of pUR400. *Mol. Microbiol.* 2, 1–8. doi:10.1111/j.1365-2958.1988.tb00001.x
- Schmid, K., Ebner, R., Jahreis, K., Lengeler, J., and Titgemeyer, F. (1991). A sugar-specific porin, ScrY, is involved in sucrose uptake in enteric bacteria. *Mol. Microbiol.* 5, 941–950. doi:10.1111/j.1365-2958.1991.tb00769.x
- Schmid, K., Schupfner, M., and Schmitt, R. (1982). Plasmid-mediated uptake and metabolism of sucrose by *Escherichia coli* K-12. *J. Bacteriol.* 151, 68–76.
- Schüle, K., Andersen, C., and Benz, R. (1995). The deletion of 70 amino acids near the N-terminal end of the sucrose-specific porin ScrY causes its functional similarity to LamB in vivo and in vitro. *Mol. Microbiol.* 17, 757–767. doi:10.1111/j.1365-2958.1995.mm1_17040757.x
- Schüle, K., and Benz, R. (1990). Lamb (maltoporin) of *Salmonella typhimurium*: isolation, purification and comparison of sugar binding with LamB of *Escherichia coli*. *Mol. Microbiol.* 4, 625–632. doi:10.1111/j.1365-2958.1990.tb00631.x
- Schüle, K., Schmid, K., and Benz, R. (1991). The sugar-specific outer membrane channel ScrY contains functional characteristics of general diffusion pores and substrate-specific porins. *Mol. Microbiol.* 5, 2233–2241. doi:10.1111/j.1365-2958.1991.tb02153.x
- Smart, O. S., Breed, J., Smith, G. R., and Sansom, M. (1997). A novel method for structure-based prediction of ion channel conductance properties. *Biophys. J.* 72, 1109. doi:10.1016/S0006-3495(97)78760-5
- Smart, O. S., Neduvellil, J. G., Wang, X., Wallace, B., and Sansom, M. S. (1996). Hole: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* 14, 354–360. doi:10.1016/S0263-7855(97)00009-X
- Szmecman, S., and Hofnung, M. (1975). Maltose transport in *Escherichia coli* K-12: involvement of the bacteriophage lambda receptor. *J. Bacteriol.* 124, 112–118.
- Tessier, M. B., DeMarco, M. L., Yongye, A. B., and Woods, R. J. (2008). Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Mol. Simul.* 34, 349–364. doi:10.1080/08927020701710890
- Wang, Y.-F., Dutzler, R., Rizkallah, P. J., Rosenbusch, J. P., and Schirmer, T. (1997). Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.* 272, 56–63. doi:10.1006/jmbi.1997.1224
- Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P., and Marrink, S. J. (2015). Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* 11, 2144–2155. doi:10.1021/acs.jctc.5b00209
- Weiss, M., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J., et al. (1991). The structure of porin from *Rhodobacter capsulatus* at 1.8 Å resolution. *FEBS Lett.* 280, 379–382. doi:10.1016/0014-5793(91)80336-2
- Welte, W., Nestel, U., Wacker, T., and Diederichs, K. (1995). Structure and function of the porin channel. *Kidney Int.* 48, 930–940. doi:10.1038/ki.1995.374
- Wiese, A., Reiners, J. O., Brandenburg, K., Kawahara, K., Zähringer, U., and Seydel, U. (1996). Planar asymmetric lipid bilayers of glycosphingolipid or lipopolysaccharide on one side and phospholipids on the other: membrane potential, porin function, and complement activation. *Biophys. J.* 70, 321–329. doi:10.1016/S0006-3495(96)79573-5
- Wiese, A., Schröder, G., Brandenburg, K., Hirsch, A., Welte, W., and Seydel, U. (1994). Influence of the lipid matrix on incorporation and function of LPS-free porin from *Paracoccus denitrificans*. *Biochim. Biophys. Acta* 1190, 231–242. doi:10.1016/0005-2736(94)90079-5

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 © 2016 Sun, Bertelshofer, Greiner and Böckmann. 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.



GroPBS: Fast Solver for Implicit Electrostatics of Biomolecules

Franziska Bertelshofer^{1,2*}, Liping Sun², Günther Greiner¹ and Rainer A. Böckmann^{2*}

¹ Computer Graphics Group, Department of Computer Science, University Erlangen-Nürnberg, Erlangen, Germany,

² Computational Biology, Department of Biology, University Erlangen-Nürnberg, Erlangen, Germany

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

Zhaoyong Yang,
Chinese Academy of Medical
Sciences, China
Daehee Lee,
Korea Research Institute of
Bioscience and Biotechnology,
South Korea

*Correspondence:

Franziska Bertelshofer
franziska.bertelshofer@fau.de;
Rainer A. Böckmann
rainer.boeckmann@fau.de

Specialty section:

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

Received: 11 September 2015

Accepted: 30 October 2015

Published: 17 November 2015

Citation:

Bertelshofer F, Sun L, Greiner G and
Böckmann RA (2015)
GroPBS: Fast Solver for Implicit
Electrostatics of Biomolecules.
Front. Bioeng. Biotechnol. 3:186.
doi: 10.3389/fbioe.2015.00186

Knowledge about the electrostatic potential on the surface of biomolecules or biomembranes under physiological conditions is an important step in the attempt to characterize the physico-chemical properties of these molecules and, in particular, also their interactions with each other. Additionally, knowledge about solution electrostatics may also guide the design of molecules with specified properties. However, explicit water models come at a high computational cost, rendering them unsuitable for large design studies or for docking purposes. Implicit models with the water phase treated as a continuum require the numerical solution of the Poisson–Boltzmann equation (PBE). Here, we present a new flexible program for the numerical solution of the PBE, allowing for different geometries, and the explicit and implicit inclusion of membranes. It involves a discretization of space and the computation of the molecular surface. The PBE is solved using finite differences, the resulting set of equations is solved using a Gauss–Seidel method. It is shown for the example of the sucrose transporter ScrY that the implicit inclusion of a surrounding membrane has a strong effect also on the electrostatics within the pore region and, thus, needs to be carefully considered, e.g., in design studies on membrane proteins.

Keywords: electrostatics, Poisson–Boltzmann equation, finite-difference method, molecular surface, membranes

1. INTRODUCTION

Electrostatic interactions govern the physical–chemical interactions in and between biomolecules (Perutz, 1978). A quantitative description of electrostatic energies (Warshel et al., 2006) is required both for a thorough understanding of biomolecular systems, e.g., of membrane-embedded ion channels or of pK_a changes during enzymatic function and in protein design, e.g., the design of novel protein folds or in the search for high-affinity ligands.

Coulombic forces are modulated by the environment (Warshel et al., 2006), i.e., in case of soluble proteins by water and ions and possibly other proteins, and additionally by phospholipids in the case of membrane proteins. In atomistic molecular dynamics (MD) simulations (Karplus and McCammon, 2002), this environment is treated explicitly for a proper description, in particular, of the local electrostatics, e.g., in water-mediated hydrogen bonds. Computationally less demanding models introduced uniform dielectric constants for both the protein and the solvent (Tanford and Kirkwood, 1957), a distance-dependent dielectric constant accounting for electrostatic shielding within the solvent (Brooks et al., 1983), or modeled the solvent using an explicit grid of Langevin dipoles (Warshel and Levitt, 1976).

The electrostatic contribution to free energy differences between two states of a biomolecular system – e.g., two proteins bound to each other vs. the two proteins separated in space – is in many cases difficult to directly access; however, it may be computed from ensembles of microscopic structures. The simulation-based free energy perturbation (FEP) approach or the thermodynamic

integration (TI) method is frequently used to analyze free energy difference of biomolecular states. The computational cost of such methods is, however, at variance with the need for high throughput, e.g., in protein design or protein–ligand docking and in the pK_a analysis of titratable sites as well. The latter problem requires the accurate estimation of the free energy differences between protonated and de-protonated states of all titratable groups in proteins, either on crystal structures or on trajectories obtained from molecular dynamics simulations to better grasp the protein flexibility (Narzi et al., 2008a). Protonation or pK changes are, e.g., important during the function of enzymes (Narzi et al., 2008a), or in protein–ligand binding (Narzi et al., 2008b; Onufriev and Alexov, 2013).

Due to the computational efficiency required to tackle the above problems, the solvation free energy is usually addressed in implicit models by solving the Poisson–Boltzmann equation (PBE) for the different states [see, e.g., Ullmann and Bombarda (2014)]. Different values for the dielectric continuum inside the protein have been used in the literature. The states may be crystal structures, modeled structures, snapshots from molecular dynamics simulations [MM/PBSA (Kollman et al., 2000)], or structural ensembles generated from crystal structures [CC/PBSA (Benedix et al., 2009)]. The solution of the PBE on structural ensembles should yield a more accurate solution for the solvation free energies of the studied biomolecules as it takes the flexibility of the protein into account.

A number of program packages for the numerical solution of the Poisson–Boltzmann equation have been developed in the past, namely APBS (Baker et al., 2001), UHBD (Madura et al., 1995), or DelPhi (Li et al., 2012), to mention a few. Here, we present a sequential PBE solver (*GroPBS*) that allows the accurate analysis of the electrostatic component of the solvation free energy of soluble proteins and of membrane proteins in explicit or implicit membranes. *GroPBS* is compatible with the GROMACS simulation suite and can, thus, easily be combined with simulations of biomolecules and the various biomolecular force fields available in Gromacs.

2. METHODS

2.1. Poisson–Boltzmann Equation

The analysis of the electrostatic potential Φ of a biomolecule in an implicitly treated solvent environment requires the numerical solution of the Poisson–Boltzmann equation (PBE). For the most simple case of a solute molecule in a homogeneous medium, three different domains (Holst, 1994) may be distinguished (see **Figure 1**):

1. Inside the molecule (green) (i.e., within the van der Waals radii of the atoms of the solute) Φ can be computed using the Poisson equation and Green's identity resulting in

$$\nabla^2 \Phi(x) = \sum_{i=1}^M \frac{-4\pi q_i}{\varepsilon_1} \delta(x - x_i)$$

where M is the number of atoms of the solute with point charges q_i at positions x_i .

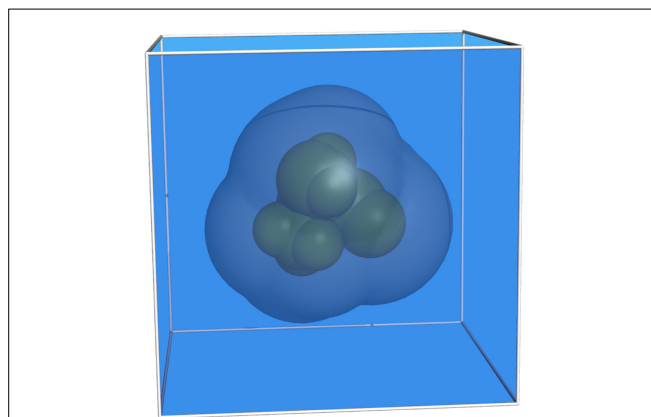


FIGURE 1 | Representation of a molecule consisting of nine atoms. The different regions are colored as follows: green, van der Waals surface (inside); blue, solvent; dark blue, ion exclusion layer.

2. Outside the molecule (light blue), the charge density of the ions in the solvent is assumed to follow a Boltzmann distribution that yields

$$\nabla^2 \Phi(x) = \kappa^2 \frac{k_B T}{e_c} \sinh \left(\frac{e_c \Phi(x)}{k_B T} \right),$$

k_B is the Boltzmann constant, e_c the elementary charge, T the temperature, and κ is the modified Debye–Hückel parameter, given by

$$\kappa^2 = \frac{8\pi N_A e_c^2 I_s}{1000 \varepsilon_3 k_B T}$$

with N_A Avogadro's number and I_s the ionic strength of the solvent.

3. In the ion exclusion layer (dark blue) around the molecule no mobile ions are present. The Poisson equation reads accordingly

$$\nabla^2 \Phi(x) = 0.$$

Combining these conditions into one equation, results in the non-linear PBE

$$-\nabla(\varepsilon(x)\nabla\Phi(x)) + \kappa^2(x) \frac{k_B T}{e_c} \sinh \left(\frac{e_c \Phi(x)}{k_B T} \right) = 4\pi\rho(x) \quad (1)$$

with the charge distribution function $\rho(x)$ of the molecule and the spatial relative dielectric function $\varepsilon(x)$ and $\kappa^2(x)$, which is 0 in the solute and the ion exclusion layer and $\kappa^2(x) = \varepsilon_3 \kappa^2$ for x in the solvent. $\varepsilon(x)$ allows for the description of the dielectric discontinuity between the molecule and the solvent and is typically chosen to adopt values between 2 and 20 inside the biomolecule and 80 outside. More complicated models for the dielectric “constant” taking smoothed boundaries into account were suggested recently by Li et al. (2013) and will be considered for future work.

The framework provided by equation (1) allows the straightforward inclusion of various regions with different dielectric properties, e.g., of membranes. The membrane is modeled as another bulk medium like the solvent but with its own dielectric constant [$\varepsilon_{\text{mem}} \approx 2 \dots 3$ (Böckmann et al., 2008)]. In a first step, the

membrane is treated as a box in the xy -plane represented by its upper and lower z -value. However, more sophisticated models like curved membranes or taking the surface structure into account may easily be implemented.

The above partial differential equation cannot be solved analytically for objects shaped more complex than, e.g., a single sphere or cylinder. Therefore, the PBE has to be solved numerically.

2.2. Finite-Difference Method

The first step in the numerical solution of the PBE is to map all physical quantities (charges at atom centers, dielectric values, etc.) onto a three-dimensional uniform grid. Such a grid allows to replace differential operators by grid value differences. This approach is facilitated by linearizing the PBE [LPBE (Holst, 1994)]; For $\sinh(x) \approx x$ the latter simplifies to

$$-\nabla(\varepsilon(x)\nabla\Phi(x)) + \kappa^2(x)\Phi(x) = 4\pi\rho(x). \quad (2)$$

Discretization of this equation yields for every grid point

$$\Phi_0 = \frac{\sum_{k=1}^6 \varepsilon_k \Phi_k + 4\pi q_0/h}{\sum_{i=k}^6 \varepsilon_k + (\kappa_0 \cdot h)^2}, \quad (3)$$

q_0 and κ_0 denote the charge and Debye-Hückel parameter at the grid point, Φ_k is the potentials at the six neighboring grid positions (in x -, y -, and z -direction), and ε_k is the dielectric values at the midpoints between Φ_0 and its neighbors Φ_k (see Figure 2) (Nicholls and Honig, 1991). h is the step size, i.e., the distance between the grid points. While charge and the Debye-Hückel parameter are given values at the grid points, one important step is to define in which medium a grid point or midpoint is located. This problem is described in more detail below.

Application of equation (3) for each grid point results in a system of N^3 linear equations where N is the grid size. This system can be reformulated as

$$\Phi = M\Phi + v, \quad (4)$$

with Φ as a vector containing the potential at all grid points and M being a sparse matrix containing zeros at positions (i, j) if i and j are not neighboring grid points and $\frac{\varepsilon_j}{\sum_{k=1}^6 \varepsilon_k + (\kappa_0 \cdot h)^2}$ at positions (i, j) else, with k denoting all neighboring grid points of i . v is a vector containing the remaining terms of the discretized LPBE.

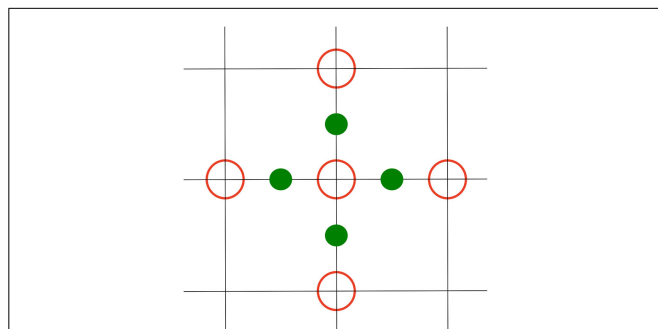


FIGURE 2 | 2D representation of the computation scheme for equation (3): The inner red circle Φ_0 is computed using the outer red circles Φ_k , and the green circles ε_k .

2.3. Successive Over-Relaxation

Several alternatives were suggested for the treatment of the grid boundaries: setting the electrostatic potential at the boundaries to 0, application of distance-dependent quasi-Coulombic boundary conditions, or periodic extensions of the system box in one or more directions.

The set of linear equations (4) is iteratively solved using methods like Jacobi or Gauss-Seidel (Demmel, 1997). Here, we used in a first serial implementation a successive over-relaxation (SOR) for Gauss-Seidel yielding the iteration rule

$$\Phi^{(n+1)} = \omega\Phi^{new} + (1 - \omega)\Phi^{(n)}, \quad (5)$$

n is the iteration step, Φ^{new} is computed using equation (3). $\omega > 1$ is the relaxation parameter. In a sequential implementation, every newly computed grid point value is immediately used for computing further grid points in the same iteration step – in contrast to the Jacobi method that renders the Gauss-Seidel approach much faster. Since the convergence rate of the Gauss-Seidel method depends on the spectral radius, i.e., the largest eigenvalue of M , ω should be chosen in a way it makes the spectral radius smaller. It can be shown that the optimal value is given by

$$\omega = \frac{2}{1 + \sqrt{1 - \lambda_N}},$$

λ_N is the spectral radius. The spectral radius of M can be computed using the Connected-Moments Expansion (Cioslowski, 1987; Nicholls and Honig, 1991).

2.4. Solvation Free Energy

The grid-based electrostatic potential is used to compute the solvation free energy of the system. This is achieved by summing the product of the potential and the charge at each grid point, followed by subtraction of the corresponding energy as obtained for the potential using a uniform dielectric inside and outside of the solute. This approach eliminates the self-energy terms that are physically not meaningful. The disadvantage of this approach is the required double computation of the electrostatic potential.

A different approach can be used if the molecular surface is known (see below). The reaction field effects due to a dielectric boundary are replaced by the computation of the induced charges at this boundary (Rocchia et al., 2002). The solvation energy G_W may then be calculated applying Coulomb's law for the induced and the real charges. Formalizing this approach yields the following equation:

$$G_W = 0.5 \cdot \sum_{b \in \text{boundary}} \left(\sum_{p \in \text{grid}} \frac{q_p}{\text{dist}(b_s, p)} \right) \cdot \left(\frac{3h}{2\pi} \left(\Phi_b - \frac{1}{6} \sum_{k=1}^6 \Phi_k \right) - q_b \right), \quad (6)$$

b denotes the grid points at the dielectric boundary, p all grid points (these can be reduced to all charged grid points, as for all other points the term is 0) and b_s is the points at the molecular surface obtained by projecting the boundary grid points to the surface.

2.5. Optimizations

Some simplifications of equation (3) are possible (Nicholls and Honig, 1991): first, most grid points do not hold a point charge, i.e., the term $4\pi q_0/h$ is equal to 0. Second, most grid points are not found at a dielectric boundary, i.e., all neighboring midpoints are located in only one medium. These modifications lead to a quite simple 7-point stencil (modified by salt if present)

$$\underbrace{\Phi_0 = \frac{1}{6} \sum_{k=1}^6 \Phi_k}_{\text{without salt}} \quad \text{or} \quad \underbrace{\Phi_0 = \frac{\sum_{k=1}^6 \Phi_k}{6 + \left(\frac{(\kappa_0 \cdot h)^2}{\epsilon_0} \right)}}_{\text{with salt}} \quad (7)$$

In each iteration step of the SOR first this stencil is used, followed by analysis of the correction for dielectric discontinuities and charges if necessary. This approach is also a first step for a parallelization as the uniform stencil in equation (7) suits to be applied in parallel.

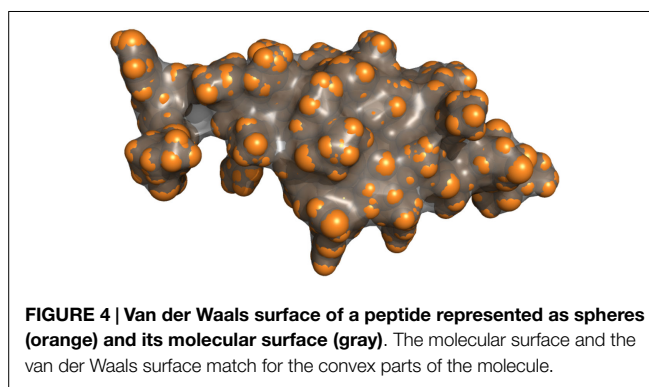
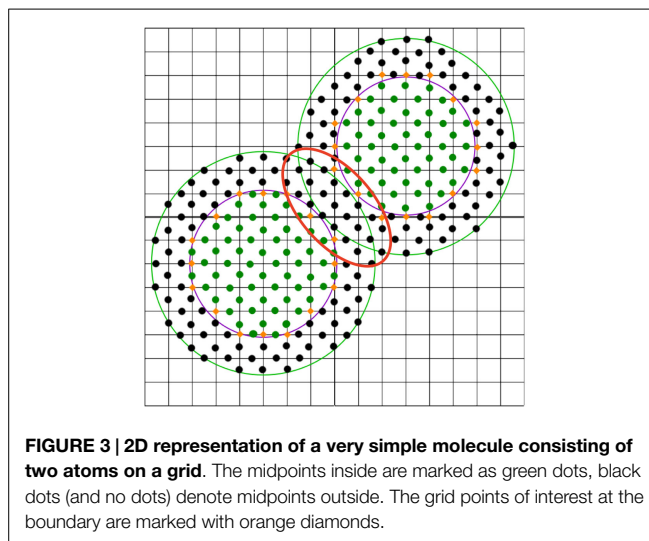
2.6. Extraction of Molecular Surface

In order to define which grid points lie inside and outside of the solute, the assignment of dielectric constants to the midpoints and, in particular for the computation of the free energy in a more sophisticated way (see above), the solute surface needs to be defined. The molecular surface is determined by its implicit description from the atom positions and radii of the solute. This surface is defined as the contact surface between the van der Waals surface and the surface of a spherical probe representing the solvent.

The first step is to determine a van der Waals map by mapping the atoms onto the grid. Every midpoint is categorized as inside or outside of the solute. This is done by examining all midpoints inside a box surrounding each atom and comparing their distance to the atom center with the radius of the atom. The same approach is used to determine which grid points are in solution, which is important for knowing which grid points' stencil has to be modified by salt. Additionally, the grid points are classified as internal points if all surrounding midpoints are inside, external points if all surrounding midpoints are outside, or boundary points if some midpoints are in solution and some are not.

Taking into account the probe radius, some of these points have to be reclassified as sketched in the red marked region in **Figure 3**. This is done by iteratively examining the midpoints that surround the boundary points. For each such point, the distance to the solvent accessible surface is computed. This distance is defined as the distance from the van der Waals surface (violet line in **Figure 3**) extended by the radius of the probe, e.g., by 1.4 Å for water (one has to take into account that these extended atoms can overlap, see green line in **Figure 3**). If this distance is smaller than the probe radius, the midpoint remains outside; otherwise, it will be turned into an inside midpoint. Next, the grid points are reclassified using the new midpoints. This is repeated until no new boundary grid points are produced.

The actual surface points are finally constructed by projecting the boundary grid points onto the molecular surface either directly by moving them along the line connecting the grid point



and the nearest atom center or by projecting them first onto the closest point of the solvent accessible surface and then back on the molecular surface in a similar way (Rocchia et al., 2002).

Figure 4 shows the result of the surface computation for a small peptide. In the convex parts, the molecular surface corresponds to the van der Waals surface (brightly colored parts), whereas they differ in concave parts and cavities.

3. RESULTS

3.1. Evaluation of Current Program

The approach described in the Section “Methods” has been implemented in a sequential program using C++. As input separate files containing the atom positions (“pdb,” Protein Data Bank format) and force field parameters, like in the program Delphi (Li et al., 2012) can be used as well as the Gromacs (Pronk et al., 2013) input file format (“tpr”) for atom positions, Lennard Jones parameters for the atom sizes, and partial atomic charges. The leading biomolecular force fields for molecular modeling such as CHARMM, GROMOS, OPLS, or Amber are supported by Gromacs. This combination with the widely used Gromacs simulation package considerably simplifies the usage and enhances the applicability of the presented PBE solver in combination with biomolecular simulations, e.g., in the analysis of protein–protein binding free energies using the MM/PBSA approach.

TABLE 1 | Possible parameters.

in(tp, path) in(pdb, path) in(siz, path) in(crg, path) in(sph, path)	Gromacs input file Protein positions Protein sizes Protein charges Positions in pdb format	Can also be given in the command line Alternative to tpr Necessary with pdb, optional with tpr Necessary with pdb, optional with tpr Optional, to compute the potential Along a path or a specific postilions
Filling Spacing gridS	Percentage of box that is filled with protein Spacing (in Å) between two grid points Grid size, i.e., number of grid points in each direction (odd)	Default: 80% Default: 1 Å Default: is computed Two of these three parameters Can be chosen
rmsc maxc maxit	Convergence criterion Convergence criterion Maximal iteration	Default: 0.0001 Default: 0.0001 Default: 500
epsIn epsOut salt temp	Dielectric constant inside the molecule Dielectric constant in the solution Concentration in moles/liter Temperature	Default: 2.0 Default: 80.0 Default: 0.0 Default: 273.15
bc = {1,2} pb = xyz	Boundary conditions Periodic boundary conditions in x-, y-, z-direction	1: 0-boundary, 2: quasi-Coulombic Default: false
nomem mem = zmin, zmax, eps	No membrane input Minimal and maximal spread of membrane In z-direction and its dielectric constant	Default: no information Default: no information
tprmem = res, atm, eps	Residue and atoms to be considered membrane in tpr Its dielectric constant	Default: no information

The above-described input files and additional parameters like grid size, percentage of filling, probe radius, etc. are listed in one parameter file that serves as input. **Table 1** contains all currently available parameters.

In the following, some of these parameters are explained in more detail.

3.1.1. Membranes

Membranes are modeled as a slab in the *xy*-plane defined in extent by their minimal and maximal *z*-value. There are several possibilities to define the membrane extension:

- If the exact values are known beforehand, the `mem = zmin,zmax,eps` option in the parameter file can be used.
- These values can also be entered in the interactive mode after the *z*-extent of the molecule is reported.
- If a membrane is included in the “tpr” file, with the `tprmem = res,atm,eps` option in the parameter file (or in the interactive mode) the associated residue names and atoms (e.g., POPC as residuum and P as atom) can be used to specify the membrane. The chosen atoms are separated into upper and lower parts of the membrane and the according center of mass is used as *zmin* and *zmax*.

3.1.2. Boundary conditions

There are three possibilities to specify boundary conditions. These have to be fixed, as the grid points at the grid’s boundary do not have enough neighbors to be computed directly. The first and easiest way is to set the boundary grid points to 0. Another

possibility is to approximate the potential at the boundary using quasi-Coulombic dipole conditions:

$$\Phi(x) = \frac{c_+ \cdot \exp(\frac{-d_+}{\lambda})}{d_+ \cdot \epsilon_{sol}} + \frac{c_- \cdot \exp(\frac{-d_-}{\lambda})}{d_- \cdot \epsilon_{sol}},$$

where ϵ_{sol} is the dielectric constant of the solvent, c_+ and c_- are the total positive and negative charges, d_+ and d_- are the distance of the grid point to the center of the positive and negative charges, and λ denotes the Debye length.

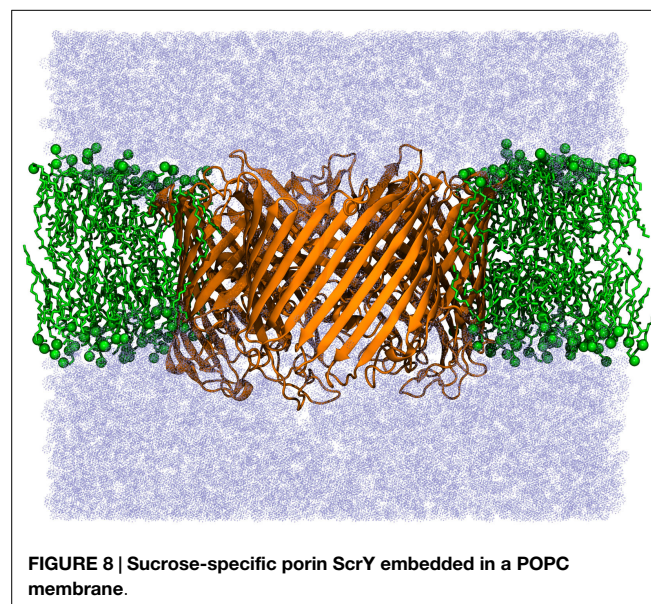
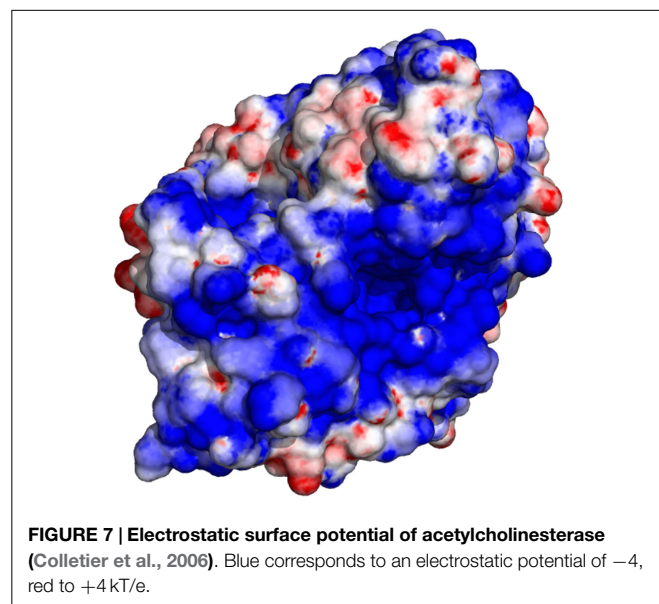
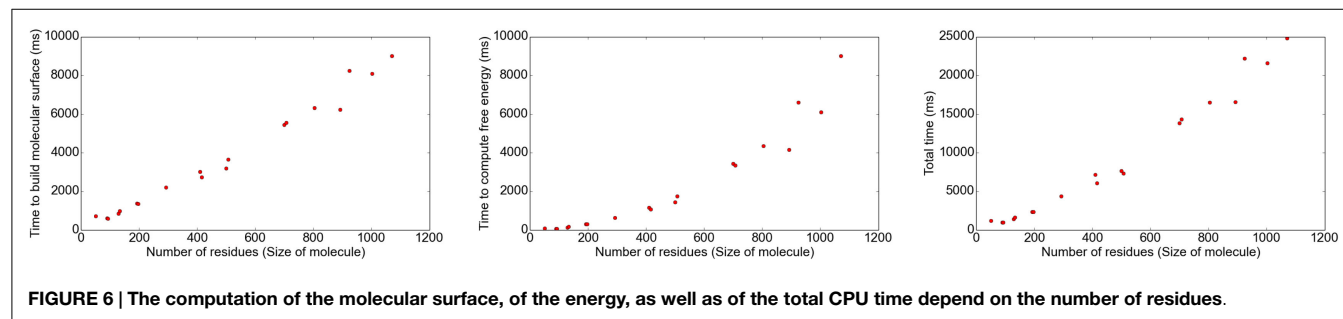
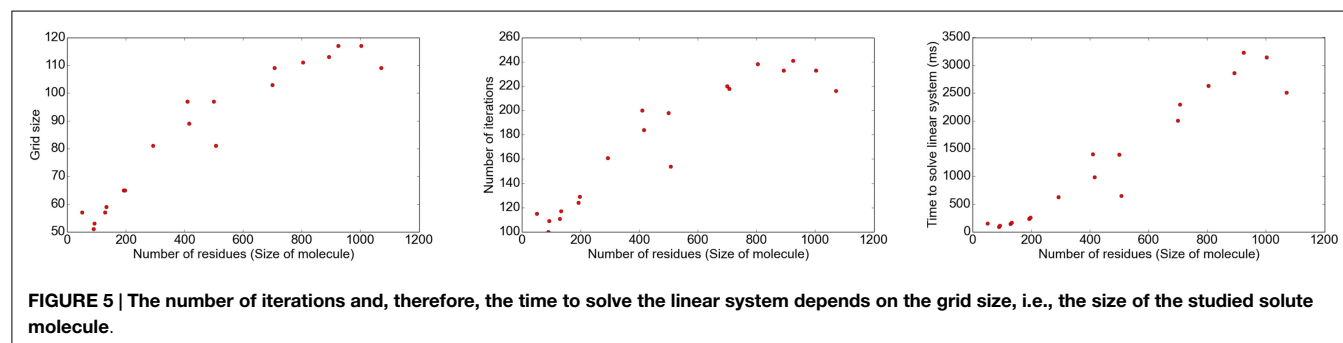
Alternatively, periodic boundary conditions can be used in each grid dimension separately. Then for the missing neighbors, the corresponding points at the opposite side of the grid are used. This provides the opportunity to model infinite boxes.

3.1.3. Convergence Criterion

There are three possible ways to define the convergence of the iterative procedure. The first is to set a threshold on the root mean squared change (rmsc), defined as

$$rmsc^{(i)} = \sqrt{\frac{1}{M} \sum_x (\Phi^{(i)}(x) - \Phi^{(i-1)}(x))^2},$$

M is the number grid points and *i* is the iteration count. The rmsc measures the mean differences in the potential between two iterations. The option maxc limits the maximal change in the potential between two iterations. The third option allows to limit the number of iterations without regarding the convergence at all. Of course, these criteria can be combined, stopping the iteration process as soon as one criterion is fulfilled.



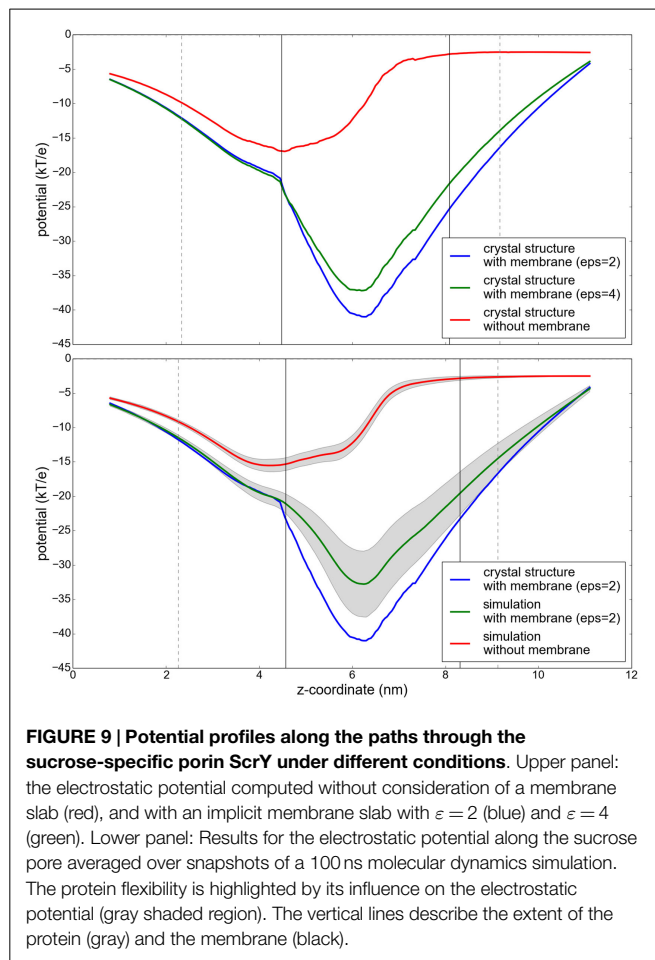
The implementation was tested using molecules of different sizes, ranging from small proteins with 49 residues to proteins as large as 1,070 residues (1,260–16,260 atoms, respectively). The grid size was chosen such that the filling rate was $\approx 80\%$ using a grid size of 1 Å. Different results regarding the correlation between the number of atoms and grid size, number of iterations, and time consumption for different parts of the program are shown in **Figures 5** and **6**.

The results show that the number of iterations and, therefore, the time to solve the linear system strongly depends on the number of grid points. The surface computation and energy calculation, however, depend on the size of the molecule.

The electrostatic potential computed using GroPBS may be mapped onto the surface of biomolecules using, e.g., PyMOL (Schrödinger, 2010). As an example, **Figure 7** shows the surface potential of acetylcholinesterase.

3.2. Influence of an Implicit Membrane

In order to evaluate the influence of the low dielectric of a lipid membrane on the electrostatic potential of an embedded membrane protein, the potential was compared for the sucrose-specific porin ScrY (Forst et al., 1998) in different environments. **Figure 8** shows this porin embedded in a POPC bilayer. The electrostatic



potential was analyzed along a path through the pore of each monomer of the ScrY homo-trimer. This path was obtained using the program hole (Smart et al., 1993). For the solution of the PBE, the grid was chosen such that the filling rate was $\approx 80\%$ using a step size of 0.5 Å and periodic boundary conditions in lateral direction (membrane plane) and dipole boundary conditions normal to the membrane. The dielectric constants were set to 4.0 inside the protein and 80.0 in the solvent phase. The dielectric of the surrounding implicit membrane was varied to study the influence of the membrane on the central, membrane-distal pore region. The PBE was solved every 0.5 ns of a 100 ns simulation, excluding the initial equilibration period of 20 ns. The electrostatic potential was averaged over the trajectory.

The obtained electrostatic potential through the ScrY pore (see Figure 9, upper panel) is drastically decreased if an implicit

membrane is included in the calculations. While the shape of the potential is similar, the minimum is shifted by approximately 1.5 nm, from the membrane interfacial region to the interior of the pore. The additional inclusion of flexibility by averaging the potential along a molecular dynamics trajectory results in an increase of the potential by up to 7 kT/e (Figure 9, lower panel).

4. DISCUSSION AND FUTURE WORK

A new program for the numerical solution of the Poisson-Boltzmann equation around biological macromolecules is presented (GroPBS). Apart from soluble proteins, GroPBS may as well be used to analyze the electrostatic potential of integral membrane proteins. The low-dielectric membrane environment may be modeled implicitly or explicitly. Additionally, GroPBS is shown to efficiently perform such computations both on pdb files and using the Gromacs input file format. This significantly simplifies the fast calculation of, e.g., the solvation free energies of biomolecules for different force fields or on ensembles of structures obtained from molecular dynamics simulations.

On the example of the sucrose-specific porin ScrY, we show that the inclusion of a membrane may have a substantial influence also on the potential inside the protein, and thus should not be neglected in PB calculations of membrane proteins.

In a subsequent step, GroPBS will be parallelized for multi-core architectures, and in particular, GPUs to enable, e.g., the fast analysis of changes in pK_a values on the fly during biomolecular simulations. Parallelization may be easily achieved by the so-called checkerboard ordering in the update of the electrostatic potential (Adams and Ortega, 1982).

The program will be made available free of charge on the following website: www.biotechnik.nat.uni-erlangen.de/research/boeckmann/downloads/GroPBS.

AUTHOR CONTRIBUTIONS

Research was designed by FB and RB, performed by FB and LS, and the manuscript was written by all.

ACKNOWLEDGMENTS

We acknowledge support by the Deutsche Forschungsgemeinschaft within the Research Training Group grant No. 1962/1, Dynamic Interactions at Biological Membranes – From Single Molecules to Tissue (FB, LS, and RB). This work was also supported by the Emerging Field Initiative Synthetic Biology of the Friedrich-Alexander University of Erlangen-Nürnberg. Liping Sun was supported by the China Scholarship Council (CSC).

REFERENCES

- Adams, L., and Ortega, J. (1982). "A multi-color SOR method for parallel computation," in *Proceedings of the International Conference on Parallel Processing, ICPP'82* (Bellaire: IEEE), 53–56.
- Baker, N. A., Sept, D., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi:10.1073/pnas.181342398
- Benedix, A., Becker, C. M., de Groot, B. L., Caflisch, A., and Böckmann, R. A. (2009). Predicting free energy changes using structural ensembles. *Nat. Methods* 6, 3–4. doi:10.1038/nmeth0109-3

- Böckmann, R. A., De Groot, B. L., Kakorin, S., Neumann, E., and Grubmüller, H. (2008). Kinetics, statistics, and energetics of lipid membrane electroporation studied by molecular dynamics simulations. *Biophys. J.* 95, 1837–1850. doi:10.1529/biophysj.108.129437
- Brooks, R. B., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4, 187–217. doi:10.1002/jcc.540040211
- Cioslowski, J. (1987). Connected moments expansion: a new tool for quantum many-body theory. *Phys. Rev. Lett.* 58, 83. doi:10.1103/PhysRevLett.58.83

- Colletier, J. P., Fournier, D., Greenblatt, H. M., Stojan, J., Sussman, J. L., Zaccai, G., et al. (2006). Structural insights into substrate traffic and inhibition in acetylcholinesterase. *EMBO J.* 25, 2746. doi:10.1038/sj.emboj.7601175
- Demmel, J. W. (1997). *Applied Numerical Linear Algebra*. Philadelphia: Siam.
- Forst, D., Welte, W., Wacker, T., and Diederichs, K. (1998). Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Mol. Biol.* 5, 37–46. doi:10.1038/nsb0198-37
- Holst, M. J. (1994). *The Poisson–Boltzmann Equation: Analysis and Multilevel Numerical Solution*. PhD thesis, University of Illinois (UIUC), Caltech.
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi:10.1038/nsb0902-646
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897. doi:10.1021/ar000033j
- Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., et al. (2012). DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* 5:9. doi:10.1186/2046-1682-5-9
- Li, L., Li, C., Zhang, Z., and Alexov, E. (2013). On the dielectric “constant” of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* 9, 2126–2136. doi:10.1021/ct400065j
- Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. D., Ilin, A., et al. (1995). Electrostatics and diffusion of molecules in solution – simulations with the University of Houston Brownian Dynamics. *Comput. Phys. Commun.* 91, 57–95. doi:10.1016/0010-4655(95)00043-F
- Narzi, D., Siu, S. W. I., Stirnimann, C. U., Grimshaw, J. P. A., Glockshuber, R., Capitani, G., et al. (2008a). Evidence for proton shuffling in a thioredoxin-like protein during catalysis. *J. Mol. Biol.* 382, 978–986. doi:10.1016/j.jmb.2008.07.061
- Narzi, D., Winkler, K., Saidowski, J., Misselwitz, R., Ziegler, A., Böckmann, R. A., et al. (2008b). Molecular determinants of MHC class I complex stability: shaping antigenic features through long-range electrostatic interactions. *J. Biol. Chem.* 283, 23093–23103. doi:10.1074/jbc.M710234200
- Nicholls, A., and Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comp. Chem.* 12, 435–445. doi:10.1002/jcc.540120405
- Onufriev, A. V., and Alexov, E. (2013). Protonation and pK changes in protein–ligand binding. *Q. Rev. Biophys.* 46, 181–209. doi:10.1017/S0033583513000024
- Perutz, M. F. (1978). Electrostatics effects in proteins. *Science* 201, 1187–1191. doi:10.1126/science.694508
- Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi:10.1093/bioinformatics/btt055
- Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002). Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comp. Chem.* 23, 128–137. doi:10.1002/jcc.1161
- Schrödinger, L. L. C. (2010). *The PyMOL Molecular Graphics System, Version 1.3r1*.
- Smart, O. S., Goodfellow, J. M., and Wallace, B. (1993). The pore dimensions of gramicidin A. *Biophys. J.* 65, 2455. doi:10.1016/S0006-3495(93)81293-1
- Tanford, C., and Kirkwood, J. G. (1957). Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79, 5333–5339. doi:10.1021/ja01577a001
- Ullmann, G. M., and Bombarda, E. (2014). “Continuum electrostatics analysis of proteins,” in *Protein Modelling*, ed. G. Naray-Szabo (Switzerland: Springer), 135–163.
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249. doi:10.1016/0022-2836(76)90311-9
- Warshel, A., Sharma, P. K., Kato, M., and Parson, W. W. (2006). Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta* 1764, 1647–1676. doi:10.1016/j.bbapap.2006.08.007

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 Bertelshofer, Sun, Greiner and Böckmann. 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.



Model-Based Design of Biochemical Microreactors

Tobias Elbinger¹, Markus Gahn¹, Maria Neuss-Radu^{1*}, Falk M. Hante², Lars M. Voll³,
Günter Leugering² and Peter Knabner¹

¹ Chair of Applied Mathematics 1, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ² Chair of Applied Mathematics 2, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ³ Chair of Biochemistry, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

Xinyao Liu,
SABIC, Saudi Arabia
Jesus Picó,
Universitat Politècnica de València,
Spain

*Correspondence:

Maria Neuss-Radu
maria.neuss-radu@math.fau.de

Specialty section:

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

Received: 17 September 2016

Accepted: 28 January 2016

Published: 15 February 2016

Citation:

Elbinger T, Gahn M, Neuss-Radu M,
Hante FM, Voll LM, Leugering G and
Knabner P (2016) Model-Based
Design of Biochemical Microreactors.
Front. Bioeng. Biotechnol. 4:13.
doi: 10.3389/fbioe.2016.00013

Mathematical modeling of biochemical pathways is an important resource in Synthetic Biology, as the predictive power of simulating synthetic pathways represents an important step in the design of synthetic metabolons. In this paper, we are concerned with the mathematical modeling, simulation, and optimization of metabolic processes in biochemical microreactors able to carry out enzymatic reactions and to exchange metabolites with their surrounding medium. The results of the reported modeling approach are incorporated in the design of the first microreactor prototypes that are under construction. These microreactors consist of compartments separated by membranes carrying specific transporters for the input of substrates and export of products. Inside the compartments of the reactor multienzyme complexes assembled on nano-beads by peptide adapters are used to carry out metabolic reactions. The spatially resolved mathematical model describing the ongoing processes consists of a system of diffusion equations together with boundary and initial conditions. The boundary conditions model the exchange of metabolites with the neighboring compartments and the reactions at the surface of the nano-beads carrying the multienzyme complexes. Efficient and accurate approaches for numerical simulation of the mathematical model and for optimal design of the microreactor are developed. As a proof-of-concept scenario, a synthetic pathway for the conversion of sucrose to glucose-6-phosphate (G6P) was chosen. In this context, the mathematical model is employed to compute the spatio-temporal distributions of the metabolite concentrations, as well as application relevant quantities like the outflow rate of G6P. These computations are performed for different scenarios, where the number of beads as well as their loading capacity are varied. The computed metabolite distributions show spatial patterns, which differ for different experimental arrangements. Furthermore, the total output of G6P increases for scenarios where microcompartmentation of enzymes occurs. These results show that spatially resolved models are needed in the description of the conversion processes. Finally, the enzyme stoichiometry on the nano-beads is determined, which maximizes the production of glucose-6-phosphate.

Keywords: biochemical microreactor, multienzymes complexes, spatio-temporal mathematical model, numerical simulation, PDE-constrained optimization, model-based design

1. INTRODUCTION

One of the greatest challenges in biology is to understand the fundamental principles on how evolution has selected networks to fulfill specific functional needs in the control of metabolism or transcription. Synthetic biology approaches may help to shed light on such principles by identifying modular functional units of a network and uncover how units can be linked together to yield new function (Bashor et al., 2010). There has already been early success in engineering simple regulatory circuits that recapitulate some of the behaviors of natural regulatory circuits, like, e.g., circuits that regulate gene expression oscillations, bistable switches, or circuits that perform combinatorial logic operations, see Boyle and Silver (2009) and Purnick and Weiss (2009) for reviews. However, synthetic biology approaches are not limited to this field. An other important area of synthetic biology is in the development of synthetic organelles, which host metabolic processes, and which are able to communicate with the outside environment *via* transport processes over semipermeable membranes, which can either be built from natural constituents or from synthetic polymers. Such robust bioreactors can be, e.g., applied in biotechnology or drug delivery for the production of bioactive ingredients (Roodbeen and van Hest, 2009; Marguet et al., 2013).

A fundamental principle in the development of bioreactors is compartmentation. Hereby, the strategy is to mimic cellular organization, see, e.g., Roodbeen and van Hest (2009). The presence of compartments (organelles) inside living cells allows for a better regulatory control over the biological processes that occur inside these compartments, e.g., pathways competing for intermediates can occur in spatially separated compartments and can be regulated differently. Furthermore, microcompartmentation by means of metabolic channeling prevents the loss of intermediates and minimizes competing cross-reactions. However, clearcut experimental evidence demonstrating the importance of metabolic channeling for metabolic flux *in vivo* is lacking. Mimicking the natural situation, nanoreactors can be built by encapsulating enzymes in vesicular compartments [as summarized in Peters et al. (2012)], and in addition microcompartmentation of enzymes can be achieved with the help of synthetic protein scaffolds (Chen and Silver, 2012). Microreactors based on microfluidic devices carrying out enzymatic processes, reviewed in Nomura et al. (2004) and Asanomi et al. (2011), have also been into the focus of research in the past years. Several methods used to immobilize enzymes are available, like, e.g., immobilization of enzymes on magnetic microparticles. An approach in the development of microreactors for biosynthesis is the creation of fluidic assay-based microreactors with membrane-bounded subcompartments, carrying out biochemical conversion, and allowing exchange of substrates and products between individual subcompartments.

In our paper, we are focusing on the model-based design of biochemical microreactors able to carry out enzymatic reactions and to exchange metabolites between individual subcompartments. The bioreactor is built according to reported model predictions and is based on a microfluidic system consisting of chambers separated by porous walls, which are interspersed with biomembranes (Figure 1). The membranes carry specific transporters for input of substrates and export of products. Inside the chambers magnetic

nano-beads are present that are positioned by an outer magnetic field.

Multienzyme complexes assembled on these nano-beads by peptide adapters are used to carry out metabolic reactions. The nano-beads allow a maximal enzyme concentration, which depends on the beads' surface. Furthermore, it is possible to provide beads with a given enzyme stoichiometry. We exploit the microcompartmentation offered by the immobilized enzymes on the bead surface to address the question, if the spatial proximity of the individual enzymes and their stoichiometry has an impact on the productivity of the microreactor. In Figure 1, a microreactor consisting of an array of compartments is illustrated.

Our goal is to describe the spatio-temporal dynamics of metabolite concentrations involved in the biosynthetic conversions, and to optimize the microreactor, in order to increase the accumulation of the final product. To achieve this goal, we have developed a mathematical model describing the ongoing metabolic processes in spatial resolution. By applying a model with spatial resolution, we can assess, if the spatial arrangement of the individual enzymes and their stoichiometry has an impact on the productivity of the microreactor. Hence, our model will be able to predict, if metabolic channeling plays a role in the described *in vitro* assembly. Our model consists of a system of diffusion equations together with boundary and initial conditions. The boundary conditions model the exchange of metabolites with the neighboring chambers and the reactions at the surface of the nano-beads carrying the multienzyme complexes. For the numerical simulation of the mathematical model, efficient and accurate approaches are developed. These allow the computation of the spatio-temporal distribution of the metabolite concentrations inside the compartments and the flux of products through the export boundaries. Here, different computational scenarios can be considered including different numbers and loading capacities of nano-beads, different enzyme stoichiometries on the surface of the beads, as well as the distribution of enzymes both inside the fluid and on the beads' surfaces. Comparing the metabolic flux through the system for those scenarios allow to test the hypothesis that microcompartmentation of enzymes increases the efficiency of the metabolic pathway. Based on the mathematical model and the simulation approach, the optimal design of the microreactor is performed.

As a proof-of-concept scenario, we choose a synthetic pathway for the conversion of sucrose to glucose-6-phosphate. This metabolic pathway is localized in one chamber of microreactor allowing the import of sucrose and the export of glucose-6-phosphate. Based on the mathematical model, several investigations are performed. First, the spatio-temporal distributions of the metabolite concentrations, as well as application relevant quantities like the production rate of the metabolites and outflow rate of G6P are computed. These computations are performed for different scenarios, where the number of beads as well as their loading capacity are varied. Furthermore, for the calculations, we consider two different hexokinases, namely HsHK2 and SchK2, and we suppose sucrose and ATP to be present in surplus quantities. Finally, for the mentioned scenarios, we determine the stoichiometry of enzyme concentrations on the nano-beads which maximizes the production of glucose-6-phosphate.

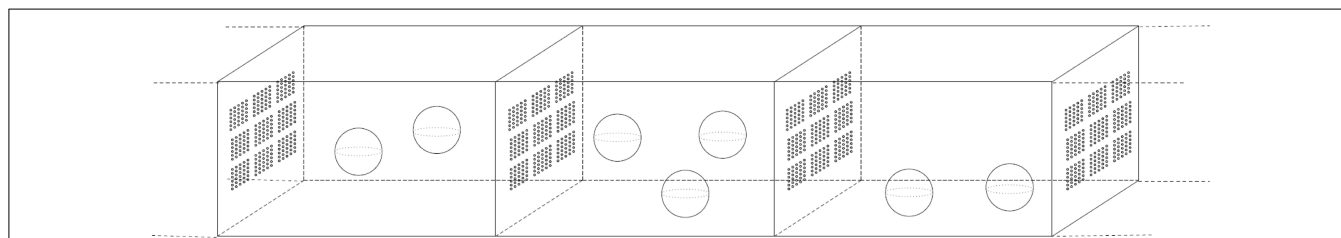


FIGURE 1 | A biochemical microreactor consisting of an array of compartments separated by membranes carrying specific transporters for input of substrates and export of products. This microreactor is currently under construction and the modeling results reported here are being utilized to influence the conceptual design of the microreactor.

2. MATERIALS AND METHODS

2.1. The Mathematical Model

In this section, we set up a mathematical model describing the conversion of sucrose (S) to glucose-6-phosphate (G6P). This metabolic pathway is carried out in a microreactor consisting of a chamber separated from the surrounding medium by membranes carrying transport proteins for the input of sucrose and export of glucose-6-phosphate. Nano-beads loaded with multienzyme complexes are distributed inside the chamber. The enzymatic reactions constituting the metabolic pathway, as well as the corresponding enzymes and metabolites are listed in **Table 1**.

The layout of the microreactor is given in **Figure 2**. We denote the reactor chamber by $\Omega_c \subset \mathbb{R}^n$ ($n = 2$ or $n = 3$). The space inside the chamber occupied by nano-beads, which are balls with diameter d , is denoted by Ω_b , whereas the remainder, representing the domain occupied by the bulk solution, is denoted by $\Omega := \Omega_c \setminus \Omega_b$. We assume that Ω_b is strictly included in Ω_c , this means that the beads do not touch the walls of the chamber. The number of beads in the reactor is denoted by n_b . The boundary $\partial\Omega$ of the domain Ω is decomposed into $\partial\Omega_c$, the boundary of the chamber, and $\Gamma_b := \partial\Omega_b$, the reactive surface of the nano-beads. Furthermore, the boundary $\partial\Omega_c$ of the chamber consists of Γ_i , Γ_e , and Γ_0 , i.e.,

$$\partial\Omega_c = \overline{\Gamma_i} \cup \overline{\Gamma_e} \cup \overline{\Gamma_0},$$

and the three sets are pairwise disjoint. The sets Γ_i , respectively, Γ_e represent the boundary parts where metabolites are transported into, respectively, out of the chamber Ω_c . These import/export boundaries have a complex geometric structure. They consist of fenestrations lined with lipid-membranes carrying transporters for the exchange of metabolites. On the boundary Γ_i the sucrose/proton cotransporters are distributed, whereas Γ_e contains the transporters for the exchange of glucose-6-phosphate and inorganic phosphate. In our model, the microscopic structure of the boundaries is taken into account in an averaged (homogenized) way, which makes the model amenable for numerical calculations. More precisely, we assign to each of the boundaries Γ_i , Γ_e an effective permeability for the transported metabolites denoted by θ^i , respectively, θ^e ($\theta^i, \theta^e \in [0,1]$). These permeabilities can be calculated by an averaging approach (homogenization), assuming that the pores of the transporters are very small compared to the dimension of the reactor chamber and occur in a large number, and that the transporters are uniformly distributed within the lipid-membrane. A sketch illustrating the idea behind the homogenization approach is given in **Figure 3**.

TABLE 1 | Metabolic reactions and the corresponding enzymes and metabolites.

	Reactions	Enzymes
(0)	$S_e + H_e^+ \rightleftharpoons S + H^+$	S-transporter
(1)	$S \rightarrow G + F$	Invertase (inv)
(2)	$G + \text{ATP} \rightarrow \text{G6P} + \text{ADP}$	Hexokinase (hk)
(3)	$F + \text{ATP} \rightarrow \text{F6P} + \text{ADP}$	Hexokinase
(4)	$\text{F6P} \rightleftharpoons \text{G6P}$	Phosphoglucose isomerase (pgi)
(5)	$\text{G6P} + \text{P}_i \rightleftharpoons \text{G6P}_e + \text{P}_i$	G6P-transporter

Metabolites: S, sucrose; H^+ , protons; G, glucose; F, fructose; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; P_i , inorganic phosphate. The index e identifies metabolite located outside the chamber.

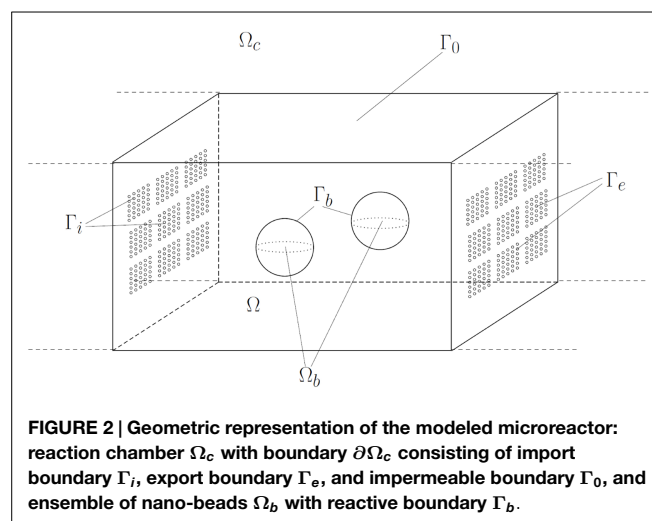


FIGURE 2 | Geometric representation of the modeled microreactor: reaction chamber Ω_c with boundary $\partial\Omega_c$ consisting of import boundary Γ_i , export boundary Γ_e , and impermeable boundary Γ_0 , and ensemble of nano-beads Ω_b with reactive boundary Γ_b .

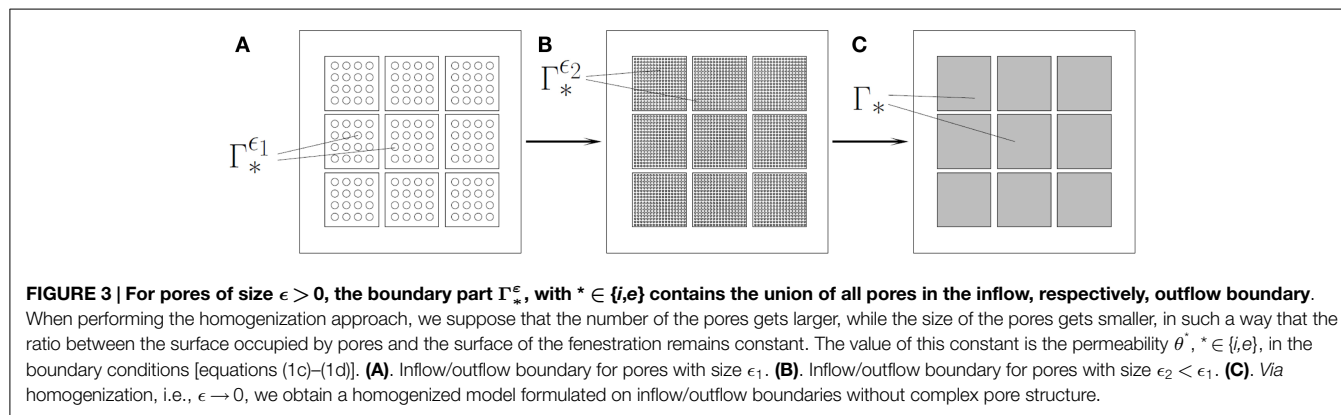
The spatio-temporal dynamics of the concentrations y_j , $j = 1, \dots, m$ of metabolites involved in the metabolic pathway is governed by a system of reaction–diffusion equations of the form:

$$\partial_t y_j(t, x) - D_j \Delta y_j(t, x) = R_j^\Omega(y(t, x), \lambda) \quad \text{for } (t, x) \in (0, T) \times \Omega, \quad (1a)$$

together with the boundary conditions

$$-D_j \nabla y_j(t, x) \cdot \nu(x) = -R_j^b(y(t, x), \lambda) \quad \text{for } (t, x) \in (0, T) \times \Gamma_b, \quad (1b)$$

$$-D_j \nabla y_j(t, x) \cdot \nu(x) = -\theta^e R_j^e(y(t, x)) \quad \text{for } (t, x) \in (0, T) \times \Gamma_e, \quad (1c)$$



$$\begin{aligned}
 -D_j \nabla y_j(t, x) \cdot \nu(x) &= -\theta^i R_j^i(y(t, x)) & \text{for } (t, x) \in (0, T) \times \Gamma_i, \\
 & & (1d) \\
 -D_j \nabla y_j(t, x) \cdot \nu(x) &= 0 & \text{for } (t, x) \in (0, T) \times \Gamma_0, \\
 & & (1e)
 \end{aligned}$$

and the initial condition

$$y_j(0, x) = y_j^0(x) \quad \text{for } x \in \Omega. \quad (1f)$$

Here, ν denotes the outer unit normal on $\partial\Omega$ with respect to Ω . Equation (1a) describes diffusion with diffusivity D_j and enzymatic reactions with kinetics R_j^Ω for the metabolite number j . In case that enzymes are localized on beads (and thus no reactions are carried out in the bulk domain), the terms R_j^Ω are equal to zero. The boundary conditions involve the quantity $-D_j \nabla y_j \cdot \nu$ which describes the normal component of the diffusive flux of the j -th metabolite at the boundary of the domain Ω . Thus, condition (1e) models an impermeable boundary, where the normal flux is equal to zero. Conditions (1c)–(1d) model the flux of metabolites through the import and export boundary, respectively. This flux is proportional to the permeability of the boundary, and the kinetics of the corresponding transport protein. Finally, condition (1b) describes the normal flux of metabolites at the boundary of the beads, generated by enzymatic reactions at the beads' surface. We emphasize that the reaction rates R_j^b and R_j^Ω depend on an additional parameter $\lambda = (\lambda_1, \dots, \lambda_{n_e}) \in \mathbb{R}^{n_e}$. This parameter describes the stoichiometry of enzymes on the beads and in the bulk, and thus for $i = 1, \dots, n_e$, where n_e denotes the number of enzyme species involved in the reactions, holds

$$\lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n_e} \lambda_i = 1. \quad (2)$$

(The case $\lambda_{i^*} = 1$ for some $i^* \in \{1, \dots, n_e\}$ means that all binding sites on the bead surface are occupied by the enzyme i^* , whereas the case $\lambda_i = \frac{1}{n_e}$ for all $i \in \{1, \dots, n_e\}$ means that all enzymes occupy the same amount of binding sites).

The existence and uniqueness of positive solutions for the model (1) can be shown by arguments similar to those in Gahn et al. (under review)¹. We also emphasize that the model (1), valid

for one conversion chamber, can easily be extended to model a multicompartiment microreactor. This is done by adding transmission conditions at the interfaces separating the compartments, which model the exchange of metabolites between neighboring compartments.

In this paper, we want to investigate the interplay between the metabolic processes at the beads and the export of the product glucose-6-phosphate. We assume that during the conversion processes sucrose and ATP concentrations can be regarded as constant, since these two metabolite species are present in excess. As a consequence, we drop reaction (0). The equation for ADP can be neglected, since ADP is just a product of irreversible reactions and therefore has no direct effect on the reaction rates in equation (1). Finally, we assume that the concentrations of G6P_e and Pi_e (glucose-6-phosphate and inorganic phosphate outside the chamber) are constant. This is motivated by the fact that G6P_e may be consumed in a following reaction, and Pi_e may be delivered at a desirable rate in the space outside the chamber. Hence, we can drop the equations for G6P_e and Pi_e. The constant values of the concentrations mentioned above are given by the corresponding initial concentrations (Table 3). The scenario taking into account these assumptions shall be referred to as the sucrose excess scenario (SE-scenario).

The reactions relevant for the SE-scenario are reactions (1)–(5) (Table 1). These are in general multisubstrate enzymatic reactions. Their reaction mechanisms and the corresponding reaction kinetics are displayed in Table 2A. Note that in Table 2, the concentrations of metabolites are denoted with brackets (e.g., for sucrose, we use [S] instead of y_S). This is chosen in order to keep the notation clear.

The unknowns of the model in the SE-scenario are the following metabolite concentrations: $y_G, y_F, y_{F6P}, y_{Pi}, y_{G6P}$. For each unknown, a reaction–diffusion equation of type (1a) complemented by boundary conditions of type (1b)–(1e), and initial conditions holds. The reaction terms occurring in the equations are denoted by $R_G^\Omega, \dots, R_{G6P}^\Omega$, whereas the fluxes at the boundaries $\Gamma_b, \Gamma_e, \Gamma_i$ are given by reaction terms denoted by $R_G^b, \dots, R_{G6P}^b, R_G^e, \dots, R_{G6P}^e$, and R_G^i, \dots, R_{G6P}^i . The precise form of these reaction terms, in case when enzymes are distributed on beads, i.e., $R_j^\Omega = 0$, is given in Table 2B. We emphasize that, when the metabolite $j, j = G, \dots, G6P$ participates in different reactions, the

¹Gahn, M., Neuss-Radu, M., and Knabner, P. (2015). Homogenization of reaction-diffusion processes in a two-component porous medium with nonlinear flux conditions at the interface. *SIAM J. Appl. Math.* (Under Review).

TABLE 2 | (A) Metabolic reactions relevant for the sucrose excess scenario and the corresponding reaction kinetics; reaction mechanisms: (1) irreversible Michaelis–Menten; (2), (3) irreversible bi–bi ordered; (4) reversible Michaelis–Menten; (5) bi–bi ping pong. See, e.g., Segel (1975) for an overview on reaction mechanisms for multisubstrate enzymatic reactions; (B) reaction terms in equation (1) corresponding to the concentrations y_j , $j = G, \dots, G6P$, for the SE-scenario with enzymes distributed on beads, i.e., $R_j^\Omega = 0$.

(A)		
Reaction	Reaction rate	
(1)	$r_{\text{inv}}(S) = \frac{\lambda_{\text{inv}} v_{\text{max}}^{\text{inv}} [S]}{K_{\text{MS}} + [S]}$	
(2)	$r_{\text{hk}}^G(G, \text{ATP}) = \frac{\lambda_{\text{hk}} v_{\text{max}}^{\text{hk}} [G][\text{ATP}]}{K_{\text{mG}} K_{\text{mATP}} + K_{\text{mATP}} [G] + K_{\text{mG}} [\text{ATP}] + [G][\text{ATP}]}$	
(3)	$r_{\text{hk}}^F(F, \text{ATP}) = \frac{\lambda_{\text{hk}} v_{\text{max}}^{\text{hk}} [F][\text{ATP}]}{K_{\text{mF}} K_{\text{mATP}} + K_{\text{mATP}} [F] + K_{\text{mF}} [\text{ATP}] + [F][\text{ATP}]}$	
(4)	$r_{\text{pgi}}^F(F6P, G6P) = \frac{\lambda_{\text{pgi}} v_{\text{max}}^{\text{pgi},f} [F6P]}{K_{\text{mF6P}} + [F6P]} - \frac{\lambda_{\text{pgi}} v_{\text{max}}^{\text{pgi},b} [G6P]}{K_{\text{mG6P}} + [G6P]}$	
(5)	$r_{\text{G6P}}(G6P, \text{Pi}_e, G6P_e, \text{Pi}) = \frac{v_{\text{max}}^{\text{G6P},f} \left([G6P][\text{Pi}_e] - \frac{[G6P_e][\text{Pi}]}{K_{\text{eq}}} \right)}{[G6P][\text{Pi}_e] + K_{\text{mPi}_e} [G6P] + K_{\text{mG6P}_e} [\text{Pi}] \left(1 + \frac{[\text{Pi}]}{K_{\text{Pi}}} \right) + \dots}$ $\dots \frac{v_{\text{max}}^{\text{G6P},b}}{K_{\text{eq}}} \left(K_{\text{mPi}} [G6P] \left(1 + \frac{[G6P]}{K_{\text{G6P}}} \right) + [\text{Pi}] \left(K_{\text{mG6P}_e} + [G6P_e] \right) \right)$	
(B)		
Species j	R_j^b	R_j^e
G	$r_{\text{inv}}(S) - r_{\text{hk}}^G(G, \text{ATP})$	0
F	$r_{\text{inv}}(S) - r_{\text{hk}}^F(F, \text{ATP})$	0
F6P	$r_{\text{hk}}^F(F, \text{ATP}) - r_{\text{pgi}}^F(F6P, G6P)$	0
Pi	0	$r_{\text{G6P}}(G6P, \text{Pi}_e, G6P_e, \text{Pi})$
G6P	$r_{\text{hk}}^G(G, \text{ATP}) + r_{\text{pgi}}^F(F6P, G6P)$	$-r_{\text{G6P}}(G6P, \text{Pi}_e, G6P_e, \text{Pi})$

reaction term R_j^b is given as a sum of all relevant reaction kinetics. If reactions take place also in the bulk, the reaction terms R_j^Ω , for $j = G, \dots, G6P$, have the same structure as R_j^b with potentially different parameters. Finally, we mention that in the SE-scenario the inflow boundary Γ_i is impermeable, thus the reaction terms R_G^i, \dots, R_{G6P}^i are set to zero.

The values v_{max}^i for an enzyme i , $i \in \{\text{inv}, \text{hk}, \text{pgi}\}$ in Table 2, can be calculated by $v_{\text{max}}^i = k_{\text{cat}}^i [E]_0$, with a turnover number k_{cat}^i given in Table 3, and the concentration $[E]_0$ of occupied binding sites on a bead. The enzymatic activity for the enzyme i , $i \in \{\text{inv}, \text{hk}, \text{pgi}\}$ on each bead is then given by $\lambda_i v_{\text{max}}^i$, where the vector $\lambda = (\lambda_{\text{inv}}, \lambda_{\text{hk}}, \lambda_{\text{pgi}})$ describes the enzymes stoichiometry on the bead. For the simulations, we choose $\lambda = \lambda^E := (\lambda_{\text{inv}}^E, \lambda_{\text{hk}}^E, \lambda_{\text{pgi}}^E)$, where λ^E satisfies

$$\begin{aligned} k_{\text{cat}}^{\text{inv}} \lambda_{\text{inv}}^E &= k_{\text{cat}}^{\text{hk}} \lambda_{\text{hk}}^E = k_{\text{cat}}^{\text{pgi}} \lambda_{\text{pgi}}^E, \\ \lambda_{\text{inv}}^E + \lambda_{\text{hk}}^E + \lambda_{\text{pgi}}^E &= 1. \end{aligned} \quad (3)$$

This enzyme stoichiometry leads to equal enzymatic activity for all enzyme species. The values of λ^E are given in Table 3. Please note that in Section 2.3 optimal values for the parameter λ are computed.

TABLE 3 | Parameter values for the SE-scenario, corresponding to two different hexokinases HsHK2 and ScHK2.

Parameter	Value	Parameter	Value
K_{MS}	$0.9 \frac{\text{mol}}{\text{m}^2}$	$k_{\text{cat}}^{\text{pgi}}$	646.61s^{-1}
K_{mG}	$0.052 \frac{\text{mol}}{\text{m}^2}$ (HsHK2), $0.12 \frac{\text{mol}}{\text{m}^2}$ (ScHK2)	$v_{\text{max}}^{\text{inv}, \text{full}}$	$6.57 \cdot 10^{-5} \frac{\text{mol}}{\text{sm}}$
K_{mATP}	$0.5 \frac{\text{mol}}{\text{m}^2}$ (HsHK2), $0.1 \frac{\text{mol}}{\text{m}^2}$ (ScHK2)	$v_{\text{max}}^{\text{hk}, \text{full}}$	$10.62 \cdot 10^{-7} \frac{\text{mol}}{\text{sm}}$ (HsHK2), $2.03 \cdot 10^{-8} \frac{\text{mol}}{\text{sm}}$ (ScHK2)
K_{mF}	$11.4 \frac{\text{mol}}{\text{m}^2}$ (HsHK2), $0.33 \frac{\text{mol}}{\text{m}^2}$ (ScHK2)	$v_{\text{max}}^{\text{pgi}, \text{full}}$	$12.57 \cdot 10^{-6} \frac{\text{mol}}{\text{sm}}$
K_{mF6P}	$0.19 \frac{\text{mol}}{\text{m}^2}$	$v_{\text{max}}^{\text{G6P}, f}$	$3.23 \cdot 10^{-8} \frac{\text{mol}}{\text{sm}}$
K_{mG6P}	$0.5 \frac{\text{mol}}{\text{m}^2}$	$v_{\text{max}}^{\text{G6P}, b}$	$3.23 \cdot 10^{-8} \frac{\text{mol}}{\text{sm}}$
K_{mG6PT}	$0.7 \frac{\text{mol}}{\text{m}^2}$	λ^E	(0.0147, 0.91, 0.0753) (HsHK2)
K_{mG6Pe}	$0.7 \frac{\text{mol}}{\text{m}^2}$	λ^E	(0.0003, 0.9981, 0.0016) (ScHK2)
K_{mPi}	$1.1 \frac{\text{mol}}{\text{m}^2}$	θ^e	0.15
K_{mPi_e}	$1.1 \frac{\text{mol}}{\text{m}^2}$	$[G6P_e]$	$1 \frac{\text{mol}}{\text{m}^2}$
K_{fG6P}	$0.9 \frac{\text{mol}}{\text{m}^2}$	$[P]_e$	$10 \frac{\text{mol}}{\text{m}^2}$
K_{fPi}	$0.1 \frac{\text{mol}}{\text{m}^2}$	$[ATP]$	$10 \frac{\text{mol}}{\text{m}^2}$
K_{eq}	$0.5 \frac{\text{mol}}{\text{m}^2}$	$[S]$	$50 \frac{\text{mol}}{\text{m}^2}$
$\lambda_{\text{cat}}^{\text{hk}}$	54.63s^{-1} (HsHK2), 1.06s^{-1} (ScHK2)	d	10^{-5}m
$k_{\text{cat}}^{\text{inv}}$	3379.60s^{-1}	$[E]_{\text{full}}$	$1.94 \cdot 10^{-8} \frac{\text{mol}}{\text{m}}$

These parameters correspond to the 2-dimensional case and their units of measurements were adapted to this case. Approaches for the experimental determination of the k_{cat} -values can be found in Gao and Leary (2003), Gloy et al. (2005), Lin et al. (2009), Lafraya et al. (2011), and Somarowthu et al. (2011). We use the values below to calculate the v_{max} -values. The initial values y_j^0 in equation (6) for the species $j = G, F, F6P, G6P, \text{Pi}$ are equal to zero. For the diffusion-coefficients, D_j belonging to these species we use the value $D_j = 5.5 \cdot 10^{-10} \frac{\text{m}^2}{\text{s}}$.

2.2. Numerical Methods

We use a finite element method in order to find an approximation to the solution of equation (1) on a fixed time interval $[0, T]$ and with a given initial state.

For the numerical accuracy and the computational complexity of an implementation, also in view of the optimization, it is crucial to choose a suitable discretization. For our problem, we use lowest order Raviart–Thomas elements (Raviart and Thomas, 1977). On a first glance, linear finite elements seem superior to Raviart–Thomas elements due to a lower number of unknowns and a higher order of convergence. However, for our problem, it turns out that the use of linear finite elements leads to solutions with negative concentrations when the triangulation \mathcal{T}_h is not extremely fine. More precisely, for $n_b = 1$ we investigated two scenarios, and compared the number of unknowns that are needed for each discretization, to obtain realistic results. For $\Omega_c = (0.50 \mu\text{m})^2$, we need 3200 degrees of freedom per species in order to obtain realistic results for linear finite elements, whereas Raviart–Thomas elements only need 993 degrees of freedom. For $\Omega_c = (0.300 \mu\text{m})^2$ we need 122,544 degrees of freedom per species in order to obtain realistic results for linear finite elements, whereas Raviart–Thomas elements only need 38,256 degrees of freedom. The computational complexity for Raviart–Thomas

elements can be further reduced by hybridization, in this case only 607 degrees of freedom for the smaller domain and 23,016 degrees of freedom for the larger domain are needed.

The resulting space-discrete system is stiff, therefore, we use the implicit Euler method for the time integration. The resulting finite dimensional non-linear problems are solved with Newton's method. Whenever the non-linear solver can not find a non-negative solution, the time step is rejected and a smaller time step size is used (until the non-linear solver can find a non-negative solution or the time step size is less than a predefined minimal time step size). We emphasize that decreasing the time step size was not necessary on the meshes used for spatial discretization (see Supplementary Material) with Raviart–Thomas elements and a time step size of 0.25 s.

In order to decrease the computational complexity of problem (1), arising from the number of species, we use the scheme presented in Kräutle and Knabner (2005) with minor modifications: instead of finding the basis of the image of the stoichiometric matrix S on the right hand side of the partial differential equation, we find the basis of the image of the matrix $(S_b|S_e)$, where S_x refers to the stoichiometric matrix of the reactions on Γ_x , $x \in \{b, e\}$. The decoupled problem consists of 6 coupled species, whereas the original problem consists of 11 coupled species. In the SE-scenario, the decoupling scheme does not decrease the computational complexity, since the resulting stoichiometric matrix has full rank.

The numerical scheme has been implemented using the software package M++ that is based on the data structure in Wieners (2005). The algorithm uses MPI parallelization and can therefore use an arbitrary number of processors for computation. This is particularly useful for problems involving many species.

2.3. Optimization Methods

Besides setting up a mathematical model for the microreactor to describe the temporal and spatial distribution of enzymes and reactants, our goal also was to determine optimal parameters for enhanced performance of the microreactor based on this model. We demonstrate here how to succeed in a systematical way using derivative-based non-linear optimization techniques. Exemplary, we focus on how to determine the real parameters λ modeling the stoichiometry of the enzymes loaded on the beads entering in the non-linear terms in equation (1). We then discuss extensions of such methods to other parameters such as the discrete number of beads n_b , the combinatorial problem of which specific enzymes should be used for the metabolic pathway and continuous geometry parameters such as the bead diameters or locations, the total activity of the transporter proteins, the shape and size of the chamber itself, etc.

The criterion for the performance of the reactor will be the total outflow of a desired product, say $y_{j_*}(\lambda)$ on Γ_e , over a fixed production time horizon $[0, T]$,

$$\begin{aligned} J(\lambda) &= \int_0^T \int_{\Gamma_e} -D_{j_*} \nabla y_{j_*}(t, x) \cdot \nu(x) d\sigma(x) dt \\ &= \int_0^T \int_{\Gamma_e} -\theta^e R_{j_*}^e(y(t, x)) d\sigma(x) dt, \end{aligned} \quad (4)$$

where we have used equation (1c) for $j = j_*$ in equation (4). For our sucrolytic chamber, we have $j_* = j_{G6P}$. By optimal parameters, we mean that any feasible choice close to the optimal one does not lead to a higher total outflow predicted by the model.

While derivative-based non-linear optimization techniques are conceptually well-known, we emphasize that due to the large-scale, non-linear system of equations to be solved for each direct simulation of the model, the main challenge is an efficient computation of the derivative of the reduced cost function

$$\hat{J}(\lambda) = (y(\lambda)), \quad (5)$$

where $y(\lambda)$ denotes the numerical solution of the system [equations (1a)–(1f)] as a function of the parameter λ . We will exploit that the dimension of the parameter space n_p is typically small compared to the dimension obtained from discretization of the infinite-dimensional state space for y in the model and will therefore follow a sensitivity-based approach to compute the derivative [see, e.g., Hinze et al. (2009)]. Letting e_i , $i = 1, \dots, n_p$, being a bases of the parameter space this means that we can compute the derivative $J'(\lambda)$ as

$$J'(\lambda) = \sum_{i=1}^l J_y(y(\lambda)) \delta_{e_i} y, \quad (6)$$

where the sensitivity $\delta_{e_i} y = y'(\lambda) e_i$ is given by the solution $(\tilde{y}_1, \dots, \tilde{y}_m)$ of the following linearized problem

$$\begin{aligned} \partial_t \tilde{y}_j(t, x) - D_j \Delta \tilde{y}_j(t, x) &= \partial_y R_j^\Omega(y(t, x), \lambda) \tilde{y}_j \\ &\quad - \partial_\lambda R_j^\Omega(y(t, x), \lambda) e_i \quad \text{for } (t, x) \in (0, T) \times \Omega, \end{aligned} \quad (7a)$$

together with the boundary conditions

$$\begin{aligned} -D_j \nabla \tilde{y}_j(t, x) \cdot \nu(x) &= -\partial_y R_j^b(y(t, x), \lambda) \tilde{y}_j \\ &\quad + \partial_\lambda R_j^b(y(t, x), \lambda) e_i \quad \text{for } (t, x) \in (0, T) \times \Gamma_b, \end{aligned} \quad (7b)$$

$$\begin{aligned} -D_j \nabla \tilde{y}_j(t, x) \cdot \nu(x) &= \\ &= -\theta^e (R_j^e)'(y(t, x)) \tilde{y}_j \quad \text{for } (t, x) \in (0, T) \times \Gamma_e, \end{aligned} \quad (7c)$$

$$\begin{aligned} -D_j \nabla \tilde{y}_j(t, x) \cdot \nu(x) &= \\ &= -\theta^i (R_j^i)'(y(t, x)) \tilde{y}_j \quad \text{for } (t, x) \in (0, T) \times \Gamma_i, \end{aligned} \quad (7d)$$

$$-D_j \nabla \tilde{y}_j(t, x) \cdot \nu(x) = 0 \quad \text{for } (t, x) \in (0, T) \times \Gamma_0, \quad (7e)$$

and the initial condition

$$\tilde{y}_j(0, x) = 0 \quad \text{for } x \in \Omega, \quad (7f)$$

where ∂_y and ∂_λ denote partial derivatives and $(\cdot)'$ [e.g., $(R_j^i)'$] denotes total derivatives of the corresponding kinetic functions, respectively, and where y is the solution of equations (1a)–(1f). The solution of this linearized problem can be computed simultaneously with the direct simulation using the same discretization method as described in Section 2.2. Compared to the evaluation of the objective function (5), the additional computational effort to obtain the derivative is solving n_p linear equation systems in each time step. Instead of using the linearized problem (7), we can

also obtain a derivative from an adjoint problem, see again, e.g., Hinze et al. (2009). The latter approach is more efficient when n_p is large.

With the derivative at hand, we can then solve the reduced optimization problem subject to further parameter constraints for instance with primal-dual interior-point algorithms using a quasi-Newton approximation of the Hessian. These methods are known to have both good theoretical and practical behavior, see, e.g., Forsgren et al. (2002). Given some initial parameter $\lambda^{(0)}$, they compute a sequence of parameters $\lambda^{(k)}$, $k = 1, 2, 3, \dots$, converging to an optimal λ until for some k^* first order conditions or stationarity in the objective function are satisfied within a tolerance tol_X or tol_{fun} , respectively. We set $\lambda^* = \lambda^{(k^*)}$, $J^* = J(y(\lambda^*))$, and $J^0 = J(y(\lambda^{(0)}))$.

In order to determine an optimal stoichiometry λ for our sucrolytic chamber, we have the three parameters λ_{inv} , λ_{hk} , and λ_{pgi} , but we may eliminate for example λ_{pgi} by the condition (2). This yields $n_p = 2$, together with a linear inequality constraint $\lambda_{\text{inv}} + \lambda_{\text{hk}} \leq 1$ and box constraints $\lambda_{\text{inv}}, \lambda_{\text{hk}} \in [0, 1]$. For our numerical results in Section 3.2 for the SE-scenario, we have used the interior-point algorithm implemented in the Optimization Toolbox in MATLAB (2014) with a BFGS quasi-Newton approximation of the hessian, where equations (5) and (6) are computed using the methods described in Section 2.2. Again we use the software package M++ for our implementation. As initial parameter $\lambda^{(0)}$, we either use the stoichiometry from equal activity λ^E given by equation (3) or the uniform stoichiometry $\lambda^{(0)} = \lambda^U = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ depending on which has a larger value $J^{(0)}$. Furthermore, we use $\text{tol}_X = 1.0e-06$ and $\text{tol}_{\text{fun}} = 1.0e-04$ for all our optimization runs.

The derivative computation and hence the above method can directly be adapted to other real parameters entering the kinetic functions such as, e.g., the permeability θ . Combinatorial problems from (additional) discrete parameters such as the number of beads can be solved using enumeration or, approximately, by derivative-based optimization techniques based on relaxation methods, even for time dependent parameters as in Hante and Sager (2013). Continuously varying geometry parameters such as bead diameters or shape and size of the chamber Ω_c can be handled similarly by shape derivatives as in Leugering et al. (2011). For the SE-scenario, we enumerate optimal λ^* for various combinations of number of beads n_b , different choices of hexokinases in the sucrolytic pathway and different time horizons T .

3. RESULTS

We apply our simulation and optimization methods to the situation of the microreactor in the SE-scenario for realistic parameters. Hereby, we consider two-dimensional quadratic reaction chambers $\Omega_c = (0, L)^2$ with $L = 500$ and $L = 3000 \mu\text{m}$, and different numbers of uniformly distributed beads, each of diameter of $10 \mu\text{m}$. The outflow boundary is $\Gamma_e = \{L\} \times (0, L)$. The rest of the chambers' boundary is impermeable. The parameters of the mathematical model are given in Table 3. The final time T varies between 0.5 and 4 h. This is appropriate, due to the fact that the outflow-membrane Γ_e is in general unstable and bursts after several hours. We stress that, due to the need of small time step

sizes and the ratio of bead and chamber size, larger time horizons and chamber sizes lead to a higher computational complexity.

For our numerical investigations, we consider two types of hexokinase, namely HsHK2 and ScHK2. We also have two stoichiometry values which play a role in our calculations: the uniform stoichiometry $\lambda^U = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and the equal activity stoichiometry λ^E described in equation (3). More precisely, λ^E is used for the numerical simulations, whereas either of the two is used as initial parameter in the optimization, depending on which leads to a larger value of the total glucose-6-phosphate outflow.

3.1. Simulation Results

We consider a quadratic chamber Ω_c with a length of $3000 \mu\text{m}$, a fixed run time of $T = 3$ h, a fixed enzyme stoichiometry $\lambda = \lambda^E$ and a number of beads $n_b \in \{0^2, 1^2, 2^2, \dots, 6^2, 12^2, 18^2, \dots, 60^2\}$, where $n_b = 0$ refers to a scenario with enzymes distributed in the bulk solution.

The meshes used for spatial discretization consist of between 15,190 triangles and 22,941 edges for 1 bead, and 93,184 triangles and 151,592 edges for 2916 beads. Detailed information about the meshes is available in the Supplementary Material. We used a time step size of 0.25 s. All calculations have been repeated for $T = 1.5$ h on uniform refined meshes and half of the time step size. No significant differences between the results on the coarse and on the fine meshes were observed.

First results are concerned with the numerical computation of the spatial and temporal distribution of the metabolites. Figure 4 shows the distribution of G6P for a scenario with 1296 beads and the hexokinase HsHK2 after 3 h for completely and partially loaded beads.

Next results are concerned with the computation of quantities, which are particularly relevant for our microreactor, namely the export of G6P at time $t \in [0, T]$, i.e., the outflow rate at time t given by

$$\int_{\Gamma_e} -D_{\text{G6P}} \nabla y_{\text{G6P}}(t, x) \cdot \nu(x) d\sigma(x), \quad (8)$$

and the production rate for the metabolites at the beads at time t given by

$$\int_{\Gamma_b} D_j \nabla y_j(t, x) \cdot \nu(x) d\sigma(x), \quad (9)$$

for $j = \text{G}, \dots, \text{G6P}$. In computing these quantities, we consider the following arrangements:

- (S.1) Increase the number of beads n_b , where beads are completely loaded.
- (S.2) The total number of enzymes in the domain is fixed (and given by the number of binding sites on one completely loaded bead) and enzymes are distributed uniformly on n_b beads or in bulk. The latter case, we characterize by $n_b = 0$.

Both arrangements are simulated for the two hexokinases HsHK2 and ScHK2, and the beads are distributed periodically in the bulk, like, e.g., in Figure 4. For the arrangement (S.2), we also consider the case without beads where all enzymes are distributed in the bulk. Then the reaction terms R_j^{Ω} in equation (1a) are equal to R_j^b from Table 2. In (S.1), the maximal velocity $v_{\text{max}}^i = k_{\text{cat}}^i [E]_0$ corresponding to enzyme i , $i \in \{\text{inv}, \text{hk}, \text{pgi}\}$ is calculated with

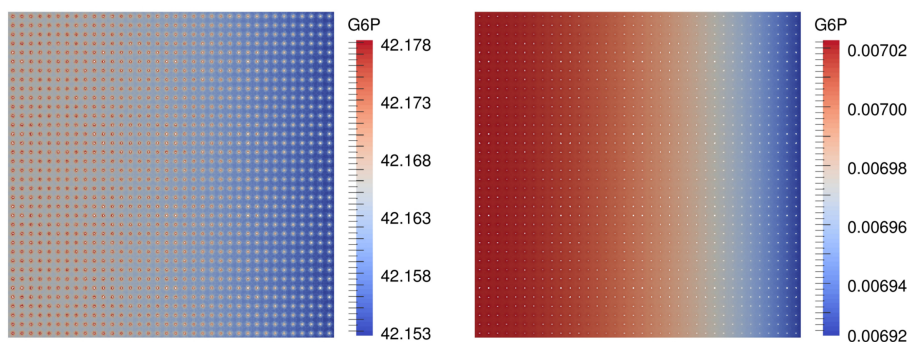


FIGURE 4 | G6P concentration in $\frac{\text{mol}}{\text{m}^2}$ in a scenario with $n_b = 1296$ beads and the hexokinase HsHK2, for a completely loaded beads (left) and partially loaded beads, where the total number of enzymes distributed to the beads corresponds to that on one completely loaded bead (right), after 3 h.

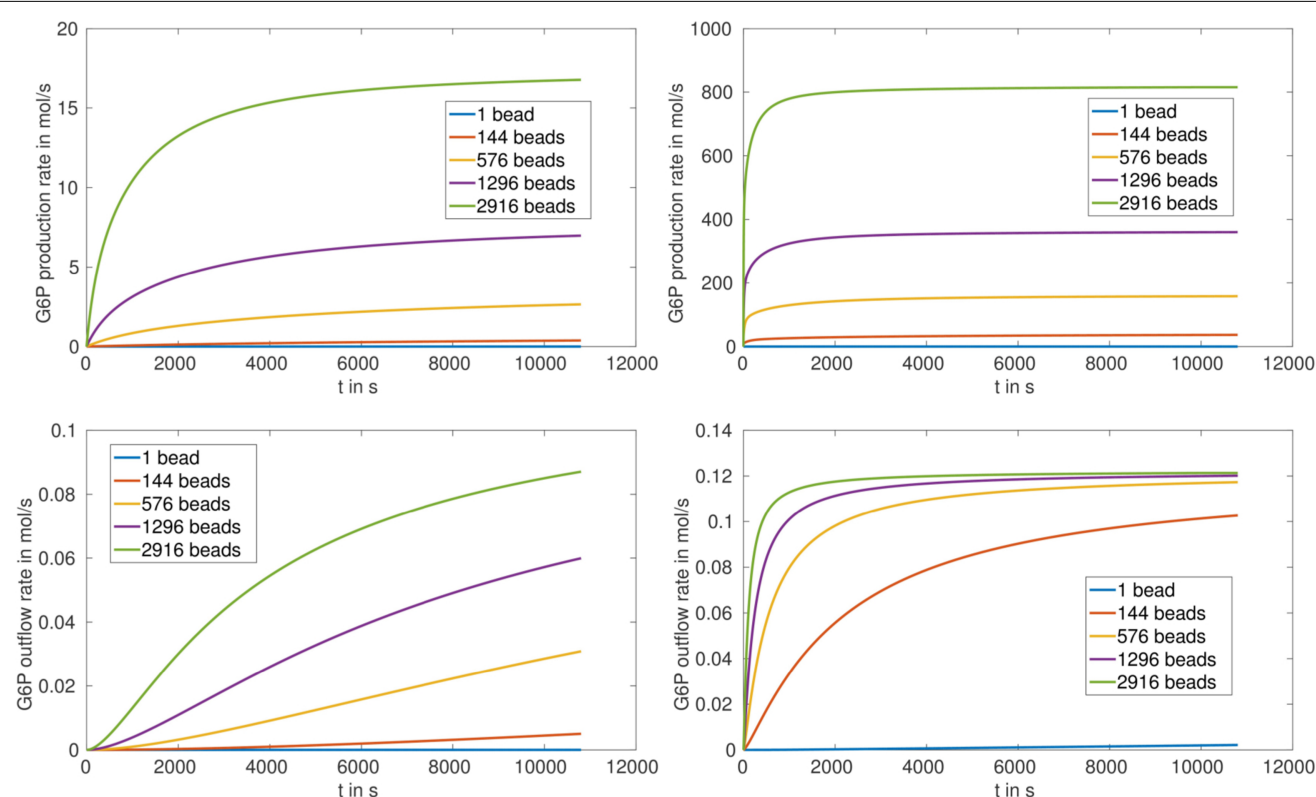


FIGURE 5 | G6P production rate (top) and outflow rate (bottom) in $\frac{\text{mol}}{\text{s}}$ for arrangement (S.1) with $n_b \in \{1, 144, 576, 1296, 2916\}$ and the enzymes SchHK2 (left) and HsHK2 (right).

$[E]_0 = [E]_{\text{full}}$ where $[E]_{\text{full}}$ is the concentration of binding sites on one bead. These values are denoted by $v_{\text{max}}^{i,\text{full}}$, and are given in Table 3. In (S.2), we fix the total concentration of enzyme to be given by $[E]_{\text{full}}$. However, this concentration is now distributed on n_b beads. Thus, the values for v_{max}^i , $i \in \{\text{inv}, \text{hk}, \text{pgi}\}$ are now obtained by dividing the values used in (S.1) by the number of beads n_b .

The production and outflow rate for G6P are plotted in Figure 5 for (S.1) and Figure 6 for (S.2). For fully loaded beads the production rate increases with increasing number of beads, and reaches saturation after approximately 1000 s in the case of HsHK2 and

after approximately 10,800 s in the case of SchHK2 (Figure 5, upper pictures). Concerning the outflow rate, in case of HsHK2 (Figure 5 right bottom) an upper bound is asymptotically reached, the value of which seems to be the same for all $n_b \geq 576$ (simulations with longer runtime suggest that the value seems to be the same for $n_b \geq 16$).

In case of partially loaded beads corresponding to arrangement (S.2), the production rate of G6P at the beads up to a time of 3 h first decreases with increasing number of beads and then stays nearly constant (Table 4; Figure 6, upper figures). Please note, that for the arrangement (S.2) the outflow rates are not reaching

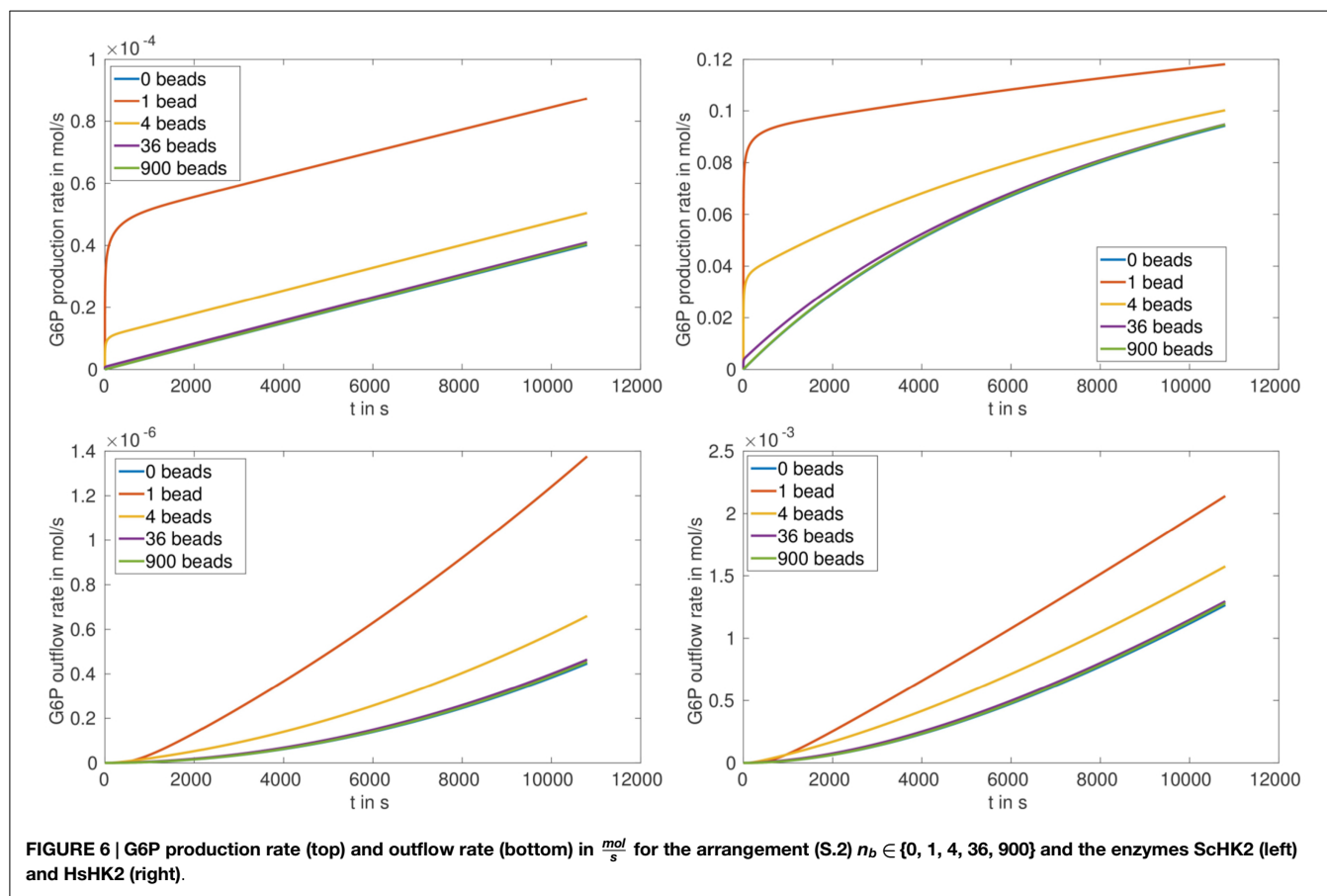


FIGURE 6 | G6P production rate (top) and outflow rate (bottom) in $\frac{\text{mol}}{\text{s}}$ for the arrangement (S.2) $n_b \in \{0, 1, 4, 36, 900\}$ and the enzymes SchHK2 (left) and HsHK2 (right).

TABLE 4 | Production and outflow rates at $T = 3\text{h}$ for scenario (S.2) with the hexokinase HsHK2 and $n_b \in \{1^2, 2^2, 6^2, 18^2, 30^2, 42^2, 54^2, 0\}$, and times until half of these rates are reached.

	n_b	1	4	36	324	900	1764	2916	0
Rates after 3 h (in $\frac{\text{mol}}{\text{s}}$)	Production	0.12	0.10	0.095	0.094	0.095	0.095	0.095	0.094
	Outflow	0.0021	0.0016	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
Time until half of the rate above is reached (in seconds)	Production	2	1494	3460	3603	3606	3600	3590	3618
	Outflow	5959	6457	7025	7090	7092	7090	7086	7096

saturation, which implies that the transporters activity are not limiting the outflow for the chosen enzyme concentration $[E]_{\text{full}}$. We also remark that for the chosen stoichiometry λ^E , in case of arrangement (S.2), the production rate of G6P for HsHK2 is by a factor 10^3 to 10^4 larger than for SchHK2, i.e., in (S.2) the selected hexokinase has a greater influence on the conversion process as in (S.1).

In order to quantify how fast production and outflow rate for various number of beads increase in time in the scenario (S.2) with the hexokinase HsHK2, in Table 4, we give the values of these rates at $T = 3\text{h}$, and the times the reactor needs to reach half of these values.

3.2. Optimization Results

We aim at finding the optimal stoichiometry λ^* for the SE-scenario. Since our focus is on computing a variety of different scenarios with a reasonable computational effort, we have chosen

the smaller chamber $\Omega_c = (0.500\text{ }\mu\text{m})^2$ for these experiments. The meshes used for the spatial discretization of this chamber consist of between 820 triangles and 1270 edges for 1 bead, and 2394 triangles and 3883 edges for 64 beads. Further information about the meshes is available in the Supplementary Material.

For our study, we consider the following arrangements:

- (O.1) For a fixed runtime of $T = 1\text{h}$, we increase the number of completely loaded beads in the chamber with $n_b \in \{1, 4, 9, 16, 25, 36, 49, 64\}$.
- (O.2) For a sacrolytic chamber with one completely loaded bead, we optimize the G6P outflow for different runtimes $T \in \{0.5, 1, 1.5, 2, \dots, 4\}$ hours.

We study these arrangements for the hexokinases SchHK2 and HsHK2.

For the enzyme SchHK2, we find in case of arrangement (O.1) that the optimal stoichiometry for a runtime of 1 h is between 3

TABLE 5 | Optimal parameters λ^* for the SE-scenario with ScHK2 (top) and HsHK2 (bottom), for a fixed runtime $T = 3600$, and different numbers n_b of beads.

Enzyme	n_b	$\lambda^{(0)}$	J^0	k^*	λ_{inv}^*	λ_{hk}^*	λ_{pgi}^*	J^*	$\frac{J^* - J^0}{J^0}$
ScHK2	1	λ^U	0.228066	16	0.120404	0.815447	0.064149	0.446978	0.9599
ScHK2	4	λ^U	0.968913	11	0.084901	0.860564	0.054536	2.090510	1.1576
ScHK2	9	λ^U	2.144680	19	0.063959	0.895103	0.040938	4.686620	1.1852
ScHK2	16	λ^U	3.652360	14	0.052440	0.912839	0.034722	7.750300	1.1220
ScHK2	25	λ^U	5.385780	21	0.045398	0.923280	0.031322	10.90060	1.0240
ScHK2	36	λ^U	7.248180	18	0.040294	0.930381	0.029326	13.90760	0.9188
ScHK2	49	λ^U	9.160180	15	0.036641	0.935141	0.028218	16.66100	0.8189
ScHK2	64	λ^U	11.06320	17	0.033702	0.938735	0.027563	19.12850	0.7290
HsHK2	1	λ^E	8.649300	8	0.091012	0.908980	0.000008	13.12310	0.5172
HsHK2	4	λ^E	21.54300	12	0.070280	0.929719	0.000001	25.56560	0.1867
HsHK2	9	λ^E	28.92340	16	0.061383	0.938615	0.000002	31.59840	0.0925
HsHK2	16	λ^E	32.96630	20	0.056368	0.943629	0.000003	34.76250	0.0545
HsHK2	25	λ^E	35.34790	25	0.053100	0.946886	0.000015	36.61110	0.0357
HsHK2	36	λ^E	36.85580	26	0.050721	0.949259	0.000019	37.78430	0.0252
HsHK2	49	λ^E	37.86820	22	0.048855	0.951144	0.000001	38.57600	0.0187
HsHK2	64	λ^E	38.58160	26	0.047409	0.952590	0.000002	39.13760	0.0144

TABLE 6 | Optimal parameters λ^* for the SE-scenario with ScHK2 (top) and HsHK2 (bottom), for one bead and different runtimes T on a mesh with $n_h = 820$.

Enzyme	T	$\lambda^{(0)}$	J^0	k^*	λ_{inv}^*	λ_{hk}^*	λ_{pgi}^*	J^*	$\frac{J^* - J^0}{J^0}$
ScHK2	1800	λ^U	0.053486	17	0.138824	0.808964	0.052212	0.100827	0.8851
ScHK2	3600	λ^U	0.228066	16	0.120404	0.815447	0.064149	0.446978	0.9599
ScHK2	5400	λ^U	0.524945	12	0.109005	0.826746	0.064249	1.060920	1.0210
ScHK2	7200	λ^U	0.941239	11	0.100778	0.837288	0.061934	1.944520	1.0659
ScHK2	9000	λ^U	1.473430	13	0.094578	0.845965	0.059457	3.092240	1.0987
ScHK2	10,800	λ^U	2.117810	19	0.089829	0.852873	0.057298	4.495210	1.1226
ScHK2	12,600	λ^U	2.870590	19	0.085547	0.859325	0.055128	6.142920	1.1400
ScHK2	14,400	λ^U	3.727980	16	0.082116	0.864621	0.053264	8.024060	1.1524
HsHK2	1800	λ^E	2.246900	12	0.104737	0.895244	0.000019	4.001410	0.7809
HsHK2	3600	λ^E	8.649300	8	0.091012	0.908980	0.000008	13.12310	0.5172
HsHK2	5400	λ^E	17.80560	8	0.083651	0.916269	0.000080	24.82340	0.3941
HsHK2	7200	λ^E	28.77910	9	0.078851	0.921146	0.000003	38.03850	0.3217
HsHK2	9000	λ^E	31.53360	10	0.075383	0.924615	0.000002	52.22680	0.6562
HsHK2	10,800	λ^E	54.15470	9	0.072366	0.927633	0.000001	67.09490	0.2389
HsHK2	12,600	λ^E	67.98330	9	0.070322	0.929672	0.000006	82.46000	0.2129
HsHK2	14,400	λ^E	82.34300	9	0.068190	0.931809	0.000001	98.20600	0.1926

and 12% for invertase, between 82 and 94% for hexokinase and between 3 and 6% for phosphoglucose isomerase. With an increasing number of beads, the proportion of ScHK2 increases and the proportions of invertase and phosphoglucose isomerase decrease. The exact results are listed in **Table 5**, where for each number of beads n_b , the optimal stoichiometry λ^* as well as the corresponding total outflow J^* are displayed. From the arrangement (O.2), we find that the optimal stoichiometry for one bead is between 8 and 13% for invertase, between 80 and 87% for hexokinase and between 5 and 6% for phosphoglucose isomerase, again with an increase in the proportion of binding sites loaded with ScHK2 in expense of proportion for ScINV, but with an almost constant proportion of binding sites loaded with ScPGI when the runtime is successively increased. The exact results are listed in **Table 6**, where for each run time T , the optimal stoichiometry λ^* as well as the corresponding total outflow J^* are displayed.

In the presence of the hexokinase HsHK2, we find from the first arrangement that the optimal stoichiometry for a runtime of 1 h is between 5 and 9% for invertase, between 91 and 95% for hexokinase and approximately 0% for the phosphoglucose

isomerase. With an increasing number of beads, the proportion of hexokinase increases and the proportion for invertase decreases. The exact results are listed in **Table 5**. From the arrangement (O.2), we find that the optimal stoichiometry for one bead is between 7 and 10% for invertase and between 90 and 93% for the hexokinase again with an increase in the proportion of binding sites loaded with HsHK2 in expense of the proportions for ScINV. As in arrangement (O.1), the optimal proportion of phosphoglucose isomerase is approximately 0%. The exact results are again listed in **Table 6**. We stress that the optimization results in **Table 5** correspond to local maxima.

4. DISCUSSION

In this paper, we developed mathematical techniques for the simulation and optimization of metabolic processes in biological microreactors with membrane-bounded subcompartments. The results from our models will be incorporated into the design of a microreactor and will be validated by experiments with this microreactor. The microreactor is based on a microfluidic

system, and consists of chambers separated by membranes which carry specific transporters for input of substrates and export of products. Inside the chambers, magnetic nano-beads carrying multienzyme-complexes are distributed, and immobilized by an outer magnetic field.

We have set up a mathematical model describing the spatio-temporal dynamics of the metabolite concentrations involved in the assembled metabolic pathway, and develop efficient and accurate approaches for the numerical simulation of this model. Furthermore, we show that model-based optimization for such systems is feasible by the methods presented. These approaches are applied to the proof-of-concept microreactor carrying out the synthetic pathway for the conversion of sucrose to glucose-6-phosphate. The obtained results shed light on the following important questions linked to the design and functionality of microreactors.

4.1. Are Spatially Resolved Models Necessary in the Description of Biochemical Microreactors?

The simulation of the distribution of metabolites, e.g., that of G6P in the presence of the enzyme HsHK2 (Figure 4), shows spatial patterns, which differ for different experimental arrangements. In case of completely loaded beads (Figure 4, left) the highest concentration of G6P can be found around the beads. If the beads are just partially loaded (Figure 4, right), a concentration gradient across the chamber prevails. In this case, the total number of enzymes distributed to the beads corresponds to the number of enzymes on one completely loaded bead. These spatial patterns in the metabolite distributions confirm the demand of spatially resolved models in the description of metabolic processes carried out in synthetic microreactors and, ultimately in living cells, which are more complex in architecture. Furthermore, the limiting effect of diffusion also shows that spatial effects have to be taken into account. This is visible in arrangement (S.2), where in the initial phase of the simulation the production rate of G6P at the beads decreases with increasing number of beads (Figure 6, upper figures); however, the outflow rate of G6P is higher for a large number of beads. This results from the fact that for high number of beads there are more beads close to the outflow boundary Γ_e , and consequently G6P diffusion to the boundary Γ_e needs less time. We mention, however, that for longer runtimes, the higher production rate of G6P for lower numbers of beads is predominating, and the outflow rates of G6P are higher for less number of beads (Figure 6, lower pictures).

4.2. Has Spatial Organization and Microcompartmentation of Enzymes the Potential to Influence the Efficiency of Metabolic Pathways?

Comparing the outflow rate of the product G6P for a fixed total enzyme quantity which is distributed on different numbers of beads, it turns out that the production rate can be increased with decreasing number of beads (Table 4; Figure 6, lower figures), especially, assembling all enzymes on one bead leads to the maximal outflow of G6P. This remains true also after comparing the

upper scenarios with the scenario, in which enzymes are distributed in the bulk fluid within the reactor chamber. This last scenario turns out to lead to the lowest G6P outflow. For the scenario with one bead and the hexokinase HsHK2, the outflow of G6P is approximately 60% higher than the total outflow for the scenario with enzymes distributed within the bulk fluid with the same hexokinase (Table 4). These results suggest that microcompartmentation of enzymes increases the total output of G6P, not only in the model but also *in vitro*.

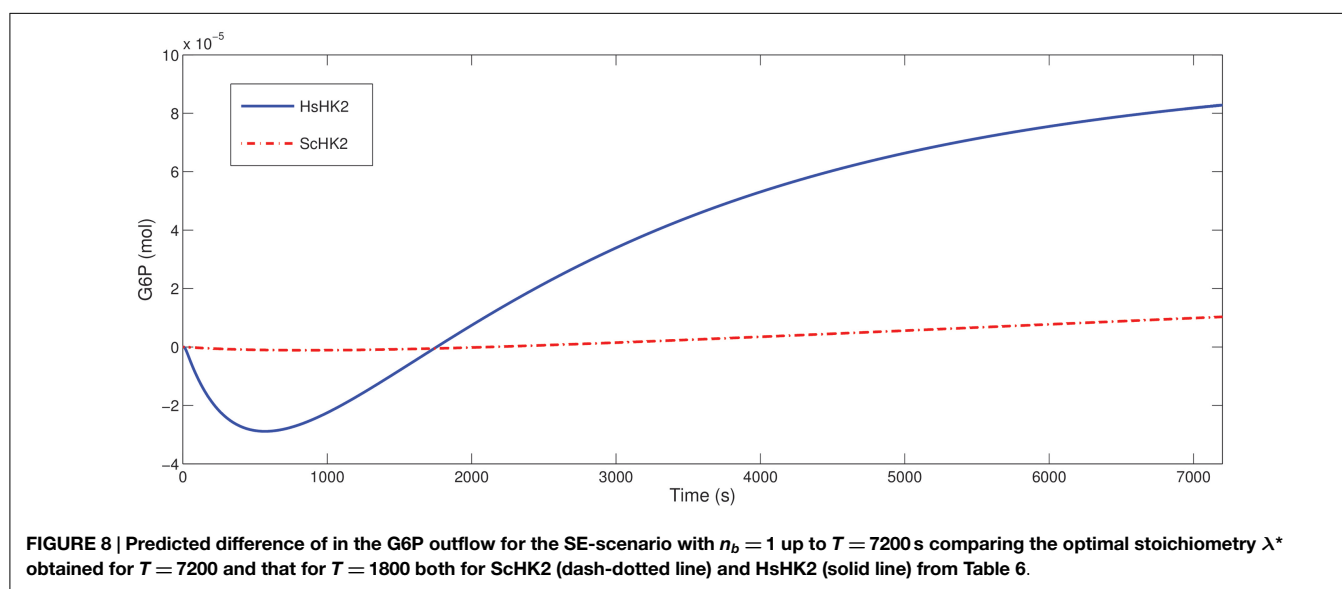
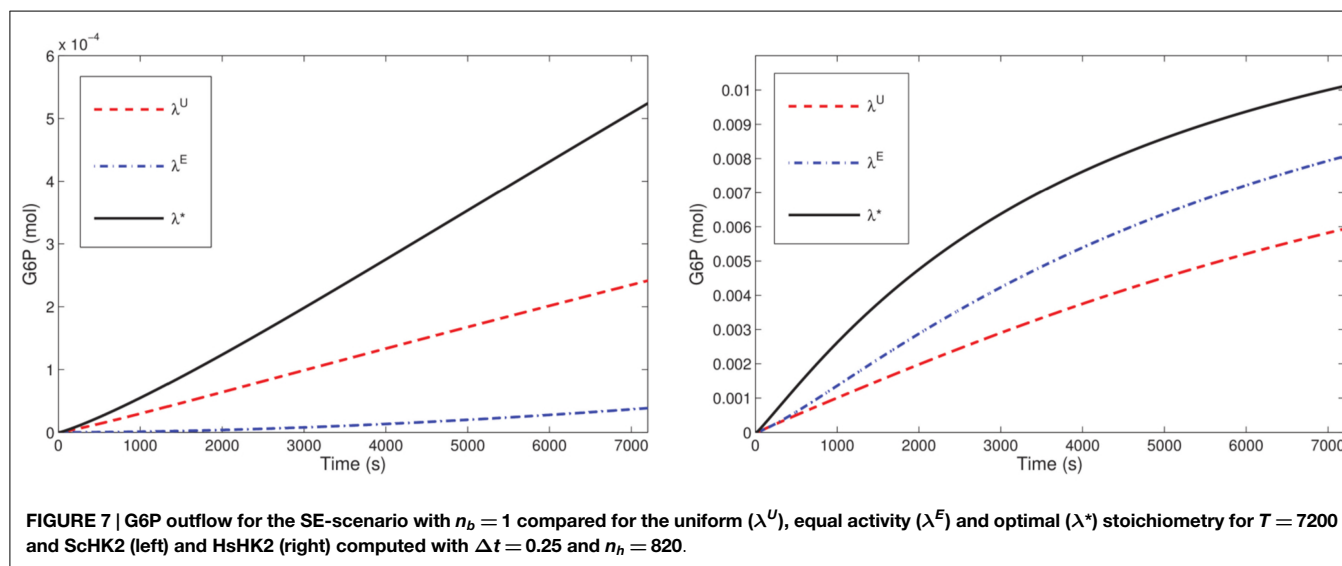
4.3. What Are the Limitations for the Productivity of the Modeled (Synthetic) Membrane-Bounded Microreactors?

The simulation results for the arrangement (S.1) (with completely loaded beads) show that the outflow rates for G6P reach saturation, in spite of the fact that the activity inside the chamber increases with increasing number of beads. This can directly be seen for the hexokinase HsHK2 (Figure 5, right bottom), whereas for ScHK2 this effect does not occur up to the run time of 3 h (Figure 5, left bottom). However, our computations show that at later time, this limiting effect also occurs for this hexokinase. This suggests that the limiting factor of the G6P outflow is the activity of the transporters in the membrane Γ_e . Finally, we mention that for the arrangement (S.2) the outflow rates are not reaching saturation, which implies that the activity of the transporters is not limiting the outflow for the chosen enzyme concentration $[E]_{\text{full}}$.

4.4. Can Optimal Design Help to Improve the Yield of a Biosynthetic Pathway?

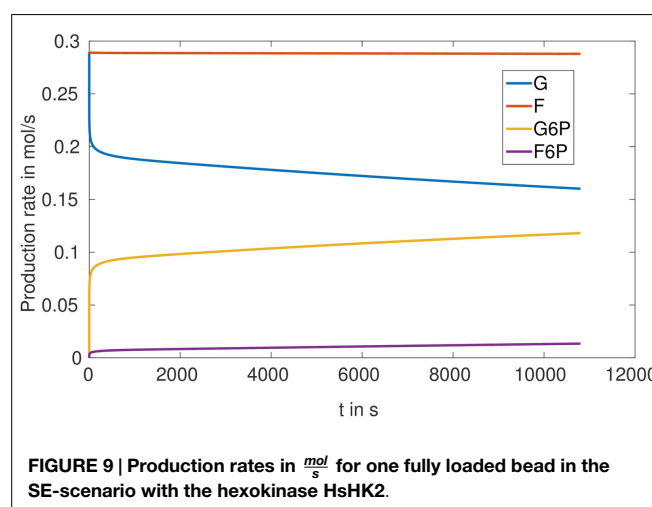
Our optimization results for the microreactor in the SE-scenario with the hexokinase ScHK2, show that after a run time $T = 3600$ s, the total outflow of G6P, denoted by J^* , for the optimal stoichiometry λ^* is on average twice as much as the total outflow J^0 for the uniform stoichiometry λ^U , and even much more than for the stoichiometry λ^E , e.g., 2746% in case of one bead. For one bead, this improvement is also visualized in Figure 7 (left). When run time is successively increased (for a reactor with one bead) starting from $T = 1800$ s, again the optimal stoichiometries enhance the total outflow of the desired product G6P on average twofold compared to the uniform stoichiometry and even much more compared to the stoichiometry λ^E . The comparison of the G6P outflow over time for stoichiometries computed for different runtimes shows that a larger proportion of binding sites loaded with ScINV leads to a higher G6P outflow at the initial phase while a larger proportion of binding sites loaded with ScHK2 leads to a higher G6P outflow on the long run (Figure 8).

Optimization of the microreactor setup with the hexokinase HsHK2 suggests that the phosphoglucose isomerase can be eliminated for the run times considered in the optimization scenarios. Furthermore, in contrast to the results for ScHK2, the degree of improvement of the optimal stoichiometries is between 50% (for one bead) and 1% (for 64 beads) compared to the stoichiometry computed from equal activity and slightly higher compared to the uniform stoichiometry. For one bead, the improvement is also visualized in Figure 7 (right). When time is successively increased (for a reactor with one bead) from $T = 1800$ –14,400 s, the optimal stoichiometries now enhance the total outflow of G6P. This



enhancement is 20% up to 78% compared to the stoichiometry computed from equal activity and slightly more compared to the uniform stoichiometry, e.g., 147% for $T = 1800$. The observation obtained for ScHK2 concerning the role of the invertase and hexokinase for the initial phase and on the long run is now even more pronounced (Figure 8).

To understand the optimization result obtained for the proportion of phosphoglucose isomerase (the enzyme catalyzing the reaction $G6P \rightleftharpoons F6P$), in case of the microreactor with HsHK2, namely that this enzyme should be better eliminated, we simulated the production rates for all metabolites, for one completely loaded bead (Figure 9). We see that, with respect to our considered time interval, the production rates for G and F increase very fast to a maximum-level around $0.3 \frac{\text{mol}}{\text{s}}$. The reaction rates in Table 2 imply that the conversion of S into F is constant, since the concentration of S is constant. Furthermore, Figure 9 shows that the production rate of F6P increases in time, but the production



rate of F remains nearly constant. Hence, we conclude that there is a metabolic flux from G6P to F6P, what contradicts the reactor to obtain a maximal outflow of G6P, and should be avoided by considering a small activity of phosphoglucose isomerase.

We stress that none of the reported maxima can be guaranteed to be global optima. Nevertheless, concerning the quality of the maxima, we note that all search paths from $\lambda^{(0)}$ to λ^* transverse a large part of the parameter space and are associated with a significant decrease in the costs. Moreover, we have tested various initial points $\lambda^{(0)}$ for the scenario with one bead, HsHK2 and $T = 3600$. In all cases, we have observed convergence to the same λ^* . We conclude that the reported local maxima are reasonable candidates for an optimized stoichiometry for the respective microreactor.

Finally, we emphasize that the simulation and optimization approaches developed in this work can be repeated with minor adaptations for more complex biochemical microreactors and for other optimization parameters, like, e.g., other real parameters entering the reaction kinetics.

AUTHOR CONTRIBUTIONS

MN-R and LV gave substantial contributions to the Abstract and Introduction. MG and MN-R gave substantial contributions to

the conception of the Derivation of the Mathematical Model, see Section 2.1. TE and PK gave substantial contributions to the conception of the Numerical Analysis and Simulations, see Section 2.2 and 3.1. FH and GL gave substantial contributions to the conception of the Optimization Methods and Results, see Section 2.3 and 3.2. LV gave substantial contributions to the acquisition, analysis, and interpretation of data for the work. All authors contributed essentially in the discussion of the results, see Section 4. Also, all authors have revisited the work critically for important intellectual content, approved the version to be published, and agreed to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

The work of the first and the second author was supported by the Emerging Fields Initiative (EFI) for Synthetic Biology at the Friedrich-Alexander-Universität Erlangen-Nürnberg.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2016.00013>

REFERENCES

- Asanomi, Y., Yamaguchi, H., Miyazaki, M., and Maeda, H. (2011). Enzyme-immobilized microfluidic process reactors. *Molecules* 16, 6041–6059. doi:10.3390/molecules16076041
- Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. (2010). Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu. Rev. Biophys.* 39, 515. doi:10.1146/annurev.biophys.050708.133652
- Boyle, P. M., and Silver, P. A. (2009). Harnessing nature's toolbox: regulatory elements for synthetic biology. *J. R. Soc. Interface* 6(Suppl. 4), S535–S546. doi: 10.1098/rsif.2008.0521.focus
- Chen, A. H., and Silver, P. A. (2012). Designing biological compartmentalization. *Trends Cell Biol.* 22, 662–670. doi:10.1016/j.tcb.2012.07.002
- Forsgren, A., Gill, P. E., and Wright, M. H. (2002). Interior methods for nonlinear optimization. *SIAM Rev.* 44, 525–597. doi:10.1137/S0036144502414942
- Gao, H., and Leary, J. A. (2003). Multiplex inhibitor screening and kinetic constant determinations for yeast hexokinase using mass spectrometry based assays. *J. Am. Soc. Mass Spectrom.* 14, 173–181. doi:10.1016/S1044-0305(02)00867-X
- Gloyn, A., Odili, S., Zelen, D., Buettger, C., Castleden, H., Steele, A., et al. (2005). Insights into the structure and regulation of glucokinase from a novel mutation (v62m), which causes maturity-onset diabetes of the young. *J. Biol. Chem.* 280, 14105–14113. doi:10.1074/jbc.M413146200
- Hante, F. M., and Sager, S. (2013). Relaxation methods for mixed-integer optimal control of partial differential equations. *Comput. Optimization Appl.* 55, 197–225. doi:10.1007/s10589-012-9518-3
- Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. (2009). *Optimization with PDE Constraints, Volume 23 of Mathematical Modelling: Theory and Applications*. New York: Springer.
- Kräutle, S., and Knabner, P. (2005). A new numerical reduction scheme for fully coupled multicomponent transport-reaction problems in porous media. *Water Resour. Res.* 41, 09414. doi:10.1029/2004WR003624
- Lafraya, A., Sanz-Aparicio, J., Polaina, J., and Marin-Navarro, J. (2011). Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. *Appl. Environ. Microbiol.* 77, 6148–6157. doi:10.1128/AEM.05032-11
- Leugering, G., Novotny, A. A., Perla Menzala, G., and Sokolowski, J. (2011). On shape optimization for an evolution coupled system. *Appl. Math. Optimization* 64, 441–466. doi:10.1007/s00245-011-9148-7
- Lin, H.-Y., Kao, Y.-H., Chen, S.-T., and Meng, M. (2009). Effects of inherited mutations on catalytic activity and structural stability of human glucose-6-phosphate isomerase expressed in *Escherichia coli*. *Biochim. Biophys. Acta* 1794, 315–323. doi:10.1016/j.bbapap.2008.11.004
- Marguet, M., Bonduelle, C., and Lecommandoux, S. (2013). Multicompartmentalized polymeric systems: towards biomimetic cellular structure and function. *Chem. Soc. Rev.* 42, 512–529. doi:10.1039/c2cs35312a
- MATLAB and Statistic Toolbox Release. (2014). *The Mathworks, Inc.*, Natick, MA, USA.
- Nomura, A., Shin, S., Mehdi, O. O., and Kauffmann, J.-M. (2004). Preparation, characterization, and application of an enzyme-immobilized magnetic microreactor for flow injection analysis. *Anal. Chem.* 76, 5498–5502. doi:10.1021/ac049489v
- Peters, R. J., Louzao, I., and van Hest, J. C. (2012). From polymeric nanoreactors to artificial organelles. *Chem. Sci.* 3, 335–342. doi:10.1039/C2SC00803C
- 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
- Raviart, P. A., and Thomas, J. M. (1977). A mixed finite element method for second-order elliptic problems. *Math. Aspects Finite Elem. Method* 606, 292–315. doi:10.1007/BFb0064470
- Roodbeen, R., and van Hest, J. (2009). Synthetic cells and organelles: compartmentalization strategies. *Bioessays* 31, 1299–1308. doi:10.1002/bies.200900106
- Segel, I. H. (1975). *Enzyme Kinetics*. New York: John Wiley & Sons, Inc.
- Somarowthu, S., Brodtkin, H., D'Aquino, J., Ringe, D., Ondrechen, M., and Beuning, P. (2011). A tale of two isomerases: compact versus extended active sites in ketosteroid isomerase and phosphoglucose isomerase. *Biochemistry* 50, 9283–9295. doi:10.1021/bi201089v
- Wieners, C. (2005). “Distributed point objects. a new concept for parallel finite elements,” in *Domain Decomposition Methods in Science and Engineering, Volume 40 of Lecture Notes in Computational Science and Engineering*, eds T. Barth, M. Griebel, D. Keyes, R. Nieminen, D. Roose, T. Schlick, et al. (Berlin, Heidelberg: Springer), 175–182.

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 © 2016 Elbinger, Gahn, Neuss-Radu, Hante, Voll, Leugering and Knabner. 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.



Underpinning starch biology with *in vitro* studies on carbohydrate-active enzymes and biosynthetic glycomaterials

Ellis C. O'Neill¹ and Robert A. Field^{2*}

¹ Department of Plant Sciences, University of Oxford, Oxford, UK, ² Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, UK

OPEN ACCESS

Edited by:

Lars Matthias Voll,
Friedrich-Alexander-University
Erlangen-Nuremberg, Germany

Reviewed by:

Weiwen Zhang,
Tianjin University, China
M. Kalim Akhtar,
University of Edinburgh, UK

*Correspondence:

Robert A. Field,
Department of Biological Chemistry,
John Innes Centre, Norwich Research
Park, Norwich NR4 7UH, UK
rob.field@jic.ac.uk

Specialty section:

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

Received: 03 July 2015

Accepted: 24 August 2015

Published: 07 September 2015

Citation:

O'Neill EC and Field RA (2015)
Underpinning starch biology with
in vitro studies on carbohydrate-
active enzymes and biosynthetic
glycomaterials.
Front. Bioeng. Biotechnol. 3:136.
doi: 10.3389/fbioe.2015.00136

Starch makes up more than half of the calories in the human diet and is also a valuable bulk commodity that is used across the food, brewing and distilling, medicines and renewable materials sectors. Despite its importance, our understanding of how plants make starch, and what controls the deposition of this insoluble, polymeric, liquid crystalline material, remains rather limited. Advances are hampered by the challenges inherent in analyzing enzymes that operate across the solid-liquid interface. Glyconanotechnology, in the form of glucan-coated sensor chips and metal nanoparticles, present novel opportunities to address this problem. Herein, we review recent developments aimed at the bottom-up generation and self-assembly of starch-like materials, in order to better understand which enzymes are required for starch granule biogenesis and metabolism.

Keywords: starch, glucans, phosphorylase, self-assembly, nanotechnology, synthetic biology

Introduction

Starch makes the highest calorific contribution to the human diet (U.S. Department of Agriculture, 2012) and is a bulk commodity, with a global market of billions of tons per year (Ellis et al., 1998). We cannot meet the ever-increasing demand for starch-based products solely by further commitment of agricultural land to conventional starch-based crops – as well as increases in crop yield, new ways to produce modified starch-based materials are also needed (Diouf, 2009). In order to achieve this goal, the production of modified starches *in planta* is an attractive alternative to post harvest modification (Jobling, 2004) – i.e., moving the diversification of starch functionality from the chemical plant into the crop plant. To this end, better understanding of the enzymes involved in starch metabolism and those used in industrial starch modification is needed.

Starch Structure and Metabolism

Starch granules comprise linear α -1,4-glucans with periodic α -1,6-branches, giving long chains capable of wrapping around each other to form double helical arrangements, which stack side by side, forming alternating layers of highly ordered liquid crystalline lamella interspersed with amorphous regions (**Figure 1**) (Waigh et al., 1998). This self-organizing nanostructure makes the surface of a starch granule highly resistant to enzymatic attack, requiring specialized enzymes to initiate degradation.

In order to engineer precisely defined starch structures, we need to understand the metabolic enzymes involved, insights into which are provided by continued research in the model plant

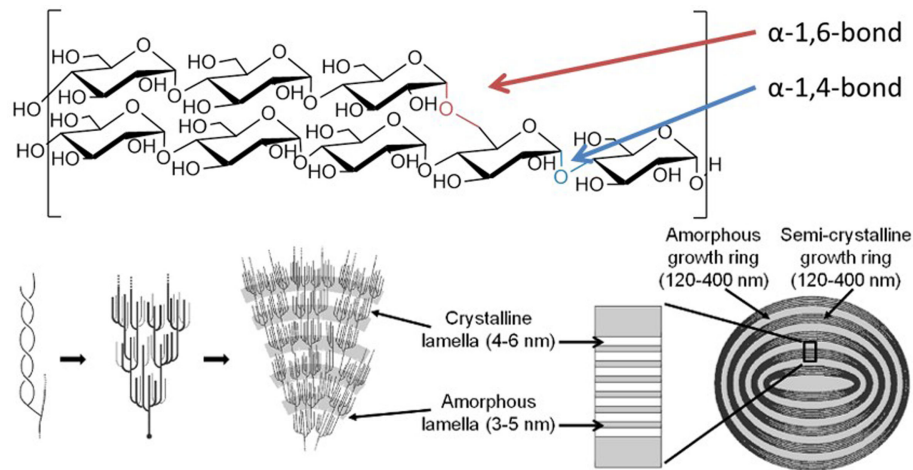


FIGURE 1 | Structure of the starch granule. Starch granules are composed of α -1,4-glucans with some α -1,6-branches, placing them parallel to each other, allowing double helix formation. The branches are specifically placed, giving regions of branching, known as amorphous lamella, and regions of exclusively linear chains which form crystalline lamella, and these lamella form defined growth rings. Specific enzymes are needed to synthesize and degrade this highly ordered insoluble structure. Coultate (2002) – Reproduced by permission of The Royal Society of Chemistry.

Arabidopsis thaliana (Santelia and Zeeman, 2011). Multiple enzymes are required to create the correct morphology of the starch granule, and the specific role of individual enzyme isoforms is not well-understood (Ball and Morell, 2003; Tetlow, 2011). For example, the correct *synthesis* of granular starch counter-intuitively requires two *degradative* isoamylases (Bustos et al., 2004). Storage starch is formed in the amyloplasts of specialist storage organs, such as the grain or tuber, while transitory starch is stored in the leaf chloroplasts, as a carbon source for photosynthetic cells at night. The two plastids use different suites of enzymes to attack the starch granule (Smith et al., 2005; Smirnova et al., 2015). In the storage organs, the granule is enzymatically hydrolyzed (Radchuk et al., 2009), ultimately releasing glucose; in photosynthetic organs, the transitory starch granule is initially phosphorylated by specific glucan kinases, before being broken down into short maltooligosaccharides (Fettker et al., 2009). There are several isoforms of each class of enzyme involved in starch synthesis and breakdown, but differences in their action are generally interpreted in terms of the expression of the gene and protein in question, rather than of the enzyme activity *per se*. The complication of enzymatic reactions taking place on an insoluble starch substrate accounts for these crude approximations, a situation that needs to be addressed in order to inform plant engineering/synthetic biology studies.

Current Status on the Enzymatic Degradation of Insoluble Starch

The enzymology of the starch granule actually takes place at a solid-liquid interface, rather than solely in the solution phase used in most experiments. This can have profound effects on the reaction; for example, the reaction rate of granule-bound starch synthase is enhanced on crystalline amylopectin (Edwards et al., 1999). There are few enzyme studies involving crystalline maltodextrins (Hejazi et al., 2009) or purified starch granules

(Edner et al., 2007), but even these studies measured discreet product formation, not breakdown of the starch granule. Electron microscopy clearly shows that the initial attack on the surface alters the granule by disrupting the insoluble packing of the glucan chains (Sun and Henson, 1990), undermining endpoint kinetic assessments. Direct monitoring of purified starch granules during degradation, either using electron microscopy (Planchot et al., 1995) or synchrotron radiation (Tawil et al., 2011), provides snapshots of these processes during the degradation of the granule. Clearly such experiments provide information on structures, but not on reaction rate as such.

In Vivo Modification of Starch

The source of starch can have a marked impact on physico-chemical properties that are important for industrial uses. For instance, cereal starch is virtually free of phosphate, while potato starch has high phosphate content, with increased viscosity and decreased crystallization, factors important in the paper industry (Blennow et al., 2003). For commercial applications, starches may need to be chemically modified – by fragmentation, oxidation, or esterification – although this is costly and potentially hazardous. Production of fit-for-purpose modified starches *in planta* is therefore an attractive alternative (Jobling, 2004) and has been achieved in a number of ways. For instance, the gelatinization properties of the starch can be controlled by using genetic engineering to alter the ratio of linear amylose to branched amylopectin (Visser et al., 1991; Jobling et al., 2002). Removal of the glucan water dikinase, which adds phosphate to starch chains during degradation, from potato tubers produced a starch that was resistant to degradation during storage (Lorberth et al., 1998). By replacing the plant disproportionating enzyme, which transfers glucans between chains during starch degradation, with its bacterial homolog the complex soluble heteroglycan could be essentially bypassed during transitory starch degradation (Ruzanski et al., 2013). It has also been possible to engineer a vaccine-displaying

starch by fusing starch-binding proteins to known antigens in algal chloroplasts (Dauvillée et al., 2010). However, it is not easy to predict outcomes at the whole plant level due to poor knowledge of individual enzyme functions (Stanley et al., 2011), coupled with compensatory mechanisms and genetic redundancy or partially overlapping enzymatic capabilities.

In Vitro Synthesis and Modification of Starches

A number of approaches have been investigated to produce natural and non-natural starch-like materials *in vitro*. To simulate the branch point of a starch granule, specific glucans could be linked synthetically using “click chemistry,” producing known length chains with known positioning of the branch point (Marmuse et al., 2005; Nepogodiev et al., 2007). Longer unbranched amylose polymers can be synthesized enzymatically in solution, either by the debranching of amylopectin with isoamylase (Harada et al., 1972) or by the extension of acceptor glucans with amylosucrase or glucan phosphorylase (Yanase et al., 2007). Amylosucrase has been used to synthesize dendritic nanoparticles, based on a glycogen core (Putaux et al., 2006); in combination with branching enzyme, a highly branched glucan can be accessed from sucrose (Grimaud et al., 2013). Amylomaltases, which transfer glucans from one chain onto another, can also be used to generate fluorinated glucans (Tantanarat et al., 2012) and to modify starches in order to make thermoreversible gels (Kaper et al., 2005). Interestingly, phosphorylases also display some promiscuity toward the donor substrate and they have been used to prepare a range of modified glucans (O'Neill and Field, 2015). 2-Deoxy-maltooligosaccharides (Klein et al., 1982) have been synthesized from D-glucal in the presence of Pi, which can then be phosphorylated to synthesize 2-deoxy- α -glucose-1-phosphate. Deoxy- and fluoro-glucose moieties can also be transferred onto glycogen by glucan phosphorylase, but in low yield (Withers, 1990). Alternative sugar-1-phosphates can be utilized by glucan phosphorylase, including those derived from xylose (Nawaji et al., 2008b), mannose (Evers and Thiem, 1997), glucosamine (Nawaji et al., 2008a), *N*-formyl-glucosamine (Kawazoe et al., 2010), and glucuronic acid (Umegatani et al., 2012), although the products were all isolated after a single residue extension, indicating that these sugars cannot be bound in the acceptor site of the enzyme for further extension. Under certain conditions, and with careful choice of phosphorylase, further extension has been achieved to make a polyglucosamine, analogous to the α -stereoisomer of chitosan (Kadokawa et al., 2015). By addition of both glucosamine and glucuronic acid onto a glycogen core, pH-responsive amphoteric hydrogels could be synthesized (Takata et al., 2015).

Plant phosphorylases, with their lack of allosteric regulation (Fukui et al., 1982), have proven more useful than the mammalian equivalent in the synthesis of long chain amylose derivatives, for instance, by extension of glucan immobilized on chitosan (Kaneko et al., 2007) or on polystyrene (Loos and Müller, 2002), or by twining polysaccharides around a hydrophobic core to assemble a macromolecular complex, such as amylose-wrapped lipid (Gelders et al., 2005).

Engineering Insoluble Starch Surfaces for Enzymatic Analysis

When amylose is synthesized in solution, it self-assembles into crystalline structures (Buleon et al., 2007). Using electron microscopy, this type of crystalline material can be seen after extension of glycogen particles using amylosucrase (Putaux et al., 2006) and X-ray diffraction shows that it has assembled into head-to-tail, B-type starch (Potocki-Veronese et al., 2005). Many physiologically important enzymatic reactions occur on surfaces and their analysis is far from trivial. A range of techniques have been deployed to address this point, including mass spectrometry, radioactivity, and fluorescence-based assays (Gray et al., 2013). Quartz crystal microbalance technology has also been used to monitor real-time extension of amylopectin by phosphorylase (Murakawa et al., 2007) and to provide associated kinetic information (Nishino et al., 2004). The recent development of plant oligosaccharide microarrays also enables high throughput analysis of carbohydrate-active enzymes (Pedersen et al., 2012).

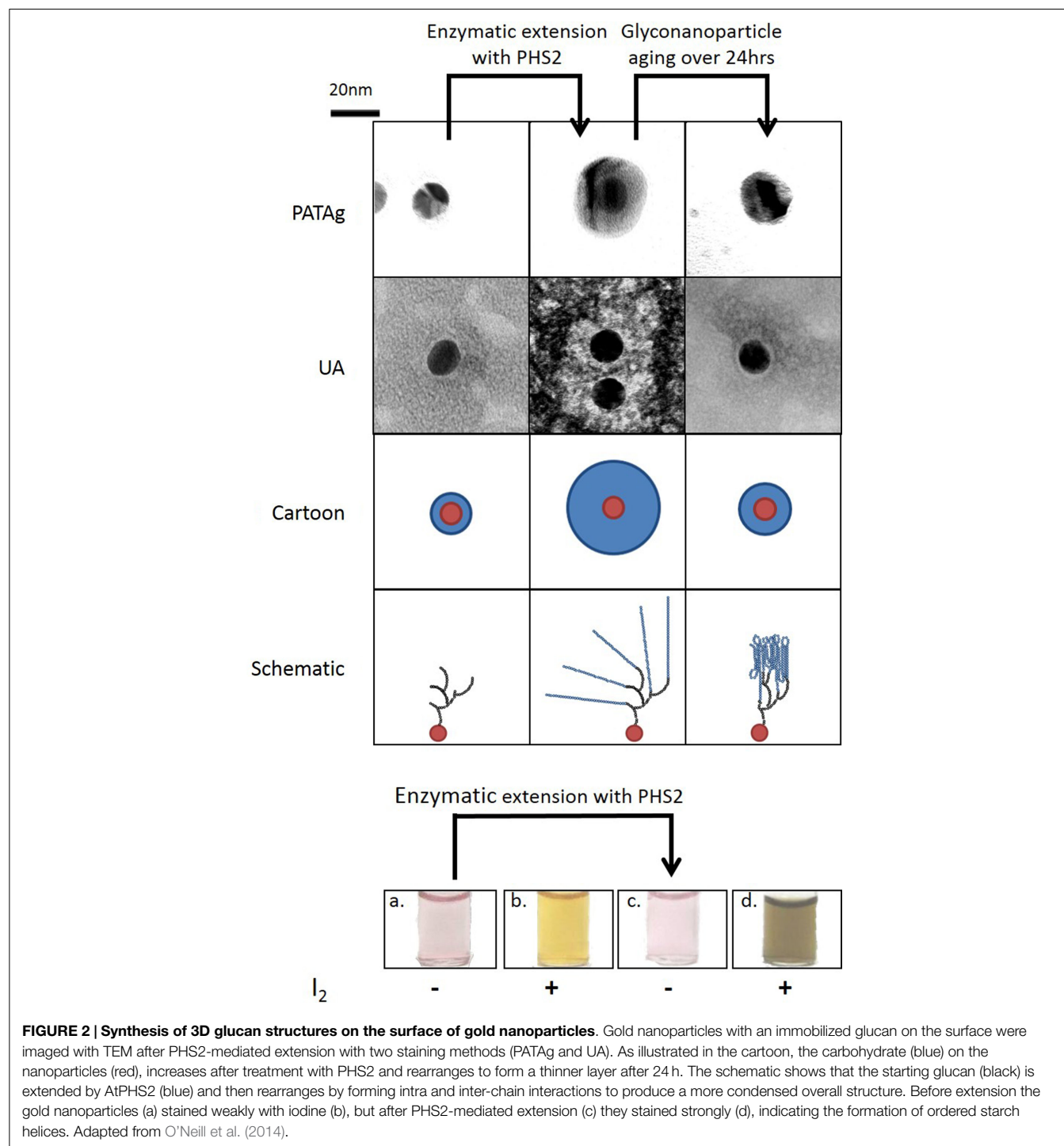
While it has been possible to detect glycosylation reactions on surfaces using surface plasmon resonance (SPR), this technique has typically been deployed to assess non-catalytic lectin binding to sugars (Karamanska et al., 2008). Direct measurement of *trans*-glycosylation reactions on SPR sensors has been achieved (Cle et al., 2008, 2010), but the enzymes to which this technique has been applied use soluble substrates and act in solution. Any kinetic analysis has to therefore take into account the unnatural nature of the surface-immobilized substrates. Starch-active enzymes would benefit markedly from studies using immobilized substrates, which mimic the insoluble surface upon which they naturally act.

The plant phosphorylase, PHS2, can rapidly synthesize α -1,4-glucans in solution and, by immobilizing their acceptor oligosaccharide substrates, insoluble glucan surfaces could be developed (O'Neill et al., 2014). The SPR surfaces behaved in a manner dependent on the density of the immobilized glucan. When relatively dilute, the glucan surface behaved in the same way as in solution; when a denser surface was used, the material produced took on the pattern of enzyme resistance seen in natural granular starch. These results indicate that the amyloglucan polymer created by PHS2 on the high density surface formed a macromolecular architecture, which may be crystalline, rendering the surface resistant to enzymatic digestion. These surfaces could be used to assay proteins that interact with the starch granule to ascertain binding kinetics in classical SPR experiments. Starch-degrading enzymes are normally assayed in solution, which is not a good analog of their natural insoluble granule substrate. Classic enzyme kinetics cannot be utilized to study reactions on an insoluble surface and care will be needed to differentiate between the binding of soluble enzymes to the insoluble surface or the reaction itself as the rate limiting step. This new surface technology offers the prospect of a more informative assay to provide kinetic information on starch surface degradation.

While a quantifiable 2D system is relevant for kinetic analysis of starch-active enzymes, a third dimension is required to represent the true spatial arrangement of the starch granule. The uniform

shape and size and defined physical and chemical properties of gold nanoparticles make them useful materials for the study of biological interactions (Saha et al., 2012). Gold nanoparticles can be simultaneously used to display carbohydrates and to provide visual output of interactions (Marin et al., 2015), using the same photophysical effects exploited in conventional SPR studies, to give a change in color from red to purple. Glucan primers immobilized on gold nanoparticles could be extended substantially

by PHS2 (O'Neill et al., 2014). The resulting nanoparticle-based glucan surfaces were subject to aging, with the glucan layer appearing to reorganize and assemble into a much more tightly packed structure over the course of ~12 h (**Figure 2**). This type of glyconanoparticle has potential for the analysis of enzymes that naturally act on starch granules, and may serve as models for bottom-up, *in vitro* synthetic biological approaches to biocompatible inorganic surfaces.



Conclusion

Reflecting the global need for sustainability and a drive toward environmentally benign manufacturing practices, new ways to produce bulk commodities, such as starch, are key. In order to achieve this goal, better information about how nature handles (bio)chemistry across the liquid–solid interface is necessary. Here, we have highlighted the potential of bottom-up synthetic biology approaches to produce starch-like surfaces in a format suitable for both structural analysis and real-time kinetic assessment of enzyme action thereon. These glyconanotechnology methods will

help to pave the way to a more complete understanding of natural starch metabolism and may be used to inform understanding and exploitation of the natural processes, including the recapitulation of starch production in non-starch producing organisms.

Acknowledgments

The authors gratefully acknowledge support from the UK BBSRC Institute Strategic Programme Grant on Understanding and Exploiting Metabolism (MET) [BB/J004561/1] and the John Innes Foundation.

References

- Ball, S. G., and Morell, M. K. (2003). From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu. Rev. Plant Biol.* 54, 207–233. doi:10.1146/annurev.arplant.54.031902.134927
- Blennow, A., Bay-Smidt, A. M., Leonhardt, P., Bandholm, O., and Madsen, M. H. (2003). Starch paste stickiness is a relevant native starch selection criterion for wet-end paper manufacturing. *Starch. Stärke* 55, 381–389. doi:10.1002/star.200300169
- Buleon, A., Veronese, G., and Putaux, J. L. (2007). Self-association and crystallization of amylose. *Aust. J. Chem.* 60, 706–718. doi:10.1071/ch07168
- Bustos, R., Fahy, B., Hylton, C. M., Seale, R., Nebane, N. M., Edwards, A., et al. (2004). Starch granule initiation is controlled by a heteromultimeric isoamylase in potato tubers. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2215–2220. doi:10.1073/pnas.0305920101
- Cle, C., Gunning, A. P., Syson, K., Bowater, L., Field, R. A., and Bornemann, S. (2008). Detection of transglucosidase-catalyzed polysaccharide synthesis on a surface in real time using surface plasmon resonance spectroscopy. *J. Am. Chem. Soc.* 130, 15234–15235. doi:10.1021/ja805264w
- Cle, C., Martin, C., Field, R. A., Kuzmic, P., and Bornemann, S. (2010). Detection of enzyme-catalyzed polysaccharide synthesis on surfaces. *Biocatal. Biotransformation* 28, 64–71. doi:10.3109/10242420903388744
- Coulter, T. P. (2002). *Food: The Chemistry of its Components*. London: Royal Society of Chemistry.
- Dauvillée, D., Delhay, S., Gruyer, S., Slomianny, C., Moretz, S. E., d'Hulst, C., et al. (2010). Engineering the chloroplast targeted malarial vaccine antigens in *Chlamydomonas* starch granules. *PLoS ONE* 5:e15424. doi:10.1371/journal.pone.0015424
- Diouf, J. (2009). FAO's director-general on how to feed the world in 2050. *Popul. Dev. Rev.* 35, 837–839. doi:10.1111/j.1728-4457.2009.00312.x
- Edner, C., Li, J., Albrecht, T., Mahlow, S., Hejazi, M., Hussain, H., et al. (2007). Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial β -amylases. *Plant Physiol.* 145, 17–28. doi:10.1104/pp.107.104224
- Edwards, A., Borthakur, A., Bornemann, S., Venail, J., Denyer, K., Waite, D., et al. (1999). Specificity of starch synthase isoforms from potato. *Eur. J. Biochem.* 266, 724–736. doi:10.1046/j.1432-1327.1999.00861.x
- Ellis, R. P., Cochrane, M. P., Dale, M. F. B., Duffus, C. M., Lynn, A., Morrison, I. M., et al. (1998). Starch production and industrial use. *J. Sci. Food Agric.* 77, 289–311. doi:10.1002/(SICI)1097-0010(199807)77:3<289::AID-JSFA38(3.0.CO;2-D
- Evers, B., and Thiem, J. (1997). Further syntheses employing phosphorylase. *Bioorg. Med. Chem.* 5, 857–863. doi:10.1016/S0968-0896(97)00031-X
- Fettke, J., Hejazi, M., Smirnova, J., Hochel, E., Stage, M., and Steup, M. (2009). Eukaryotic starch degradation: integration of plastidial and cytosolic pathways. *J. Exp. Bot.* 60, 2907–2922. doi:10.1093/jxb/erp054
- Fukui, T., Shimomura, S., and Nakano, K. (1982). Potato and rabbit muscle phosphorylases – comparative studies on the structure, function and regulation of regulatory and non-regulatory enzymes. *Mol. Cell. Biochem.* 42, 129–144. doi:10.1007/BF00238507
- Gelders, G. G., Goesaert, H., and Delcour, J. A. (2005). Potato phosphorylase catalyzed synthesis of amylose-lipid complexes. *Biomacromolecules* 6, 2622–2629. doi:10.1021/bm0502011
- Gray, C. J., Weissenborn, M. J., Evers, C. E., and Flitsch, S. L. (2013). Enzymatic reactions on immobilised substrates. *Chem. Soc. Rev.* 42, 6378–6405. doi:10.1039/C3CS60018A
- Grimaud, F., Lancelon-Pin, C., Rolland-Sabaté, A., Roussel, X., Laguerre, S., Viksø-Nielsen, A., et al. (2013). *In vitro* synthesis of hyperbranched α -glucans using a biomimetic enzymatic toolbox. *Biomacromolecules* 14, 438–447. doi:10.1021/bm301676c
- Harada, T., Misaki, A., Akai, H., Yokobayashi, K., and Sugimoto, K. (1972). Characterization of *Pseudomonas* isoamylase by its actions on amylopectin and glycogen: comparison with *Aerobacter pullulanase*. *Biochim. Biophys. Acta* 268, 497–505. doi:10.1016/0005-2744(72)90345-2
- Hejazi, M., Fettke, J., Paris, O., and Steup, M. (2009). The two plastidial starch-related dikinases sequentially phosphorylate glucosyl residues at the surface of both the A- and B-type allomorphs of crystallized maltodextrins but the mode of action differs. *Plant Physiol.* 150, 962–976. doi:10.1104/pp.109.138750
- Jobling, S. (2004). Improving starch for food and industrial applications. *Curr. Opin. Plant Biol.* 7, 210–218. doi:10.1016/j.pbi.2003.12.001
- Jobling, S. A., Westcott, R. J., Tayal, A., Jeffcoat, R., and Schwall, G. P. (2002). Production of a freeze-thaw-stable potato starch by antisense inhibition of three starch synthase genes. *Nat. Biotechnol.* 20, 295–299. doi:10.1038/nbt0302-295
- Kadokawa, J., Shimohigoshi, R., Yamashita, K., and Yamamoto, K. (2015). Synthesis of chitin and chitosan stereoisomers by thermostable α -glucan phosphorylase-catalyzed enzymatic polymerization of α -D-glucosamine 1-phosphate. *Org. Biomol. Chem.* 13, 4336–4343. doi:10.1039/C5OB00167F
- Kaneko, Y., Matsuda, S., and Kadokawa, J. (2007). Chemoenzymatic syntheses of amylose-grafted chitin and chitosan. *Biomacromolecules* 8, 3959–3964. doi:10.1021/bm701000t
- Kaper, T., Talik, B., Ettema, T. J., Bos, H., van der Maarel, M. J. E. C., and Dijkhuizen, L. (2005). Amylomaltase of *Pyrobaculum aerophilum* IM2 produces thermoreversible starch gels. *Appl. Environ. Microbiol.* 71, 5098–5106. doi:10.1128/aem.71.9.5098-5106.2005
- Karamanska, R., Clarke, J., Blixt, O., MacRae, J., Zhang, J., Crocker, P., et al. (2008). Surface plasmon resonance imaging for real-time, label-free analysis of protein interactions with carbohydrate microarrays. *Glycoconj. J.* 25, 69–74. doi:10.1007/s10719-007-9047-y
- Kawazoe, S., Izawa, H., Nawaji, M., Kaneko, Y., and Kadokawa, J. (2010). Phosphorylase-catalyzed N-formyl- α -glucosaminylation of maltooligosaccharides. *Carbohydr. Res.* 345, 631–636. doi:10.1016/j.carres.2010.01.001
- Klein, H. W., Palm, D., and Helmreich, E. J. M. (1982). General acid-base catalysis of α -glucan phosphorylases – stereospecific glucosyl transfer from D-glucal is a pyridoxal 5'-phosphate and ortho-phosphate (arsenate) dependent reaction. *Biochemistry* 21, 6675–6684. doi:10.1021/bi00269a010
- Loos, K., and Müller, A. H. E. (2002). New routes to the synthesis of amylose-block-polystyrene rod-coil block copolymers. *Biomacromolecules* 3, 368–373. doi:10.1021/bm0156330
- Lorberth, R., Ritte, G., Willmitzer, L., and Kossmann, J. (1998). Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nat. Biotechnol.* 16, 473–477. doi:10.1038/nbt0598-473
- Marin, M. J., Schofield, C. L., Field, R. A., and Russell, D. A. (2015). Glyconanoparticles for colorimetric bioassays. *Analyst* 140, 59–70. doi:10.1039/C4AN01466A
- Marmuse, L., Nepogodiev, S. A., and Field, R. A. (2005). “Click chemistry” en route to pseudo-starch. *Org. Biomol. Chem.* 3, 2225–2227. doi:10.1039/B504293C
- Murakawa, A., Mori, T., and Okahata, Y. (2007). Direct monitoring of both phosphorolysis and elongation of amylopectin catalyzed by phosphorylase on a 27-MHz quartz-crystal microbalance. *Chem. Lett.* 36, 312–313. doi:10.1246/cl.2007.312

- Nawaji, M., Izawa, H., Kaneko, Y., and Kadokawa, J. (2008a). Enzymatic α -glucosaminylation of maltooligosaccharides catalyzed by phosphorylase. *Carbohydr. Res.* 343, 2692–2696. doi:10.1016/j.carres.2008.08.013
- Nawaji, M., Izawa, H., Kaneko, Y., and Kadokawa, J. I. (2008b). Enzymatic synthesis of α -D-xylosylated malto-oligosaccharides by phosphorylase-catalyzed xylosylation. *J. Carbohydr. Chem.* 27, 214–222. doi:10.1080/07328300802105340
- Nepogodiev, S. A., Dedola, S., Marmuse, L., de Oliveira, M. T., and Field, R. A. (2007). Synthesis of triazole-linked pseudo-starch fragments. *Carbohydr. Res.* 342, 529–540. doi:10.1016/j.carres.2006.09.026
- Nishino, H., Murakawa, A., Mori, T., and Okahata, Y. (2004). Kinetic studies of AMP-dependent phosphorolysis of amylopectin catalyzed by phosphorylase b on a 27 MHz quartz-crystal microbalance. *J. Am. Chem. Soc.* 126, 14752–14757. doi:10.1021/ja046582k
- O'Neill, E. C., and Field, R. A. (2015). Enzymatic synthesis using glycoside phosphorylases. *Carbohydr. Res.* 403, 23–37. doi:10.1016/j.carres.2014.06.010
- O'Neill, E. C., Rashid, A. M., Stevenson, C. E. M., Hetru, A.-C., Gunning, A. P., Rejzek, M., et al. (2014). Sugar-coated sensor chip and nanoparticle surfaces for the *in vitro* enzymatic synthesis of starch-like materials. *Chem. Sci.* 5, 341–350. doi:10.1039/c3sc51829a
- Pedersen, H. L., Fangel, J. U., McCleary, B., Ruzanski, C., Rydahl, M. G., Ralet, M.-C., et al. (2012). Versatile high-resolution oligosaccharide microarrays for plant glycobiology and cell wall research. *J. Biol. Chem.* 287, 39429–39438. doi:10.1074/jbc.M112.396598
- Planchot, V., Colonna, P., Gallant, D. J., and Bouchet, B. (1995). Extensive degradation of native starch granules by alpha-amylase from *Aspergillus fumigatus*. *J. Cereal Sci.* 21, 163–171. doi:10.1016/0733-5210(95)90032-2
- Potocki-Veronese, G., Putaux, J.-L., Dupeyre, D., Albenne, C., Remaud-Siméon, M., Monsan, P., et al. (2005). Amylose synthesized *in vitro* by amylosucrase: morphology, structure, and properties. *Biomacromolecules* 6, 1000–1011. doi:10.1021/bm049326g
- Putaux, J. L., Potocki-Veronese, G., Remaud-Simeon, M., and Buleon, A. (2006). α -D-Glucan-based dendritic nanoparticles prepared by *in vitro* enzymatic chain extension of glycogen. *Biomacromolecules* 7, 1720–1728. doi:10.1021/bm050988v
- Radchuk, V. V., Borisjuk, L., Sreenivasulu, N., Merx, K., Mock, H.-P., Rolletschek, H., et al. (2009). Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain. *Plant Physiol.* 150, 190–204. doi:10.1104/pp.108.133520
- Ruzanski, C., Smirnova, J., Rejzek, M., Cockburn, D., Pedersen, H. L., Pike, M., et al. (2013). A Bacterial glucanotransferase can replace the complex maltose metabolism required for starch to sucrose conversion in leaves at night. *J. Biol. Chem.* 288, 28581–28598. doi:10.1074/jbc.M113.497867
- Saha, K., Agasti, S. S., Kim, C., Li, X., and Rotello, V. M. (2012). Gold nanoparticles in chemical and biological sensing. *Chem. Rev.* 112, 2739–2779. doi:10.1021/cr2001178
- Santelia, D., and Zeeman, S. C. (2011). Progress in *Arabidopsis* starch research and potential biotechnological applications. *Curr. Opin. Biotechnol.* 22, 271–280. doi:10.1016/j.copbio.2010.11.014
- Smirnova, J., Fernie, A. R., and Steup, M. (2015). "Starch degradation," in *Starch: Metabolism and Structure*, ed. Y. Nakamura (Springer), 239–290. doi:10.1007/978-4-431-55495-0
- Smith, A. M., Zeeman, S. C., and Smith, S. M. (2005). Starch degradation. *Annu. Rev. Plant Biol.* 56, 73–98. doi:10.1146/annurev.arplant.56.032604.144257
- Stanley, D., Rejzek, M., Naested, H., Smedley, M., Otero, S., Fahy, B., et al. (2011). The role of α -glucosidase in germinating barley grains. *Plant Physiol.* 155, 932–943. doi:10.1104/pp.110.168328
- Sun, Z., and Henson, C. A. (1990). Degradation of native starch granules by barley α -glucosidases. *Plant Physiol.* 94, 320–327. doi:10.1104/pp.94.1.320
- Takata, Y., Yamamoto, K., and Kadokawa, J. (2015). Preparation of pH-responsive amphoteric Glycogen hydrogels by α -glucan phosphorylase-catalyzed successive enzymatic reactions. *Macromol. Chem. Phys.* 216, 1415–1420. doi:10.1002/macp.201500151
- Tantanarat, K., Rejzek, M., O'Neill, E., Ruzanski, C., Hill, L., Fairhurst, S. A., et al. (2012). An expedient enzymatic route to isomeric 2-, 3- and 6-monodeoxy-monofluoro-maltose derivatives. *Carbohydr. Res.* 358, 12–18. doi:10.1016/j.carres.2012.05.026
- Tawil, G., Jamme, F., Refregiers, M., Vikso-Nielsen, A., Colonna, P., and Buleon, A. (2011). *In situ* tracking of enzymatic breakdown of starch granules by synchrotron UV fluorescence microscopy. *Anal. Chem.* 83, 989–993. doi:10.1021/ac1027512
- Tetlow, I. J. (2011). Starch biosynthesis in developing seeds. *Seed Sci. Res.* 21, 5–32. doi:10.1017/S0960258510000292
- U.S. Department of Agriculture, A.R.S. (2012). "Energy intakes: Percentages of energy from protein, carbohydrate, fat, and alcohol, by gender and age," in *What We Eat in America* (NHANES), 2009–2010. Available at: http://www.ars.usda.gov/main/site_main.htm?modecode=80-40-05-30
- Umegatani, Y., Izawa, H., Nawaji, M., Yamamoto, K., Kubo, A., Yanase, M., et al. (2012). Enzymatic α -glucuronylation of maltooligosaccharides using α -glucuronic acid 1-phosphate as glycosyl donor catalyzed by a thermostable phosphorylase from *Aquifex aeolicus* VF5. *Carbohydr. Res.* 350, 81–85. doi:10.1016/j.carres.2011.12.022
- Visser, R. G. F., Somhorst, I., Kuipers, G. J., Ruys, N. J., Feenstra, W. J., and Jacobsen, E. (1991). Inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs. *Mol. Gen. Genet.* 225, 289–296. doi:10.1007/bf00269861
- Waigh, T. A., Perry, P., Riekel, C., Gidley, M. J., and Donald, A. M. (1998). Chiral side-chain liquid-crystalline polymeric properties of starch. *Macromolecules* 31, 7980–7984. doi:10.1021/ma971859c
- Withers, S. G. (1990). The enzymic synthesis and N.M.R. characterisation of specifically deoxygenated and fluorinated glycogens. *Carbohydr. Res.* 197, 61–73. doi:10.1016/0008-6215(90)84130-m
- Yanase, M., Takaha, T., and Kuriki, T. (2007). Developing and engineering enzymes for manufacturing amylose. *J. Appl. Glycosci.* 54, 125–131. doi:10.5458/jag.54.125

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 O'Neill and Field. 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.



Compartmentalization and Transport in Synthetic Vesicles

Christine Schmitt^{1†}, Anna H. Lippert^{2†,‡}, Navid Bonakdar², Vahid Sandoghdar² and Lars M. Voll^{1*}

¹ Division of Biochemistry, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ² Max-Planck-Institute for the Science of Light, Erlangen, Germany

OPEN ACCESS

Edited by:

Shota Atsumi,
University of California Davis, USA

Reviewed by:

Sebastien Lecommandoux,
University of Bordeaux, France
Oscar Ces,
Imperial College London, UK

*Correspondence:

Lars M. Voll
lars.voll@fau.de

[†]Christine Schmitt and
Anna H. Lippert contributed equally.

*Present address:

Anna H. Lippert,
Department of Chemistry, University
of Cambridge, Cambridge, UK

Specialty section:

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

Received: 16 November 2015

Accepted: 11 February 2016

Published: 29 February 2016

Citation:

Schmitt C, Lippert AH, Bonakdar N,
Sandoghdar V and Voll LM (2016)
Compartmentalization and Transport
in Synthetic Vesicles.
Front. Bioeng. Biotechnol. 4:19.
doi: 10.3389/fbioe.2016.00019

Nanoscale vesicles have become a popular tool in life sciences. Besides liposomes that are generated from phospholipids of natural origin, polymersomes fabricated of synthetic block copolymers enjoy increasing popularity, as they represent more versatile membrane building blocks that can be selected based on their specific physicochemical properties, such as permeability, stability, or chemical reactivity. In this review, we focus on the application of simple and nested artificial vesicles in synthetic biology. First, we provide an introduction into the utilization of multicompartimented vesosomes as compartmentalized nanoscale bioreactors. In the bottom-up development of protocells from vesicular nanoreactors, the specific exchange of pathway intermediates across compartment boundaries represents a bottleneck for future studies. To date, most compartmented bioreactors rely on unspecific exchange of substrates and products. This is either based on changes in permeability of the coblock polymer shell by physicochemical triggers or by the incorporation of unspecific porin proteins into the vesicle membrane. Since the incorporation of membrane transport proteins into simple and nested artificial vesicles offers the potential for specific exchange of substances between subcompartments, it opens new vistas in the design of protocells. Therefore, we devote the main part of the review to summarize the technical advances in the use of phospholipids and block copolymers for the reconstitution of membrane proteins.

Keywords: liposomes, vesosomes, block copolymers, reconstitution techniques, porins, metabolite transporters, membrane transport, compartmentalized bioreactors

INTRODUCTION

Compartmentalization is a key feature of eukaryotic cells to spatially separate distinct biochemical processes from each other. Lipid bilayer membranes serve as impermeable barriers that effectively separate subcellular compartments. This (i) enables the simultaneous operation of metabolic pathways that utilize the same intermediates and (ii) allows for the adjustment of specific reaction conditions inside individual organelles. In order to translate this natural principle of biological organization, the

Abbreviations: DOPC, 1,2-dioleoyl-*sn*-glycero-3-phosphocholine; DPPC, 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine; PB-*b*-PEO, polybutadiene-*b*-poly(ethylene oxide); PEG-*b*-PSBA, poly(ethylene glycol)-*b*-poly(styrene boronic acid); PEO-PPO-PEO, poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide); PEtOz-PDMS-PEtOz, poly(2-ethyl-2-oxazoline)-*b*-poly(dimethylsiloxane)-*b*-poly(2-ethyl-2-oxazoline); PMOXA-PDMS-PMOXA, poly(2-methyloxazoline)-poly(dimethylsiloxane)-poly(2-methyl-oxazoline); PS-*b*-PDMAEMA, polystyrene-*b*-poly(*N,N*-dimethylaminoethyl methacrylate); PS-*b*-PIAT, polystyrene-poly(L-isocyanoalanine(2-thiophen-3-yl-ethyl)amide), aka polystyrene-*b*-poly(3-(isocyano-lalanyl-amino-ethyl)-thiophene); PVFc-*b*-P2VP, polyvinylferrocene-*b*-poly(2-vinylpyridine).

use of membranes to encapsulate chemical reactions has attracted interest in the design of synthetic systems. Lipids or block copolymers are commonly used to build membranes in synthetic systems, in so-called liposomes or polymersomes, respectively.

Liposomes consist of a shell of amphiphilic lipid species, such as phospholipids, that encapsulate an aqueous solution. The lipids are arranged in a bilayer with the polar head groups of the two leaflets facing toward the inside and the outside aqueous phase and the hydrophobic tails of the phospholipids facing toward each other (**Figure 1B**). Based on the number of membrane layers, vesicles are called unilamellar or multilamellar. For biotechnological applications, the use of unilamellar vesicles is desirable, and these vesicle species can be defined according to their size, ranging from small unilamellar vesicles (SUVs having a diameter between 25 and 100 nm) to large unilamellar vesicles (LUVs with a diameter between 100 nm and 1 μm) and giant unilamellar vesicles (GUVs being larger than 1 μm up to 100 μm in diameter). Polymersomes consisting of amphiphilic block copolymers are of rising interest for compartmentalization in synthetic systems due to their increased mechanical stability and low membrane permeability compared to liposomes. The ability to (i) encapsulate specific cargo, (ii) trigger cargo release by external stimuli, and (iii) the possibility to incorporate particular membrane transport proteins are distinct features that depend on liposome and polymersome composition of artificial vesicles.

THE POTENTIAL OF LIPOSOMES, POLYMERSOMES, AND VESOSOMES – AN OVERVIEW

Both liposomes and polymersomes have become popular as vectors for targeted and tailored drug delivery and for the application in biochemical microreactors. Due to the amphiphilic nature of the lipid or polymer building blocks, a spontaneous assembly into vesicles occurs in aqueous environments (Discher and Eisenberg, 2002). The phase transition temperature represents an important parameter for the choice of specific lipid or polymer building blocks for drug delivery purposes. At the phase transition temperature, lipids and polymers are transformed from a liquid crystalline phase to a gel phase, which leads to maximal bilayer permeability (Van Hoogevest et al., 1984), and hence, to the release of cargo from the lumen of the vesicles. Based on the choice of lipid or block copolymer, the discharge of cargo from artificial vesicles can also be achieved by pH shifts, or redox potential, which we will discuss later in this article. However, the use of artificial vesicles for drug delivery is out of focus of this review, and we would like to refer the interested reader to recent reviews focusing on this topic (Ohya et al., 2011; Lee and Feijen, 2012; Khan et al., 2015; Thambi et al., 2016).

For vesicle formation, a variety of lipids with different properties is available (Marsh, 2012), which can either be used separately or as mixtures. Polymersomes made of amphiphilic block copolymers came into focus because of their potential for functionalization and increased mechanical stability compared to liposomes (Bermudez et al., 2002). Polyethylene glycol (PEG) and polystyrene (PS)-based block copolymers are widely used

to produce polymersomes for all kinds of applications. In addition, polypeptide-based polymersomes have become increasingly popular for biomedical applications, which are not only due to their biodegradability and high tissue compatibility but also based on their ability to change aggregation state and permeability in response to environmental stimuli [as recently reviewed by Zhao et al. (2014)]. Membrane thickness of copolymer-derived polymersomes predominantly depends on the length of the hydrophobic block (Smart et al., 2008). However, not only the chain length of the individual hydrophilic and hydrophobic blocks in diblock and triblock copolymers but also the length ratio of the hydrophilic and hydrophobic segments were found to represent an important parameter for membrane permeability and stiffness (Rodríguez-García et al., 2011). Copolymers that combine a low molecular weight with high hydrophobicity were found to preferably arrange into GUVs (Rodríguez-García et al., 2011). For a more in depth view on the use of polymersomes as vesicle scaffolds in biotechnology, please see recent reviews on the topic (e.g., Lee and Feijen, 2012; Zhao et al., 2014).

Besides simple, single-compartment vesicles, the formation of multicompartmentalized vesicular systems was engineered in the last years to allow the encapsulation of distinct cargos in different vesicular compartments. Various approaches were investigated for this purpose, such as the encapsulation of smaller vesicles into larger vesicles, so-called vesosomes (Walker et al., 1997; Bolinger et al., 2008; Marguet et al., 2012; Paleos et al., 2012). Vesosomes represent nested vesicles that harbor multiple compartments of different sizes encapsulated in each other without a direct connection between the individual compartment boundaries (**Figures 1A,B,D**).

VESICLES AND VESOSOMES AS COMPARTMENTALIZED NANOREACTORS

Compartmentation of enzymatic reactions represents the basic biochemical principle of all living systems. The compartmentation of cellular metabolism has many advantages: the cell can be protected against toxic intermediates formed in one compartment, cellular subcompartments offer optimal reaction conditions for subsets of enzymes, compartmentation can avoid competition for substrates by different metabolic pathways, and it permits the differential regulation of isoenzymes within distinct compartments. Therefore, it seems reasonable to attempt to design multicompartmentalized synthetic systems that are confined by lipid or polymer bilayers, which provide optimal reaction conditions for different enzyme species. Pioneering work in the field was obtained when Meier and coworkers reported an enzymatic reaction inside a PMOXA-PDMS-PMOXA triblock copolymer vesicle in 2000. Passive diffusion of the substrate ampicillin into the polymersome was mediated by the reconstituted bacterial porin OmpF, the encapsulated enzyme β -lactamase subsequently hydrolyzed the substrate before the product ampicillinoic acid diffused out again through OmpF (Nardin et al., 2000, 2001).

To date, vesosomes have successfully been applied to generate compartmentalized biochemical systems, e.g., in the co-factor-dependent enzymatic formation of the fluorescent dye resorufin

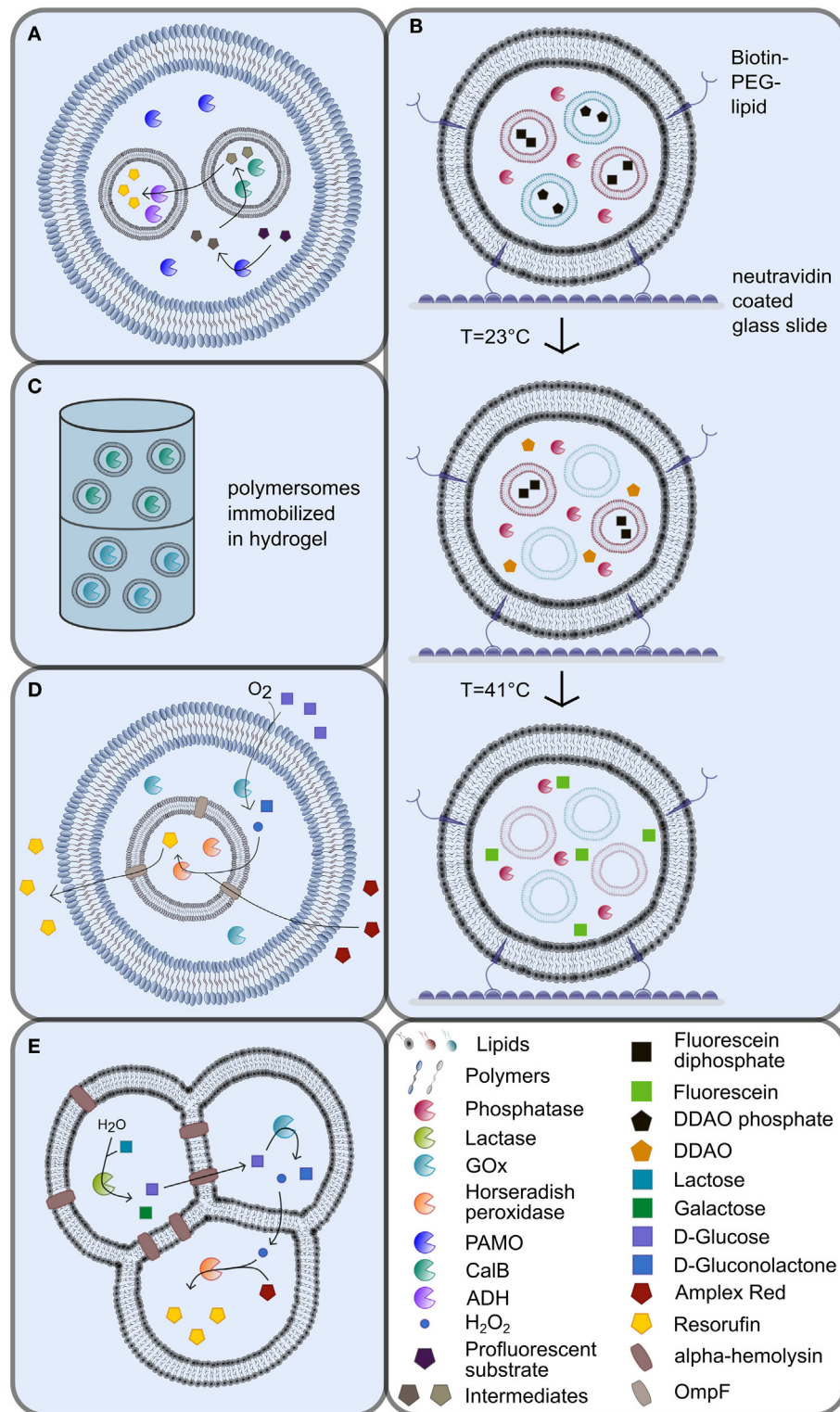


FIGURE 1 | Multicompartmented artificial vesicles as bioreactors. (A) Vesosomes consisting of inner and outer coblock polymersomes with different physicochemical properties [adapted from (Peters et al., 2014)]. **(B)** Immobilization of biotinylated lipid vesosomes via the interaction with neutravidin in nanofluidic reactors [adapted from (Bolinger et al., 2008)]. The release of liposome contents is triggered by specific, consecutive temperature shifts. **(C)** Continuous-flow polymersome reactor with immobilized polymersomes in hydrogel (De Hoog et al., 2010). **(D)** Vesosomes using the porin OmpF as shuttle system [adapted from (Siti et al., 2014)]. **(E)** Multicompartment liposomes generated by the phase transfer technique [adapted from (Elani et al., 2014)]. For explanations, please refer to the manuscript text.

from a profluorescent substrate (Peters et al., 2014). To this end, the two different enzyme species *Candida antarctica* lipase B (CalB) and alcohol dehydrogenase (ADH) were encapsulated separately into intrinsically porous sub-micrometer-sized PS-*b*-PIAT polymersome subcompartments, which, in turn, were engulfed by PB-*b*-PEO GUVs (Figure 1A). In such a vesosome assembly, the lumen of the PB-*b*-PEO vesicle resembles an artificial cytosol, while the PS-*b*-PIAT vesicle lumen corresponds to artificial organelles (Figure 1A). In the initial step, the substrate was converted by NADPH-dependent phenylacetone monooxygenase (PAMO) into an ester in the artificial cytosol of the vesosomes, before the ester intermediate diffused into the CalB containing subcompartments, where it was subsequently hydrolyzed to a primary alcohol. After diffusion of the alcohol product out of the first subcompartment into the ADH containing subcompartments, the alcohol was oxidized in a NAD⁺-dependent reaction into an aldehyde, which then produced the fluorescent dye resorufin by spontaneous beta-elimination (Figure 1A). Likewise, other cascade reactions have also been established in polymersomes using glucose oxidase (GOx), horse radish peroxidase (HRP), and CalB (Vriezema et al., 2007; Kuiper et al., 2008).

Even complex cellular processes, such as the synthesis of ATP, could be achieved by the coupled activity of bacteriorhodopsin and F₀F₁-ATP synthase in synthetic vesicles. An H⁺ gradient was built up by bacteriorhodopsin in a light-dependent manner and this H⁺ gradient was subsequently utilized by ATP synthase to convert ADP and P_i to ATP. These two membrane-associated proteins have successfully been reconstituted into amphiphilic triblock copolymer PEO-*b*-PDMS-*b*-PEO polymersomes leading to ATP synthesis (Choi and Montemagno, 2005). This example nicely demonstrates the potential of synthetic systems to mimic complex cellular functions.

The potential to immobilize vesicular systems can also be exploited for the application as nanoreactors in nanofluidic devices, since the provision of substrate to and the harvest of product from immobilized vesicular compartments is much easier compared to open reaction systems. To this end, vesosomes have been immobilized on neutravidin-coated glass surfaces in nanoreactor systems through the integration of biotin-PEG-lipids into the outer vesosome bilayer [Figure 1B; Bolinger et al. (2008)]. In these systems, the inner SUVs consisted of lipids with different phase transition temperatures compared to the outer SUVs. The inner SUVs were loaded with the profluorescent dyes dichlorodimethylacridinone phosphate or fluorescein diphosphate, while the outer compartment was loaded with alkaline phosphatase (AP). The sequential, temperature-triggered release of the substrates from the encapsulated SUVs drove the conversion of the substrates by AP in the outer compartment in two distinct, consecutive steps (Figure 1B). The produced fluorescent products dichlorodimethylacridinone and fluorescein, respectively, were still trapped inside the outer lipid bilayer.

Alternatively, artificial vesicles have been immobilized in alginate capsules or hydrogels (De Hoog et al., 2010; Ullrich et al., 2015). A “continuous-flow polymersome reactor” was constructed by the immobilization of CalB and GOx loaded polymersomes in a hydrogel (Figure 1C) (De Hoog et al., 2010). The substrate was added on top of the reactor in this setup, while the

product was collected at the bottom (Figure 1C). Since enzyme leakage from immobilized polymersomes was more than four times lower compared to free enzyme, the total enzyme activity required for nanoreactors can be decreased, once the proteins are encapsulated into polymersomes (De Hoog et al., 2010). These two examples illustrate the advantages of vesicular functional units in nanoreactor assemblies.

CONTROLLED RELEASE OF CARGO FROM LIPOSOMES, POLYMERSOMES, AND VESOSOMES BY PHYSICOCHEMICAL TRIGGERS

An important feature in the construction of vesicle-based nanoreactors is the design of the vesicle shell by the choice of lipids, polymers, or a mixture of both. Based on the chosen lipid or block copolymer, a specific release of cargo by external stimuli can be achieved following a physical or chemical trigger that alters membrane permeability. Before we focus on membrane transport proteins for the specific exchange of solutes between vesicle compartments, we would like to briefly summarize the advances in the use of block copolymers that permit a triggered release of solutes.

The release of ABTS²⁻ by repeated thermal stimuli and the subsequent conversion to ABTS¹⁻ by laccase inside the alginate capsules was observed by Ullrich et al. (2015). The heat stimulus was applied either by heating above the phase transition temperature of DPPC in a water bath or by subjecting encapsulated superparamagnetic iron oxide nanoparticles to radiofrequency to cause heat emission.

In addition, stimuli- and cargo-selective content release was achieved in dual stimuli-responsive polymersomes with two kinds of cargo (Staff et al., 2014). On the one hand, the polymer vesicles consisted of the redox- and pH-responsive polymer polyvinylferrocene-*b*-poly(2-vinylpyridine) (PVFc-*b*-P2VP) or the pH- and temperature-responsive polymer polystyrene-*b*-poly(*N,N*-dimethylaminoethyl methacrylate) (PS-*b*-PDMAEMA). On the other hand, the cargos dimethyldodecylamine (DDA) and diphenyl disulfide (DPDS) were selectively switchable from the water-insoluble to the soluble form by pH change or H₂O₂ redox trigger, respectively. The substances were stored in the water-insoluble form inside the polymersomes and were then specifically released by an external stimulus that allowed the passage of DPDS (by oxidation) or DDA (by pH change) across the membrane.

Finally, vesosomes were composed of lipids with different phase transition temperatures were used to trigger the successive mixing of the contents inside the vesosome by a suite of specific temperature changes (Bolinger et al., 2004, 2008). Likewise, the incorporation of stimuli-responsive polymers into polymersomes, such as the sugar and pH-responsive PEG-*b*-PSBA block copolymer or pH-responsive non-ionic amphiphilic triblock copolymers such as PEO-PPO-PEO, can lead to partial permeabilization of these vesicles. Pore-like structures are formed upon applying an appropriate external stimulus, but the vesicles are not disrupted (Binder, 2008; Kim et al., 2009).

THE USE OF TRANSPORT PROTEINS FOR EXCHANGE OF SUBSTRATE BETWEEN VESICULAR SUBCOMPARTMENTS

While the release of cargo by triggered permeabilization of the bilayer is accompanied by at least partial or temporal loss of compartmentalization, a more controlled discharge of cargo from vesicles can be achieved by the integration of membrane proteins into the vesicle membrane. This can either be achieved by the reconstitution of unspecific diffusion pores, such as porins, or by the integration of substrate-specific transporters. This strategy avoids the increase in bilayer permeability by external stimuli but instead enables specific substrate flow across compartment boundaries. A frequently used protein for this purpose is the *Escherichia coli* porin OmpF (outer membrane protein F), which is a trimeric integral membrane protein that enhances the passive diffusion of small hydrophilic molecules (Cowan et al., 1995).

The integration of OmpF in a lipid bilayer encapsulating β -lactamase inside the vesicle lead to hydrolysis of the externally added substrate ampicillin and yielded the product ampicillinoic acid (Graff et al., 2001). The product was first detected inside the vesicle before its accumulation in the medium occurred, which was also facilitated by OmpF. Since the diffusion across the membrane bilayer represented the bottleneck of the reaction, a higher substrate concentration was necessary to achieve comparable activity of encapsulated enzymes compared to free enzymes. Recently, OmpF was also reconstituted into an ABA triblock copolymer bilayer that served as the inner compartment of vesosomes that encapsulated horseradish peroxidase (Figure 1D; Siti et al., 2014). The semi-permeable outer compartment was built of PS-PIAT diblock copolymers and contained GOx as well as the inner ABA polymersomes. After glucose and Amplex Red were added to the outside solution, they diffused into the outer compartment, where glucose was oxidized by GOx to produce H_2O_2 (Figure 1D). Hydrogen peroxide then diffused into the inner compartment, where it subsequently oxidized Amplex Red in the presence of horseradish peroxidase to yield the fluorescent end-product Resorufin (Figure 1D). Most importantly, this study by Siti et al. (2014) showed an increased reaction rate in the presence of OmpF, which enhanced diffusion into the inner reaction compartment.

The constituents of the heptameric protein α -hemolysin can self-assemble in membrane bilayers to form pores, which makes it an attractive target for the use in artificial membranes. Elani et al. (2013, 2014) incorporated α -hemolysin into DOPC bilayers of multicompartment vesicle networks by phase transfer of water-in-oil droplets (Figures 1E and 2E). To proof functionality of the pore protein in a two-compartment system, the Ca^{2+} -sensitive dye Fluo-4 was encapsulated in one compartment and Ca^{2+} in the other compartment. Only those vesicular systems with α -hemolysin in the internal bilayer showed an increase in fluorescence, while no fluorescence was detectable in vesicle systems that did not contain α -hemolysin (Elani et al., 2013). A spatially segregated reaction setup was established in a three-compartment system with α -hemolysin pores connecting the first

compartment with the second and with the third compartment as well as with the surrounding (Figure 1E; Elani et al., 2014). Each reaction step was performed in a single compartment: in the first compartment, lactose was hydrolyzed to glucose and galactose by lactase, and glucose was then oxidized to gluconolactone in the second compartment *via* GOx, thereby producing hydrogen peroxide (Figure 1E). The diffusion of glucose from the first into the second compartment was conferred by α -hemolysin, while lipid bilayers are permeable to hydrogen peroxide to allow for the diffusion of hydrogen peroxide from the second into the third compartment (Figure 1E). Finally, hydrogen peroxide initiated the oxidation of Amplex Red by horseradish peroxidase in the third compartment to yield the fluorescent product resorufin. No increase in fluorescence was detectable in vesicular systems without α -hemolysin (Elani et al., 2014).

SPECIFIC TRANSPORT PROTEINS AS TOOLS IN ARTIFICIAL VESICLE SYSTEMS

The chapters above deal with the unspecific release of cargo and substrates from vesicles *via* physicochemical triggers or by unspecific porins such as OmpF or α -hemolysin. To enable specific transport of cargo and substrates, appropriate transport proteins can be reconstituted into liposome or polymersome membranes. The incorporation of specific transport proteins into artificial membranes of liposomes, polymersomes, or vesosomes allows to conceptualize much more tightly controlled vesicle-based bioreactors. Therefore, we devote the rest of this review on recent advances in the reconstitution of membrane proteins.

RECONSTITUTION OF TRANSMEMBRANE PROTEINS IN LIPID SYSTEMS

The studies on reconstitution of transmembrane proteins into a membrane environment mainly focus on the work with liposomes (Kahya et al., 2001; Montes et al., 2007; Kaneda et al., 2009; Aimon et al., 2011; Dezi et al., 2013; Hansen et al., 2013; Liu et al., 2013) but have also been successfully applied to polymersomes (Meier et al., 2000; Choi and Montemagno, 2005; Nallani et al., 2011; Martino et al., 2012).

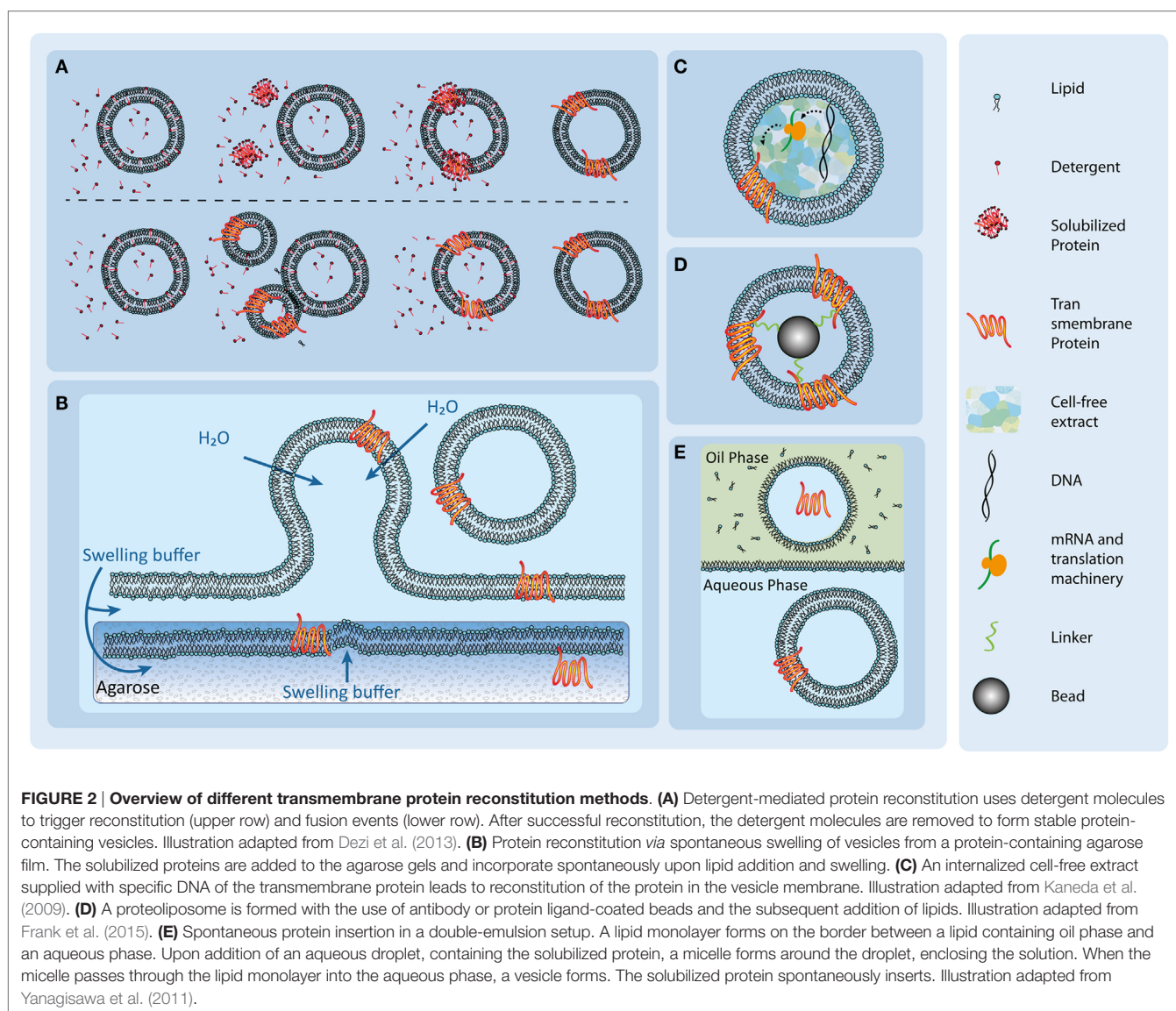
There are multiple ways to reconstitute transmembrane proteins in artificial membrane systems. The basic problem behind transmembrane protein reconstitution is the nature of these proteins. Since transport proteins are anchored in the hydrophobic core of the cell membrane, they have a hydrophobic nature. This hydrophobicity aggravates their extraction as well their insertion from and into membrane systems. The cell itself circumvents this problem employing several strategies (Wickner and Lodish, 1985; Gutensohn et al., 2006). One of these is the so-called Sec pathway found in bacteria and eukaryotes (Economou, 1999; Rapoport et al., 2004). Hydrophobic protein parts are co-translationally recognized by a signal recognition particle, which leads to a translocation of the protein translation machinery to the endoplasmic reticulum (ER) and the co-translational insertion of the protein

into the cell membrane *via* the Sec apparatus. Another strategy involves the post-translational insertion of membrane proteins. After protein translation in the cytoplasm of eukaryotic cells, the protein is delivered, unfolded, inserted, and refolded in the target membrane, i.e., the TIC TOC (Gutensohn et al., 2006; Andr  s et al., 2010) complex in the chloroplast envelope or the TOM TIM (Bauer et al., 2000) complex in mitochondria. In addition, certain transmembrane proteins exhibit spontaneous insertion into the membrane, such as predominantly cytochromes (Wickner and Lodish, 1985). It is debated whether helical hairpin motifs can mediate this spontaneous insertion of proteins into membranes (Engelman and Steitz, 1981). All in all there is a huge variety and complexity of mechanisms the cell uses to insert membrane proteins into the target membrane as well as the heterogeneity of membrane proteins involved. Similarly, there are a couple of methods available for the functional reconstitution of membrane proteins, which need to be tested for individual proteins of interest.

DETERGENT-MEDIATED RECONSTITUTION

Since the solubilization process of membrane proteins, the removal from their natural environment, is usually performed in the presence of detergent, the detergent-mediated reconstitution is one of the most common strategies for protein reconstitution (**Figure 2A**). Here, liposomes are formed *via* extrusion (Torchilin and Weissig, 2003) or sonification (Torchilin and Weissig, 2003), and after a presolubilization step, the solubilized proteins are subsequently added to the liposome preparation, which eventually leads to the incorporation of the proteins into the membranes.

After the pioneering work of Kagawa and Racker (1971), the detergent-mediated reconstitution has been successfully reported on the reconstitution of various transmembrane proteins in vesicles up to sizes larger than 1 μm (Rigaud et al., 1988; Steinberg-Yfrach et al., 1998; Seddon et al., 2004; Dezi et al., 2013). The successfully reconstituted proteins includes



porins, such as FhuA (Dezi et al., 2013), transporters such as bacteriorhodopsin (Steinberg-Yfrach et al., 1998), the Glucose-6-P/P antiporter (Kammerer et al., 1998), H⁺-ATPase (Steinberg-Yfrach et al., 1998), Ca²⁺-ATPase (Steinberg-Yfrach et al., 1998), CFOF1-ATPase (Steinberg-Yfrach et al., 1998), and channels, such as the voltage-gated potassium channel (Ruta et al., 2003), BmrC/BmrD, a bacterial heterodimeric ATP-binding cassette efflux transporter (Dezi et al., 2013), as well as receptor proteins such as GPCRs (Ishihara et al., 2005).

With the advantage of very low external influences on the membrane protein, the detergent-mediated reconstitution is a suitable method for many proteins. The reconstitution of the membrane proteins is either achieved through direct incorporation of solubilized membrane proteins or, in order to form larger proteoliposomes, through detergent-mediated fusion of vesicles. In both cases, the vesicles are often pre-solubilized by detergent prior to the addition of proteins. The detergent-to-lipid ratio needs to be meticulously adjusted to achieve a fusogenic liposome state that ranges between detergent saturation and solubilization of the vesicle. For detergents, the relationship between critical micelle concentration and lipid-to-detergent ratio is given by (Ollivon et al., 2000; Rigaud and Lévy, 2003).

$D_{\text{total}} = D_{\text{water}} + R_{\text{eff}} \cdot L$, where D_{total} represents the total detergent concentration, D_{water} provides the monomeric detergent concentration in water, i.e., the cmc determined in the presence of lipids, while L represents the lipid concentration and R_{eff} the lipid-to-detergent ratio specific for the solubilization state. D_{water} and R_{eff} are constants that need to be determined for the detergent of choice (Ollivon et al., 2000; Rigaud and Lévy, 2003).

With increasing detergent concentration, the amount of incorporated detergent molecules in the vesicles increases, leading to the fragile state of saturation (Ollivon et al., 2000; Rigaud and Lévy, 2003). At this stage, further addition of detergent leads to vesicle shrinkage until full solubilization occurs (Ollivon et al., 2000; Rigaud and Lévy, 2003). The incorporation rates are dependent on the protein and detergent used (Rigaud and Lévy, 2003), and the state of solubilization also represents an important factor for the reconstitution rate and the orientation of the reconstituted protein (Rigaud et al., 1995; Rigaud and Lévy, 2003). It was shown that some proteins incorporate better in a state of low detergent incorporation, while other need complete saturation for an efficient reconstitution (Rigaud et al., 1995). The reconstitution efficiency thereby depends also on the type of detergent and protein used (Rigaud et al., 1995; Dezi et al., 2013). The method has to be tuned for the protein of interest to achieve the best fusion rate; therefore, there are various protocols available to achieve reconstitution but no specific protocol is applicable to all membrane proteins (Ollivon et al., 2000; Rigaud and Lévy, 2003).

After protein reconstitution, the detergent molecules need to be removed for stable vesicles to form (Ollivon et al., 2000; Rigaud and Lévy, 2003). The removal method and its efficiency are thereby dependent on the type of detergent (Rigaud et al., 1995; Rigaud and Lévy, 2003). Detergents with a high cmc, such as CHAPS, chaps, cholate, and octyl glucoside, generally form small micelles, which makes them easy to remove *via* dialysis or gel filtration (Rigaud et al., 1995; Rigaud and Lévy, 2003).

Detergents with a lower cmc, which form larger micelles, are barely removable by gel filtration or dialysis. Here, the removal can be done using detergent-adsorbent beads. These detergents include Triton-X 100 (Rigaud et al., 1995; Rigaud and Lévy, 2003). Since there are no general protocols available for protein solubilization or reconstitution, they are still accounted for as bottle neck processes (Rigaud et al., 1995; Rigaud and Lévy, 2003; Yanagisawa et al., 2011).

The detergent-mediated reconstitution can also be combined with liposome fabrication methods, such as the double-emulsion approach (Pautot et al., 2003; Yanagisawa et al., 2011). To this end, the solubilized potassium channel protein KcsA was added in the vesicles as well in as the external aqueous phase (Yanagisawa et al., 2011). In both cases, protein reconstitution was detected and the protein was functional (Yanagisawa et al., 2011). During this study, it was observed that not every lipid composition favored the insertion of the target protein. In the case of KcsA, there was no reconstitution detectable when PC lipids were used (Yanagisawa et al., 2011). It was also observed that the reconstitution of KcsA was oriented with the directionality depending on the outside or inside configuration of the protein and the size or the intracellular and extracellular domains (Yanagisawa et al., 2011). KcsA and alpha-hemolysin were successfully reconstituted both from the outer aqueous phase as well as from the inner aqueous solution (Takiguchi et al., 2011; Yanagisawa et al., 2011).

For the reconstitution of KcsA, the detergent DDM was used and due to its high dilution, the detergent concentration was far below the cmc. Nevertheless, detergent molecules can remain in the membranes after successful reconstitution. Since remaining detergent might alter lipid composition and consequently protein behavior, this approach might be unsuitable for certain studies.

DIRECT INCORPORATION OF MEMBRANE PROTEINS INTO VESICLES IN CELL-FREE SYSTEMS

Another approach to reconstitute transmembrane proteins is to encapsulate the protein translation machinery in vesicles (Figure 2C). The use of cell-free extract to reconstitute transmembrane proteins in liposomes or polymersomes has been successfully reported by various groups (Kalmbach et al., 2007; Murtas et al., 2007; Goren and Fox, 2008; Liguori et al., 2008a,b; Kaneda et al., 2009; Katzen et al., 2009; Kuruma et al., 2009; Maeda et al., 2012; Martino et al., 2012; Liu et al., 2013).

The addition of cDNA coding for the protein of interest allows specific protein synthesis, and it has been shown that membrane proteins can be functionally synthesized and incorporated in the surrounding compartment membrane. Synthesized membrane proteins include pores (Shimizu et al., 2001), channels (Liguori et al., 2008a,b), transporters (Liguori et al., 2008a,b), and receptors (Junge et al., 2011) from eukaryotic and prokaryotic origin (Zanders, 2005). The reconstituted membrane proteins cover sizes from 15 kDa, as the mechanosensitive heptamer protein channel MscL (Madin et al., 2000), up to the 114-kDa transporter MdtB (Zanders, 2005; Liguori et al., 2008a). Also the synthesis of presecretory and integral membrane proteins requiring

SecA-dependent translocation, for example, proteins with large periplasmic regions, such as FtsQ, or presecretory proteins, such as OmpA or MtlA (Kim et al., 2006), was reported. An overview of the successful use of cell-free systems is given in Zanders (2005).

There are various cell-free extracts available, but the most commonly used extracts are the whole-wheat germ extract, the *E. coli* extract and the PURESYSTEM. The PURESYSTEM was developed by Kuruma et al. (2008) and includes the chaperone-free *E. coli* translation machinery assembled from purified recombinant components. This defined environment might be beneficial for the functional characterization of proteins, with protein yields of around 6 µg/ml at relative high cost (Kuruma et al., 2008). The wheat germ extract shows stable protein expression for weeks (Zanders, 2005) but is considerably more labor intensive in preparation (Berrier et al., 2004; Sawasaki et al., 2004) than the production of the *E. coli* extract. The *E. coli* assay can be prepared in approximately 1 day (Swartz, 2006; Hovijitra et al., 2009) and shows similar efficiencies as the wheat germ extract with protein yields of around 1–6 mg/ml (Ishihara et al., 2005; Kuruma et al., 2008). While the reaction time of the *E. coli* and the wheat germ extract take between 6 and 24 h for the latter, the reaction time of 2 h is considerably shorter for the PURESYSTEM (Kuruma et al., 2008).

All methods have been used to synthesize and reconstitute troublesome proteins exceeding 100 kDa in size as well as membrane proteins (PURESYSTEM: Kim et al., 2006, wheat germ: Schwarz et al., 2010, and *E. coli*: Madin et al., 2000). Post-translational modifications (Kaiser et al., 2008), such as phosphorylation, prenylation, and glycosylation, as well as the formation of disulfide bonds can be achieved by adding the corresponding enzymes (Kalmbach et al., 2007).

The general workflow to achieve protein synthesis in cell-free systems can be summarized as follows: (i) identify the best expression vector compatible with the cell-free expression system (Kuruma et al., 2008). (ii) If immunological detection or fusion with reporter proteins is desired, N-terminal tags showed to generate a higher yield. (iii) After a small-scale optimization step to identify the best expression conditions in expression tests and (iv) a scale-up step, (v) the proteoliposomes can be purified. The estimated time scale from vector design to protein synthesis is around 15 days to 1 month.

An interesting feature of the cell-free systems is the compatibility of the expression systems with some detergents (Madin et al., 2000; Goren and Fox, 2008). A wide range of non-ionic or zwitterionic detergents, Triton X-100, Tween 20, Brij 58p, *n*-dodecyl β-D-maltoside, and CHAPS, were compatible with cell-free synthesis, allowing the expression of proteins in the presence of detergents, while *n*-octyl β-D-glucoside and deoxycholate had an inhibitory effect on protein yield (Madin et al., 2000).

The incorporation of cell-free extract expression systems into vesicles can be achieved using various liposome formation techniques, including natural swelling (Nomura et al., 2003; Kaneda et al., 2009), double-emulsion (Noireaux and Libchaber, 2004; Maeda et al., 2012; Liu et al., 2013), and microfluidic (Martino et al., 2012) approaches.

A drawback of this reconstitution method is the introduction of the complete translation machinery into the lumen of vesicles,

which might introduce unwanted complexity to the synthetic system. Also necessary post-translational modifications which might be necessary for the formation of a fully functional protein may not occur, unless the responsible enzymes, if known, are added to the cell-free extract.

REHYDRATION OF PROTEIN-CONTAINING AGAROSE

For their reconstitution, solubilized membrane proteins can be dissolved in warm, molten agarose gels (Figure 2B) (Hansen et al., 2013, 2015; Gutierrez and Malmstadt, 2014). In this technique, precipitation of the detergents is prevented at a dilution below the cmc. The gel is spread on a coverslip and partially dehydrated. Since agarose retains a high water content (Horger et al., 2009), it is hypothesized that the proteins do not denature (Hansen et al., 2013). Subsequently, lipid droplets are deposited on top of the gel. Under a stream of nitrogen, the solvent of the droplets is evaporated and the gel can be rehydrated using a protein compatible buffer (Hansen et al., 2013). Upon rehydration of the lipids, protein-containing liposomes form from the surface of the protein-containing agarose. The proteins successfully used in this approach so far were aquaporin-Z, bacteriorhodopsin, and SoPIP2 (Hansen et al., 2013), as well as the glucose transporter GLUT1 (Hansen et al., 2015) and the human serotonin receptor 5-HT1A (Gutierrez and Malmstadt, 2014).

This method is relatively easy but requires a comparably large amount of protein. Moreover, it has been reported that liposomes grown on agarose gels contain agarose in the membrane as well as in the interior (Horger et al., 2009). This introduces a change in the mechanical properties (Lira et al., 2014) of the vesicles, which might introduce artifacts in protein diffusion as well as in enzyme kinetics.

USING PROTEOLIPOBEADS FOR THE RECONSTITUTION OF MEMBRANE PROTEINS

Transmembrane protein-coated beads can be applied for the reconstitution of membrane proteins into lipid bilayers (Mirzabekov et al., 2000; Frank et al., 2015). The beads are coated with streptavidin together with a tag or antibody (Figure 2D). The tag allows the purification of the target proteins from detergent-containing cell lysates (Mirzabekov et al., 2000) as well as the coating of the bead with the solubilized transmembrane protein. After addition of detergent-solubilized lipids, the lipids cluster around the protein. The addition of biotinylated lipids, which bind to the streptavidin on the bead, supports the stable and saturated formation of a lipid bilayer around the bead (Mirzabekov et al., 2000). This approach was successfully applied to reconstitute the G-coupled receptor protein CCR5, a seven transmembrane helix protein, into liposomes in the native confirmation and in a uniformly oriented fashion. The detergent is then removed *via* a dialysis step. Here, the non-ionic maltoside detergent cymal was used (Mirzabekov et al., 2000). Throughout the process, the use of

paramagnetic beads facilitates buffer changes and the fabrication process (Mirzabekov et al., 2000). A disadvantage of the method is the fixed position of the transmembrane proteins since they are anchored on the bead surface. This reconstitution method is therefore not universally applicable.

PARTIAL DRYING OF LIPOSOMES

Another method to achieve protein reconstitution into cell-sized scale vesicles of 1 μm diameter is based on protein-containing bilayers (Girard et al., 2004; Aimon et al., 2011; Fenz et al., 2014). These layers are created by partially drying protein-containing small liposomes. Then, various methods can be employed to form liposomes, including electroformation or swelling (Girard et al., 2004; Aimon et al., 2011; Fenz et al., 2014). A drawback of these techniques is the inevitable rupture and partial drying of the protein-containing vesicles. This is accompanied by the risk of protein denaturation as well as the need for vesicle formation in protein compatible buffers, since mostly high-salt conditions are still required. Electroformation offers a wide range of parameters to fine-tune the vesicle formation process, with a number of protocols available for the use of physiological buffers (Pott et al., 2008). Changes in the electric field (amplitude and frequency), the duration of the protocol, and the swelling buffer can be applied to influence vesicle formation. However, the process of vesicle formation is not yet fully understood (Pott et al., 2008) and side effects, such as lipid-peroxidation (Zhou et al., 2007), or impact of the electric field on the proteins are often not assessable.

Gel-assisted swelling techniques offer a comparably limited range of parameters to improve vesicle production. The use of buffers and lipids involved, as well as the substrate, PVA (Weinberger et al., 2013), agarose (Horger et al., 2009), or others, dominate vesicle size and yield. The processes of vesicle formation in physiological buffers are still not fully understood and require further investigation. The method of proteoliposome formation in the size range greater than one micron is therefore still limited by the available protocols.

PEPTIDE-INDUCED FUSION

Membrane protein incorporation into larger vesicles can also be achieved by the fusion of vesicles. Besides detergent-mediated fusion (as already discussed above), other, non-detergent-mediated fusion techniques have been elaborated, including the work of Kahya et al. (2001). Here, the fusogenic peptide WAE has been used to initiate vesicle fusion. Liposomes were formed out of DOPC:chol/PE-PDP (3.5:1.5:0.25), and the WAE peptide was subsequently covalently attached to the vesicles (Pécheur et al., 1997, 1999; Kahya et al., 2001). Larger vesicles, with positively charged lipids, a mixture of DOPC:DOPE:SAINT-2 (10:3:1.3), were used as peptide target. Under these conditions, fusion events were observed and the method was successfully used to reconstitute bacteriorhodopsin (Kahya et al., 2001) as well as a complex of the seven-helix photoreceptor NpSRII and its cognate transducer NpHtrII, with the latter containing two transmembrane α -helices and a large cytoplasmic domain (Kriegsmann et al., 2009).

MECHANICAL AND SPONTANEOUS INSERTION OF MEMBRANE PROTEINS

In other reconstitution techniques, solubilized proteins are added to the already formed liposomes. Defects in the liposome membrane are either induced by sonification pulses (Rigaud and Lévy, 2003), electrical pulses (Rigaud and Lévy, 2003), or freeze-thawing steps (Kammerer et al., 1998; Rigaud and Lévy, 2003). The drawback of these techniques is that mechanical stress is imposed on the solubilized proteins, which can lead to denaturation or low reconstitution rates (Rigaud and Lévy, 2003).

Some classes of proteins, e.g., cytochromes, bacteriorhodopsin and F_0F_1 -ATPases, and porins (Elani et al., 2013, 2014), show spontaneous incorporation into lipid bilayers without the addition of any detergent. Nevertheless, a certain lipid composition of mostly acidic lipids (Eytan and Broza, 1978), as well as vesicles of small size (Eytan and Broza, 1978; Eytan, 1982), is required for spontaneous insertion of these transmembrane proteins. This process has been examined more closely. Jain and Zakim (1987) have revealed that defects in the membrane associated with amphiphilic contaminants as cholesterol, short-chain lipids, and others are facilitating spontaneous insertion.

CONCLUSION

Artificial vesicles are a versatile resource to establish compartmentation in synthetic biochemical nanoreactors. To this end, compartmentation can either be achieved by immobilizing artificial liposome or polymersome vesicles to the matrix of microfluidic reactors or by generating nested or concatenated vesicular systems such as vesosomes.

In order to maximize yield of biochemical processes inside vesicular nanoreactors, it is indispensable to control the exchange of substrates and products across the membranes between the reactor compartments. Technically, it is possible to reconstitute entire metabolic pathways into liposomes or polymersomes, and by the use of vesosomes or immobilized vesicles, it seems even feasible to mimic the natural compartmentation of these pathways in synthetic systems. However, the exchange of substances across membrane boundaries needs to be highly specific to make compartmented reconstituted biochemical pathways work.

Exchange of substances across membranes can either be accomplished by utilizing lipid and coblock polymers that change permeability in response to physicochemical triggers such as heat, redox potential, or pH. However, only unspecific mixing of contents can be achieved by altering the permeability of the membrane boundary. The incorporation of unspecific protein pores into the membranes of vesicle compartmented reactors, e.g., porins or α -hemolysin, allows for a more selective exchange of low molecular weight substances but still permits the passage of a variety of intermediates that are similar in charge and/or structure. Ultimately, the reconstitution of specific membrane transporters that only allow the passage of individual substrates is necessary for the functional reconstitution of compartmented biochemical pathways. As outlined in this review, diverse

approaches can be undertaken to achieve successful reconstitution of membrane transport proteins into artificial membranes. Since the optimal conditions for successful reconstitution are quite specific for each individual transport protein and can hardly be transferred to other candidates, we have taken care to summarize the available techniques for membrane transporter reconstitution. We believe that the integration of metabolite transporters into vesicle-based nanoreactors will largely advance bottom-up approaches in the development of compartmented protocells in the future.

REFERENCES

- Aimon, S., Manzi, J., Schmidt, D., Poveda Larrosa, J. A., Bassereau, P., and Toombes, G. E. (2011). Functional reconstitution of a voltage-gated potassium channel in giant unilamellar vesicles. *PLoS ONE* 6:e25529. doi:10.1371/journal.pone.0025529
- Andr s, C., Agne, B., and Kessler, F. (2010). The TOC complex: preprotein gateway to the chloroplast. *Biochim. Biophys. Acta* 1803, 715–723. doi:10.1016/j.bbamcr.2010.03.004
- Bauer, M. F., Hofmann, S., Neupert, W., and Brunner, M. (2000). Protein translocation into mitochondria: the role of TIM complexes. *Trends Cell Biol.* 10, 25–31. doi:10.1016/S0962-8924(99)01684-0
- Bermudez, H., Brannan, A. K., Hammer, D. A., Bates, F. S., and Discher, D. E. (2002). Molecular weight dependence of polymersome membrane structure, elasticity, and stability. *Macromolecules* 35, 8203–8208. doi:10.1021/ma020669l
- Berrier, C., Park, K. H., Abes, S., Bibonne, A., Betton, J. M., and Ghazi, A. (2004). Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. *Biochemistry* 43, 12585–12591. doi:10.1021/bi049049y
- Binder, W. H. (2008). Polymer-induced transient pores in lipid membranes. *Angew. Chem. Int. Ed. Engl.* 47, 3092–3095. doi:10.1002/anie.200800269
- Bolinger, P.-Y., Stamou, D., and Vogel, H. (2004). Integrated nanoreactor systems: triggering the release and mixing of compounds inside single vesicles. *J. Am. Chem. Soc.* 126, 8594–8595. doi:10.1021/ja049023u
- Bolinger, P.-Y., Stamou, D., and Vogel, H. (2008). An integrated self-assembled nanofluidic system for controlled biological chemistries. *Angew. Chem. Int. Ed. Engl.* 47, 5544–5549. doi:10.1002/anie.200801606
- 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
- Cowan, S. W., Garavito, R. M., Jansonius, J. N., Jenkins, J. A., Karlsson, R., K nig, N., et al. (1995). The structure of OmpF porin in a tetragonal crystal form. *Structure* 3, 1041–1050. doi:10.1016/S0969-2126(01)00240-4
- De Hoog, H. P. M., Arends, I. W. C. E., Rowan, A. E., Cornelissen, J. J. L. M., and Nolte, R. J. M. (2010). A hydrogel-based enzyme-loaded polymersome reactor. *Nanoscale* 2, 709–716. doi:10.1039/b9nr00325h
- Dezi, M., Di Cicco, A., Bassereau, P., and Levy, D. (2013). Detergent-mediated incorporation of transmembrane proteins in giant unilamellar vesicles with controlled physiological contents. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7276–7281. doi:10.1073/pnas.1303857110
- Discher, D. E., and Eisenberg, A. (2002). Polymer vesicles. *Science* 297, 967–973. doi:10.1126/science.1074972
- Economou, A. (1999). Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol.* 7, 315–320. doi:10.1016/S0966-842X(99)01555-3
- Elani, Y., Gee, A., Law, R. V., and Ces, O. (2013). Engineering multi-compartment vesicle networks. *Chem. Sci.* 4, 3332–3338. doi:10.1039/c3sc51164b
- Elani, Y., Law, R. V., and Ces, O. (2014). Vesicle-based artificial cells as chemical microreactors with spatially segregated reaction pathways. *Nat. Commun.* 5, 5305. doi:10.1038/ncomms6305
- Engelman, D. M., and Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* 23, 411–422. doi:10.1016/0092-8674(81)90136-7
- Eytan, G. D. (1982). Use of liposomes for reconstitution of biological functions. *Biochim. Biophys. Acta* 694, 185–202. doi:10.1016/0304-4157(82)90024-7
- Eytan, G. D., and Broza, R. (1978). Role of charge and fluidity in the incorporation of cytochrome oxidase into liposomes. *J. Biol. Chem.* 253, 3196–3202.
- Fenz, S. F., Sachse, R., Schmidt, T., and Kubick, S. (2014). Cell-free synthesis of membrane proteins: tailored cell models out of microsomes. *Biochim. Biophys. Acta* 1838, 1382–1388. doi:10.1016/j.bbamem.2013.12.009
- Frank, P., Siebenhofer, B., Hanzer, T., Geiss, A. F., Schadauer, F., Reiner-Rozman, C., et al. (2015). Proteo-lipobeads for the oriented encapsulation of membrane proteins. *Soft Matter* 11, 2906–2908. doi:10.1039/c4sm02646b
- Girard, P., P cr aux, J., Lenoir, G., Falson, P., Rigaud, J. L., and Bassereau, P. (2004). A new method for the reconstitution of membrane proteins into giant unilamellar vesicles. *Biophys. J.* 87, 419–429. doi:10.1529/biophysj.104.040360
- Goren, M. A., and Fox, B. G. (2008). Wheat germ cell-free translation, purification, and assembly of a functional human stearyl-CoA desaturase complex. *Protein Expr. Purif.* 62, 171–178. doi:10.1016/j.pep.2008.08.002
- Graff, A., Winterhalter, M., and Meier, W. (2001). Nanoreactors from polymer-stabilized liposomes. *Langmuir* 17, 919–923. doi:10.1021/la001306m
- Gutensohn, M., Fan, E., Frielingsdorf, S., Hanner, P., Hou, B., Hust, B., et al. (2006). Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. *J. Plant Physiol.* 163, 333–347. doi:10.1016/j.jplph.2005.11.009
- Gutierrez, M. G., and Malmstadt, N. (2014). Human serotonin receptor 5-HT 1A preferentially segregates to the liquid disordered phase in synthetic lipid bilayers. *J. Am. Chem. Soc.* 136, 13530–13533. doi:10.1021/ja507221m
- Hansen, J. S., Elbing, K., Thompson, J. R., Malmstadt, N., and Lindkvist-Petersson, K. (2015). Glucose transport machinery reconstituted in cell models. *Chem. Commun.* 51, 2316–2319. doi:10.1039/c4cc08838g
- Hansen, J. S., Thompson, J. R., H lix-Nielsen, C., and Malmstadt, N. (2013). Lipid directed intrinsic membrane protein segregation. *J. Am. Chem. Soc.* 135, 17294–17297. doi:10.1021/ja409708e
- Horger, K. S., Estes, D. J., Capone, R., and Mayer, M. (2009). Films of agarose enable rapid formation of giant liposomes in solutions of physiologic ionic strength. *J. Am. Chem. Soc.* 131, 1810–1819. doi:10.1021/ja805625u
- Hovijitra, N. T., Wu, J. J., Peaker, B., and Swartz, J. R. (2009). Cell-free synthesis of functional aquaporin Z in synthetic liposomes. *Biotechnol. Bioeng.* 104, 40–49. doi:10.1002/bit.22385
- Ishihara, G., Goto, M., Saeki, M., Ito, K., Hori, T., Kigawa, T., et al. (2005). Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein Expr. Purif.* 41, 27–37. doi:10.1016/j.pep.2005.01.013
- Jain, M. K., and Zakim, D. (1987). The spontaneous incorporation of proteins into preformed bilayers. *Biochim. Biophys. Acta* 906, 33–68. doi:10.1016/0304-4157(87)90004-9
- Junge, F., Haberkstock, S., Roos, C., Stefer, S., Proverbio, D., D tsch, V., et al. (2011). Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *New Biotechnol.* 28, 262–271. doi:10.1016/j.nbt.2010.07.002
- Kagawa, Y., and Racker, E. (1971). Partial resolution of the enzymes catalyzing oxidative phosphorylation XXV. Reconstitution of vesicles catalyzing ³²Pi-adenosine triphosphate exchange. *J. Biol. Chem.* 246, 5477–5487.
- Kahya, N., P cheur, E. I., de Boeij, W. P., Wiersma, D. A., and Hoekstra, D. (2001). Reconstitution of membrane proteins into giant unilamellar vesicles via peptide-induced fusion. *Biophys. J.* 81, 1464–1474. doi:10.1016/S0006-3495(01)75801-8
- Kaiser, L., Graveland-Bikker, J., Steuerwald, D., Vanberghem, M., Herlihy, K., and Zhang, S. (2008). Efficient cell-free production of olfactory receptors: detergent

AUTHOR CONTRIBUTIONS

CS, AL, and LV wrote the article. NB, VS, and LV conceptualized the article.

FUNDING

The authors would like to acknowledge generous funding of the EFI SynBio program by the emerging fields initiative of the FAU Erlangen-N rnberg.

- optimization, structure, and ligand binding analyses. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15726–15731. doi:10.1073/pnas.0804766105
- Kalmbach, R., Kalmbach, R., Chizhov, I., Schumacher, M. C., Friedrich, T., Bamberg, E., et al. (2007). Functional cell-free synthesis of a seven helix membrane protein: *in situ* insertion of bacteriorhodopsin into liposomes. *J. Mol. Biol.* 371, 639–648. doi:10.1016/j.jmb.2007.05.087
- Kammerer, B., Fischer, K., Hilpert, B., Schubert, S., Gutensohn, M., Weber, A., et al. (1998). Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. *Plant Cell* 10, 105–117. doi:10.2307/3870632
- Kaneda, M., Nomura, S. M., Ichinose, S., Kondo, S., Nakahama, K., Akiyoshi, K., et al. (2009). Direct formation of proteo-liposomes by *in vitro* synthesis and cellular cytosolic delivery with connexin-expressing liposomes. *Biomaterials* 30, 3971–3977. doi:10.1016/j.biomaterials.2009.04.006
- Katzen, F., Peterson, T. C., and Kudlicki, W. (2009). Membrane protein expression: no cells required. *Trends Biotechnol.* 27, 455–460. doi:10.1016/j.tibtech.2009.05.005
- Khan, I., Khan, M., Umar, M. N., and Oh, D. H. (2015). Nanobiotechnology and its applications in drug delivery system: a review. *IET Nanobiotechnol.* 9, 396–400. doi:10.1049/iet-nbt.2014.0062
- Kim, K. T., Cornelissen, J. J. L. M., Nolte, R. J. M., and van Hest, J. C. M. (2009). A polymersome nanoreactor with controllable permeability induced by stimuli-responsive block copolymers. *Adv. Mater.* 21, 2787–2791. doi:10.1002/adma.200900300
- Kim, T.-W., Keum, J. W., Oh, I. S., Choi, C. Y., Park, C. G., and Kim, D. M. (2006). Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.* 126, 554–561. doi:10.1016/j.jbiotec.2006.05.014
- Kriegsmann, J., Gregor, I., von der Hocht, I., Klare, J., Engelhard, M., Enderlein, J., et al. (2009). Translational diffusion and interaction of a photoreceptor and its cognate transducer observed in giant unilamellar vesicles by using dual-focus FCS. *ChemBiochem* 10, 1823–1829. doi:10.1002/cbic.200900251
- Kuiper, S. M., Nallani, M., Vriezema, D. M., Cornelissen, J. J., van Hest, J. C. M., Nolte, R. J., et al. (2008). Enzymes containing porous polymersomes as nano reaction vessels for cascade reactions. *Org. Biomol. Chem.* 6, 4315–4318. doi:10.1039/b811196k
- Kuruma, Y., Nishiyama, K., Shimizu, Y., Müller, M., and Ueda, T. (2008). Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins. *Biotechnol. Prog.* 21, 1243–1251. doi:10.1021/bp049553u
- Kuruma, Y., Stano, P., Ueda, T., and Luisi, P. L. (2009). A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta* 1788, 567–574. doi:10.1016/j.bbame.2008.10.017
- Lee, J. S., and Feijen, J. (2012). Polymersomes for drug delivery: design, formation and characterization. *J. Control. Release* 161, 473–483. doi:10.1016/j.jconrel.2011.10.005
- Liguori, L., Marques, B., and Lenormand, J.-L. (2008a). A bacterial cell-free expression system to produce membrane proteins and proteoliposomes: from cDNA to functional assay. *Curr. Protoc. Protein Sci.* 5, 22. doi:10.1002/0471140864.ps0522s54
- Liguori, L., Marques, B., Villegas-Mendez, A., Rothe, R., and Lenormand, J.-L. (2008b). Liposomes-mediated delivery of pro-apoptotic therapeutic membrane proteins. *J. Control. Release* 126, 217–227. doi:10.1016/j.jconrel.2007.12.004
- Lira, R. B., Dimova, R., and Riske, K. A. (2014). Giant unilamellar vesicles formed by hybrid films of agarose and lipids display altered mechanical properties. *Biophys. J.* 107, 1609–1619. doi:10.1016/j.bpj.2014.08.009
- Liu, Y.-J., Hansen, G. P. R., Venancio-Marques, A., and Baigl, D. (2013). Cell-free preparation of functional and triggerable giant proteoliposomes. *ChemBiochem* 14, 2243–2247. doi:10.1002/cbic.201300501
- Madin, K., Sawasaki, T., Ogasawara, T., and Endo, Y. (2000). A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 559–564. doi:10.1073/pnas.97.2.559
- Maeda, Y. T., Nakada, T., Shin, J., Uryu, K., Noireaux, V., and Libchaber, A. (2012). Assembly of MreB filaments on liposome membranes: a synthetic biology approach. *ACS Synth. Biol.* 1, 53–59. doi:10.1021/sb200003v
- Marguet, M., Edembe, L., and Lecommandoux, S. (2012). Polymersomes in polymersomes: multiple loading and permeability control. *Angew. Chem. Int. Ed. Engl.* 124, 1199–1202. doi:10.1002/ange.201106410
- Marsh, D. (2012). Thermodynamics of phospholipid self-assembly. *Biophys. J.* 102, 1079–1087. doi:10.1016/j.bpj.2012.01.049
- Martino, C., Kim, S. H., Horsfall, L., Abbaspourrad, A., Rosser, S. J., Cooper, J., et al. (2012). Protein expression, aggregation, and triggered release from polymersomes as artificial cell-like structures. *Angew. Chem. Int. Ed. Engl.* 51, 6416–6420. doi:10.1002/anie.201201443
- Meier, W., Nardin, C., and Winterhalter, M. (2000). Reconstitution of channel proteins in (polymerized) ABA triblock copolymer membranes. *Angew. Chem. Int. Ed. Engl.* 39, 4599–4602. doi:10.1002/1521-3773(20001215)39:24<4599::AID-ANIE4599>3.0.CO;2-Y
- Mirzabekov, T., Kontos, H., Farzan, M., Marasco, W., and Sodroski, J. (2000). Paramagnetic proteoliposomes containing a pure, native, and oriented seven-transmembrane segment protein, CCR5. *Nat. Biotechnol.* 18, 649–654. doi:10.1038/76501
- Montes, L.-R., Alonso, A., Goñi, F. M., and Bagatolli, L. A. (2007). Giant unilamellar vesicles electroformed from native membranes and organic lipid mixtures under physiological conditions. *Biophys. J.* 93, 3548–3554. doi:10.1529/biophysj.107.116228
- Murtas, G., Kuruma, Y., Bianchini, P., Diaspro, A., and Luisi, P. L. (2007). Protein synthesis in liposomes with a minimal set of enzymes. *Biochem. Biophys. Res. Commun.* 363, 12–17. doi:10.1016/j.bbrc.2007.07.201
- Nallani, M., Andreasson-Ochsner, M., Tan, C. W., Sinner, E. K., Wisantoso, Y., Geifman-Shochat, S., et al. (2011). Proteopolymersomes: *in vitro* production of a membrane protein in polymersome membranes. *Biointerphases* 6, 153–157. doi:10.1116/1.3644384
- Nardin, C., Thoeni, S., Widmer, J., Winterhalter, M., and Meier, W. (2000). Nanoreactors based on (polymerized) ABA-triblock copolymer vesicles. *Chem. Commun.* 2000, 1433–1434. doi:10.1039/b004280n
- Nardin, C., Widmer, J., Winterhalter, M., and Meier, W. (2001). Amphiphilic block copolymer nanocontainers as bioreactors. *Eur. Phys. J. E* 4, 403–410. doi:10.1007/s101890170095
- Noireaux, V., and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17669–17674. doi:10.1073/pnas.0408236101
- Nomura, S. M., Nomura, S. M., Tsumoto, K., Hamada, T., Akiyoshi, K., Nakatani, Y., et al. (2003). Gene expression within cell-sized lipid vesicles. *ChemBiochem* 4, 1172–1175. doi:10.1002/cbic.200300630
- Ohya, Y., Takahashi, A., and Nagahama, K. (2011). Biodegradable polymeric assemblies for biomedical materials. *Adv. Polym. Sci.* 247, 65–114. doi:10.1007/12_2011_160
- Ollivon, M., Lesieur, S., Grabielle-Madellmont, C., and Paternostre, M. (2000). Vesicle reconstitution from lipid-detergent mixed micelles. *Biochim. Biophys. Acta* 1508, 34–50. doi:10.1016/S0304-4157(00)00006-X
- Paleos, C. M., Tsiourvas, D., and Sideratou, Z. (2012). Preparation of multicompartment lipid-based systems based on vesicle interactions. *Langmuir* 28, 2337–2346. doi:10.1021/la2027187
- Pautot, S., Frisken, B. J., and Weitz, D. A. (2003). Production of unilamellar vesicles using an inverted emulsion. *Langmuir* 19, 2870–2879. doi:10.1021/la026100v
- Pécheur, E. I., Hoekstra, D., Sainte-Marie, J., Maurin, L., Bienvenüe, A., and Philippot, J. R. (1997). Membrane anchorage brings about fusogenic properties in a short synthetic peptide. *Biochemistry* 36, 3773–3781. doi:10.1021/bi9622128
- Pécheur, E. I., Sainte-Marie, J., Bienvenüe, A., and Hoekstra, D. (1999). Peptides and membrane fusion: towards an understanding of the molecular mechanism of protein-induced fusion. *J. Membr. Biol.* 167, 1–17. doi:10.1007/s002329900466
- Peters, R. J., Marguet, M., Marais, S., Fraaije, M. W., van Hest, J. C. M., and Lecommandoux, S. (2014). Cascade reactions in multicompartmentalized polymersomes. *Angew. Chem. Int. Ed. Engl.* 126, 150–154. doi:10.1002/anie.201308141
- Pott, T., Bouvrais, H., and Méléard, P. (2008). Giant unilamellar vesicle formation under physiologically relevant conditions. *Chem. Phys. Lipids* 154, 115–119. doi:10.1016/j.chemphyslip.2008.03.008
- Rapoport, T. A., Goder, V., Heinrich, S. U., and Matlack, K. E. S. (2004). Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* 14, 568–575. doi:10.1016/j.tcb.2004.09.002
- Rigaud, J.-L., and Lévy, D. (2003). Reconstitution of membrane proteins into liposomes. *Meth. Enzymol.* 372, 65–86. doi:10.1016/S0076-6879(03)72004-7
- Rigaud, J.-L., Paternostre, M. T., and Bluzat, A. (1988). Mechanisms of membrane protein insertion into liposomes during reconstitution procedures involving the

- use of detergents. 2. Incorporation of the light-driven proton pump bacteriorhodopsin. *Biochemistry* 27, 2677–2688. doi:10.1021/bi00408a007
- Rigaud, J.-L., Pitard, B., and Levy, D. (1995). Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim. Biophys. Acta* 1231, 223–246. doi:10.1016/0005-2728(95)00091-V
- Rodríguez-García, R., Mell, M., López-Montero, I., Netzel, J., Hellweg, T., and Monroy, F. (2011). Polymersomes: smart vesicles of tunable rigidity and permeability. *Soft Matter* 7, 1532–1542. doi:10.1039/c0sm00823k
- Ruta, V., Jiang, Y., Lee, A., Chen, J., and MacKinnon, R. (2003). Functional analysis of an archaeobacterial voltage-dependent K⁺ channel. *Nature* 422, 180–185. doi:10.1038/nature01473
- Sawasaki, T., Hasegawa, Y., Morishita, R., Seki, M., Shinozaki, K., and Endo, Y. (2004). Genome-scale, biochemical annotation method based on the wheat germ cell-free protein synthesis system. *Phytochemistry* 65, 1549–1555. doi:10.1016/j.phytochem.2004.04.023
- Schwarz, D., Daley, D., Beckhaus, T., Dötsch, V., and Bernhard, F. (2010). Cell-free expression profiling of *E. coli* inner membrane proteins. *Proteomics* 10, 1762–1779. doi:10.1002/pmic.200900485
- Seddon, A. M., Curnow, P., and Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* 1666, 105–117. doi:10.1016/j.bbmem.2004.04.011
- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802
- Siti, W., de Hoog, H.-P., Fischer, O., Shan, W. S., Tomczak, N., Nallani, M., et al. (2014). An intercompartmental enzymatic cascade reaction in channel-equipped polymersome-in-polymersome architectures. *J. Mater. Chem. B* 2, 2733–2737. doi:10.1039/c3tb21849j
- Smart, T., Lomas, H., Massignani, M., Flores-Merino, M. V., Perez, L. R., and Battaglia, G. (2008). Block copolymer nanostructures. *Nano Today* 3, 38–46. doi:10.1016/S1748-0132(08)70043-4
- Staff, R. H., Gallei, M., Landfester, K., and Crespy, D. (2014). Hydrophobic nanocontainers for stimulus-selective release in aqueous environments. *Macromolecules* 47, 4876–4883. doi:10.1021/ma501233y
- Steinberg-Yfrach, G., Rigaud, J. L., Durantini, E. N., Moore, A. L., Gust, D., and Moore, T. A. (1998). Light-driven production of ATP catalysed by F₀F₁-ATP synthase in an artificial photosynthetic membrane. *Nature* 392, 479. doi:10.1038/33116
- Swartz, J. (2006). Developing cell-free biology for industrial applications. *J. Ind. Microbiol. Biotechnol.* 33, 476–485. doi:10.1007/s10295-006-0127-y
- Takiguchi, K., Negishi, M., Tanaka-Takiguchi, Y., Homma, M., and Yoshikawa, K. (2011). Transformation of actoHMM assembly confined in cell-sized liposome. *Langmuir* 27, 11528–11535. doi:10.1021/la2016287
- Thambi, T., Park, J. H., and Lee, D. S. (2016). Stimuli-responsive polymersomes for cancer therapy. *Biomater. Sci.* 4, 55–69. doi:10.1039/C5BM00268K
- Torchilin, V., and Weissig, V. (2003). *Liposomes: A Practical Approach*. London, UK: Oxford University Press.
- Ullrich, M., Hanuš, J., and Štěpánek, F. (2015). Remote control of enzymatic reaction in compartmentalized microparticles: a system for the delivery of unstable actives. *Chem. Eng. Sci.* 125, 191–199. doi:10.1016/j.ces.2014.06.020
- Van Hoogevest, P., de Gier, J., and de Kruijff, B. (1984). Determination of the size of the packing defects in dimyristoylphosphatidylcholine bilayers, present at the phase transition temperature. *FEBS Lett.* 171, 160–164. doi:10.1016/0014-5793(84)80479-2
- Vriezema, D. M., Garcia, P. M. L., Sancho Oltra, N., Hatzakis, N. S., Kuiper, S. M., Nolte, R. J. M., et al. (2007). Positional assembly of enzymes in polymersome nanoreactors for cascade reactions. *Angew. Chem. Int. Ed. Engl.* 119, 7522–7526. doi:10.1002/ange.200701125
- Walker, S. A., Kennedy, M. T., and Zasadzinski, J. A. (1997). Encapsulation of bilayer vesicles by self-assembly. *Nature* 387, 61–64. doi:10.1038/387061a0
- Weinberger, A., Tsai, F. C., Koenderink, G. H., Schmidt, T. F., Itri, R., Meier, W., et al. (2013). Gel-assisted formation of giant unilamellar vesicles. *Biophys. J.* 105, 154–164. doi:10.1016/j.bpj.2013.05.024
- Wickner, W. T., and Lodish, H. F. (1985). Multiple mechanisms of protein insertion into and across membranes. *Science* 230, 400–407. doi:10.1126/science.4048938
- Yanagisawa, M., Iwamoto, M., Kato, A., Yoshikawa, K., and Oiki, S. (2011). Oriented reconstitution of a membrane protein in a giant unilamellar vesicle: experimental verification with the potassium channel KcsA. *J. Am. Chem. Soc.* 133, 11774–11779. doi:10.1021/ja2040859
- Zanders, E. D. (2005). *Chemical Genomics: Reviews and Protocols*. New York, USA: Humana Press.
- Zhao, L., Li, N., Wang, K., Shi, C., Zhang, L., and Luan, Y. (2014). A review of polypeptide-based polymersomes. *Biomaterials* 35, 1284–1301. doi:10.1016/j.biomaterials.2013.10.063
- Zhou, Y., Berry, C. K., Storer, P. A., and Raphael, R. M. (2007). Peroxidation of polyunsaturated phosphatidyl-choline lipids during electroformation. *Biomaterials* 28, 1298–1306. doi:10.1016/j.biomaterials.2006.10.016

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 © 2016 Schmitt, Lippert, Bonakdar, Sandoghdar and Voll. 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.



Metabolomics, standards, and metabolic modeling for synthetic biology in plants

Camilla Beate Hill^{1*}, Tobias Czauderna², Matthias Klapperstück², Ute Roessner¹ and Falk Schreiber^{2,3}

¹ School of BioSciences, University of Melbourne, Parkville, VIC, Australia, ² Faculty of Information Technology, Monash University, Clayton, VIC, Australia, ³ Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany

OPEN ACCESS

Edited by:

Lars Matthias Voll,
Friedrich-Alexander-University
Erlangen-Nuremberg, Germany

Reviewed by:

Weiwen Zhang,
Tianjin University, China
Holger Hesse,
Freie Universität Berlin, Germany
Esteban Marcellin,
The University of Queensland,
Australia

*Correspondence:

Camilla Beate Hill
camilla.hill@unimelb.edu.au

Specialty section:

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

Received: 30 July 2015

Accepted: 05 October 2015

Published: 21 October 2015

Citation:

Hill CB, Czauderna T,
Klapperstück M, Roessner U and
Schreiber F (2015) Metabolomics,
standards, and metabolic modeling
for synthetic biology in plants.
Front. Bioeng. Biotechnol. 3:167.
doi: 10.3389/fbioe.2015.00167

Life on earth depends on dynamic chemical transformations that enable cellular functions, including electron transfer reactions, as well as synthesis and degradation of biomolecules. Biochemical reactions are coordinated in metabolic pathways that interact in a complex way to allow adequate regulation. Biotechnology, food, biofuel, agricultural, and pharmaceutical industries are highly interested in metabolic engineering as an enabling technology of synthetic biology to exploit cells for the controlled production of metabolites of interest. These approaches have only recently been extended to plants due to their greater metabolic complexity (such as primary and secondary metabolism) and highly compartmentalized cellular structures and functions (including plant-specific organelles) compared with bacteria and other microorganisms. Technological advances in analytical instrumentation in combination with advances in data analysis and modeling have opened up new approaches to engineer plant metabolic pathways and allow the impact of modifications to be predicted more accurately. In this article, we review challenges in the integration and analysis of large-scale metabolic data, present an overview of current bioinformatics methods for the modeling and visualization of metabolic networks, and discuss approaches for interfacing bioinformatics approaches with metabolic models of cellular processes and flux distributions in order to predict phenotypes derived from specific genetic modifications or subjected to different environmental conditions.

Keywords: metabolic engineering, synthetic biology, metabolic modeling, systems biology, metabolomics

INTRODUCTION

One of the greatest challenges for scientists is to understand the genetics, physiology, and biochemistry of plants and the interaction of genes with the environment in order to provide strategies to manipulate these processes to improve plant growth and performance and to prevent diseases. There have been immense advances in the past decades allowing scientists to sequence genomes of organisms addressing many questions, such as how the genome determines a plant's response to any environmental stimuli. Simultaneously, the development of analytical technologies allows us to take a comprehensive and unbiased glance at the gene products, such as gene transcripts (mRNA), proteins, lipids, and metabolites. The post-genomics era was born with the establishment of transcriptomics, proteomics, lipidomics, and metabolomics, each

with their associated computational advances providing the path for data analysis, visualization, and integration to establish the relationships between the genome and gene products under certain conditions.

Metabolites are synthesized by enzyme-catalyzed reactions in any living cell. They are important for the maintenance and survival of cells, most importantly for energy storage and provision, and they also contribute to building and maintaining the cell's structural components. Metabolites and their functionalities are indispensable in the interaction of a cell with the environment, and it has been argued that the metabolome of any biological system represents the final “read-out” of the expression of many genes in that system in a particular situation, and reflecting gene \times environment relationships (Hill et al., 2014).

In comparison to transcripts or proteins, metabolites have a vast range of different chemical structures with an astonishing array of different functional groups that lead to differences in their physical and chemical properties, such as solubility, reactivity, stability, and polarity (Trethewey, 2004). This sheer diversity presents challenges to assay these compounds in a multiparallel fashion. Firstly, a number of different solvent extraction procedures need to be utilized to extract metabolites efficiently from any given plant tissue. In addition, no single analytical approach is capable of detecting and quantifying such chemical diversity; therefore, a range of different approaches (more detail below) has to be employed to analyze as many metabolites as possible. Today, metabolomics is considered as the science combining modern and sophisticated analytical instrumentation for metabolite detection and quantification with appropriate computational and statistical approaches to extract, mine, and interpret metabolomics data.

Metabolomics is now also becoming an important tool for biotechnological and metabolic engineering approaches, which aim to manipulate biochemical pathways to enhance the accumulation of compounds of interest (Dromms and Styczynski, 2012). Since metabolomics can provide a more complete picture of the biological system studied, it has been argued that it can be applied to identify metabolite markers that indicate a particular phenotype [e.g., level of target compound(s)] to allow the assessment of the successes of engineering steps to provide further guidance for future engineering strategies (Beckles and Roessner, 2011). Historically, metabolic engineers have used the analysis of the levels of the target compound(s) and potentially a few closely related metabolites to define metabolic engineering strategies. However, the potential metabolomics offers, which measures hundreds of metabolites rather than just a few, has only been rarely explored in metabolic engineering approaches, particularly in plants (Rios-Esteva and Lange, 2007; Fernie and Morgan, 2013). The unbiased and broad approach of metabolomics helps to assess how plants maintain energy, carbon, and nutrient resources and provides information how these resources may potentially be redirected into the synthesis of the desired metabolites, therefore allowing smart engineering strategies to be developed.

In the following, we review current metabolomics technologies, information resources for metabolomics, as well as computational analysis and modeling approaches with a focus on plant related research.

TOOLS AND TECHNOLOGIES TO STUDY PLANT METABOLISM

Analytical Technologies

The plant metabolome, compared to the metabolome of other organisms, is represented by a particularly vast variety of chemical structures with an enormous diversity of chemical and physical properties (Villas-Boas et al., 2007). In the past decade, researchers have developed and validated a number of complementary analytical approaches to extract, separate, detect, and quantify this diversity. Different solvent extraction procedures may need to be employed to cover the range of polarity of metabolites (Dias et al., 2012). However, most routine metabolite extractions are based on a methanol/water/chloroform biphasic extraction, which captures a large complement of the plant metabolome.

Once metabolites are extracted, the complex mixtures need to be separated allowing individual detection and quantification of compounds. The polarity of metabolites also influences the choice of the separation approach. In liquid chromatography (LC)-based separations, researchers now commonly use two types of separation chemistries, such as C18 reverse phase, which separates the more hydrophobic complement of a metabolite extract, such as many secondary metabolites and lipids, and hydrophilic interaction chromatography, which is better suited for the polar metabolites (Callahan et al., 2009; Hill and Roessner, 2015). In addition, many metabolites are either positively or negatively charged molecules, therefore molecules need to be ionized using both positive and negative ionization mode (Beckles and Roessner, 2011; Hill et al., 2014). There are a number of different ionization techniques available, such as electrospray ionization or atmospheric pressure ionization, again each better suited for a particular subclass of metabolites. However, the most common technique used in LC-MS-based metabolomics is electrospray ionization, which allows reliable ionization of thousands of compounds (Hill et al., 2014).

An alternative separation technique is gas chromatography (GC), which is known for its superior separation power and reproducibility (Dias et al., 2015). The Metabolomics Standards Initiative (MSI; Fiehn et al., 2007) has developed minimal reporting standards for metabolomics data, and strategies to further enhance reproducibility, experimental and data standardization are continuously developed (Allwood et al., 2009). However, GC requires compounds to be volatile to be amenable for analysis. Most metabolites are not volatile and therefore require chemical derivatization to make them volatile. This limits the utility of GC-based separation in metabolomics applications; however, GC coupled to MS still remains the “work horse” in metabolomics due to its reproducibility and its ease of use (Hill and Roessner, 2013).

Mass spectrometry (MS) is the most commonly used detector in metabolomics approaches. The power of MS is that it can distinguish the size of the ionized molecule (by determining the mass to charge ratio of each detected ion) and allows the determination of the number of individual ions detected. The mass to charge ratio determined by MS is used for compound identification and the number of ions detected can be related back to initial concentration of the molecule in the metabolite extract. Another

advantage of MS is that it can fragment an ion once (MS/MS) or multiple times (MS^n), a feature used for structural elucidation of unknown compounds. MS also has excellent capabilities to detect and distinguish isotopic patterns of each ion under analysis. This is particularly important for stable isotope labeling experiments used for metabolic flux analysis. MS analysis is able to determine enrichments in stable isotope labels in all the analyzed compounds and therefore allows determining metabolic fluxes through particular pathways. Most commonly ^{13}C labeled metabolic precursors are used, and the distribution and enrichment of the label across pathways of interest are quantified (O'Grady et al., 2012; Chokkathukalam et al., 2014). ^{13}C -labeling patterns can also be detected by nuclear magnetic resonance spectroscopy (NMR), which also represents an important platform for untargeted, non-destructive metabolite profiling (Eisenreich and Bacher, 2007).

Metabolomics Data Analysis and Visualization

Metabolic data give us a snapshot of the current state of an organism (Hill and Roessner, 2013). It represents the outcome of a preceding gene expression profile, which influences the activity of pathways, transport processes, as well as production and consumption of metabolites. The resulting metabolic profile can be used to classify organisms, for example, by genotype or treatment. Furthermore, these profiles enable comparative analysis between selected treatments or genotypes (Hill et al., 2013a,b, 2015) and to obtain information about metabolites with most or least changes as a result of changes in the gene expression profile.

There are two general methods to analyze metabolic data, which can also be combined. The first, analytical method uses commonly known statistics and clustering algorithms, the second method implies the use of networks to visualize spatial and temporal properties of the data. Performing data statistics in the scope of networks for visualization purposes represents the combination of both methods.

Scientists can choose from a broad repository of statistical methods with respect to the objective at hand. Methods such as frequency distribution, analysis of variance (ANOVA), min–max, and Pearson's and Spearman-rank correlation are examples for univariate data analysis. Principal component analysis (PCA), partial least squares regression (PLS), and multivariate analysis of variance (MANOVA) are commonly used for multivariate data analysis. Self-organizing maps (SOM), support vector machines (SVM), and k-means are popular methods for cluster analysis. Generated results can be visualized using different kinds of diagrams such as plots, histograms, cluster diagrams, and heat maps.

To perform the above-mentioned analytical methods, several software tools can be used. There is a distinction between low-level tools to provide the actual set of algorithms, which are in turn used by high-level tools to provide user-friendly application of those algorithms. **Table 1** shows a selection of frequently used software tools.

Graphical representations of metabolic pathways and networks have been used for long time to represent knowledge about metabolic processes, and with the availability of pathways in databases,

TABLE 1 | Software tools for metabolomics data analysis.

Name	Reference	URL
HIGH LEVEL		
MetaboAnalyst	Xia et al. (2015)	http://www.metaboanalyst.ca/
XCMS	Gowda et al. (2014)	https://xcmsonline.scripps.edu/
MetATT (Metabolomics tool for analyzing two-factor and time-series data)	Xia et al. (2011)	http://metatt.metabolomics.ca/MetATT/
metaP-Server	Kastenmüller et al. (2011)	http://metap.helmholtz-muenchen.de/metap2/
LOW LEVEL		
Matlab (PLS toolbox, msalign, etc.)	MathWorks (2012)	http://ch.mathworks.com/products/matlab/
R-packages (AmsRPM, apLCMS, metabolomics, muma, etc.)	R Development Core Team (2008), Kirchner et al. (2007), Yu et al. (2009), Gaude et al. (2013)	http://www.r-project.org/

TABLE 2 | Network visualization software tools that support metabolic data.

Name	Reference	URL
MetScape Plugin for Cytoscape	Karnovsky et al. (2012)	http://metscape.ncibi.org/
MetaMapp (MS data in context of metabolic networks using Cytoscape)	Barupal et al. (2012)	http://metamapp.fiehnlab.ucdavis.edu
MAVEN (LC-MS)	Clasquin et al. (2012)	http://maven.princeton.edu
VANTED	Rohn et al. (2012)	http://www.vanted.org
Pathomx	Fitzpatrick et al. (2014)	http://pathomx.org/

several tools have been developed to visualize metabolic data in the context of networks. These tools mostly support tasks such as data mapping and network analysis but often also try to help with layout and exploration of data. The latter touches the field of visual analytics for metabolic information (Kerren and Schreiber, 2012, 2014). **Table 2** shows a selection of tools supporting visualization of metabolic data as diagrams or heat maps in the context of biological networks.

As a result of applying common procedures of extracting and measuring metabolic abundance and concentration, usually metabolomic data possesses no spatial information. With the advent of modern methods such as imaging MS (Kaspar et al., 2011; Miura et al., 2012), additional spatial information can be gathered and displayed using 2D (Rohn et al., 2011) or 3D immersive techniques (Sommer et al., 2011).

Figure 1 shows an exemplary metabolic network in SBGN style (see section Standards for Systems and Synthetic Biology). Time series data for different genotypes and treatments have been mapped on the tricarboxylic acid (TCA) cycle pathway. The left figure not only shows the data using box plots but also a Pearson's correlation analysis for starch production (yellow). Red- and blue-colored elements have a positive or negative correlation, respectively. For further investigation, single elements can be

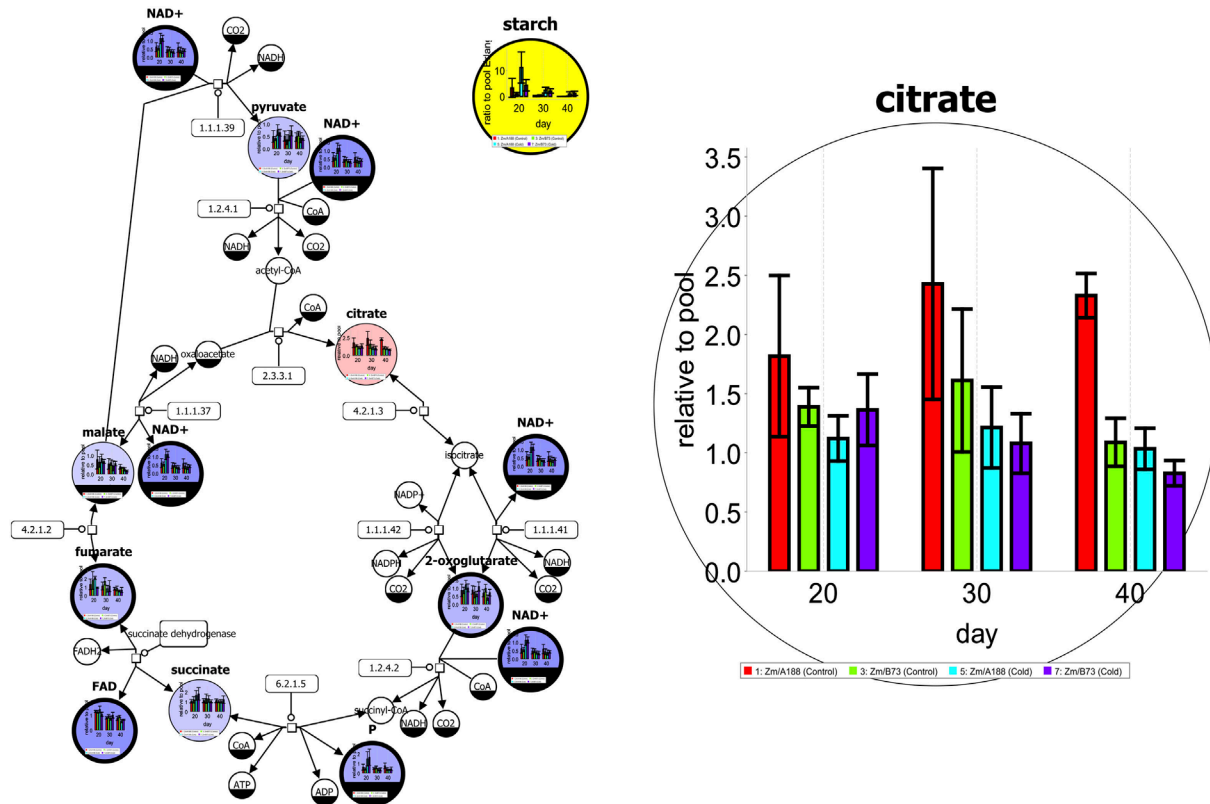


FIGURE 1 | A metabolic network of the tricarboxylic acid (TCA) cycle in SBGN style with experimental data from Optimas-DW (Colmsee et al., 2012) [produced using Vanted (Junker et al., 2006)]. More details are given in the text.

enlarged to support additional exploration of data as shown on the right side.

Databases and Repositories for Metabolites and Metabolism

Given the tools introduced before, metabolite and pathway databases and repositories are valuable resources that can be used to manage, explore, and export knowledge about metabolites and reactions in meaningful ways. They thereby deliver an encyclopedia on metabolic information as well as a base for integration of complex data into metabolic pathways in the context of graphical pathway representations (e.g., data about compound levels, reaction flow, enzyme activity, and gene expression, see Metabolomics Data Analysis and Visualization). Often these databases and repositories also provide or allow building metabolic models which can be analyzed and simulated using mathematical modeling techniques, see Section “Methods for Metabolic Modeling in Plants.” A typical example of information provided by databases is shown in Figure 2.

A range of pathway databases and repositories are currently available; an overview is available from the Pathguide resource (Bader et al., 2006). In Table 3, we provide a summary of important databases and repositories especially for plant research. It should be noted that in addition to general metabolic pathway databases such as Reactome, KEGG (Kyoto Encyclopedia of

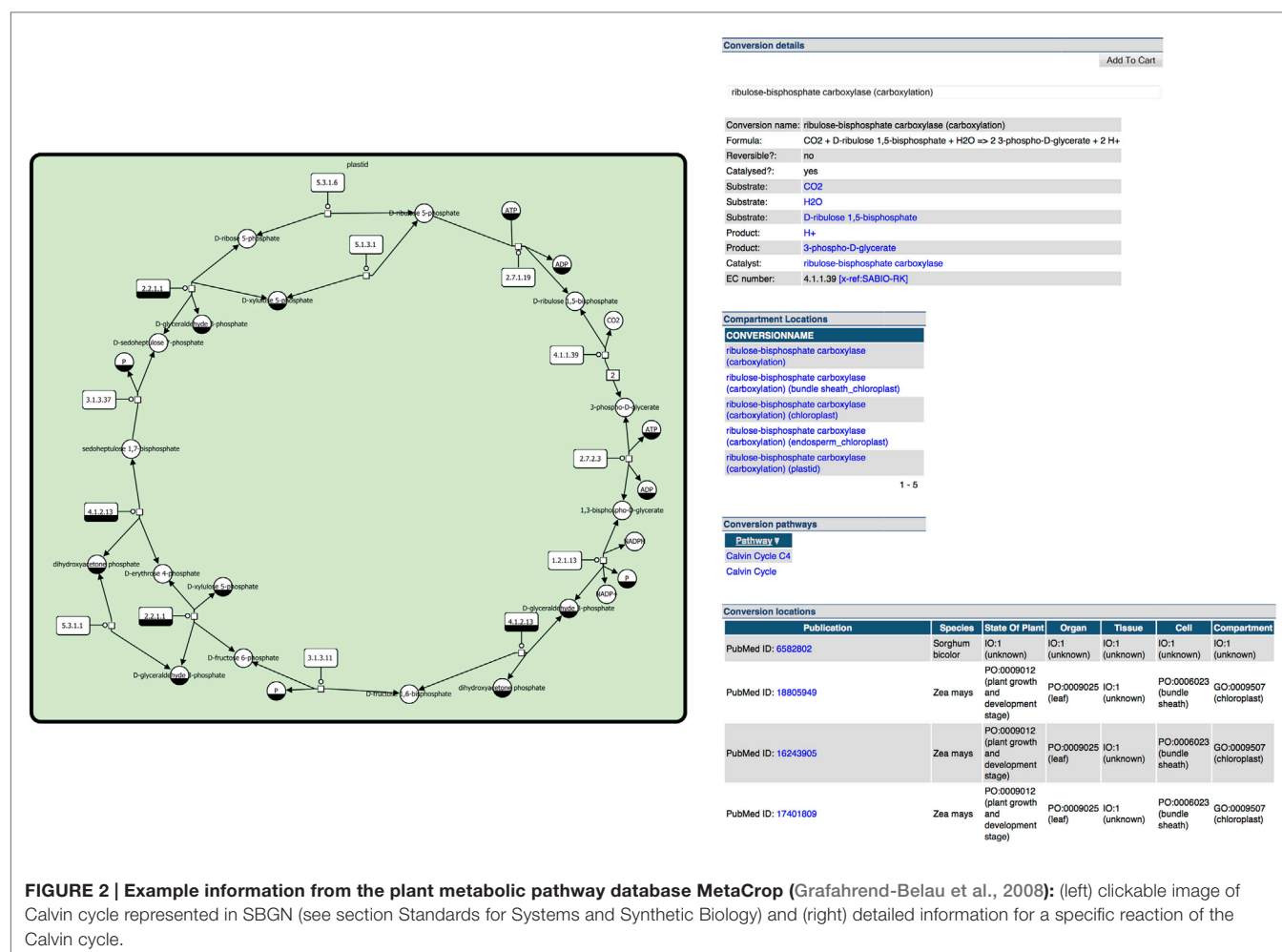
Genes and Genomes), and PANTHER Pathway, there are also general plant metabolic pathway databases (see Table 3) and several species-specific plant metabolic pathway databases available, for example, for *Arabidopsis* AraCyc (Mueller et al., 2003) and MetNetDB (Yang et al., 2005).

Examples for multispecies plant metabolic databases are:

- *PlantCyc*, which contains curated information on pathways, metabolites, reactions, genes, and enzymes,
- *Arabidopsis Reactome*, which has been extended from the initial representation of *Arabidopsis* pathways to further plant species and contains curated information about pathways, experimental evidence, literature citations, and further pathways imported from KEGG and AraCyc, and
- *MetaCrop*, which focuses on crop plant species with agro-nomical importance and contains curated information about pathways, metabolites, reactions, enzymes (including kinetic information), location, and developmental stage.

CURRENT STATUS OF ENGINEERING SYNTHETIC METABOLIC NETWORKS

For a long time, the use of biological organisms for the production of chemicals was limited by the repertoire of biosynthetic pathways naturally present in these organisms. Microbial



low-molecular-weight metabolic end products have long been used as commodity chemicals for medical applications: for example, antibiotics, such as penicillin, or immunosuppressive drugs, such as cyclosporine, have been discovered in 1929 and 1972, respectively (Drews, 2000). Microbes have also been used to produce chemical compounds via biotransformation, a process in which the compound of interest is produced by the microorganism through enzymatic conversion of an external substrate added to the microbial culture medium, used, for example, for the production of acrylamide by nitrile-assimilating bacteria (Asano et al., 1982).

With the advent of recombinant DNA technology in the early 1990s, it was possible to engineer specific genes in biological organisms, which has significantly reduced the time required for mutagenesis and selection of desirable traits. Genetic engineering made it possible to use heterologous hosts for the production of chemical compounds that are not naturally present in the organism. The emergence of the clustered, regularly interspaced, short palindromic repeat (CRISPR) and related technologies that use targeted genome editing via engineered nucleases are the latest developments to introduce alterations of genome sequences and gene expression, which can be ultimately used to also introduce modifications to existing metabolic pathways and to transfer

novel traits in agricultural crops (Shan et al., 2013; Sander and Joung, 2014).

The development of new sophisticated genomic sequencing and other enabling technologies for synthetic biology facilitated the production of naturally present chemicals at levels that made extraction economically feasible, and the field of metabolic engineering began to emerge. Metabolic engineering is defined as the targeted modification of metabolic pathways of biological organisms for metabolite overproduction or the improvement of cellular properties (Lessard, 1996). Since the last decade, significant progress was made to engineer the metabolism of plants to produce specific lipids, secondary metabolites, derivatives of complex natural products, and even vaccines (Mortimer et al., 2012). Many recent studies show that it is often not sufficient to modify existing metabolic pathways, but rather it is required to design metabolic pathways *de novo* from other plants or bacteria. Editing or redesigning existing plant metabolic networks is a challenging task that will benefit from advances in targeted genome modification, tissue-, cell-, and organelle-specific gene expression, the controlled expression of multigene pathways, and improvements in analytical technologies (as described in Analytical Technologies) as well as computational analysis and modeling methods (as described in Metabolomics Data Analysis

TABLE 3 | A summary of important databases and repositories for plant research.

Database	Exchange (download) formats	Reference	URL
Metabolite databases (provide information about metabolites/compounds such as names, chemical structures, molecular weight, occurrence in pathways, EC number, and mass spectrum references)			
ChEBI	XML, SDF, Tab-delimited	De Matos et al. (2009)	http://www.ebi.ac.uk/chebi/
GMD	MS formats	Kopka et al. (2005)	http://gmd.mpimp-golm.mpg.de/
KEGG compound	Jmol, MDL/MOL, KCF, KegDraw	Kanehisa et al. (2014)	http://www.genome.jp/kegg/compound/
PubChem	XML, SDF, SMILES	Bolton et al. (2008)	http://www.ncbi.nlm.nih.gov/pccompound
Reaction databases (provide information about reactions and enzymes such as names, reaction diagrams, reaction mechanisms, enzymatic parameters, occurrence in pathways, and links to encoding genes)			
BRENDA	SBML, Fasta, CVS	Scheer et al. (2010)	http://www.brenda-enzymes.org/
ExPASy-enzyme	–	Gasteiger et al. (2003)	http://enzyme.expasy.org/
KEGG enzyme/KEGG reaction	–	Kanehisa et al. (2014)	http://www.genome.jp/kegg/reaction/
Rhea	BioPAX, Tab-delimited	Morgat et al. (2015)	http://www.ebi.ac.uk/rhea/
Sabio-RK	SBML	Rojas et al. (2007)	http://sabio.villa-bosch.de/
Pathway databases (provide information about plant-specific metabolic pathways such as names, involved reactions, metabolites and enzymes, and pathway structure)			
KEGG pathway	KGML, BioPAX	Kanehisa et al. (2014)	http://www.genome.jp/kegg/pathway.html
MetaCrop	SBML, SBGN-ML	Schreiber et al. (2012)	http://metacrop.ipk-gatersleben.de/
MetaCyc	SBML, BioPAX	Krieger et al. (2004), Caspi et al. (2006)	http://metacyc.org/
PANTHER pathway	SBML, BioPAX, SBGN-ML	Mi et al. (2013)	http://www.pantherdb.org/pathway/
PlantCyc	SBML, BioPAX		http://www.plantcyc.org/
Reactome	SBML, BioPAX	Croft et al. (2014)	http://www.reactome.org/

and Visualization, Databases and Repositories for Metabolites and Metabolism, and Computational Approaches for Metabolic Engineering).

The ultimate goal of synthetic biology is the efficient design of biological systems (Heinemann and Panke, 2006). In this section, we will discuss the current status of engineering synthetic metabolic networks using recent examples for synthetic biology endeavors in plants: engineering of synthetic metabolic networks

of plant lipids to provide an alternative and sustainable source of nutrients (see Metabolic Engineering of Plant Lipids to Provide an Alternative and Sustainable Source of Nutrients) and for the production of fuels from renewable resources (see Metabolic Engineering of Plant Lipids for the Production of Fuels from Renewable Resources), and plant secondary metabolites including alkaloids and lignins (see Metabolic Engineering of Plant Secondary Metabolites).

Metabolic Engineering of Plant Lipids to Provide an Alternative and Sustainable Source of Nutrients

Plant oils are a major component of human diets, comprising as much as 25% of average caloric intake (Broun et al., 1999). However, certain fatty acids such as omega-3 long-chain polyunsaturated fatty acids (ω 3 LC-PUFA) are present predominantly in fish and have important functions for human health, as deficiencies in these fatty acids can increase the risk or severity of cardiovascular and inflammatory diseases (Abeywardena and Patten, 2011). Until recently, the chemical composition of plant oils was constrained by the repertoire of naturally present lipid biosynthetic pathways. Novel opportunities have emerged to tailor the composition of plant-derived lipids so that they are optimized with respect to food functionality and human dietary needs. For example, Petrie et al. (2010, 2012) have recently described metabolic engineering of ω 3 LC-PUFA in plants: after inserting seven biosynthesis genes of the docosahexaenoic acid (DHA) biosynthesis pathway from microalgae into the genome of *Arabidopsis thaliana*, they were able to obtain ω 3 LC-PUFA levels in seeds similar to that observed in bulk fish oil. If applied to oilseed crops such as *Brassica napus*, this technology could potentially form the basis of a plant-based sustainable source to complement the existing marine fish oil supply.

Metabolic Engineering of Plant Lipids for the Production of Fuels from Renewable Resources

Fossil fuels are the primary source of many industrial products, but reserves are decreasing rapidly and are non-renewable, and their widespread use has contributed to environmental problems arising from increased CO₂ levels in the atmosphere (Le Quéré et al., 2009). Currently, biologically derived fuels from plant oils represent one of the main strategies to provide renewable and sustainable source material that can potentially substitute fossil fuels in some industrial applications. Among the many proposed solutions, algal biofuels are seen as one of the most promising: algal biomass is less resistant to conversion into simple sugars than plant biomass due to lack of lignin, and there is no issue arising from the food versus feed dilemma as no farmland has to be diverted for the production of biofuels (Daroch et al., 2013). Over the last few years, progress was made in bioethanol production through fermentation from algal feedstock (Kim et al., 2012) as well as biodiesel production from algal oils (Singh and Dhar, 2011). Increasingly, research efforts are focusing to metabolically engineer lipid pathways to increase lipid accumulation without

compromising growth (Trentacoste et al., 2013). Although previously *A. thaliana* mutants of lipid catabolism were found to be linked with impaired growth (Graham, 2008), Trentacoste et al. (2013) demonstrated that disrupting lipid catabolism via the knockdown of a multifunctional lipase/phospholipase/acyl-transferase in the microalgae *Thalassiosira pseudonana* led to an increased lipid accumulation without compromised algal growth. Further elucidation of lipid metabolism has the potential to lead to new strategies to engineer improved algal strains for their fuel molecules.

Metabolic Engineering of Plant Secondary Metabolites

Plant secondary metabolites, such as alkaloids, flavonoids, terpenes, and phenylpropanoids (Hill et al., 2014), are considered to be non-essential for normal growth and development but play important roles in plant defense against pathogens and other environmental stresses. Additionally, plant secondary metabolites are of great interest to pharmaceutical industries, as they often have beneficial medicinal effects on humans. For example, many plant alkaloids are currently in medical use, such as atropine derived from the nightshade *Atropa belladonna*, morphine from the opium poppy *Papaver somniferum*, and quinine from the *Cinchona* tree (Roberts and Wink, 1998). Recent progress has been made in the metabolic engineering of morphine, a medically important benzyloquinoline alkaloid: Rungtaphan et al. (2012) reengineered a codeine *O*-demethylase mutant that selectively demethylates codeine instead of both codeine and thebaine, as is common in the wild-type morphinan biosynthesis pathway. The integration of this highly selective mutant enzyme into commercial poppy plants as part of a future metabolic engineering effort has the potential to increase yields of morphine and codeine.

The phenylpropanoid pathway is conserved in all terrestrial plants and is responsible for the biosynthesis of many compounds that are involved in plant cell wall structure and integrity, water transport, and plant defense. They are required for the biosynthesis of lignins, aromatic natural polymers in secondary cell walls derived from the oxidative polymerization of monolignols. Decreasing or altering lignin structure provides enhanced cell wall digestibility and can greatly increase the utilization of lignin itself or cell wall polysaccharides. Due to the importance of lignin in agriculture and industry, the genes participating in lignin biosynthesis have been identified and modified in many plant species including switchgrass (Fu et al., 2011), *A. thaliana* (Gallego-Giraldo et al., 2011), and sugarcane (Jung et al., 2012). In a recent study, Zhang et al. (2012) were able to manipulate lignification in *A. thaliana* without compromising plant growth by introducing an artificial enzyme that esterifies the para-hydroxyl of phenols. The modified 4-*O*-methyl lignin monomers deprive the products of participation in oxidative dehydrogenation, leading to a decreased level of available monolignols for lignin polymerization and thus to depressed lignin biosynthesis. Further metabolic engineering efforts are currently underway to integrate this artificial enzyme into poplar with the potential to manipulate lignin levels.

COMPUTATIONAL APPROACHES FOR METABOLIC ENGINEERING

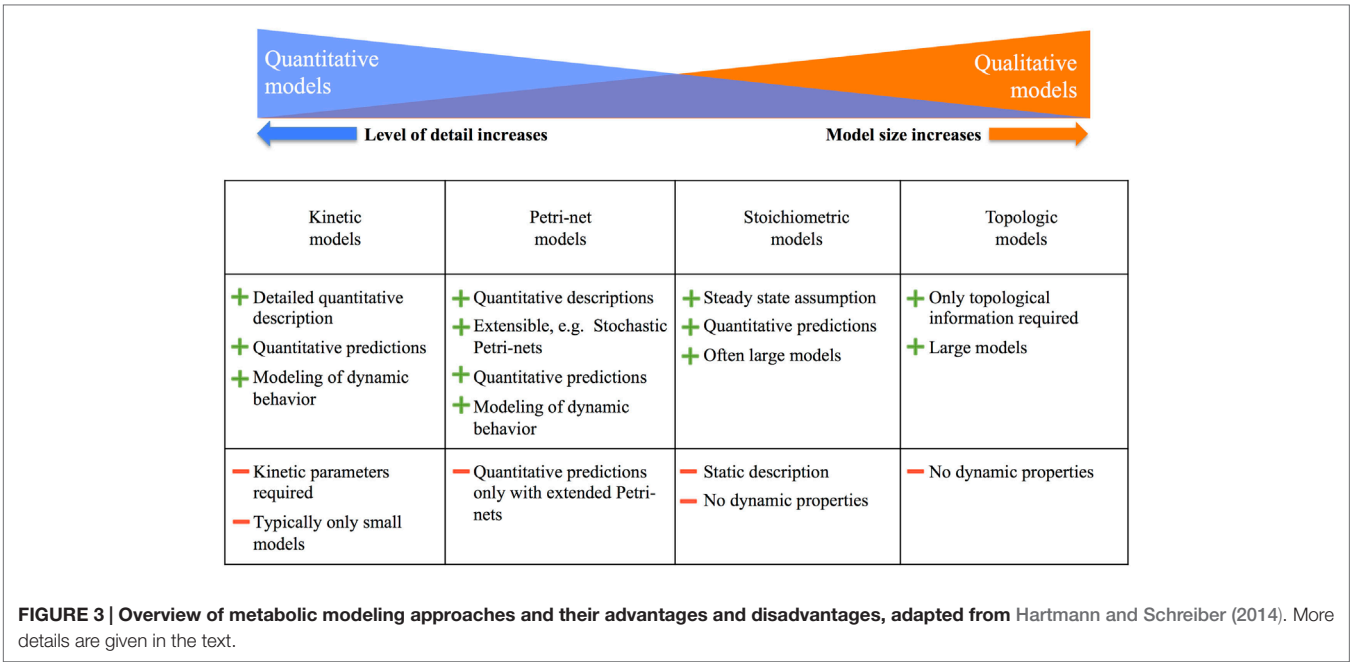
A variety of different methods and approaches to collect experimental data can be used to quantify metabolites and other components of regulatory networks in plants, such as metabolomics (see Analytical Technologies). Computational modeling is an important tool for metabolic engineering as it facilitates the integration and analysis of experimental datasets to quantify metabolic fluxes and model metabolic networks. Section “Methods for Metabolic Modeling in Plants” presents a brief overview of modeling approaches. There are many tools and databases available for computational modeling and therefore a standardized exchange of models is highly relevant; Section “Standards for Systems and Synthetic Biology” provides an introduction to major standards in systems in synthetic biology.

Methods for Metabolic Modeling in Plants

Several approaches have been developed to qualitatively and quantitatively model and simulate metabolic systems *in silico*. This ranges from topological analysis of network models (which looks at the interconnections between metabolites) to stoichiometric models (where constraints can be applied to define the potential metabolic flux state space or which can be analyzed using Petri-nets) to detailed kinetic models (which model changes of metabolite concentration over time). A current review (Baghalian et al., 2014) discusses the different modeling approaches, modeling software, and metabolic models of several plants in detail. A particular challenge in plants compared to prokaryotic cells is the number of different compartments, which needs to be considered in metabolic models. In addition, plants have a greater complexity of metabolic pathways and especially a large number of special pathways for secondary metabolites.

For an overview, **Figure 3** summarizes the major modeling approaches and their advantages and disadvantages. The most detailed are kinetic models, which allow for a comprehensive quantitative description and prediction of metabolic fluxes. However, in plants, these models are common only in the size of 10–20 reactions. As we move further to the right in **Figure 3**, the model size increases, but the level of detailed descriptions and predictions decreases. As the other extreme, topological models allow covering the complete metabolism in plants, but predictions are restricted to qualitative information such as reachability of metabolites.

As detailed in (Baghalian et al., 2014), plant-specific computational models of metabolism can be used for different purposes such as predicting the behavior of the metabolism under different conditions, analyzing the effect of mutations, and investigating the effect of changes due to manipulation of the metabolic system, for example, via the introduction of new metabolic pathways. Computational models usually allow investigating effects much faster and cheaper than by running wet laboratory experiments (Rohwer, 2012). In addition, using a computational model can often generate a set of alternative strategies (Copeland et al., 2012). Finally, computational models allow integrating additional data such as transcriptomics and proteomics data sets, which together with bioinformatics approaches can support a better understanding of metabolic behavior in plants (Töpfer et al., 2012, 2013).



Standards for Systems and Synthetic Biology

This section presents a short introduction to major standards in systems and synthetic biology related to software infrastructure (see also **Table 4**). Software infrastructure plays an important role in systems biology research (Kitano, 2002), in particular in supporting standardized exchange of information between different tools and databases. The major standards are Systems Biology Markup Language (SBML), CellML, Systems Biology Graphical Notation (SBGN), and Synthetic Biology Open Language (SBOL), Language (SBOL), see (Schreiber et al., 2015) for detailed specifications.

Systems Biology Markup Language

Systems Biology Markup Language is a machine-readable format for representation and exchange of computational models in systems biology. It can represent models of metabolism, signal transduction, and gene regulation. The main goals of SBML are (1) sharing and publication of models, (2) reusability of models, and (3) surviving of models beyond the lifetime of the software used to create them. The SBML web page currently lists more than 270 software applications that support SBML, and thousands of SBML-encoded models are available from public repositories such as BioModels including Path2Models (Büchel et al., 2013; Chelliah et al., 2013).

An SBML model consists of hierarchical lists of conceptual elements: (1) species (biological entities taking part in reactions), (2) compartments (physical containers for species), and (3) reactions (transformation, transport, or binding processes occurring over time). For the analysis and simulation of a model, more properties need to be defined such as stoichiometries, rate laws, local and global parameters, as well as units on quantities. A formal

TABLE 4 | Overview of major standards in systems and synthetic biology.

Name	Reference	URL
SBML	Hucka et al. (2003)	http://sbml.org/
CellML	Cuellar et al. (2003)	http://www.cellml.org/
SBGN	Le Novère et al. (2009)	http://sbgn.org/
SBOL	Galdzicki et al. (2014)	http://sbolstandard.org/

description of SBML can be found in the detailed specification (Hucka et al., 2015).

CellML

CellML is a machine-readable format for representation, publication, and sharing of mathematical models of cellular function. In comparison to SBML, the focus of CellML is on the representation of a variety of models such as models of biological pathways, electrophysiological models, and mechanical models. The CellML web page lists a couple of software tools which support the CellML format and the development of models. The CellML Model Repository (Lloyd et al., 2008) contains several hundred models including a subrepository providing SVPs (Standard Virtual Biological Parts) for the composition of synthetic biology models (Cooling et al., 2010).

A CellML model description consists of components and lists of connections between the components. A component contains at least one variable and mathematical equations describing its behavior. Connections are mappings of variables between components enabling information exchange between them. Components and connections can be imported from an existing model, CellML allows reusing of parts of other models. A detailed description of CellML can be found in the specification (Cuellar et al., 2006).

Systems Biology Graphical Notation

Systems Biology Graphical Notation is a standard for the graphical representation of processes and networks studied in systems biology. Three SBGN languages (Process Description, PD; Entity Relationship, ER; and Activity Flow, AF) allow for the representation of different aspects of biological systems at different levels of detail as SBGN maps, thus providing corresponding views on the underlying biological system. The PD language (Moodie et al., 2015) describes biological entities and processes between these entities, the ER language (Sorokin et al., 2015) focuses on interactions between biological entities, and the AF language (Mi et al., 2015) depicts information flow between biological activities. An example of an SBGN PD map is shown in **Figure 1**.

The standardization of graphical representations helps to exchange biological knowledge more efficiently and accurately between different research communities, industry, and other players in systems biology. Several databases already provide maps in SBGN, e.g., BioModels Database including Path2Models (Büchel et al., 2013; Chelliah et al., 2013), MetaCrop (Schreiber et al., 2012), PANTHER Pathway (Mi et al., 2013), Reactome (Croft et al., 2014), and RIMAS (Junker et al., 2010). The SBGN web page lists more than 20 software tools that support creating, editing, and viewing of SBGN maps, some of these tools allow to visualize SBML models in SBGN PD.

Synthetic Biology Open Language

Synthetic Biology Open Language is a data format for sharing and exchanging synthetic biology designs. It allows synthetic biologists to provide an unambiguous description of a design in a hierarchical and fully annotated form with the goal to improve designing, building, testing, and dissemination of synthetic biology designs. For the visualization of synthetic biology designs in SBOL, SBOL Visual (Synthetic Biology Open Language Visual) has been developed. It is a graphical notation allowing depiction of the structure of a design using glyphs to specify genetic parts, devices, modules, and systems.

The SBOL web page currently lists more than 20 software applications supporting SBOL and SBOL Visual. Some of these applications allow the generation of SBML models from synthetic biology designs in SBOL (Roehner et al., 2015) as well as the creation of synthetic biology designs in SBOL by automatically generating DNA sequences from annotated SBML and CellML models (Misirli et al., 2011). More detailed information

about SBOL and SBOL Visual can be found in the specifications (Quinn et al., 2013; Bartley et al., 2015).

CONCLUSION

The total number of metabolites in the plant kingdom is estimated to be between 100,000 and 200,000 and can be highly variable depending on the physiological and environmental conditions as well as the genetic background of the plant (Hill et al., 2013a,b). Such great metabolic diversity holds great promise for expanding our repertoire of known beneficial plant compounds, as many metabolic pathways and regulatory mechanisms are still awaiting discovery. Reaching significant benchmarks toward attaining these goals will be possible with better analytical tools. More accurate representations of metabolite identities and quantities will require analytical instruments and improved techniques for sample extraction and data analysis. The engineering of synthetic metabolic networks of plants will require further advances in targeted genome modification such as the application of the CRISPR/Cas system, as well as tissue-, cell-, and organelle-specific gene expression, and the controlled expression of multigene pathways. The development of methods for measuring metabolic flux directly, the quantification of metabolites in individual plant compartments, and the analysis of metabolites and activities between compartments *in vivo* will be very important next steps to further enhance the predictive capabilities of existing metabolic models. Continuous development of more user-friendly software, databases, languages, and computer models that incorporate and interpret complex information will be crucial to handle the acquired data and to aid interpretation in a biological context. We are just at the beginning of a new area of synthetic biology in plants based on metabolomics and metabolic modeling.

ACKNOWLEDGMENTS

UR and CH are funded through an Australian Research Council Future Fellowship program and are also grateful to the Victorian Node of Metabolomics Australia, which is funded through Bioplatforms Australia Pty Ltd, a National Collaborative Research Infrastructure Strategy (NCRIS), 5.1 biomolecular platforms and informatics investment, and coinvestment from the Victorian State government and The University of Melbourne. The authors are grateful for financial support from EFI SynBio.

REFERENCES

- Abeywardena, M. Y., and Patten, G. S. (2011). Role of ν_3 longchain polyunsaturated fatty acids in reducing cardio-metabolic risk factors. *Endocr. Metab. Immune Disord. Drug Targets* 11, 232–246. doi:10.2174/187153011796429817
- Allwood, J. W., Erban, A., de Koning, S., Dunn, W. B., Luedemann, A., Lommen, A., et al. (2009). Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics* 5, 479–496. doi:10.1007/s11306-009-0169-z
- Asano, Y., Yasuda, T., Tani, Y., and Yamada, H. (1982). A new enzymatic method of acrylamide production. *Agric. Biol. Chem.* 46, 1183–1189. doi:10.1271/bbb1961.46.1183
- Bader, G. D., Cary, M. P., and Sander, C. (2006). Pathguide: a pathway resource list. *Nucleic Acids Res.* 34, D504–D506. doi:10.1093/nar/gkt1076
- Baghalian, K., Hajirezaei, M. R., and Schreiber, F. (2014). Plant metabolic modeling: achieving new insight into metabolism and metabolic engineering. *Plant Cell* 26, 3847–3866. doi:10.1105/tpc.114.130328
- Bartley, B., Beal, J., Clancy, K., Misirli, G., Roehner, N., Oberortner, E., et al. (2015). Synthetic Biology Open Language (SBOL) Version 2.0.0 *J. Int. Bioinformatics* 12, 272. doi:10.2390/biecoll-jib-2015-272
- Barupal, D. K., Haldiya, P. K., Wohlgenuth, G., Kind, T., Kothari, S. L., Pinkerton, K. E., et al. (2012). MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 13:99. doi:10.1186/1471-2105-13-99
- Beckles, D., and Roessner, U. (2011). “Plant metabolomics – applications and opportunities for agricultural biotechnology,” in *Plant Biotechnology and Agriculture: Prospects for the 21st Century*, eds Altmann A. and Hasegawa P. M. (London, UK: Elsevier), p. 67–80.

- Bolton, E. E., Wang, Y., Thiessen, P. A., and Bryant, S. H. (2008). PubChem: integrated platform of small molecules and biological activities. *Ann. Rep. Comput. Chem.* 4, 217–241. doi:10.1016/s1574-1400(08)00012-1
- Broun, P., Gettner, S., and Somerville, C. (1999). Genetic engineering of plant lipids. *Ann. Rev. Nutr.* 19, 197–216. doi:10.1146/annurev.nutr.19.1.197
- Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., et al. (2013). Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* 7:116. doi:10.1186/1752-0509-7-116
- Callahan, D. L., Souza, D. D., Bacic, A., and Roessner, U. (2009). Profiling of polar metabolites in biological extracts using diamond hydride-based aqueous normal phase chromatography. *J. Sep. Sci.* 32, 2273–2280. doi:10.1002/jssc.200900171
- Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., et al. (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 34(Suppl. 1), D511–D516. doi:10.1093/nar/gkj128
- Chelliah, V., Laibe, C., and Le Novère, N. (2013). “BioModels database: a repository of mathematical models of biological processes,” in *Silico Systems Biology (Methods in Molecular Biology)*, ed. Schneider M. V. (New York, US: Humana Press), 189–199.
- Chokkathukalam, A., Kim, D. H., Barrett, M. P., Breitling, R., and Creek, D. J. (2014). Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis* 6, 511–524. doi:10.4155/bio.13.348
- Clasquin, M. F., Melamud, E., and Rabinowitz, J. D. (2012). LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* 14, 11. doi:10.1002/0471250953.bi1411s37
- Colmsee, C., Mascher, M., Czauderna, T., Hartmann, A., Schlüter, U., Zellerhoff, N., et al. (2012). OPTIMAS-DW: a comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize. *BMC Plant Biol.* 12:245. doi:10.1186/1471-2229-12-245
- 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
- Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., et al. (2012). Computational tools for metabolic engineering. *Metab. Eng.* 14, 270–280. doi:10.1016/j.ymben.2012.03.001
- Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., et al. (2014). The reactome pathway knowledgebase. *Nucleic Acids Res.* 42, D472–D477. doi:10.1093/nar/gkt1102
- Cuellar, A. A., Lloyd, C. M., Nielsen, P. F., Bullivant, D. P., Nickerson, D. P., and Hunter, P. J. (2003). An overview of CellML 1.1, a biological model description language. *Simulation* 79, 740–747. doi:10.1177/0037549703040939
- Cuellar, A. A., Nielsen, P. F., Halstead, M., Bullivant, D. P., Nickerson, D. P., Hedley, W., et al. (2006). *CellML 1.1 Specification*. Available at: http://www.cellml.org/specifications/cellml_1.1/index_html
- Daroch, M., Geng, S., and Wang, G. (2013). Recent advances in liquid biofuel production from algal feedstocks. *Appl. Energy* 102, 1371–1381. doi:10.1016/j.apenergy.2012.07.031
- De Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., et al. (2009). Chemical entities of biological interest: an update. *Nucleic Acids Res.* 38, D249–D254. doi:10.1093/nar/gkp886
- Dias, D. A., Hill, C. B., Jayasinghe, N. S., Atieno, J., Sutton, T., and Roessner, U. (2015). Quantitative profiling of polar primary metabolites of two chickpea cultivars with contrasting responses to salinity. *J. Chromatogr. B* 1000, 1–13. doi:10.1016/j.jchromb.2015.07.002
- Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi:10.3390/metabo2020303
- Drews, J. (2000). Drug discovery: a historical perspective. *Science* 287, 1960–1964. doi:10.1126/science.287.5460.1960
- Dromms, R. A., and Styczynski, M. P. (2012). Systematic applications of metabolomics in metabolic engineering. *Metabolites* 2, 1090–1122. doi:10.3390/metabo2041090
- Eisenreich, W., and Bacher, A. (2007). Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry* 68, 2799–2815. doi:10.1016/j.phytochem.2007.09.028
- Fernie, A. R., and Morgan, J. A. (2013). Analysis of metabolic flux using dynamic labelling and metabolic modelling. *Plant Cell Environ.* 36, 1738–1750. doi:10.1111/pce.12083
- Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., et al. (2007). The metabolomics standards initiative (MSI). *Metabolomics* 3, 175–178. doi:10.1007/s11306-007-0070-6
- Fitzpatrick, M. A., McGrath, C. M., and Young, S. P. (2014). Pathomx: an interactive workflow-based tool for the analysis of metabolomic data. *BMC Bioinformatics* 15:396. doi:10.1186/s12859-014-0396-9
- Fu, C., Mielenz, J. R., Xiao, X., Ge, Y., Hamilton, C. Y., Rodriguez, M., et al. (2011). Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3803–3808. doi:10.1073/pnas.1100310108
- 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
- Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L. A., and Dixon, R. A. (2011). Salicylic acid mediates the reduced growth of lignin down-regulated plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20814–20819. doi:10.1073/pnas.1117873108
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi:10.1093/nar/gkg563
- Gaude, E., Chignola, F., Spiliotopoulos, D., Spitaleri, A., Ghitti, M., Garcia-Manteiga, J. M., et al. (2013). muma, an R package for metabolomics univariate and multivariate statistical analysis. *Curr. Metabolomics* 1, 180–189. doi:10.2174/2213235x11301020005
- Gowda, H., Ivanisevic, J., Johnson, C. H., Kurczyk, M. E., Benton, H. P., Rinehart, D., et al. (2014). Interactive XCMS online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* 86, 6931–6939. doi:10.1021/ac500734c
- Grafarend-Belau, E., Weise, S., Koschützki, D., Scholz, U., Junker, B. H., and Schreiber, F. (2008). MetaCrop: a detailed database of crop plant metabolism. *Nucleic Acids Res.* 36(Suppl. 1), D954–D958. doi:10.1093/nar/gkm835
- Graham, I. A. (2008). Seed storage oil mobilization. *Annu. Rev. Plant Biol.* 59, 115–142. doi:10.1146/annurev.arplant.59.032607.092938
- Hartmann, A., and Schreiber, F. (2014). Integrative analysis of metabolic models – from structure to dynamics. *Front. Bioeng. Biotechnol.* 2:91. doi:10.3389/fbioe.2014.00091
- Heinemann, M., and Panke, S. (2006). Synthetic biology – putting engineering into biology. *Bioinformatics* 22, 2790–2799. doi:10.1002/9780470437988.ch11
- Hill, C. B., Bacic, A., and Roessner, U. (2014). “LC-MS profiling to link metabolic and phenotypic diversity in plant mapping populations,” in *Mass Spectrometry in Metabolomics*, ed. Raftery D. (New York, NY: Springer), 29–41.
- Hill, C. B., Jha, D., Bacic, A., Tester, M., and Roessner, U. (2013a). Characterization of ion contents and metabolic responses to salt stress of different *Arabidopsis AtHKT1;1* genotypes and their parental strains. *Mol. Plant* 6, 350–368. doi:10.1093/mp/sss125
- Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Bacic, A., Langridge, P., et al. (2013b). Whole-genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment. *Plant Physiol.* 162, 1266–1281. doi:10.1104/pp.113.217851
- Hill, C. B., and Roessner, U. (2013). “Metabolic profiling of plants by GC–MS,” in *The Handbook of Plant Metabolomics*, eds Weckwerth W. and Kahl G. (Hoboken, NJ: John Wiley & Sons), 1–23.
- Hill, C. B., and Roessner, U. (2015). “Advances in high-throughput LC-MS analysis for plant metabolomics,” in *Advanced LC-MS Applications for Metabolomics*, ed. De Vooght-Johnson R. (London, UK: Future Science Ltd), 58–71.
- Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Langridge, P., Bacic, A., et al. (2015). Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. *Plant Sci.* 233, 143–154. doi:10.1016/j.plantsci.2015.01.008
- Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., Schaff, J. C., et al. (2015). The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core. *J. Int. Bioinformatics* 12, 266. doi:10.2390/biecoll-jib-2015-266
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531. doi:10.1093/bioinformatics/btg015
- Jung, J. H., Fouad, W. M., Vermerris, W., Gallo, M., and Altpeter, F. (2012). RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel

- production from lignocellulosic biomass. *Plant Biotechnol. J.* 10, 1067–1076. doi:10.1111/j.1467-7652.2012.00734.x
- Junker, A., Hartmann, A., Schreiber, F., and Bäumlein, H. (2010). An engineer's view on regulation of seed development. *Trends Plant Sci.* 15, 303–307. doi:10.1016/j.tplants.2010.03.005
- Junker, B. H., Klukas, C., and Schreiber, F. (2006). VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7:109. doi:10.1186/1471-2105-7-109
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi:10.1093/nar/gkt1076
- Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., et al. (2012). Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* 28, 373–380. doi:10.1093/bioinformatics/btr661
- Kaspar, S., Peukert, M., Svatos, A., Matros, A., and Mock, H. P. (2011). MALDI-imaging mass spectrometry – an emerging technique in plant biology. *Proteomics* 11, 1840–1850. doi:10.1002/pmic.201000756
- Kastenmüller, G., Römisch-Margl, W., Wägele, B., Altmaier, E., and Suhre, K. (2011). metaP-server: a web-based metabolomics data analysis tool. *Biomed Res. Int.* 2011, 839862. doi:10.1155/2011/839862
- Kerren, A., and Schreiber, F. (2012). “Towards the role of interaction in visual analytics,” in *Proc. Winter Simulation Conference – WSC (Berlin: IEEE)*, Vol. 420, 1–13. doi:10.1109/WSC.2012.6465208
- Kerren, A., and Schreiber, F. (2014). ‘Why integrate InfoVis and SciVis? An example from systems biology’. *IEEE Comput. Graph. Appl.* 34, 6–10. doi:10.1109/mcg.2014.122
- Kim, J., Um, B.-H., and Kim, T. (2012). Bioethanol production from micro-algae, *Schizocytium* sp., using hydrothermal treatment and biological conversion. *Korean J. Chem. Eng.* 29, 209–214. doi:10.1007/s11814-011-0169-3
- Kirchner, M., Saussen, B., Steen, H., Steen, J. A. J., and Hamprecht, F. A. (2007). amsrpm: robust point matching for retention time alignment of LC/MS data with R. *J. Stat. Softw.* 18, 1–12. doi:10.18637/jss.v018.i04
- Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. doi:10.1126/science.1069492
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2005). GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* 21, 1635–1638. doi:10.1093/bioinformatics/bti236
- Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., et al. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32(Suppl. 1), D438–D442. doi:10.1093/nar/gkh100
- Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., et al. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741. doi:10.1038/nbt.1558
- Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nat. Geosci.* 2, 831–836. doi:10.1038/ngeo689
- Lessard, P. (1996). Metabolic engineering: the concept coalesces. *Nat. Biotechnol.* 14, 1654–1655. doi:10.1038/nbt1296-1654
- Lloyd, C. M., Lawson, J. R., Hunter, P. J., and Nielsen, P. F. (2008). The CellML model repository. *Bioinformatics* 24, 2122–2123. doi:10.1093/bioinformatics/btn390
- MathWorks. (2012). *MATLAB and Statistics Toolbox Release*. Natick, MA: The MathWorks, Inc.
- Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D377–D386. doi:10.1093/nar/gks1118
- Mi, H., Schreiber, F., Moodie, S., Czauderna, T., Demir, E., Haw, R., et al. (2015). Systems Biology Graphical Notation: activity flow language level 1 version 1.2. *J. Int. Bioinformatics* 12, 265. doi:10.2390/biecoll-jib-2015-265
- Misirli, G., Hallinan, J. S., Yu, T., Lawson, J. R., Wimalaratne, S. M., Cooling, M. T., et al. (2011). Model annotation for synthetic biology: automating model to nucleotide sequence conversion. *Bioinformatics* 27, 973–979. doi:10.1093/bioinformatics/btr048
- Miura, D., Fujimura, Y., and Wariishi, H. (2012). In situ metabolomic mass spectrometry imaging: recent advances and difficulties. *J. Proteomics* 75, 5052–5060. doi:10.1016/j.jprot.2012.02.011
- Moodie, S., Le Novère, N., Demir, E., Mi, H., and Villéger, A. (2015). Systems Biology Graphical Notation: process description language level 1 version 1.3. *J. Int. Bioinformatics* 12, 263. doi:10.1038/npre.2009.3721.1
- Morgat, A., Axelsen, K. B., Lombardot, T., Alcántara, R., Aimo, L., Zerara, M., et al. (2015). Updates in RHEA – a manually curated resource of biochemical reactions. *Nucleic Acids Res.* 43, D459–D464. doi:10.1093/nar/gku961
- Mortimer, E., Maclean, J. M., Mbewana, S., Buys, A., Williamson, A. L., Hitzeroth, I. L., et al. (2012). Setting up a platform for plant-based influenza virus vaccine production in South Africa. *BMC Biotechnol.* 12:14. doi:10.1186/1472-6750-12-14
- Mueller, L. A., Zhang, P., and Rhee, S. Y. (2003). AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132, 453–460. doi:10.1104/pp.102.017236
- O'Grady, J., Schwender, J., Shachar-Hill, Y., and Morgan, J. A. (2012). Metabolic cartography: experimental quantification of metabolic fluxes from isotopic labelling studies. *J. Exp. Bot.* 63, 2293–2308. doi:10.1093/jxb/ers032
- Petrie, J. R., Shrestha, P., Liu, Q., Mansour, M. P., Wood, C. C., Zhou, X. R., et al. (2010). Rapid expression of transgenes driven by seed-specific constructs in leaf tissue: DHA production. *Plant Methods* 6, 8–13. doi:10.1186/1746-4811-6-8
- Petrie, J. R., Shrestha, P., Zhou, X. R., Mansour, M. P., Liu, Q., Belide, S., et al. (2012). Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLoS One* 7:e49165. doi:10.1371/journal.pone.0049165
- 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*.
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rios-Estapa, R., and Lange, B. M. (2007). Experimental and mathematical approaches to modeling plant metabolic networks. *Phytochemistry* 68, 2351–2374. doi:10.1016/j.phytochem.2007.04.021
- Roberts, M. F., and Wink, M. (eds) (1998). *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*. New York, USA: Springer Science & Business Media.
- Roehner, N., Zhang, Z., Nguyen, T., and Myers, C. J. (2015). Generating systems biology markup language models from the synthetic biology open language. *ACS Synth. Biol.* 4, 873–879. doi:10.1021/sb5003289
- Rohn, H., Junker, A., Hartmann, A., Grafarend-Belau, E., Treutler, H., Klapperstück, M., et al. (2012). VANTED v2: a framework for systems biology applications. *BMC Syst. Biol.* 6:139. doi:10.1186/1752-0509-6-139
- Rohn, H., Klukas, C., and Schreiber, F. (2011). Creating views on integrated multidomain data. *Bioinformatics* 27, 1839–1845. doi:10.1093/bioinformatics/btr282
- Rohwer, J. M. (2012). Kinetic modelling of plant metabolic pathways. *J. Exp. Bot.* 63, 2275–2292. doi:10.1093/jxb/ers080
- Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., et al. (2007). Storing and annotating of kinetic data. *In Silico Biol.* 7, 37–44.
- Runguphan, W., Glenn, W. S., and O'Connor, S. E. (2012). Redesign of a dioxxygenase in morphine biosynthesis. *Chem. Biol.* 19, 674–678. doi:10.1016/j.chembiol.2012.04.017
- 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
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., et al. (2010). BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* 39, D670–D676. doi:10.1093/nar/gkq1089
- Schreiber, F., Colmsee, C., Czauderna, T., Grafarend-Belau, E., Hartmann, A., Junker, A., et al. (2012). MetaCrop 2.0: managing and exploring information about crop plant metabolism. *Nucleic Acids Res.* 40, D1173–D1177. doi:10.1093/nar/gkr1004
- Schreiber, F., Bader, G. D., Golebiewski, M., Hucka, M., Kormeier, B., Le Novère, N., et al. (2015). Specifications of standards in systems and synthetic biology. *J. Int. Bioinformatics* 12, 258. doi:10.2390/biecoll-jib-2015-258
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650
- Singh, N. K., and Dhar, D. W. (2011). Microalgae as second generation biofuel. A review. *Agronomy Sustain. Dev.* 31, 605–629. doi:10.1007/s13593-011-0018-0
- Sommer, B., Dingersen, T., Gamroth, C., Schneider, S. E., Rubert, S., Krüger, J., et al. (2011). CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous membrane packing problems. *J. Chem. Inf. Model.* 51, 1165–1182. doi:10.1021/ci1003619
- Sorokin, A., Le Novère, N., Luna, A., Czauderna, T., Demir, E., Haw, R., et al. (2015). Systems Biology Graphical Notation: entity relationship language level 1 version 2. *J. Int. Bioinformatics* 12, 264. doi:10.2390/biecoll-jib-2015-264
- Töpfer, N., Caldana, C., Grimbs, S., Willmitzer, L., Fernie, A. R., and Nikoloski, Z. (2013). Integration of genome-scale modeling and transcript profiling

- reveals metabolic pathways underlying light and temperature acclimation in *Arabidopsis*. *Plant Cell* 25, 1197–1211. doi:10.1105/tpc.112.108852
- Töpfer, N., Jozefczuk, S., and Nikoloski, Z. (2012). Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. *BMC Syst. Biol.* 6:148. doi:10.1186/1752-0509-6-148
- Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Glé, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19748–19753. doi:10.1073/pnas.1309299110
- Trethewey, R. N. (2004). Metabolite profiling as an aid to metabolic engineering in plants. *Curr. Opin. Plant Biol.* 7, 196–201. doi:10.1016/j.pbi.2003.12.003
- Villas-Boas, S. G., Nielsen, J., Smedsgaard, J., Hansen, M. A., and Roessner-Tunali, U. (2007). *Metabolome Analysis: An Introduction*, Vol. 24. Hoboken, USA: John Wiley & Sons.
- Xia, J., Sinelnikov, I. V., Han, B., and Wishart, D. S. (2015). MetaboAnalyst 3.0 – making metabolomics more meaningful. *Nucleic Acids Res.* 43, W251–W257. doi:10.1093/nar/gkv380
- Xia, J., Sinelnikov, I. V., and Wishart, D. S. (2011). MetATT: a web-based metabolomics tool for analyzing time-series and two-factor datasets. *Bioinformatics* 27, 2455–2456. doi:10.1093/bioinformatics/btr392
- Yang, Y., Engin, L., Wurtele, E. S., Cruz-Neira, C., and Dickerson, J. A. (2005). Integration of metabolic networks and gene expression in virtual reality. *Bioinformatics* 21, 3645–3650. doi:10.1093/bioinformatics/bti581
- Yu, T., Park, Y., Johnson, J. M., and Jones, D. P. (2009). apLCMS – adaptive processing of high-resolution LC/MS data. *Bioinformatics* 25, 1930–1936. doi:10.1093/bioinformatics/btp291
- Zhang, K., Bhuiya, M. W., Pazo, J. R., Miao, Y., Kim, H., Ralph, J., et al. (2012). An engineered monolignol 4-O-methyltransferase depresses lignin biosynthesis and confers novel metabolic capability in *Arabidopsis*. *Plant Cell* 24, 3135–3152. doi:10.1105/tpc.112.101287

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 Hill, Czauderna, Klapperstück, Roessner and Schreiber. 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.



Photorespiratory Bypasses Lead to Increased Growth in *Arabidopsis thaliana*: Are Predictions Consistent with Experimental Evidence?

Georg Basler^{1,2†}, Anika Küken^{3†}, Alisdair R. Fernie⁴ and Zoran Nikoloski^{3*}

¹ Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, CA, USA, ² Department of Environmental Protection, Estación Experimental del Zaidín CSIC, Granada, Spain, ³ Systems Biology and Mathematical Modeling Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, ⁴ Central Metabolism Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

OPEN ACCESS

Edited by:

Patrik R. Jones,
Imperial College London, UK

Reviewed by:

Biswapriya Biswas Misra,
University of Florida, USA
Dong-Yup Lee,
National University of Singapore,
Singapore

*Correspondence:

Zoran Nikoloski
nikoloski@mpimp-golm.mpg.de

[†]Georg Basler and Anika Küken
contributed equally.

Specialty section:

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

Received: 20 October 2015

Accepted: 24 March 2016

Published: 07 April 2016

Citation:

Basler G, Küken A, Fernie AR and
Nikoloski Z (2016) Photorespiratory
Bypasses Lead to Increased Growth
in *Arabidopsis thaliana*: Are
Predictions Consistent with
Experimental Evidence?
Front. Bioeng. Biotechnol. 4:31.
doi: 10.3389/fbioe.2016.00031

Arguably, the biggest challenge of modern plant systems biology lies in predicting the performance of plant species, and crops in particular, upon different intracellular and external perturbations. Recently, an increased growth of *Arabidopsis thaliana* plants was achieved by introducing two different photorespiratory bypasses via metabolic engineering. Here, we investigate the extent to which these findings match the predictions from constraint-based modeling. To determine the effect of the employed metabolic network model on the predictions, we perform a comparative analysis involving three state-of-the-art metabolic reconstructions of *A. thaliana*. In addition, we investigate three scenarios with respect to experimental findings on the ratios of the carboxylation and oxygenation reactions of Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). We demonstrate that the condition-dependent growth phenotypes of one of the engineered bypasses can be qualitatively reproduced by each reconstruction, particularly upon considering the additional constraints with respect to the ratio of fluxes for the RuBisCO reactions. Moreover, our results lend support for the hypothesis of a reduced photorespiration in the engineered plants, and indicate that specific changes in CO₂ exchange as well as in the proxies for co-factor turnover are associated with the predicted growth increase in the engineered plants. We discuss our findings with respect to the structure of the used models, the modeling approaches taken, and the available experimental evidence. Our study sets the ground for investigating other strategies for increase of plant biomass by insertion of synthetic reactions.

Keywords: flux balance analysis, *Arabidopsis thaliana*, photorespiration, metabolic bypasses, metabolic engineering, crop optimization

INTRODUCTION

The investigation and understanding of cell metabolism has recently experienced a paradigm shift largely propelled by the development of high-throughput methods. As a result, the classical pathway-centered approach has given way to a network-driven perspective, which considers the entire set of characterized biochemical reactions. This has led to the construction of genome-scale

metabolic models (GEMs) for organisms from each of the three domains of life: archaea, bacteria, and eukarya (Schellenberger et al., 2010). While a GEM constitutes an organized and comprehensive system of knowledge about an organism, it also allows *in silico* analyses based on constraint-based methods, relying on the corresponding stoichiometric matrix representation and assumptions about cellular metabolism (e.g., operability in a steady state and reversibility of reactions). Flux balance analysis (FBA) has provided the basic framework for predicting growth and biomass yield as well as investigating reaction fluxes in a metabolic network (Varma and Palsson, 1993a,b; Lewis et al., 2012). Approaches in this, so-called, constraint-based framework usually invoke the steady-state assumption, whereby there is no change in the size of the metabolic pools. Therefore, without additional assumptions, constraint-based modeling framework does not account for changes over time [e.g., due to circadian rhythm, which in plants necessitates the switch between autotrophic and heterotrophic metabolism (Cheung et al., 2014)]. Extensions of this framework have subsequently allowed systematic investigations of network modifications (e.g., deletion and insertion of reactions from other species or under- and over-expression of gene products) directed at enhancing particular metabolic functions (Burgard et al., 2003; Pharkya et al., 2004; Ranganathan et al., 2010; Yang et al., 2011; Larhlimi et al., 2012; Lakshmanan et al., 2013).

A metabolic network that includes all known biochemical reactions of an organism may not be realistic in a particular cellular scenario (i.e., context), as mounting evidence shows that cells adapt their metabolism to the prevailing circumstance (e.g., external environment, developmental stage, cell type in multicellular organisms). In different cellular contexts, only a subset of reactions is typically active, which may lead to differences in biomass composition (Chang et al., 2011; Arnold and Nikoloski, 2014). Therefore, the shift toward reconstructing context-specific models of cell metabolism has become necessary to provide more accurate and biologically meaningful insights (Bordbar et al., 2014; Machado and Herrgård, 2014; Robaina Estévez and Nikoloski, 2014). This is of particular importance when tackling the physiology of multicellular organisms, not only to better understand tissue- or cell-specific metabolism, but also as a first step to reconstruct the metabolic network of an entire plant, whereby multiple specialized models are mutually interconnected (de Oliveira Dal'Molin et al., 2015).

Despite these recent developments, it remains questionable to what extent constraint-based methods and large-scale modeling can be used for devising metabolic engineering strategies in plants. The ideal approach to test whether or not modeling based on existing plant GEMs can provide insights into crop optimization and predict novel optimization strategies is to investigate the extent to which the predictions match observations from metabolic engineering experiments. Since photorespiratory bypasses have recently been successfully engineered and tested in the model plant *Arabidopsis* (Peterhansel et al., 2013b), they can be used to evaluate the capability of existing models to correctly predict novel synthetic engineering strategies. We, therefore, focus on the analysis of plant photorespiratory metabolism in the metabolic network context.

Photorespiratory metabolism and its experimentally investigated bypasses represent an excellent test case to investigate the potential of plant GEMs in plant metabolic engineering for two reasons. First, the existing GEMs in the model C_3 plant *Arabidopsis thaliana* (and other photosynthetic organisms) include almost all components of the photorespiratory metabolism (Arnold and Nikoloski, 2013); therefore, they provide a suitable starting point for investigating the role of photorespiration in the network context. Second, photorespiratory metabolism does not operate in isolation, but shapes the energetics of photosynthesis, compartmental reductant exchange, nitrate assimilation, one-carbon (C_1) metabolism, and redox signal transduction [for recent reviews, see Foyer et al. (2009) and Bauwe et al. (2012)]; therefore, investigating the effects of modulating this pathway necessitates adoption of a network perspective.

The plant photorespiration pathway involves 12 reactions which are partitioned among three compartments, namely chloroplast, peroxisome, and mitochondrion, with bypasses in the cytosol (**Figure 1**) (Timm et al., 2008). Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) reacts with ribulose-1,5-bisphosphate (RuBP), resulting in the enediol-enzyme complex that can in turn react with oxygen (O_2), termed *oxygenation* (Calvin, 1954; Ogren and Bowes, 1971), or carbon dioxide (CO_2), termed *carboxylation*. Upon carboxylation, one molecule of RuBP is transformed into two molecules of 3-phosphoglycerate. The oxygenation of RuBP yields one molecule of 3-phosphoglycerate and one molecule of phosphoglycolate. To mitigate the inhibition of the photosynthetic pathway by phosphoglycolate (Peterhansel et al., 2013b), this compound is recycled in a series of reactions comprising the photorespiratory pathway. For completeness, **Table 1** includes the enzyme names, Enzyme Commission (EC) numbers, the corresponding biochemical reactions together with their compartmentation and reversibility, obtained from AraCyc 9.0 (Mueller et al., 2003). From the considered reactions, photorespiration can be regarded as uptake of O_2 and phosphoglycolate-related evolution of CO_2 and ammonia (NH_3) (effectively refixed by the glutamine synthetase and the glutamate synthase).

Three photorespiratory bypasses have been experimentally investigated to date, recently reviewed for their benefits and energy balances (Peterhansel et al., 2013a,b). The bypasses of Kebeish et al. (2007) and Maier et al. (2012) were tested in *A. thaliana*, while the bypass of Carvalho et al. (2011) was examined in tobacco. Since GEMs for tobacco have not been assembled to date, we focus on investigating the first two bypasses (see **Table 2** for the corresponding lists of reactions).

The “Kebeish bypass” starts with glycolate and ends with glycerate by considering three reactions in the chloroplast, namely, glycolate dehydrogenase (EC 1.1.99.14), tartronate semialdehyde carboxylase (EC 4.1.1.47), taking glyoxylate and producing CO_2 and 2-hydroxy-3-oxopropionate, as well as 2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60), transforming 2-hydroxy-3-oxopropionate (and NADH, NADPH, and H^+) into glycerate (and NAD, NADP) (**Figure 1**). The “Maier bypass” consists of a complete glycolate catabolic cycle, including glycolate oxidase (EC 1.1.3.15), malate synthase (EC 2.3.3.9), transforming glyoxylate (and Acetyl-CoA, H_2O) into

TABLE 1 | Enzymes and reactions of the photorespiration including different compartmentation for *A. thaliana* (Foyer et al., 2009).

Enzyme name	EC	Enzymatic reaction	Compartment
Ribulose-1,5-bisphosphate carboxylase/oxygenase	4.1.1.39	$\text{RuBP} + \text{CO}_2 \rightarrow 2 \text{ 3PGA}$ $\text{RuBP} + \text{O}_2 \rightarrow 3\text{PGA} + 2\text{PG}$	(h)
Phosphoglycolate Phosphatase	3.1.3.18	$2\text{PG} + \text{H}_2\text{O} \rightarrow \text{GLC} + \text{Pi}$	(h)
Glycolate oxidase	1.1.3.15	$\text{GLC} + \text{O}_2 \rightarrow \text{GOX} + \text{H}_2\text{O}_2$	(p)
Catalase	1.11.1.6	$2 \text{H}_2\text{O}_2 \rightarrow 2 \text{H}_2\text{O} + \text{O}_2$	(p)
Glycine transaminase	2.6.1.4	$\text{Glu} + \text{GOX} \rightarrow \text{Gly} + \text{AKG}$	(p)
Glycine dehydrogenase (decarboxylating)	1.4.4.2	$\text{Gly} + \text{LPL} \rightarrow \text{amDHP} + \text{CO}_2$	(m)
Aminomethyltransferase	2.1.2.10	$\text{amDHP} + \text{THF} \rightarrow \text{DHP} + \text{M-THF} + \text{NH}_3$	(m)
Dihydropyridine dehydrogenase	1.8.1.4	$\text{DHP} + \text{NAD} \leftrightarrow \text{LPL} + \text{NADH}$	(m)
Glycine hydroxymethyltransferase	2.1.2.1	$\text{Gly} + \text{M-THF} + \text{H}_2\text{O} \leftrightarrow \text{Ser} + \text{THF}$	(m)
Serine-glyoxylate transaminase	2.6.1.45	$\text{Ser} + \text{GOX} \rightarrow \text{HPA} + \text{Gly}$	(p)
Glycerate dehydrogenase	1.1.1.29	$\text{HPA} + \text{NADH} \rightarrow \text{GA} + \text{NAD}$	(p)
Hydroxypyruvate reductase	1.1.1.81	$\text{HPA} + \text{NAD(P)H} \rightarrow \text{GA} + \text{NAD(P)}$	(p)
Glycerate kinase	2.7.1.31	$\text{GA} + \text{ATP} \rightarrow 3\text{PGA} + \text{ADP}$	(h)
Glutamine synthetase	6.3.1.2	$\text{Glu} + \text{NH}_3 + \text{ATP} \leftrightarrow \text{Gln} + \text{ADP} + \text{Pi}$	(h)
Glutamate synthase (ferredoxin dependent)	1.4.7.1	$2 \text{Glu} + \text{Fd}_{\text{ox}} \leftrightarrow \text{Gln} + \text{AKG} + \text{Fd}_{\text{red}} + 2 \text{H}^+$	(h,m)
Glutamate synthase (NADH-dependent)	1.4.1.14	$2 \text{Glu} + \text{NAD} \leftrightarrow \text{Gln} + \text{AKG} + \text{NADH} + 2 \text{H}^+$	(h,m)
NADP-malic enzyme	1.1.1.40	$\text{Mal} + \text{NADP} \rightarrow \text{Pyr} + \text{NADPH} + \text{CO}_2$	(h)
Pyruvate dehydrogenase	1.2.4.1	$\text{Pyr} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{NADH} + \text{H}^+ + \text{CO}_2$	(h)

The abbreviations of the compartment names in which the corresponding reaction takes place are as follows: (h), chloroplast; (m), mitochondrion; (p), peroxisome. The compound abbreviations are as follows: carbon dioxide (CO_2), oxygen (O_2), ribulose-1,5-bisphosphate (RuBP), 3-phosphoglycerate (3PGA), phosphoglycolate (2PG), glycolate (GLC), glyoxylate (GOX), hydrogen peroxide (H_2O_2), glutamate (Glu), glycine (Gly), α -ketoglutarate (AKG), serine (Ser), ammonia (NH_3), 3-hydroxypyruvate (HPA), glycerate (GA), glutamine (Gln), water (H_2O), phosphate (Pi), lipoylprotein (LPL), aminomethylidihydropyridine (amDHP), tetrahydrofolate (THF), 5,10-methylenetetrahydrofolate (M-THF), dihydropyridine (DHP), oxidized nicotinamide adenine dinucleotide (NAD), reduced nicotinamide adenine dinucleotide (NADH), adenosine diphosphate (ADP), adenosine triphosphate (ATP), ferredoxin (Fd). While EC 1.1.1.81 also operates in the cytosol (Timm et al., 2008), this function is included only in the peroxisome of the considered models; hence, the compartment given for this enzyme is (p).

TABLE 2 | Enzymes and reactions of the photorespiratory bypasses of Kebeish et al. (2007) and Maier et al. (2012) in the chloroplast.

Enzyme name	EC	Enzymatic reaction
Kebeish bypass		
Glycolate dehydrogenase	1.1.99.14	$\text{GLC} + \text{NAD} \rightarrow \text{GOX} + \text{NADH}$
Tartronate semialdehyde carboxylase	4.1.1.47	$2 \text{GOX} \rightarrow \text{HOP} + \text{CO}_2$
2-hydroxy-3-oxopropionate reductase	1.1.1.60	$\text{HOP} + \text{NADH} \rightarrow \text{GA} + \text{NAD}$
Maier bypass		
Glycolate oxidase	1.1.3.15	$\text{GLC} + \text{O}_2 \rightarrow \text{GOX} + \text{H}_2\text{O}_2$
Malate synthase	2.3.3.9	$\text{GOX} + \text{H}_2\text{O} + \text{Acetyl-CoA} \rightarrow \text{Mal} + \text{CoA}$
Catalase	1.11.1.16	$2 \text{H}_2\text{O}_2 \rightarrow \text{O}_2 + 2 \text{H}_2\text{O}$

Carbon dioxide (CO_2), oxygen (O_2), glycolate (GLC), glyoxylate (GOX), hydrogen peroxide (H_2O_2), glycerate (GA), 2-hydroxy-3-oxopropionate (HOP), malate (Mal), pyruvate (Pyr), coenzyme A (CoA), acetyl-coenzyme A (acetyl-CoA), oxidized nicotinamide adenine dinucleotide (NAD), and reduced nicotinamide adenine dinucleotide (NADH).

MATERIALS AND METHODS

Models and Bypasses

The analysis is based on three large-scale models of *A. thaliana* metabolism: (1) a bottom-up (i.e., experiment-driven) reconstruction, where the operability of the incorporated reactions and metabolites is ensured by starting from well-documented and

necessary biochemical pathways (Arnold and Nikoloski, 2014), (2) AraGEM, the first Arabidopsis genome-scale metabolic network, including primary metabolism of compartmentalized plant cells (de Oliveira Dal'Molin et al., 2010), and (3) a larger model that includes both primary and secondary metabolisms (Mintz-Oron et al., 2012). While the last two models come with a single biomass reaction, the smaller model of Arnold and Nikoloski considers three biomass reactions corresponding to biomass compositions that pertain to realistic and frequently examined scenarios: carbon-limiting, nitrogen-limiting, and optimal growth conditions (Kleessen et al., 2012). These biomass reactions were assembled by considering the composition of 1 g dry weight of Arabidopsis leaf under the respective conditions: growth under optimal condition reflects only, if any, light limitation under autotrophic conditions; nitrogen-limitation is based on a protocol that results in a mild but sustained restriction of growth upon restriction of nitrogen availability, while the carbon limitation is experimentally realized via short-day conditions (8:16 light–dark cycle) (Arnold and Nikoloski, 2014). The functional differences of these models have already been compared with respect to the ability to simulate photoautotrophic growth, number of blocked reactions, and flux coupling of reactions (Arnold and Nikoloski, 2014). This systematic comparative analysis indicated the suitability and added value of the bottom-up reconstruction, in which every reaction can carry flux, despite the greater degree of flux coupling, and, hence, more constrained flux space.

The photorespiratory bypasses were introduced by adding the corresponding reactions to the chloroplast of the three

models: glycolate dehydrogenase, tartronate semialdehyde carboxylase, and 2-hydroxy-3-oxopropionate reductase for the Kebeish bypass; glycolate oxidase, malate synthase, and catalase for the Maier bypass. Since AraGEM does not contain glyoxylate or hydrogen peroxide in chloroplasts, we introduced the bypasses using the corresponding compounds from the cytosol (which is equivalent to introducing a transport reaction for allowing the exchange of these metabolites between the cytosol and chloroplast). The same approach was used for 2-hydroxy-3-oxopropionate in the model by Mintz-Oron et al., which contains this metabolite only in the cytoplasm. Moreover, AraGEM and the model of Arnold and Nikoloski do not contain 2-hydroxy-3-oxopropionate in any compartment. We, therefore, added the metabolite to the chloroplasts to allow introducing the Kebeish bypass. Note that this leads to a full coupling (i.e., fixed flux ratios) between the tartronate semialdehyde carboxylase and 2-hydroxy-3-oxopropionate reductase reactions in these models, since 2-hydroxy-3-oxopropionate can only be produced and consumed by these reactions.

Flux Balance Analysis

Constraint-based modeling investigates the solution space of feasible flux distributions $v \in \mathbb{R}^n$ for n biochemical reactions in a metabolic network that is assumed to operate in a (quasi) steady-state, i.e., the concentrations $x \in \mathbb{R}^m$ of the m metabolites in the network are constant (Varma and Palsson, 1993a,b; Bordbar et al., 2014). The steady-state condition can be written as $\frac{dx}{dt} = Sv = 0$, whereby the stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ captures the stoichiometry of all reactions. Upper and lower boundaries of the flux vector $v_{\min} \leq v \leq v_{\max}$ further constrain the solution space and are used to model physiologically relevant scenarios (e.g., limit on nutrient import, reaction reversibility, and environmental conditions). The steady-state condition and the flux boundaries determine the solution space that usually contains infinitely many flux distributions, since the system of linear equations $Sv = 0$ is, in practice, underdetermined. To probe the functionality of the network, FBA (Varma and Palsson, 1993a,b; Bordbar et al., 2014) assumes that metabolic behavior is guided by some optimization principles (e.g., optimal biomass yield). Here, the objective is to maximize the flux through the biomass reaction, v_{biomass} , and the resulting optimization problem yields a linear program:

$$\begin{aligned} \max z &= v_{\text{biomass}} \\ \text{s.t.} \\ Sv &= 0 \\ v_{\min} &\leq v \leq v_{\max}. \end{aligned} \quad (1)$$

Flux Variability Analysis

The solution of the linear programming problem in Eq. 1, above, is the maximum flux value of the biomass reaction, denoted by z^* . FVA allows determining the minimum and the maximum value of flux that a given reaction can carry while ensuring maximum flux through the biomass reaction. These values can

be obtained by solving the following linear program for a given reaction i :

$$\begin{aligned} \max(\min) v_i \\ \text{s.t.} \\ Sv &= 0 \\ v_{\min} &\leq v \leq v_{\max} \\ v_{\text{biomass}} &= z^*. \end{aligned} \quad (2)$$

Flux-Sum

At steady state, the net rate for metabolite consumption and production is zero, but its turnover is not. The flux-sum ϕ_i of internal metabolite i , used as a proxy for metabolite turnover, is defined as sum over all reaction fluxes v_j around metabolite i , $\phi_i = 0.5 \sum_j |S_{ij} v_j|$ (Chung and Lee, 2009). The flux-sum at the optimum biomass z^* from Eq. 1 is determined by the linear program:

$$\begin{aligned} \max(\min) 0.5 * \sum_{j=1}^n |S_{ij} v_j| \\ \text{s.t.} \\ Sv &= 0 \\ v_{\min} &\leq v \leq v_{\max} \\ v_{\text{biomass}} &= z^*. \end{aligned} \quad (3)$$

Implementation

The models were obtained from the respective publications in which they were first analyzed. All analyses were carried out with the help of the optimization platform TOMLAB version 8.1 using the CPLEX solver with default parameter for MATLAB R2015a. All used models with the irreversible variants of the introduced bypasses are provided as a Datasheet S1 in Supplementary Material. The flux-sum was implemented with TOMLAB version 8.1 using the SNOPT solver and default parameters.

RESULTS

Our analysis includes three scenarios to compare and contrast the predictions from the models of *A. thaliana* metabolism with respect to actual biomass increase upon insertion of the two photosynthetic bypasses. In Scenario A, we used the models with the default flux boundaries and employed FBA to determine the extent of the increase in the optimal flux through the biomass reaction upon introduction of the bypass reactions (see Materials and Methods). The predictions from Scenario A may not be realistic, since the optimal biomass may not correspond to a flux distribution in which the ratio of the fluxes of the carboxylation and the oxygenation reactions catalyzed by RuBisCO is physiologically plausible. To this end, we considered Scenarios B and C whereby the optimal flux through the biomass reaction is determined with the additional constraint that $v_{\text{carb}} = \epsilon v_{\text{oxy}}$, with $\epsilon = 4$ and $\epsilon = 1.5$, respectively, for those cases from Scenario A where increase in biomass was predicted.

The values for the ratios were selected to match observations about the range for the rate of photorespiration estimated either from labeling experiments or from gas exchange data (Sharkey, 1988; Szecowka et al., 2013; Ma et al., 2014; Heise et al., 2015).

Since AraGEM and the model of Arnold and Nikoloski already include the glycolate oxidase, we introduced only the malate synthase and the catalase reaction in the chloroplast to simulate the Maier bypass. For both bypasses, we determined and compared the maximum biomass yield, obtained by FBA, with and without each of the bypasses.

In addition, by applying FVA (see Materials and Methods), we determined the interval of flux values that the reactions in the photorespiratory pathway (Table 1) can take at the optimum for each of the three scenarios. Comparison of the intervals allowed us to test the hypothesis that upon insertion of the bypass the photorespiratory flux is reduced, but not completely diverted, in comparison to the case without bypass. Validating the hypothesis would imply that the predictions from the large-scale models are in line with experimental observations suggesting a two- to fivefold reduction of photorespiratory flux (Peterhansel et al., 2013a) upon insertion of the considered bypasses. The results obtained from the three models under the different scenarios are summarized in Table 3.

Scenario A – No Constraints on the Flux Ratio of Carboxylation and Oxygenation Reactions

We first considered the core model of *A. thaliana* with the three biomass reactions for photoautotrophic growth: carbon-limiting, nitrogen-limiting, and optimal growth conditions (Arnold and

Nikoloski, 2014). The carbon-limiting scenario pertains to short-day conditions, where energy efficiency of carbon fixation may be more limiting for growth than under long-day conditions. Incidentally, this is the condition under which the advantage of both investigated bypasses was detectable (Peterhansel et al., 2013b).

Under the assumption that all of the inserted reactions are irreversible, our results showed that no biomass increase could be predicted for any of the considered environments following the insertion of the Maier bypass. By contrast, by allowing the malate synthase to act as a reversible reaction, the model predicted a very small increase (< 0.03%) over the three growth conditions (Table S1 in Supplementary Material). Reversibility of the other two reactions (glycolate oxidase and catalase) does not affect the predicted biomass yield. This finding implies that the insertion of the Maier bypass in the model of Arnold and Nikoloski had negligible effect on biomass increase. The Kebeish bypass with any configuration of reversibilities of three individual reactions did not result in increase of biomass, similar to the Maier bypass scenario above (see Table S4 in Supplementary Material, increase <0.02%).

The AraGEM model with all irreversible reactions for the Maier bypass did not result in an increase in biomass. However, the insertion of all reversible reactions led to an increase of 0.6% for the single biomass reaction that the model includes. The FVA indicates that the malate synthase carries a negative (net) flux (Table S5 in Supplementary Material). The insertion of the Kebeish bypass resulted in an increase of 0.3%, see Table S6 in Supplementary Material. However, both results pertain to the unrealistic scenario in which the RuBisCO oxygenase practically does not carry flux, which was revealed by the FVA.

TABLE 3 | Predicted percentage increase in biomass yield of the three models when introducing the Maier and Kebeish bypasses under the three analyzed scenarios.

			Arnold and Nikoloski			AraGEM	Mintz-Oron
			optimal growth	C-limiting	N-limiting		
Scenario A	Maier bypass	irrev	0	0	0	0	1.0
		MS rev	0	0	0	0.6	1.0
		rev	0	0	0	0.6	1.0
	Kebeish bypass	irrev	0	0	0	0.3	0.4
		TS rev	0	0	0	0.3	0.4
		rev	0	0	0	0.3	0.8
Scenario B	Maier bypass	irrev	0	0	0	0	1.0
		MS rev	0	0	0	2.7	1.0
		Rev	0	0	0	2.7	1.0
	Kebeish bypass	irrev	6.2	6.2	6.3	1.1	0.4
		TS rev	6.2	6.2	6.3	1.1	0.4
		rev	6.2	6.2	6.3	1.1	0.8
Scenario C	Maier bypass	irrev	0	0	0.1	0	1.0
		MS rev	0	0	0.1	1.6	1.0
		Rev	0	0	0.1	1.6	1.0
	Kebeish bypass	irrev	3.0	3.0	4.0	0.1	0.4
		TS rev	3.0	3.0	4.0	0.1	0.4
		rev	3.0	3.0	4.0	0.1	0.8

irrev indicates that all reactions of the bypass are considered irreversible; *MS rev* and *TS rev* indicate that only malate synthase and tartronate semialdehyde carboxylase, respectively, are considered reversible; *rev* indicates that all reactions of the bypass are considered reversible. This includes each configuration of reaction reversibilities, which leads to an increase in predicted growth with respect to a more constrained configuration.

We next investigated the larger model of Mintz-Oron et al., which also includes the glycolate oxidase, but only in the cytoplasm, mitochondria, and peroxisome. Upon introducing the Maier bypass, the predicted increase in the optimal biomass yield was 1.0% with every combination of reaction reversibilities. By inspecting the results from FVA, we found that the increase in biomass yield was associated, as one would anticipate, to larger flux ranges. The only reaction that had a fixed value at the optimum biomass with and without inclusion of the bypass was glutamine synthetase; its flux was more than fivefold reduced, lending support for the hypothesis of reduced photorespiration by Peterhansel et al. (2013a). These findings were invariant over all combinations of reaction reversibilities (Table S2 in Supplementary Material).

Inserting the Kebeish bypass into the model of Mintz-Oron et al. resulted in 0.4% increase of biomass with all reactions considered irreversible. Allowing for reversibility of tartronate semialdehyde carboxylase and the other two reactions irreversible resulted in a maximum increase of 0.8%. Glutamine synthetase was, again, the only reaction that had a fixed value at the optimum biomass with and without inclusion of the bypass; its flux was reduced by more than 1.2-fold (Table S3 in Supplementary Material).

Scenario B – Constrained Flux Ratio of Carboxylation and Oxygenation Reactions (3:2)

In this section, we repeated the analysis under the additional assumption that the flux ratio of RuBisCO carboxylase and oxygenase reactions was fixed to 3:2. The Maier bypass in the model of Arnold and Nikoloski did not predict an increase in biomass in any of the three environments. On the other hand, the Kebeish bypass resulted in a predicted biomass increase of 6.2% for optimal, 6.2% for carbon-limited, and 6.3% for nitrogen-limited growth conditions. We also inspected the change in flux variability ranges in the model upon insertion of the Kebeish bypass. As in Scenario A, the fluxes of the reactions in the photorespiratory pathway at the optimal biomass yield with the additional ratio constraint vary in a small range. Upon insertion of the bypass, the upper bounds for the photorespiratory flux were lowered by ~10% (Table S4 in Supplementary Material).

Scenario B with AraGEM resulted in biomass increase of 2.7% for the Maier bypass with all reversible reactions. Here too, we found that the increase in biomass yield was associated with a negative net flux for the malate synthase reaction (Table S5 in Supplementary Material). The increase for the Kebeish bypass was 1.1% with the three reversibilities of reactions in Table 3.

Scenario B with the model of Mintz-Oron et al. resulted in the same increase in biomass yield as in Scenario A, above, for the Maier bypass. Here too, we found that the increase in biomass yield was associated with larger flux ranges. The reaction catalyzed by glutamine synthetase was the only which showed a fixed value at the optimum biomass with and without inclusion of the bypass; its net flux shifted in the direction of synthesizing glutamine (more than fivefold decrease). As in Scenario A, these findings were invariant over all combinations of reaction reversibilities (Table S2 in Supplementary Material). The results

for the Kebeish bypass were largely invariant in comparison to that of Scenario A (Table S3 in Supplementary Material).

Scenario C – Constrained Flux Ratio of Carboxylation and Oxygenation Reactions (4:1)

In this section, we repeated the analysis under the assumption that the flux ratio of RuBisCO carboxylase and oxygenase reactions was fixed to 4:1. The model of Arnold and Nikoloski predicted an increase in biomass of 0.1% for all three reversibilities for the Maier bypass under nitrogen limiting growth condition. The Kebeish bypass resulted in an increase of 3.0% for optimal, 3.0% for carbon-limiting, and 3.9% for nitrogen-limiting growth conditions, accompanied by increase in flux variability bounds (Table S4 in Supplementary Material). AraGEM resulted in an increase of 0.1% with the Kebeish pathway and 1.6% for the Maier bypass, while the model of Mintz-Oron et al. resulted in the same predictions as in Scenario B for both bypasses (Tables S2, S3, S5, and S6 in Supplementary Material). As in Scenario B, above, glutamine synthetase for the model of Arnold and Nikoloski shows a constant value at the optimum biomass upon the insertion of the Kebeish pathway, which is at least 20% smaller than the values at the optimum biomass for the wild type.

Our earlier analysis demonstrated that consistent predictions can only be made for the Kebeish pathway with the three models. We also showed that this was the case for the model of Arnold and Nikoloski with biomass functions for three different scenarios. Since this model was reconstructed entirely from experimental data, lacks any blocked reactions, and gives the best quantitative match between the predictions and observations from the bypasses, the remaining analyses will consider only this model with and without the Kebeish pathway. Specifically, in the following we focus on (1) biomass increase and (2) variability of CO₂ uptake upon varying the carboxylation to oxygenation ratio as well as (3) the comparison of co-factor flux-sums (as proxies for turn-over) between the wild type and transformants (i.e., upon insertion of the bypass).

Biomass Increase at Varying Values for the Carboxylation to Oxygenation Ratio

Due to the observed differences in increase of biomass for the model of Arnold and Nikoloski upon insertion of the Kebeish bypass at two different fixed values of carboxylation to oxygenation ratios (i.e., 1.5 and 4), we expanded the range of investigated ratios to values between 1 and 100. This modeling scenario would account for differences in the internal CO₂ concentration, resulting from the bypass insertion and manifesting itself in increases of carboxylation rate (Kebeish et al., 2007; Peterhansel et al., 2013a). Considering the wild type (Figure 2A), we observed, expectedly, an increase in biomass with the increase in the ratio between the carboxylation and oxygenation. The largest increase in biomass (~8%) for the transformants is observed at the smallest tested value for the ratio of 1. In each case, the largest difference between the three biomass functions was observed for ratios from 2.5 to 5, with an increase for the nitrogen-limiting biomass that outperforms the increase for the optimal and the carbon-limiting

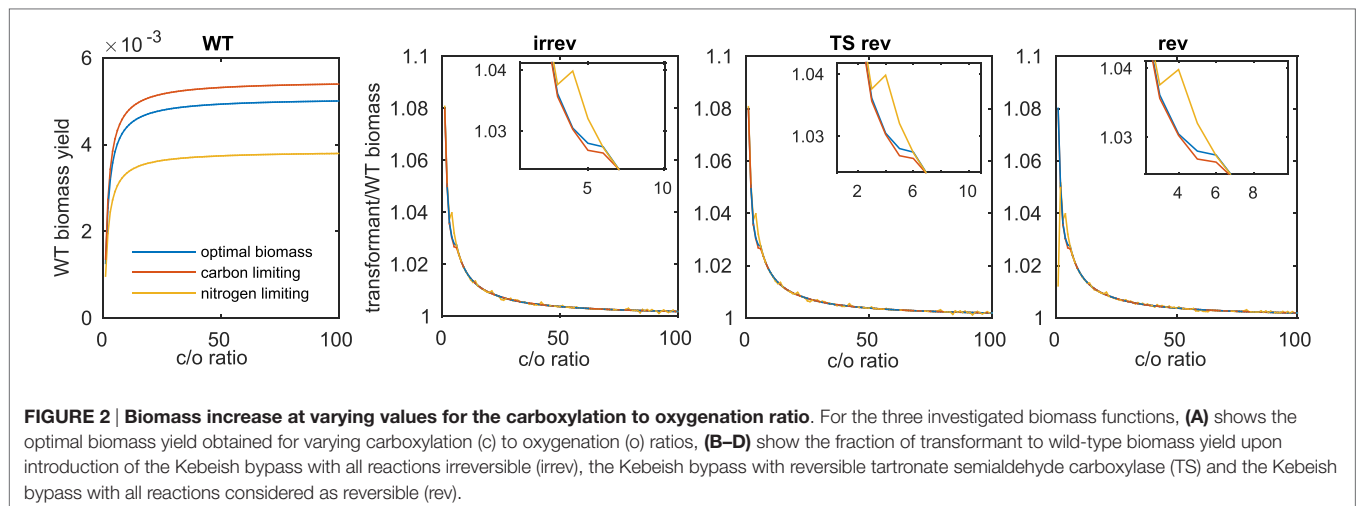


FIGURE 2 | Biomass increase at varying values for the carboxylation to oxygenation ratio. For the three investigated biomass functions, **(A)** shows the optimal biomass yield obtained for varying carboxylation (c) to oxygenation (o) ratios, **(B–D)** show the fraction of transformant to wild-type biomass yield upon introduction of the Kebeish bypass with all reactions irreversible (irrev), the Kebeish bypass with reversible tartronate semialdehyde carboxylase (TS) and the Kebeish bypass with all reactions considered as reversible (rev).

TABLE 4 | Minimum and maximum CO₂ uptake and exchange flux at optimum biomass for wild type and transformant (T), including the irreversible Kebeish pathway for carboxylation to oxygenation (c/o) ratios of 3:2 and 4:1.

Range		c/o ratio 3:2			c/o ratio 4:1		
		Wild type	Transformant	Flux change in T compared to wild type (%)	Wild type	Transformant	Flux change in T compared to wild type (%)
A	Min	41.78	44.38	6.22	73.79	76.03	3.05
	Max	41.78	44.38	6.22	73.84	76.03	2.98
B	Min	57.85	40.74	−29.57	78.08	73.03	−6.47
	Max	57.85	41.46	−28.33	80.20	73.03	−8.94
C	Min	−20.79	−2.10	−89.91	−13.55	−4.06	−70.06
	Max	−20.65	−1.24	−94.00	−12.21	−3.02	−75.28

A, flux range for CO₂ uptake from environment into cytosol; B, CO₂ exchange flux from cytosol to chloroplast; and C, ranges for CO₂ exchange flux from cytosol to the mitochondrion. Results are presented for the biomass function derived from the optimal growth conditions (Arnold and Nikoloski, 2014).

biomass functions. Interestingly, this region includes the previously tested value for the ratio of 4:1, which was derived from external gas exchange and labeling measurements [see the discussion in Heise et al. (2015)]. These observations held for the three cases of reversibilities considered in the Scenarios A–C, above (see Figures 2B–D and Table 3).

Biomass Increase Is Associated with Differences in CO₂ Uptake and Compartmentalization at Optimal Biomass Yields

In addition, we tested whether a higher CO₂ concentration in the chloroplast could explain the growth increase observed by introduction of the Kebeish bypass (Kebeish et al., 2007). To this end, we analyzed the flux variability of CO₂ import into the cytosol, the exchange of CO₂ between the cytosol and the chloroplast, as well as the exchange of CO₂ between the cytosol and the mitochondrion. We found that, upon insertion of the Kebeish bypass, the flux of CO₂ into the chloroplast is reduced by 30% under Scenario B (carboxylation/oxygenation ratio of 3:2) and 6 to 9% for Scenario C (carboxylation/oxygenation ratio of 4:1) with all reversibilities examined above (Table 4 and Table S7 in Supplementary Material). From these results, we concluded

that less CO₂ needs to be imported from the cytosol, likely due to a higher CO₂ concentration resulting from its release from the bypass. In line with a shift of CO₂ release from mitochondria to the chloroplast, we observed a decrease in CO₂ release from the mitochondrion into the cytosol. The uptake of CO₂ from the environment into the cytosol, however, was increased. The latter result is in accordance with the experimentally determined increase in the apparent rate of CO₂ assimilation in bypass transformants (Kebeish et al., 2007).

Biomass Increase Is Accompanied by Flux-Sum Differences in Co-Factors ATP, NADH, and NADPH

The flux-sum for a metabolite (see Materials and Methods) can be regarded as a measure for the overall flux through a metabolic pool at a feasible steady state (Chung and Lee, 2009). It has recently been applied in understanding changes in maize metabolism under different nitrogen conditions (Simons et al., 2014). Here, we were interested in comparing the range of the flux-sum values at the optimum biomass yield, while additionally enforcing either one of the previously considered values for the carboxylation-oxygenation ratios (3:2 and 4:1). We focused on interpreting the changes in compartment-specific flux-sum

differences in co-factors ATP, NADH, and NADPH only in the cases where the ranges of their flux-sum in the transformants do not overlap with the ranges of the flux-sum in the wild type. In other words, we interpreted a change in turn-over that was valid in every possible steady state (with the imposed constraints). Therefore, we can conclude that the observed increase in the biomass is clearly associated with the difference in flux-sum values for these metabolites.

More specifically, we obtained a constant increase in flux-sum of ATP in the chloroplast and mitochondrion across all biomass functions and for both carboxylation to oxygenation ratios (Table S8 in Supplementary Material). In fact, the flux-sum of the chloroplastic ATP increased by at least 7% in Scenario B and 3.5% in Scenario C (% is determined from the maximum flux-sum in the wild type and the minimum flux-sum in the transformant). For the mitochondrial ATP, the increase was 85% in Scenario B and 143% for the carboxylation to oxygenation ratio of 4:1. Moreover, we observed that in the transformants the flux-sum of the peroxisomal NADH is fixed to 0 at the optimal biomass. The latter implies that NADH is not consumed nor produced by any of the modeled reactions in the peroxisome. For the remaining metabolites, while we observe changes in the ranges of the flux-sum, they overlap between the transformant and the wild type, and hence no conclusive statement can be made (Töpfer et al., 2015).

DISCUSSION

The two photorespiratory bypasses considered in our study have recently been investigated by Xin et al. (2015) through kinetic modeling by considering extensions to the model of Zhu et al. (2007). However, this approach cannot be employed to provide predictions about biomass yield, since the kinetic model only considered the core of plant carbon metabolism and, thus, lacks important biomass precursors, such as amino acids and cell wall components. Importantly, the kinetic modeling approach was based on the assumption that the photorespiratory pathway is not operational upon introduction of a bypass pathway, which has not been experimentally confirmed and, thus, precludes the possibility of testing the hypothesis of a reduced photorespiration. With respect to the rate of photosynthesis, the kinetic modeling predicted an increase following insertion of the Kebeish bypass but a decrease following insertion of the Maier bypass. However, gas exchange measurements indicated that the Maier bypass enhances the photosynthetic rate when calculated per mol chlorophyll (Maier et al., 2012). In addition, the predictions based on kinetic modeling depend on the type of enzyme kinetics, the values of the respective parameters, reversibility, and number and type of effectors. These are usually unknown, but will affect the predictions from the insertion of the bypasses and the reference state (i.e., fluxes and concentrations) used in the comparison.

In contrast to kinetic modeling, the constraint-based approach relies only on the assembled stoichiometry together with the assumption that the organism optimizes a particular objective – here biomass yield. While constraint-based modeling cannot make predictions about concentrations of intermediates without additional assumptions (Töpfer et al., 2015), using this approach

one can make predictions of higher level phenotypes (e.g., biomass yield), as well as the fluxes associated with these phenotypes, based only on the metabolic reactions. To mitigate the effect of the model size and different reconstruction strategies on the predictions, we conducted a comparative analysis involving three state-of-the-art metabolic reconstructions of *A. thaliana*. We performed comparison with respect to increase in biomass yield and difference in the flux ranges for reactions involved in the photorespiratory pathway (Table 1) upon insertion of the bypasses. Moreover, unlike the kinetic modeling approach by Xin et al. (2015), we did not block photorespiratory flux, but instead left its flux unconstrained, which allowed predicting the effect of the bypasses on photorespiration with respect to optimal biomass yield. For instance, in Scenarios A and B for the model of Mintz-Oron et al., we found that the flux through the glutamine synthetase (which was constant at the optimum biomass with and without the bypass) was fivefold reduced upon introduction of the Maier bypass. Moreover, in Scenario B for the model of Arnold and Nikoloski, we found that the introduction of the Kebeish bypass lowered the upper bounds for the photorespiratory flux by 10%. Therefore, this constraint-based modeling approach allowed us to arrive at the prediction that the reduction of flux through the photorespiratory pathway, suggested by the experiments, indeed, holds.

Our results demonstrated that the constraint-based modeling approach predicts an increase in biomass upon insertion of the bypasses without the need for extensive model parameterization. Furthermore, this approach allowed us to test the effects of all possibilities for reaction reversibility in the bypass and the reactions involved in the photorespiratory pathway. However, we concluded that the qualitative match (i.e., increase in biomass yield of up to 6.22%) between the predictions and experimental evidence can be obtained only by further constraining the optimal states using experimental observations (namely, the ratio of RuBisCO carboxylation to oxygenation rates). Therefore, our findings indicate that predictions of metabolic engineering strategies are tightly bound to the reference state used, and bring to question approaches which attempt pathway engineering without considering reaction rates in the rest of the network.

In addition, our approach of simultaneous usage of *A. thaliana* metabolic models with different characteristics allows us to test the extent to which the metabolic engineering predictions may differ due to the intrinsic differences between the models. For instance, in the model of Arnold and Nikoloski, all reactions involving metabolic conversions (i.e., all reactions except those involved in maintenance, transport, import, and export processes, as well as biomass production) are annotated and there is experimental evidence for their occurrence in *A. thaliana*. By contrast, such evidence is missing for 21 and 37% of the annotated reactions in AraGEM and the model of Mintz-Oron et al. as a result of their (semi-)automated reconstruction using reaction databases. Moreover, and most importantly, the model of Arnold and Nikoloski alongside AraGEM does not include blocked reactions (i.e., reactions not carrying flux in any steady state). The percentage of blocked reactions in the model of Mintz-Oron et al. is 59%. Due to the inclusion of fewer reactions, the model of Arnold and Nikoloski is, however, less flexible (i.e., has

smaller flux variability ranges and higher coupling of reactions) than AraGEM and the model of Mintz-Oron et al. In addition, the model of Arnold and Nikoloski predicts more efficient conversion of CO₂ into biomass than AraGEM, at the cost of assimilating more photons. These increased levels of realism, discussed in detail in Arnold and Nikoloski (2014), may explain the better quantitative match between the predictions from the model of Arnold and Nikoloski upon placing physiologically meaningful constraints on the RuBisCO catalyzed reactions.

We attempted to link the increase in biomass to changes in several key molecular properties proposed to be associated with photorespiratory metabolism. To this end, we investigated the change in biomass increase as a function of an expanded range of values for the ratio of the RuBisCO carboxylation and oxygenation reactions. This allowed us to conclude that the increase in growth conferred by the bypasses may be due to the release of CO₂ in chloroplasts and the resulting increase in internal CO₂ concentration. This finding was further strengthened by simulating the changes in the CO₂ transport fluxes between compartments as well as the uptake of CO₂ from the environment, which is in line with the experimental evidence. Finally, we found changes in the flux-sum for ATP and NADH at optimum biomass, indicating that their pool sizes are associated with the predicted increase in biomass. Altogether, these molecular properties provide further evidence for the plausibility of the predictions from the constraint-based modeling of the photorespiratory bypasses in *Arabidopsis*.

Nevertheless, one question remains open: why is the predicted biomass yield much smaller than what was observed in experiments (e.g., ~17–20% for the Maier bypass [Maier et al., 2012])? The reasons may very well be related to the quality of the underlying models, the modeling strategy taken, and the available experimental evidence: the considered models already incorporate some of the reactions that are integral to the analyzed bypasses. For instance, the model of Arnold and Nikoloski includes a catalase in the chloroplast, which is deemed essential for growth (i.e., its knock-out, by restricting the flux to 0, results in zero biomass yield). Therefore, the effect of inserting the remaining reactions of the bypass may be masked. In addition, experimental evidence indicates that the reference point (i.e., enzyme activity, metabolite levels) leading to biomass values used for gaging the increase may not correspond to the absolute optimum (i.e., optimum with no additional constraints) (Maier et al., 2012). Hence, the predicted optimal growth from the original (unmodified) plant GEMs may overestimate the actual growth rate of the wild-type plants, therefore partially masking the beneficial effect of introducing the bypasses. We have partly

attempted to remedy the latter by enforcing particular flux ratios for the RuBisCO catalase and oxygenase; however, this is only one of many conceivable options. In addition, none of the experimental studies of the bypasses reported the growth rate of the plants, but instead the increase in biomass between wild type and the transgenic plant at a certain point of time. Therefore, the comparison between the growth rates (or yield) predicted by the constraint-based modeling framework cannot be directly compared with the end-point cumulative measurements from the experiments.

Further developments of this modeling framework may allow improvements in the accuracy of predictions by considering additional constraints from data (e.g., transcriptomics, proteomics, and metabolomics profiles) (Nikoloski et al., 2015). These could be complemented with easy-to-use computational tools for prediction and investigation of intervention strategies. Altogether, our study indicates that constraint-based modeling has indeed the potential to identify target reactions and pathways whose insertion could lead to increase in performance of plant species; however, our study indicates that the consideration of more physiologically meaningful constraints, rather than only simple optimization criteria, leads to predictions in line with existing experimental findings. Our study demonstrated that predictions across models of different sizes and levels of realism can provide additional reinforcement for the validity of plant-growth predictions from the constraint-based modeling framework.

AUTHOR CONTRIBUTIONS

GB and ZN designed the research. AK and GB performed the analyses. All authors interpreted the results and wrote the manuscript.

FUNDING

GB would like to acknowledge financial support by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme and the Max Kade Foundation. ZN and AF would like to acknowledge the financial support of PROMICS, Research Unit 1186 of the German Research Foundation. AK, AF, and ZN would like to thank the Max Planck Society for support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2016.00031>

REFERENCES

- Arnold, A., and Nikoloski, Z. (2013). Comprehensive classification and perspective for modelling photorespiratory metabolism. *Plant Biol. (Stuttg.)* 15, 667–675. doi:10.1111/j.1438-8677.2012.00708.x
- Arnold, A., and Nikoloski, Z. (2014). Bottom-up metabolic reconstruction of *Arabidopsis* and its application to determining the metabolic costs of enzyme production. *Plant Physiol.* 165, 1380–1391. doi:10.1104/pp.114.235358
- Bauwe, H., Hagemann, M., Kern, R., and Timm, S. (2012). Photorespiration has a dual origin and manifold links to central metabolism. *Curr. Opin. Plant Biol.* 15, 269–275. doi:10.1016/j.pbi.2012.01.008
- Bordbar, A., Monk, J. M., King, Z. A., and Palsson, B. O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* 15, 107–120. doi:10.1038/nrg3643
- Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657. doi:10.1002/bit.10803
- Calvin, M. (1954). Chemical and photochemical reactions of thioctic acid and related disulfides. *Fed. Proc.* 13, 697–711.
- Carvalho, J. D. F. C., Madgwick, P. J., Powers, S. J., Keys, A. J., Lea, P. J., and Parry, M. A. J. (2011). An engineered pathway for glyoxylate metabolism in tobacco

- plants aimed to avoid the release of ammonia in photorespiration. *BMC Biotechnol.* 11:111. doi:10.1186/1472-6750-11-111
- Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., et al. (2011). Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 7, 518. doi:10.1038/msb.2011.52
- Cheung, C. Y. M., Poolman, M. G., Fell, D. A., Ratcliffe, R. G., and Sweetlove, L. J. (2014). A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in *C₃* plants and *Crassulacean* acid metabolism leaves. *Plant Physiol.* 165, 917–929. doi:10.1104/pp.113.234468
- Chung, B. K., and Lee, D. Y. (2009). Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. *BMC Syst. Biol.* 19:3. doi:10.1186/1752-0509-3-117
- de Oliveira Dal'Molin, C., Quek, L.-E., Saa, P. A., and Nielsen, L. K. (2015). A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Front. Plant Sci.* 6:4. doi:10.3389/fpls.2015.00004
- de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M., and Nielsen, L. K. (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol.* 152, 579–589. doi:10.1186/1752-0509-3-117
- Fahnenstich, H., Scarpeci, T. E., Valle, E. M., Flügge, U.-I., and Maurino, V. G. (2008). Generation of hydrogen peroxide in chloroplasts of *Arabidopsis* overexpressing glycolate oxidase as an inducible system to study oxidative stress. *Plant Physiol.* 148, 719–729. doi:10.1104/pp.108.126789
- Foyer, C. H., Bloom, A. J., Queval, G., and Noctor, G. (2009). Photorespiratory metabolism: genes, mutants, energetics, and redox signaling. *Annu. Rev. Plant Biol.* 60, 455–484. doi:10.1146/annurev.arplant.043008.091948
- Heise, R., Fernie, A. R., Stitt, M., and Nikoloski, Z. (2015). Pool size measurements facilitate the determination of fluxes at branching points in non-stationary metabolic flux analysis: the case of *Arabidopsis thaliana*. *Front. Plant Sci.* 6:386. doi:10.1146/annurev.arplant.043008.091948
- Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., Hirsch, H.-J., Rosenkranz, R., et al. (2007). Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nat. Biotechnol.* 25, 593–599. doi:10.1038/nbt1299
- Kleessen, S., Antonio, C., Sulpice, R., Laitinen, R., Fernie, A. R., Stitt, M., et al. (2012). Structured patterns in geographic variability of metabolic phenotypes in *Arabidopsis thaliana*. *Nat. Commun.* 3, 1319. doi:10.1038/ncomms2333
- Lakshmanan, M., Mohanty, B., and Lee, D.-Y. (2013). Identifying essential genes/reactions of the rice photorespiration by in silico model-based analysis. *Rice* 6, 20. doi:10.1186/1939-8433-6-20
- Larhlmi, A., Basler, G., Grimbs, S., Selbig, J., and Nikoloski, Z. (2012). Stoichiometric capacitance reveals the theoretical capabilities of metabolic networks. *Bioinformatics* 28, i502–i508. doi:10.1093/bioinformatics/bts381
- Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305. doi:10.1038/nrmicro2737
- Ma, F., Jazmin, L., Young, D. J., and Allen, D. (2014). Isotopically non-stationary ¹³C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16967–16971. doi:10.1073/pnas.1319485111
- Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10:e1003580. doi:10.1371/journal.pcbi.1003580
- Maier, A., Fahnenstich, H., Von Caemmerer, S., Engqvist, M. K. M., Weber, A. P. M., Flügge, U.-I., et al. (2012). Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Front. Plant Sci.* 3:38. doi:10.3389/fpls.2012.00038
- Mintz-Oron, S., Meir, S., Malitsky, S., Ruppim, E., Aharoni, A., and Shlomi, T. (2012). Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 339–344. doi:10.1073/pnas.1100358109
- Mueller, L. A., Zhang, P., and Rhee, S. Y. (2003). AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132, 453–460. doi:10.1104/pp.102.017236
- Nikoloski, Z., Perez-Storey, R., and Sweetlove, L. J. (2015). Inference and prediction of metabolic network fluxes. *Plant Physiol.* 169, 1443–1455. doi:10.1104/pp.15.01082
- Ogren, W. L., and Bowes, G. (1971). Ribulose diphosphate carboxylase regulates soybean photorespiration. *Nat. New Biol.* 230, 159–160. doi:10.1038/newbio230159a0
- Peterhansel, C., Blume, C., and Offermann, S. (2013a). Photorespiratory bypasses: how can they work? *J. Exp. Bot.* 64, 709–715. doi:10.1093/jxb/ers247
- Peterhansel, C., Krause, K., Braun, H.-P., Espie, G. S., Fernie, A. R., Hanson, D. T., et al. (2013b). Engineering photorespiration: current state and future possibilities. *Plant Biol. (Stuttg.)* 15, 754–758. doi:10.1111/j.1438-8677.2012.00681.x
- Pharkya, P., Burgard, A. P., and Maranas, C. D. (2004). OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 14, 2367–2376. doi:10.1101/gr.2872004
- Ranganathan, S., Suthers, P. F., and Maranas, C. D. (2010). OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 6:e1000744. doi:10.1371/journal.pcbi.1000744
- Robaina Estévez, S., and Nikoloski, Z. (2014). Generalized framework for context-specific metabolic model extraction methods. *Front. Plant Sci.* 5:491. doi:10.3389/fpls.2014.00491
- Schellenberger, J., Park, J. O., Conrad, T. M., and Palsson, B. Ø. (2010). BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11:213. doi:10.1186/1471-2105-11-213
- Sharkey, T. D. (1988). Estimating the rate of photorespiration in leaves. *Physiol. Plant.* 73, 147–152. doi:10.1111/j.1399-3054.1988.tb09205.x
- Simons, M., Saha, R., Amiour, N., Kumar, A., Guillard, L., Clément, G., et al. (2014). Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. *Plant Physiol.* 166, 1659–1674. doi:10.1104/pp.114.245787
- Szeczowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., Huege, J., et al. (2013). Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell* 25, 694–714. doi:10.1105/tpc.112.106989
- Timm, S., Nunes-Nesi, A., Pärnik, T., Morgenthal, K., Wienkoop, S., Keerberg, O., et al. (2008). A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in *Arabidopsis*. *Plant Cell* 20, 2848–2859. doi:10.1105/tpc.108.062265
- Töpfer, N., Kleessen, S., and Nikoloski, Z. (2015). Integration of metabolomics data into metabolic networks. *Front. Plant Sci.* 6:49. doi:10.3389/fpls.2015.00049
- Varma, A., and Palsson, B. O. (1993a). Metabolic capabilities of *Escherichia coli* I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* 165, 477–502. doi:10.1006/jtbi.1993.1202
- Varma, A., and Palsson, B. O. (1993b). Metabolic capabilities of *Escherichia coli* II. Optimal growth patterns. *J. Theor. Biol.* 165, 503–522. doi:10.1006/jtbi.1993.1203
- Xin, C.-P., Tholen, D., Devloo, V., and Zhu, X.-G. (2015). The benefits of photorespiratory bypasses: how can they work? *Plant Physiol.* 167, 574–585. doi:10.1104/pp.114.248013
- Yang, L., Cluett, W. R., and Mahadevan, R. (2011). EMILiO: a fast algorithm for genome-scale strain design. *Metab. Eng.* 13, 272–281. doi:10.1016/j.ymben.2011.03.002
- Zhu, X.-G., De Sturler, E., and Long, S. P. (2007). Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiol.* 145, 513–526. doi:10.1104/pp.107.103713

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 © 2016 Basler, Küken, Fernie and Nikoloski. 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 Engineered Production of the Glucoraphanin Precursor Dihomomethionine in *Nicotiana benthamiana*

Christoph Crocoll^{1,2}, Nadia Mirza^{1,2}, Michael Reichelt³, Jonathan Gershenzon³ and Barbara Ann Halkier^{1,2*}

¹ DNRF Center DynaMo, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Frederiksberg, Denmark, ² Copenhagen Plant Science Center, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Frederiksberg, Denmark, ³ Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

OPEN ACCESS

Edited by:

Lars Matthias Voll,
Friedrich-Alexander-University
Erlangen-Nuremberg, Germany

Reviewed by:

Sixue Chen,
University of Florida, USA
Judith Becker,
Saarland University, Germany

*Correspondence:

Barbara Ann Halkier
bah@plen.ku.dk

Specialty section:

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

Received: 20 November 2015

Accepted: 01 February 2016

Published: 16 February 2016

Citation:

Crocoll C, Mirza N, Reichelt M,
Gershenzon J and Halkier BA (2016)
Optimization of Engineered
Production of the Glucoraphanin
Precursor Dihomomethionine in
Nicotiana benthamiana.
Front. Bioeng. Biotechnol. 4:14.
doi: 10.3389/fbioe.2016.00014

Glucosinolates are natural products characteristic of the Brassicales order, which include vegetables such as cabbages and the model plant *Arabidopsis thaliana*. Glucoraphanin is the major glucosinolate in broccoli and associated with the health-promoting effects of broccoli consumption. Toward our goal of creating a rich source of glucoraphanin for dietary supplements, we have previously reported the feasibility of engineering glucoraphanin in *Nicotiana benthamiana* through transient expression of glucoraphanin biosynthetic genes from *A. thaliana* (Mikkelsen et al., 2010). As side-products, we obtained fivefold to eightfold higher levels of chain-elongated leucine-derived glucosinolates, not found in the native plant. Here, we investigated two different strategies to improve engineering of the methionine chain elongation part of the glucoraphanin pathway in *N. benthamiana*: (1) coexpression of the large subunit (*LSU1*) of the heterodimeric isopropylmalate isomerase and (2) coexpression of *BAT5* transporter for efficient transfer of intermediates across the chloroplast membrane. We succeeded in raising dihomomethionine (DHM) levels to a maximum of 432 nmol g⁻¹ fresh weight that is equivalent to a ninefold increase compared to the highest production of this intermediate, as previously reported (Mikkelsen et al., 2010). The increased DHM production without increasing leucine-derived side-product levels provides new metabolic engineering strategies for improved glucoraphanin production in a heterologous host.

Keywords: dihomomethionine, glucoraphanin, glucosinolates, metabolic engineering, *Nicotiana benthamiana*

INTRODUCTION

Plants are the source of an immense diversity of natural compounds, many of which are of high value as medicine or health-promoting agents. Often these compounds are difficult or impossible to produce by chemical synthesis, and extraction from plants is the only source.

Epidemiological studies strongly indicate that dietary consumption of cruciferous vegetables (e.g., broccoli) is correlated with reduced risk of the developing cancer (Verkerk et al., 2009). These and other health-promoting effects have been associated with glucosinolates, natural products

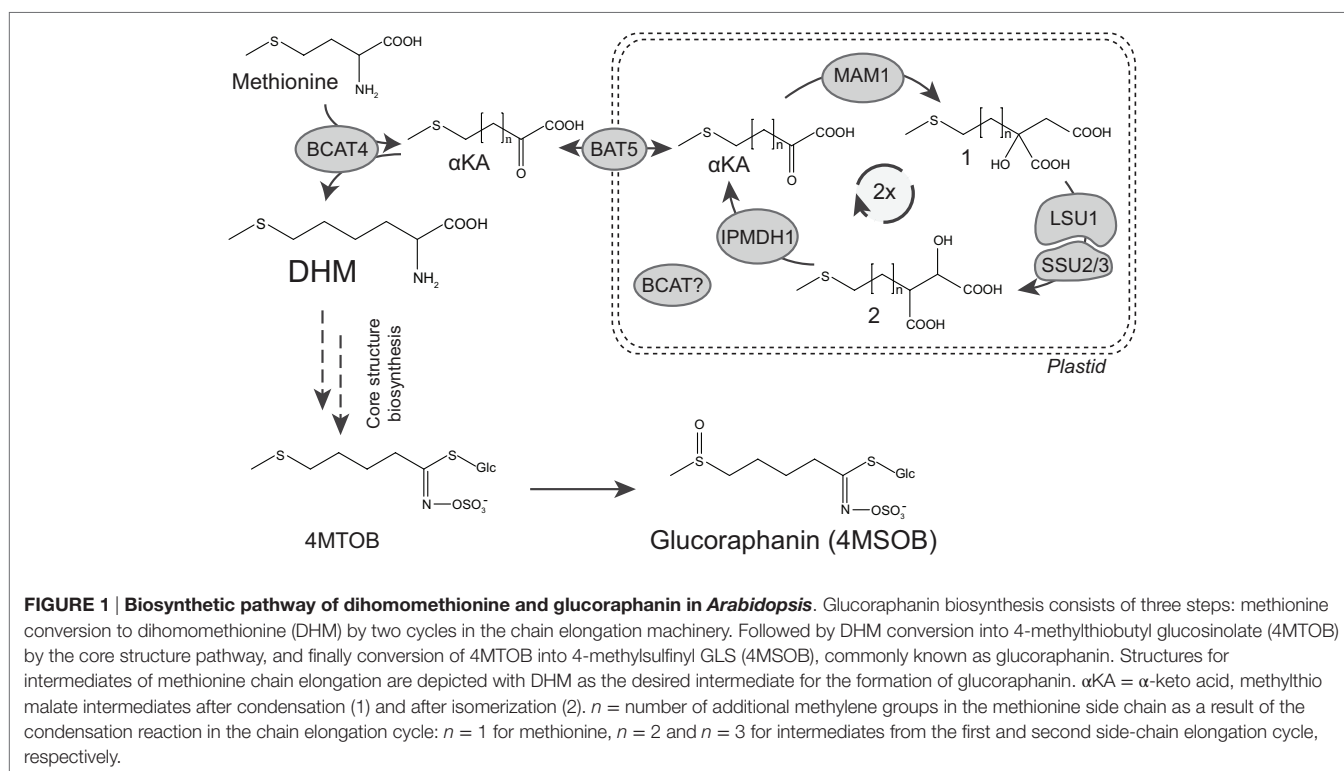
characteristic to the Brassicales order, which include vegetables such as broccoli and cabbages and the model plant *Arabidopsis thaliana* (Halkier and Gershenzon, 2006).

Substantial attention has been given to particularly the glucosinolate glucoraphanin that is present in broccoli, as it is generally thought to be the major bioactive compound associated with the cancer-preventive effects of broccoli (Traka and Mithen, 2009; Kensler et al., 2013). A recent human intervention study showed that diets with glucoraphanin-enriched broccoli resulted in retuning of cellular processes in the mitochondria to a basal level that is critical for maintaining a healthy metabolic balance (Armah et al., 2013). The health-promoting effects have resulted in a strong desire to increase the intake of glucoraphanin. The current market is based on products with unreliable amounts of glucoraphanin, if any at all. The latter has primed an interest to engineer the production of glucoraphanin into a heterologous host to obtain a stable, rich source of this product and enable intake of well-defined doses for dietary and pharmaceutical applications.

As a prerequisite for pathway engineering, all glucoraphanin biosynthetic genes have been identified in the model plant *A. thaliana* (Sonderby et al., 2010). Previously, we have engineered the six glucosinolate core pathway genes of simple indolyl- and benzyl glucosinolate-derived directly from the protein amino acids tryptophan and phenylalanine into the non-cruciferous plant *Nicotiana benthamiana* (Geu-Flores et al., 2009; Pfalz et al., 2011), and for indolylglucosinolates also into yeast (Mikkelsen et al., 2012). Engineering of the complex glucoraphanin pathway presents additional challenges as it consists of 12 biosynthetic enzymes, which are partitioned between the chloroplast and the cytosol (Halkier and Gershenzon, 2006; Sonderby et al., 2010).

Briefly, biosynthesis of glucoraphanin can be divided into three major parts (**Figure 1**). First, methionine is transaminated into a α -keto acid (α -KA) by a cytosolic branched-chain aminotransferase (BCAT4) (Schuster et al., 2006). This α -KA then enters the chloroplast where it undergoes side-chain elongation. The carbon side chain is elongated by a condensation reaction catalyzed by methylthioalkylmalate synthase (MAM) (Textor et al., 2004), followed by isomerization and oxidative decarboxylation catalyzed by an isopropylmalate isomerase (IPMI) and an isopropylmalate dehydrogenase (IPMDH), respectively (He et al., 2010, 2011). Following two cycles of chain elongation dihomomethionine (DHM) is formed, which subsequently is converted by the cytosolic, ER-associated core structure pathway to 4-methylthiobutyl (4MTOB) glucosinolate (GLS) (**Figure 1**), which finally is S-oxygenated to 4-methylsulfinylbutyl (4MSOB) glucosinolate, commonly known as glucoraphanin (**Figure 1**) (Sonderby et al., 2010).

The feasibility of metabolic engineering the complex glucoraphanin pathway was recently shown by transient expression of 10 biosynthetic genes in *N. benthamiana* (Mikkelsen et al., 2010). Though formation of ~ 50 nmol g⁻¹ fresh weight (fw) of the chain-elongated methionine-derived glucoraphanin was detected, fivefold to eightfold more of chain-elongated leucine-derived glucosinolates were detected. The latter does not accumulate in the native *Arabidopsis* but have been observed under conditions when MAM3 in the methionine chain elongation pathway was overexpressed using the 35S promoter (Field et al., 2004). Evolutionarily, the recursive methionine chain elongation pathway has evolved from the non-recursive valine to leucine chain elongation pathway in primary metabolism (Field et al.,



2004; Textor et al., 2004; Schuster et al., 2006; Binder et al., 2007; He et al., 2010; de Kraker and Gershenzon, 2011). The ability to accumulate chain-elongated leucine-derived glucosinolates upon engineering in tobacco supports the proposed promiscuity of enzymes in specialized metabolism compared with those in primary metabolism (Weng and Noel, 2012). It also indicates that the native plant has evolved mechanism(s) to prevent the formation of these leucine-derived side-products.

Homologs are described for all genes in methionine chain elongation. Coexpression analysis and knockout mutants have helped to identify specific roles of the different homologs. MAM1 was identified as the enzyme catalyzing two condensation reactions (Textor et al., 2004). For IPMDH, three homologs are known of which IPMDH1 was recently reported to be the best candidate for methionine chain elongation, whereas IPMDH2 and IPMDH3 were involved in leucine biosynthesis (LeuC) (He et al., 2009, 2011, 2013). Nevertheless, IPMDH3 was previously shown to be functional in metabolic engineering of DHM (Mikkelsen et al., 2010). The isomerization reaction catalyzed by IPMI has yet another level of complexity. IPMI is a heterodimeric enzyme consisting of a single large subunit (LSU1) that forms a catalytic active enzyme with either of three small subunits (SSU1/2/3). The small subunits define in which pathway IPMI is active. SSU2 and SSU3 are generally associated with methionine chain elongation and SSU1 with LeuC (He et al., 2010). Nevertheless, it was recently hypothesized that SSU1 was actively involved in the first two cycles of methionine chain elongation (Imhof et al., 2014). Transport of α -keto acids (formed initially by cytosolic BCAT4 and after each cycle by the chloroplastic IPMDH enzyme) across the chloroplast membranes was suggested to be performed by the bile acid transporter 5 (BAT5) (Gigolashvili et al., 2009). This was based on *bat5* knockout mutants showing a 50% reduction in methionine-derived, aliphatic glucosinolates and transport of α -keto acids into the chloroplast was impaired.

Toward our goal to establish high glucoraphanin production in a heterologous host, optimization of DHM production is essential. General means to enhance product formation in pathway engineering projects include screening for lacking enzymes, enzymes with improved properties (i.e., substrate specificity, kinetics), alleviating catalytic bottlenecks and increasing flux through the pathway, and taking compartmentalization into account (Heinig et al., 2013). Here, we report optimization of the production of the glucoraphanin precursor DHM in *N. benthamiana*. As reference for comparison, we use the highest-producing gene combination, as previously reported. The increase is obtained by optimizing the combination of biosynthetic genes used. Moreover, we provide additional evidence that BAT5 is the transporter for α -keto acids across the chloroplast membrane system.

MATERIALS AND METHODS

Plant Material

Nicotiana benthamiana plants were grown in small pots of 5.5 cm diameter in a green house at 24°C (day) and 18°C (night) with 50–60% humidity for ~3–4 weeks (to four to six leaves stage).

Cloning and Transformation

All genes were cloned into a USER compatible version of pCam-bia3300¹ plasmid by USER cloning (Nour-Eldin et al., 2006; Bitinaite et al., 2007; Geu-Flores et al., 2007). In brief, coding sequences of individual genes were amplified with single Uracil containing primers that were compatible with the USER ready-made plasmid. For PCR primers, see Table S1 in Supplementary Material. PCR products were purified (QIAquick PCR Purification Kit, Qiagen, Hilden, Germany), and 1–5 μ L of purified PCR product were subsequently mixed with 1 μ L of plasmid. The volume was adjusted to 10 μ L and after addition of 1 μ L USER Enzyme (NEB, Ipswich, MA, USA), the mix was incubated at 37 and 25°C for 30 min each. Two microliters of the USER cloning mix were added to 60 μ L of chemical competent *E. coli* DHB10 (NEB, Ipswich, MA, USA) cells by heat shock. Briefly, 10 min on ice, 90 s 42°C followed by 2 min on ice. Cells were incubated for 60 min at 37°C after addition of 250 μ L LB media. Subsequently, 100 μ L were plated on LB agar plates containing kanamycin (50 μ g mL⁻¹). Correct gene insertions were verified by sequencing (Macrogen Europe, Amsterdam, Netherlands). Constructs with correct insertions were subsequently transformed into *Agrobacterium tumefaciens* strain pGV3850 by electroporation (2 mm cuvette, 2.5 kV, 400 Ω , and 25 μ F) in a Bio-Rad GenePulser (Bio-Rad, Hercules, CA, USA). One milliliter YEP media was added, and cells were incubated for 3 h at 28°C with shaking. Subsequently, 150 μ L were plated on YEP agar plates containing antibiotics (30 μ g mL⁻¹ rifampicin and 50 μ g mL⁻¹ kanamycin).

Transient Expression by Infiltration of *Nicotiana benthamiana*

Overnight cultures of *A. tumefaciens* carrying the different gene constructs were grown in 10 mL YEP media (containing 50 μ g mL⁻¹ kanamycin and 30 μ g mL⁻¹ rifampicin) at 28°C and 220 rpm. Overnight cultures were harvested by centrifugation at 20°C for 15 min at 4500 \times g. The cell pellets were resuspended in infiltration buffer (10 mM MgCl₂, 10 mM MES, pH 5.6) containing 100 μ M acetosyringone (3,5-dimethoxy-4-hydroxyacetophenone, Sigma-Aldrich, Steinheim, Germany) and shaken at 150 rpm for 1–2 h at room temperature prior to plant infiltration. Cell densities for all cultures was adjusted to OD₆₀₀ \approx 0.21–0.25, which resulted in a final concentration for each individual construct of OD₆₀₀ \approx 0.03 for experiments with chain elongation enzymes only and OD₆₀₀ \approx 0.015 for experiments, which included chain elongation and core structure enzymes. These low OD were sufficient to ensure efficient transformation while keeping the stress levels for *N. benthamiana* leaves low. In all experiments, the suppressor protein p19 was included to reduce silencing effects (Voinnet et al., 2003). For each individual, experiment two to three leaves of four *N. benthamiana* plants (3–4 weeks old) were infiltrated with the different combinations of *A. tumefaciens* cultures harboring the different gene constructs. A maximum of seven different *A. tumefaciens* cultures were mixed. The volume of combinations with fewer than seven constructs was adjusted by addition of the respective amount of infiltration buffer.

¹<http://www.cambia.org/>

Plant Material Harvesting and Sample Preparation

Plant material was harvested 5 days after infiltration with *A. tumefaciens*. From each leaf, four leaf disks of 1 cm diameter were harvested from infiltrated areas and weighed. Amino acids were extracted with 400 μ L of 85% methanol containing norleucine (10 μ M) as internal standard (IS). Amino acid concentrations were determined by comparison to ^{13}C , ^{15}N -labeled algal amino acids described below.

Amino Acid Analysis by LC-MS

The resulting extract was diluted in a ratio of 1:10 (v:v) in water containing the ^{13}C , ^{15}N -labeled amino acid mix (Isotec, Miamisburg, OH, USA). Amino acids in the diluted extracts were directly analyzed by LC-MS/MS. The analysis method was modified from a protocol described by Jander et al. (2004). Chromatography was performed on an Agilent 1200 HPLC system (Agilent Technologies, Boeblingen, Germany). Separation was achieved on a Zorbax Eclipse XDB-C18 column (50 mm \times 4.6 mm, 1.8 μ m, Agilent Technologies, Germany). Formic acid (0.05%) in water and acetonitrile were employed as mobile phases A and B, respectively. The elution profile was 0–1 min, 3% B in A; 1–2.7 min, 3–100% B in A; 2.7–3 min 100% B, 3.1–6 min 3% B in A. The mobile phase flow rate was 1.1 mL/min. The column temperature was maintained at 25°C. The liquid chromatography was coupled to an API 5000 tandem mass spectrometer (AB Sciex, Darmstadt, Germany) equipped with a Turbospray ion source operated in positive ionization mode. The instrument parameters were optimized by infusion experiments with pure standards (amino acid standard mix, Fluka, St. Louis, MO, USA). The ionspray voltage was maintained at 5500 eV. The turbo gas temperature was set at 700°C. Nebulizing gas was set at 70 psi, curtain gas at 35 psi, heating gas at 70 psi, and collision gas at 2 psi. Multiple reaction monitoring (MRM) was used to monitor analyte parent ion \rightarrow product ion: MRMs were chosen as in Jander et al. (2004) except for Arg (m/z 175 \rightarrow 70) and Lys (m/z 147 \rightarrow 84). In addition, MRMs for homomethionine (HM, m/z 164 \rightarrow 118), DHM (m/z 178 \rightarrow 132), and S-adenosylmethionine (SAM, m/z 399 \rightarrow 136). The chain-elongated leucine products homo-leucine (HL, m/z 146 \rightarrow 100), dihomomethionine (DHL, m/z 160 \rightarrow 114), and trihomomethionine (THL, m/z 174 \rightarrow 128) were also monitored, but exact quantification was not possible for DHL and THL due to lack of reference standards. Values for DHL and THL are calculated based on the assumption of an equal response factor of 1 compared to ^{13}C , ^{15}N -labeled phenylalanine due to their similar behavior in fragmentation and ionization compared to leucine and HL. Detailed values for mass transitions can be found in Table S2 in Supplementary Material. Both Q1 and Q3 quadrupoles were maintained at unit resolution. Analyst 1.5 software (AB Sciex, Darmstadt, Germany) was used for data acquisition and processing. Linearity in ionization efficiencies were verified by analyzing dilution series of standard mixtures (amino acid standard mix, Fluka + Gln, Asn, and Trp, also Fluka). All samples were spiked with ^{13}C , ^{15}N -labeled amino acids (algal amino acids ^{13}C , ^{15}N , Isotec, Miamisburg, OH, USA) at a concentration of 10 μ g of the mix per milliliter. The concentration of the

individual labeled amino acids in the mix had been determined by classical HPLC–fluorescence detection analysis after pre-column derivatization with ortho-phthalaldehyde-mercaptoethanol using external standard curves made from standard mixtures (amino acid standard mix, Fluka + Gln, Asn, and Trp, also Fluka). Individual amino acids in the sample were quantified by the respective ^{13}C , ^{15}N -labeled amino acid IS, except for tryptophan, and asparagin: tryptophan was quantified using ^{13}C , ^{15}N -Phe applying a response factor of 0.42, asparagin was quantified using ^{13}C , ^{15}N -Asp applying a response factor of 1.0.

Statistical Analysis

Statistical analysis was performed with the SigmaPlot 12.0 statistics package (Systat Software, San Jose, CA, USA).

Accession Numbers

Sequence data from this article can be found *via* the TAIR database² under the AGI locus identifiers: *BCAT4* (At3g19710), *RBSC1A* (At1g67090), *BAT5* (At4g12030), *MAM1* (At5g23010), *IPMI-LSU1* (At4g13430), *IPMI-SSU1* (At2g43090), *IPMI-SSU2* (At2g43100), *IPMI-SSU3* (At3g58990), *IPMDH1* (At5g14200), and *IPMDH3* (At1g31180).

RESULTS AND DISCUSSION

In this study, we identified a new combination of genes for methionine chain elongation that produced the highest level of DHM in *N. benthamiana*. Additionally, we measured how these optimizations affected the formation of chain-elongated leucine-derived side-products.

Definition of a Reference Value for Optimization of DHM Production

Toward our goal of optimizing DHM production by transient expression experiments in *N. benthamiana*, we choose as reference a gene combination identical the highest-producing gene combination, as previously reported (Mikkelsen et al., 2010). In this study, genes were expressed from multi-gene constructs with two or three genes separated by 2A sequences (Mikkelsen et al., 2010). However, as we in the current study would compare multiple gene combinations, we expressed all genes from single gene constructs as this enabled us to freely combine individual genes. The previously reported gene combination for highest DHM production included a chloroplast-localized *BCAT4* together with *MAM1*, *IPMI-SSU3*, and *IPMDH3* and resulted in the formation of 51.4 ± 20.8 nmol g⁻¹ fw (Mikkelsen et al., 2010). When we expressed the same genes individually, we obtained only 14.6 ± 4.4 nmol g⁻¹ fw, which was used as reference value in this study (Table 1; Figures 2A and 3A). Several parameters can account for the discrepancy between the values published by Mikkelsen et al. (2010) and the present study. Using single gene constructs increases the number of *Agrobacterium* strains that need to be mixed for coexpression. It has also been reported that

²<http://www.arabidopsis.org>

TABLE 1 | Optimization of DHM production in *N. benthamiana*.

Combination	Genes	DHM
A1 = Reference ^a	<i>chlBCAT4</i> , <i>MAM1</i> , <i>SSU3</i> , <i>IPMDH3</i>	14.6 (±4.4)
A2	<i>chlBCAT4</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU3</i> , <i>IPMDH3</i>	312.6 (±40.2)
A3	<i>chlBCAT4</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU3</i> , <i>IPMDH1</i>	228.5 (±23.5)
A4	<i>BCAT4</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU3</i> , <i>IPMDH1</i>	41.9 (±9.2)
A5	<i>BCAT4</i> , <i>BAT5</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU3</i> , <i>IPMDH1</i>	315.0 (±54.8)
A6	<i>BCAT4</i> , <i>BAT5</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU1</i> , <i>IPMDH1</i>	340.6 (±86.5)
A7	<i>BCAT4</i> , <i>BAT5</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU2</i> , <i>IPMDH1</i>	432.2 (±70.8)
A8	<i>BCAT4</i> , <i>BAT5</i> , <i>MAM1</i> , <i>LSU1</i> , <i>SSU1</i> , <i>SSU3</i> , <i>IPMDH1</i>	400.4 (±37.3)
Ctrl	Non-infiltrated	n.d.

Data are represented as mean ± SEM in nanomole per gram fresh weight (N = 8).

DHM, dihomomethionine; fw, fresh weight; *chl*, chloroplastic signal peptide; *BCAT4*, branched-chain aminotransferase 4; *MAM1*, methylthioalkylmalate synthase 1; *LSU1*, large subunit of isopropylmalate isomerase (IPMI); *SSU*, small subunit of IPMI; *IPMDH*, isopropylmalate dehydrogenase; Ctrl, control.

^aReference to highest-producing gene combination previously reported (Mikkelsen et al., 2010) with 51.4 nmol DHM g⁻¹ fw. For values of all amino acids, see also Table S3 in Supplementary Material.

constructs containing 2A sequences for self-processing of multi-gene constructs can result in incomplete cleavage and formation of fusion proteins, which can influence the outcome of metabolic engineering from transient expression in plants (Burén et al., 2012). Other potential reasons include differences in growth conditions for the tobacco plants and differences in detection and quantification of the individual compounds by LC-MS between the two studies. In combination, the experimental and technical differences do not allow for a direct comparison of the DHM production. All calculations are based on the DHM amounts produced in the reference gene combination of the present study, which previously resulted in the highest DHM production.

Coexpression of *Arabidopsis* IPMI-LSU1 is Essential for Efficient Formation of DHM by the Methionine Chain Elongation Pathway

In the heterodimeric isopropylmalate synthase (IPMI), the large subunit (LSU1) forms a functional enzyme with one of three small subunits (SSU1–3) catalyzing the isomerization step in the methionine and valine (to leucine) chain elongation machinery (Knill et al., 2009). Previously, it was shown that endogenous LSU of tobacco was able to substitute for the *A. thaliana* LSU1 (that was not included in the gene combination) and form a functional IPMI enzyme with *A. thaliana* SSUs that resulted in production of DHM (Mikkelsen et al., 2010). Here, we show that inclusion of *Arabidopsis*' LSU1 in the gene combination resulted in a 21-fold increase of DHM production (Table 1; Figures 2A and 3A, combinations A1 and A2). Coexpression of *A. thaliana* LSU1 probably has alleviated a bottleneck and thus increased flux through the pathway. Another possibility could be better interaction between the two *A. thaliana* subunits in comparison to a heterodimer formed from *N. benthamiana* LSU and *A. thaliana* SSU. However, the high amino acid sequence identity of ~95% between *A. thaliana* LSU1 and *Nicotiana sylvestris* LSU, a close relative of *N. benthamiana*, suggests that functionality is not impaired. Recently, it was also demonstrated that *A. thaliana* SSUs to a certain extent can complement *E. coli* knockout mutants lacking the respective SSU homolog in LeuC (Imhof et al., 2014). Therefore, the increased production of DHM is most likely a result of a better ratio between large subunit to small subunit and the formation of a higher number of catalytically active IPMI heterodimers as a result of coexpressing LSU1 under the control of the same CaMV35S promoter as the other constructs.

Choice of Small Subunit in IPMI Has No Significant Influence on DHM Production

It has been reported that the three small subunits (SSU1/2/3) define in which pathway IPMI is active, with SSU2 and SSU3 being associated with methionine chain elongation and SSU1 with LeuC (He et al., 2010). More recently, it was hypothesized that SSU1 was active in the first two cycles of methionine chain elongation and that either SSU2 or SSU3 catalyzes the formation of the longer chain-elongated products (Imhof et al., 2014). Interestingly, we detected no significant difference in DHM

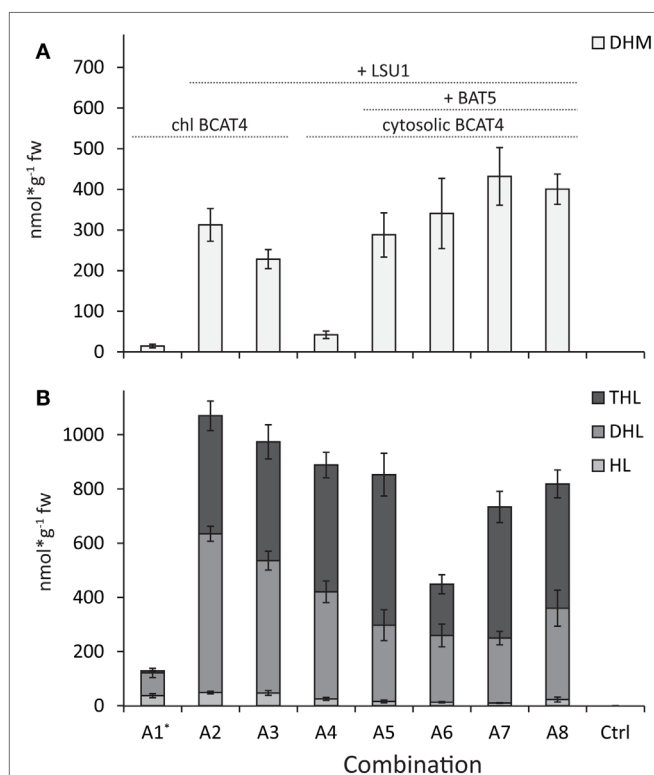


FIGURE 2 | Comparison of production of DHM and chain-elongated leucine-derived products. (A) DHM levels. **(B)** Levels of homo-leucine (HL), dihomomethionine (DHL), and trihomomethionine (THL). *BCAT4* is relocated to the chloroplast (*chl BCAT4*) in combinations A1–A3. The transporter protein *BAT5* is coexpressed in combinations A5–A8 and the large subunit (*LSU1*) of IPMI is coexpressed in all combinations except A1. Ctrl represents non-infiltrated. *Chl BCAT4* = *BCAT4* with signal peptide for relocation to chloroplast, +*LSU1* = combinations where *LSU1* was coexpressed, and +*BAT5* = combinations where *BAT5* was coexpressed. Data are represented as mean ± SEM in nanomole per gram fresh weight (N = 8).

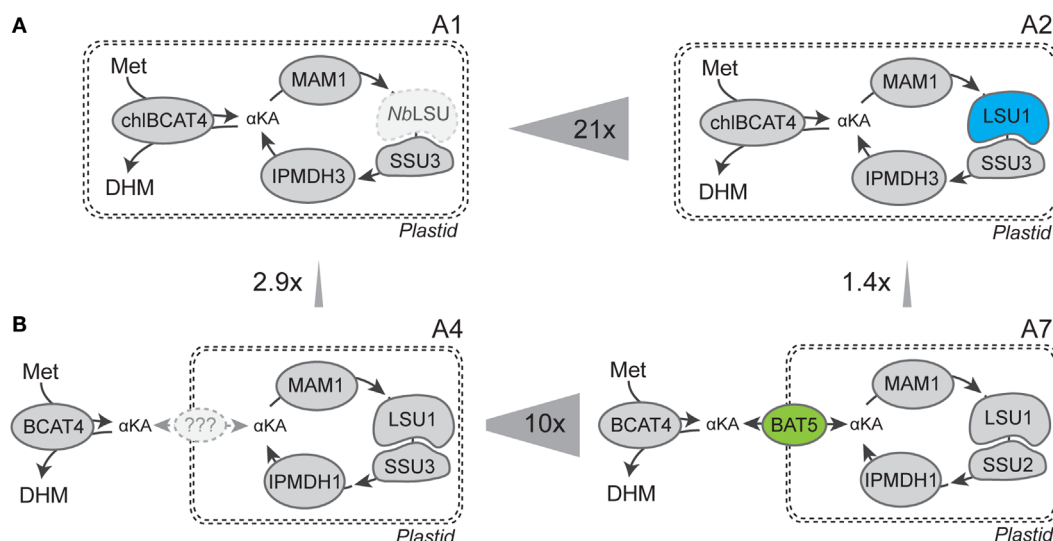


FIGURE 3 | Graphical comparison of combinations with most significant increases in DHM production. (A) *Nicotiana benthamiana* LSU (NbLSU) can complement *Arabidopsis* IPMI small subunit (SSU3) into a functional heterodimer, and co-expression of *Arabidopsis* IPMI large subunit (LSU1) increased DHM production by 21-fold. **(B)** BAT5 is an efficient transporter for translocation of intermediates between the cytosolic BCAT4 and the chloroplast-localized part of methionine chain elongation and increased DHM production by 10-fold.

production when we coexpressed the individual small subunits of IPMI with the rest of the methionine chain elongation machinery (Table 1, combinations A5–A7). Furthermore, coexpression of SSU1 with SSU3 did not result in significantly higher production of DHM (Table 1, combination A8). Coexpression of SSU2 and SSU3 was previously shown to have no positive effect on DHM production (Mikkelsen et al., 2010). Nevertheless, our data show that SSU1 has the ability to support the first two rounds of methionine chain elongation. At this point, it remains unclear if SSU1 catalyzes the first round(s) of methionine chain elongation in native *A. thaliana* plants and whether either SSU2 or SSU3 catalyzes the formation of the longer chain-elongated products, as previously suggested (Imhof et al., 2014). A previously observed sevenfold difference in DHM production between coexpression of SSU2 and SSU3 (Mikkelsen et al., 2010) was not observed. This was insofar puzzling as differences in the present study were not as dramatic and not significant for the different products upon coexpression of either of the three SSUs together with LSU1 (Figure 2; Table 1, combinations A5–A7). One possible explanation could be that coexpression of IPMI–LSU1 has alleviated effects that might be related to functional differences between the three SSUs. And even though *A. thaliana* IPMI can complement *E. coli* mutant strains deficient for the respective homologs in LeuC (*LeuC* and *LeuD*) (He et al., 2010; Imhof et al., 2014), it cannot be excluded that there are functional differences when catalyzing reactions in methionine chain elongation. Finally, it still remains unclear whether functional differences and the involvement of the different SSUs in either leucine biosynthesis or methionine chain elongation are the result of structural differences in the active sites or due to differences in

temporal and spatial expression within different tissues in the plant (He et al., 2010; Imhof et al., 2014).

Coexpression of the BAT5 Transporter Facilitates DHM Production with Cytosolic BCAT4

Bile acid transporter 5 has been proposed to translocate substrates and/or products of methionine chain elongation across the chloroplast membranes (Gigolashvili et al., 2009). The α-keto acid product of BCAT4 was proposed to be the substrate for import into the chloroplast by BAT5 as *bat5* knockout mutants show 50% reduced levels of aliphatic glucosinolates (including 4MSOB and 4MTB) and transport of α-keto acids into the chloroplast was impaired (Gigolashvili et al., 2009). As BAT5 may be critical for translocating the chain elongation products out of the chloroplast, we investigated whether inclusion of BAT5 together with cytosolic BCAT4 improves DHM production.

In experiments where BCAT4 was expressed in the cytosol without coexpression of BAT5, a massive reduction in DHM production to only 41.9 (±9.2) nmol g⁻¹ fw was observed compared to chloroplast-localized BCAT4 (Table 1; Figures 2A and 3A). This demonstrates that although methionine chain elongation is a heterologous pathway, *N. benthamiana* still has the ability to transport α-keto acid intermediates at a low level. DHM production was fully restored upon coexpression of BAT5 with the cytosolic BCAT4 (Figure 3B). DHM levels were higher than in combinations with chloroplast-localized BCAT4, though not significantly higher (Table 1). It is not known whether cytosolic BCAT4 (or other chloroplast-localized BCATs) are involved in

transaminating the final chain-elongated methionine product, and whether a chain-elongated α -keto acid is substrate for export out of the chloroplast by BAT5. Nevertheless, the fact that DHM accumulated to such high levels indicated that BCAT4 may also transaminate the chain-elongated α -keto acids into the respective chain-elongated methionine and that BAT5 is an antiporter for the different chain-lengths α -keto acids.

Differences in DHM Production with IPMDH3 Compared to IPMDH1

Three homologs exist for IPMDH, of which IPMDH3 was used for in the previous report for engineering DHM (Mikkelsen et al., 2010). A recent study linking association by coexpression suggested that IPMDH1 was the key player in methionine chain elongation while IPMDH2 and IPMDH3 were involved in LeuC (He et al., 2013). When we compared inclusion of *IPMDH1* or *IPMDH3* together with the other chain elongation genes, we observed that the production of DHM was always higher with *IPMDH3* though never significantly higher (Table 1). The variation from transient expression in tobacco made it impossible to identify significant differences. Nevertheless, association by coexpression in *Arabidopsis*, as previously reported (He et al., 2013), does not exclude enzymatic promiscuity (Weng and Noel, 2012) of the different IPMDH enzymes. This is also supported by the fact that there was no significant difference detected between the three IPMI small subunits (see above).

Presence of LSU1 Has Major Effects on Formation of Leucine-Derived Side-Products but is Independent of BCAT4 Localization

Previously, when the approximately fivefold to eightfold higher levels of leucine-derived glucosinolates were monitored (Mikkelsen et al., 2010), it was not possible to differentiate if both leucine and isoleucine were taken as substrates by the chain elongation machinery. Here, we confirmed by UHPLC-MS analysis that only leucine and not isoleucine is taken as substrate (Mirza et al., 2016). Optimization of the gene combination for DHM production also affected the formation of chain-elongated leucine-derived products (Figure 2). Similar to DHM production, chain-elongated leucine-derived products [homo-leucine (HL), dihomomethionine (DHL), and the newly detected trihomomethionine (THL)] drastically increased by including LSU1 together with the other chain elongation genes (Figure 2). In contrast to DHM, leucine-derived products were formed in similar amounts independent of whether *BCAT4* was expressed in the cytosol or the chloroplast, except for combination A1 where *LSU1* was not coexpressed (Figure 2, combinations A1 and A4; Table S3 in Supplementary Material). Especially, the formation of DHL was higher when *BCAT4* was localized to the chloroplast (Figure 2, combinations A2 and A3). Interestingly, formation of leucine-derived products – especially the longer chain-elongated products DHL and THL – was reduced in the gene combination when *SSU1* (A6) was coexpressed rather than *SSU2* (A7) or *SSU3* (A5) (Figure 2B). Also, THL was only found in combinations where

LSU1 was coexpressed, which may be related to the overall lower production in this combination or a reduced ability of tobacco LSU to support three rounds of leucine chain elongation. THL had not been described previously as a side-product from engineering of methionine chain elongation in tobacco (Mikkelsen et al., 2010) and neither from metabolic engineering of DHM in *E. coli* (Mirza et al., 2016).

Increased DHM Formation Considerably Improves DHM to Side-Product Ratio

Increased DHM formation had positive effects on the ratio of DHM to leucine-derived products. When the ratios were calculated for the sum of leucine-derived products (HL + DHL + THL) compared to DHM in the individual experiments, we detected 8.8-fold more leucine-derived products compared to DHM in the reference combination A1 as previously reported for this combination (Mikkelsen et al., 2010). All other combinations showed a more favorable ratio of DHM to chain-elongated leucine-derived products. The best ratios with almost equal amounts were found in combinations A6 and A7 (DHM:HL/DHL/THL 0.8:1) containing the complete set of genes and compartmentalization (Figures 2 and 3; Table S4 in Supplementary Material). The amounts of leucine-derived products remained largely unchanged throughout our experiments. Mainly DHL and the newly detected THL contributed to a large extent to the high amounts of leucine-derived products. The latter was drastically reduced in the gene combination including *SSU1*. Interestingly, we detected similar amounts of leucine-derived products in the combination without *BAT5* (A4), which suggests that leucine α -keto acids are present and taken up by the methionine chain elongation machinery. A similar effect was seen in experiments without coexpression of *BCAT4* (data not shown). Therefore, other solutions for further reduction of leucine-derived products may include exchange or mutation of single enzymes in the methionine chain elongation to increase affinity toward methionine and away from leucine.

In summary, our experiments provide new insights toward improving engineering of DHM production by transient expression in *N. benthamiana*. Here, we demonstrated a substantial 30-fold increase in production of DHM (432 nmol g⁻¹ fw) compared to the highest-producing gene combination previously reported and a not less impressive 9-fold increase compared to the previously highest reported DHM production levels (Mikkelsen et al., 2010). Simultaneously, the amounts of leucine-derived side-products were substantially reduced especially by re-establishing the compartmentalized organization of the methionine chain elongation in the transient expression host system. In conclusion, the optimized gene combination for production of DHM consists of five (or six) genes: *BCAT4* (*BAT5*), *MAM1*, *LSU1*, *SSU1*, and *IPMDH1*. *BAT5* is only necessary if the methionine chain elongation is expressed in plants or other chloroplast-containing organisms, such as microalgae, while in the case of a microbial host, the transport step can be omitted. Our results provide important insights for optimizing the engineering of glucoraphanin production in a heterologous host. Especially, the fact that the separation of the transamination step from the chain elongation itself improved the production of DHM mainly by

reducing formation of leucine-derived side-products has implications on metabolic engineering in microbial host organisms where such spatial separation is difficult to establish. This implies that other techniques might be required to create a similar level of spatial separation between the two parts of the chain elongation pathway. This could include creation of fusion proteins or coexpression of chaperons to create microenvironments that also could increase flux through the pathway. In addition, it should be considered to engineer pathway enzymes in a way to increase substrate specificity toward methionine and the corresponding intermediates. These steps should be considered before adding another level of complexity by engineering the core structure biosynthetic pathway on top of the methionine chain elongation pathway to ultimately create an expression system for sustainable production of glucoraphanin.

AUTHOR CONTRIBUTIONS

CC cloned the constructs for DHM biosynthesis, planned and conducted the experiments, prepared the samples for LC-MS

analysis for DHM production. NM conducted initial experiments, contributed to the discussion of results and the manuscript. MR performed analysis of amino acid and DHM production and suggested analytical improvements. JG contributed to the experimental design and made improvements to the manuscript. BH was involved in discussions about the experimental setup and results. CC and BH wrote the manuscript based on a draft written by CC.

FUNDING

This work was supported by Danish National Research Foundation (grant DNRF99) and Danish Council for Strategic Research (grant 0603-00387B).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2016.00014>

REFERENCES

- Armah, C. N., Traka, M. H., Dainty, J. R., Defernez, M., Janssens, A., Leung, W., et al. (2013). A diet rich in high-glucoraphanin broccoli interacts with genotype to reduce discordance in plasma metabolite profiles by modulating mitochondrial function. *Am. J. Clin. Nutr.* 98, 712–722. doi:10.3945/ajcn.113.065235
- Binder, S., Knill, T., and Schuster, J. (2007). Branched-chain amino acid metabolism in higher plants. *Physiol. Plant.* 129, 68–78. doi:10.1111/j.1399-3054.2006.00800.x
- Bitinaite, J., Rubino, M., Varma, K. H., Schildkraut, I., Vaisvila, R., and Vaiskunaite, R. (2007). USER™ friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.* 35, 1992–2002. doi:10.1093/nar/gkm041
- Burén, S., Ortega-Villasante, C., Ötvös, K., Samuelsson, G., Bakó, L., and Villarejo, A. (2012). Use of the foot-and-mouth disease virus 2A peptide co-expression system to study intracellular protein trafficking in *Arabidopsis*. *PLoS ONE* 7:e51973. doi:10.1371/journal.pone.0051973
- de Kraker, J. W., and Gershenzon, J. (2011). From amino acid to glucosinolate biosynthesis: protein sequence changes in the evolution of methylthioalkylmalate synthase in *Arabidopsis*. *Plant Cell* 23, 38–53. doi:10.1105/tpc.110.079269
- Field, B., Cardon, G., Traka, M., Botterman, J., Vancanney, G., and Mithen, R. (2004). Glucosinolate and amino acid biosynthesis in *Arabidopsis*. *Plant Physiol.* 135, 828–839. doi:10.1104/pp.104.039347
- Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2007). USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* 35, e55. doi:10.1093/nar/gkm106
- Geu-Flores, F., Olsen, C., and Halkier, B. (2009). Towards engineering glucosinolates into non-cruciferous plants. *Planta* 229, 261–270. doi:10.1007/s00425-008-0825-y
- Gigolashvili, T., Yatushevich, R., Rollwitz, I., Humphry, M., Gershenzon, J., and Flügge, U.-I. (2009). The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 21, 1813–1829. doi:10.1105/tpc.109.066399
- Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi:10.1146/annurev.arplant.57.032905.105228
- He, Y., Chen, B., Pang, Q., Strul, J. M., and Chen, S. (2010). Functional specification of *Arabidopsis* isopropylmalate isomerases in glucosinolate and leucine biosynthesis. *Plant Cell Physiol.* 51, 1480–1487. doi:10.1093/pcp/pcq113
- He, Y., Dai, S., Dufresne, C. P., Zhu, N., Pang, Q., and Chen, S. (2013). Integrated proteomics and metabolomics of *Arabidopsis* acclimation to gene-dosage dependent perturbation of isopropylmalate dehydrogenases. *PLoS ONE* 8:e57118. doi:10.1371/journal.pone.0057118
- He, Y., Galant, A., Pang, Q., Strul, J. M., Balogun, S. F., Jez, J. M., et al. (2011). Structural and functional evolution of isopropylmalate dehydrogenases in the leucine and glucosinolate pathways of *Arabidopsis thaliana*. *J. Biol. Chem.* 286, 28794–28801. doi:10.1074/jbc.M111.262519
- He, Y., Mawhinney, T. P., Preuss, M. L., Schroeder, A. C., Chen, B., Abraham, L., et al. (2009). A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. *Plant J.* 60, 679–690. doi:10.1111/j.1365-313X.2009.03990.x
- Heinig, U., Gutensohn, M., Dudareva, N., and Aharoni, A. (2013). The challenges of cellular compartmentalization in plant metabolic engineering. *Curr. Opin. Biotechnol.* 24, 239–246. doi:10.1016/j.copbio.2012.11.006
- Imhof, J., Huber, F., Reichelt, M., Gershenzon, J., Wiegrefe, C., Lächler, K., et al. (2014). The small subunit 1 of the *Arabidopsis* isopropylmalate isomerase is required for normal growth and development and the early stages of glucosinolate formation. *PLoS ONE* 9:e91071. doi:10.1371/journal.pone.0091071
- Jander, G., Norris, S. R., Joshi, V., Fraga, M., Rugg, A., Yu, S., et al. (2004). Application of a high-throughput HPLC-MS/MS assay to *Arabidopsis* mutant screening: evidence that threonine aldolase plays a role in seed nutritional quality. *Plant J.* 39, 465–475. doi:10.1111/j.1365-313X.2004.02140.x
- Kensler, T., Egner, P., Agyeman, A., Visvanathan, K., Groopman, J., Chen, J.-G., et al. (2013). “Keap1-Nrf2 signaling: a target for cancer prevention by sulforaphane,” in *Natural Products in Cancer Prevention and Therapy*, eds Pezzuto J. M. and Suh N. (Berlin; Heidelberg: Springer), 163–177.
- Knill, T., Reichelt, M., Paetz, C., Gershenzon, J., and Binder, S. (2009). *Arabidopsis thaliana* encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. *Plant Mol. Biol.* 71, 227–239. doi:10.1007/s11103-009-9519-5
- Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., et al. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* 14, 104–111. doi:10.1016/j.ymben.2012.01.006
- Mikkelsen, M. D., Olsen, C. E., and Halkier, B. A. (2010). Production of the cancer-preventive glucoraphanin in tobacco. *Mol. Plant.* 3, 751–759. doi:10.1093/mp/ssq020
- Mirza, N., Crocoll, C., Erik Olsen, C., and Ann Halkier, B. (2016). Engineering of methionine chain elongation part of glucoraphanin pathway in *E. coli*. *Metab. Eng.* 35, 31–37. doi:10.1016/j.ymben.2015.09.012
- Nour-Eldin, H. H., Hansen, B. G., Norholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006). Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* 34, e122. doi:10.1093/nar/gkl635

- Pfalz, M., Mikkelsen, M. D., Bednarek, P., Olsen, C. E., Halkier, B. A., and Kroymann, J. (2011). Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification. *Plant Cell* 23, 716–729. doi:10.1105/tpc.110.081711
- Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., and Binder, S. (2006). Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *Plant Cell* 18, 2664–2679. doi:10.1105/tpc.105.039339
- Sonderby, I. E., Geu-Flores, F., and Halkier, B. A. (2010). Biosynthesis of glucosinolates – gene discovery and beyond. *Trends Plant Sci.* 15, 283–290. doi:10.1016/j.tplants.2010.02.005
- Textor, S., Bartram, S., Kroymann, J., Falk, K. L., Hick, A., Pickett, J. A., et al. (2004). Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. *Planta* 218, 1026–1035. doi:10.1007/s00425-003-1184-3
- Traka, M., and Mithen, R. (2009). Glucosinolates, isothiocyanates and human health. *Phytochem. Rev.* 8, 269–282. doi:10.1007/s11101-008-9103-7
- Verkerk, R., Schreiner, M., Krumbein, A., Ciska, E., Holst, B., Rowland, I., et al. (2009). Glucosinolates in *Brassica* vegetables: the influence of the food supply chain on intake, bioavailability and human health. *Mol. Nutr. Food Res.* 53, S219–S265. doi:10.1002/mnfr.200800065
- Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33, 949–956. doi:10.1046/j.1365-313X.2003.01676.x
- Weng, J.-K., and Noel, J. P. (2012). The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320. doi:10.1101/sqb.2012.77.014787

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 © 2016 Crocoll, Mirza, Reichelt, Gershenzon and Halkier. 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.



Synthetic Peptides as Protein Mimics

Andrea Groß¹, Chie Hashimoto¹, Heinrich Sticht² and Jutta Eichler^{1*}

¹ Department of Chemistry and Pharmacy, University of Erlangen-Nuremberg, Erlangen, Germany, ² Institute of Biochemistry, University of Erlangen-Nuremberg, Erlangen, Germany

The design and generation of molecules capable of mimicking the binding and/or functional sites of proteins represents a promising strategy for the exploration and modulation of protein function through controlled interference with the underlying molecular interactions. Synthetic peptides have proven an excellent type of molecule for the mimicry of protein sites because such peptides can be generated as exact copies of protein fragments, as well as in diverse chemical modifications, which includes the incorporation of a large range of non-proteinogenic amino acids as well as the modification of the peptide backbone. Apart from extending the chemical and structural diversity presented by peptides, such modifications also increase the proteolytic stability of the molecules, enhancing their utility for biological applications. This article reviews recent advances by this and other laboratories in the use of synthetic protein mimics to modulate protein function, as well as to provide building blocks for synthetic biology.

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

Juan Manuel Pedraza,
Universidad de los Andes, Colombia
Alexander D. Frey,
Aalto University, Finland

*Correspondence:

Jutta Eichler
jutta.eichler@fau.de

Specialty section:

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

Received: 16 September 2015

Accepted: 22 December 2015

Published: 19 January 2016

Citation:

Groß A, Hashimoto C, Sticht H and
Eichler J (2016) Synthetic Peptides
as Protein Mimics.
Front. Bioeng. Biotechnol. 3:211.
doi: 10.3389/fbioe.2015.00211

Keywords: protein–protein interactions, protein mimics, peptides, structure-based design, biomaterials

INTRODUCTION

The detailed insight into the human genome does not in itself enable a comprehensive understanding of human protein function, health, and disease. In the post-genome era, an important challenge is the structural and functional analysis of the gene products, i.e., proteins. Proteins play a major role in almost all biological processes, including enzymatic reactions, structural integrity of cells, organs and tissues, cell motility, immune responses, signal transduction, and sensing. All protein-mediated biological processes are based on specific interactions between proteins and their ligands. Therefore, exploring disease-associated protein–ligand and protein–protein interactions is essential to gain insight into the molecular mechanisms underlying diseases and other phenomena, as well as for the development of novel therapeutic strategies.

Molecules that present the binding sites of proteins, which are involved in a disease-associated protein–protein interaction, are promising candidates for therapeutic intervention. Such binding site mimetic molecules can be generated either through recombinant protein synthesis or by means of chemical peptide synthesis. A specific advantage of synthetic peptides is that they can be generated as exact copies of protein fragments as well as in diverse chemical modifications, which include the incorporation of a large range of non-proteinogenic amino acids, as well as the modification of the peptide backbone. Apart from extending the chemical and structural diversity presented by peptides, such modifications also increase the proteolytic stability of the molecules, enhancing their potential as drug candidates.

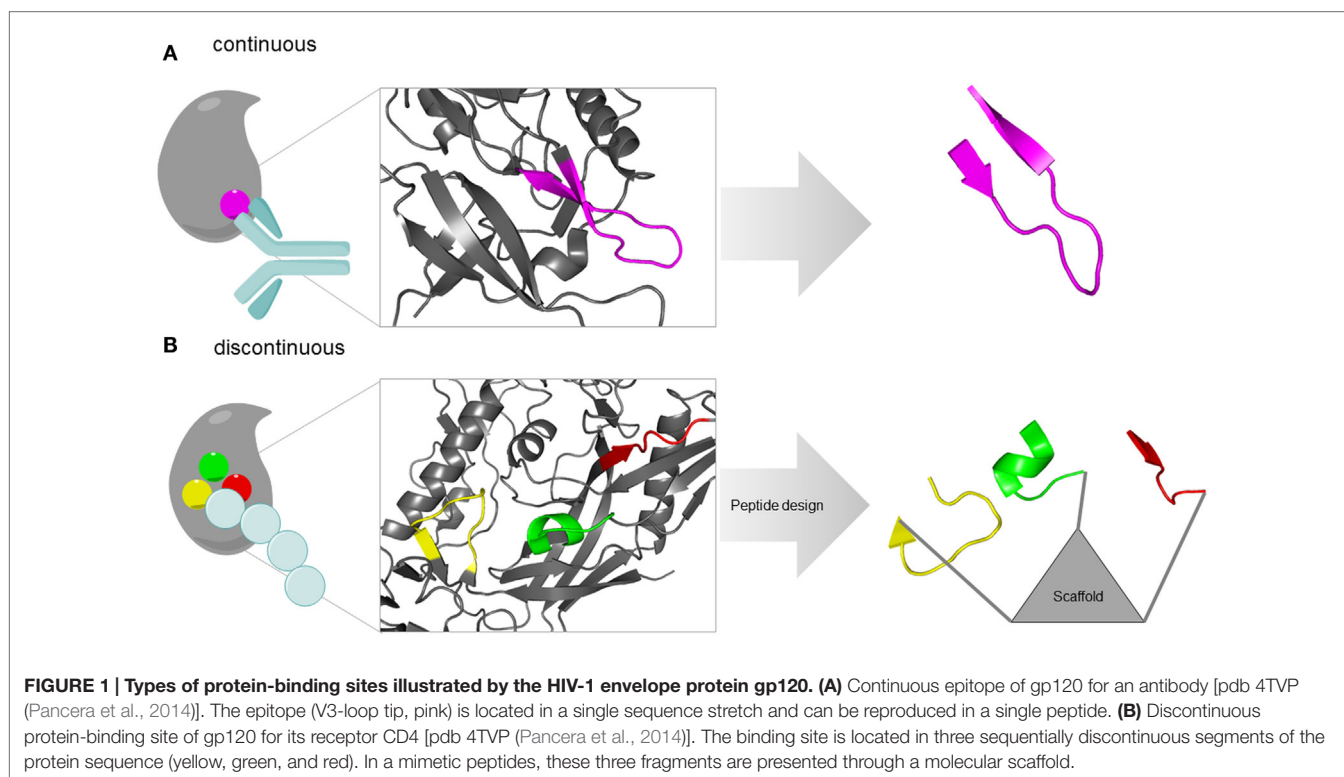
Three conceptually different approaches are available for the design of protein-binding site mimetic peptides. These approaches are based on one or more of the following information about the proteins of interest: structure, sequence, and function. In random combinatorial methods that

are based solely on protein function, such as phage display (Li and Caberoy, 2010) and synthetic peptide combinatorial libraries (Houghten et al., 1999), respectively, large populations of peptides are screened for binders to the respective partner protein, or for inhibitors of the protein–protein interaction of interest. A strategy termed peptide scanning is based on the synthesis of the entire protein sequence – or large parts of it – in the shape of short, overlapping peptides, which are then individually tested for binding to the respective partner protein (Frank, 2002), enabling the identification of protein-binding sites. The utility of this method, however, is largely limited to the identification of sequentially continuous binding sites, which are located in a protein sequence stretch of consecutive amino acids (Figure 1A). Structure-based design, finally, involves the design and generation of protein-binding site mimics based on the 3D structure of the protein–protein complexes (Eichler, 2008). This structural information enables the design and generation of mimics of continuous, as well as of sequentially discontinuous protein-binding sites, which are composed of two or more protein segments that are distant in protein sequence, but brought into spatial proximity through protein folding (Figure 1B). Mimicking such discontinuous protein-binding sites by synthetic peptides typically involves presentation of the respective protein fragments through a molecular scaffold (Figure 1B).

Here, we review strategies for the use of synthetic peptides as protein mimics. Focusing on structure-based design, the potential of such peptides as drugs against diseases, such as viral and bacterial infections, cancer, as well as autoimmune diseases, are discussed.

TOOLBOX FOR PEPTIDE SYNTHESIS: NON-PROTEINOGENIC AMINO ACIDS AND SITE-SELECTIVE LIGATION

Most current methods for the chemical synthesis of peptides utilize Merrifield's concept of solid-phase synthesis (Merrifield, 1963), which enables the synthesis of peptides and small proteins of up to 100 amino acids. A major advantage of chemical peptide synthesis, as compared to recombinant protein synthesis, is the extended set of amino acids and other building blocks that can be incorporated, which includes D-amino acids, as well as a wide range of non-proteinogenic amino acids (Figure 2). While the recombinant synthesis of proteins containing non-proteinogenic amino acids is possible only through alternative codon usage (Mehl et al., 2003), hundreds of such building blocks are commercially available for the use in chemical peptide synthesis. This opens the door to improved biological activity and peptide stability, as well as structural modifications. One possibility is the use of β - and γ -amino acids (Seebach et al., 2004), which differ from α -amino acids in having one or two additional methylene groups between the carboxy and the amino function of the amino acid (Figure 2A). Peptides composed of these amino acids are stable against proteolysis *in vitro* and *in vivo*, as well as metabolism and degradation by microbial colonies (Seebach et al., 2004; Seebach and Gardiner, 2008). On the functional side, they can act similar to the natural α -peptides, as exemplified by β - and γ -peptide agonists of naturally occurring α -peptide hormones such as somatostatin (Seebach et al., 2004; Seebach and Gardiner, 2008).



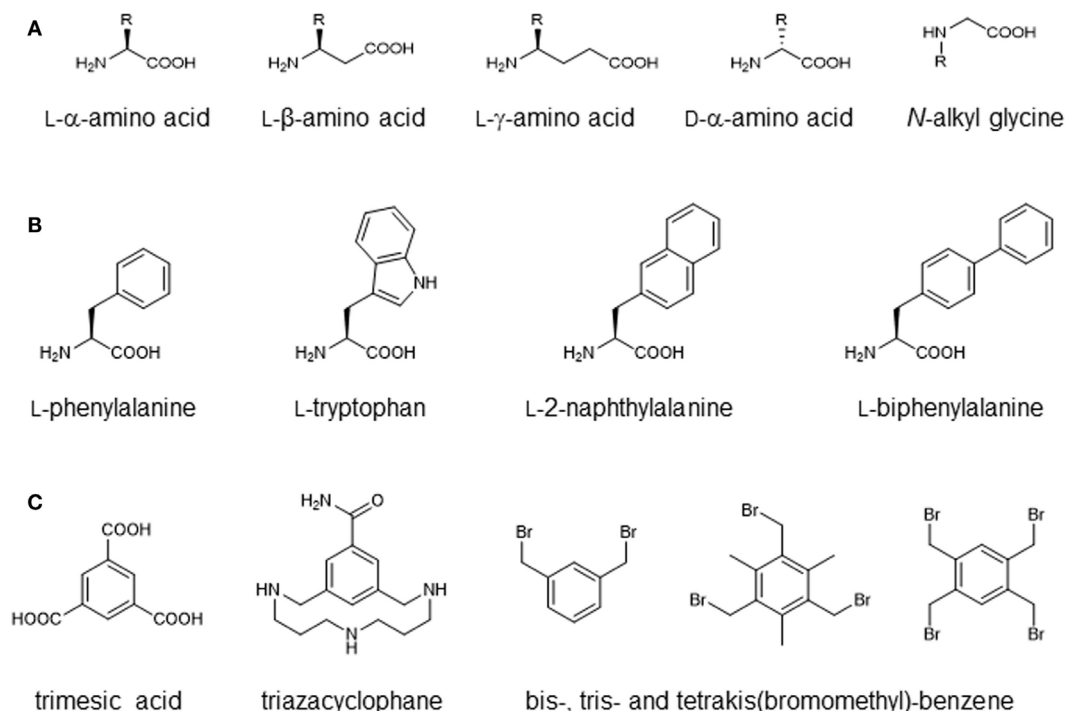


FIGURE 2 | Building blocks for chemical peptide synthesis. (A) Amino acid derivatives with modified backbone length and side-chain orientation. **(B)** Amino acid derivatives with modified aromatic side chains. **(C)** Scaffolds for multivalent or discontinuous peptide presentation.

Another possibility is the use of D-amino acids. Due to the chirality of the C α -atom, amino acids exist in two different stereoisomers (L and D). While recombinantly synthesized peptides and proteins are typically composed entirely of L-amino acids, chemical peptide synthesis can also use D-amino acids, which has been shown to increase the proteolytic stability while maintaining biological activity when D-amino acids are introduced at defined positions of an antimicrobial peptide (Hong et al., 1999b). At other positions, on the other hand, using D-amino acids instead of L-amino acids had the opposite effect due to structural damage to the peptide (Hong et al., 1999b).

Furthermore, oligomers of N-alkyl glycine monomers, termed peptoids, have been introduced as proteolytically stable peptide derivatives (Simon et al., 1992) (Figure 2A). As the amide hydrogen is missing in peptoids, the typical backbone hydrogen bonds present in proteins and peptides cannot be formed, altering the conformational preferences of these molecules. Peptoids have been used as mimics of antimicrobial peptides (termed as ampetoids) (Chongsiriwatana et al., 2008; Mojsoska et al., 2015) as well as novel therapeutics (Zuckermann and Kodadek, 2009).

In addition to alteration of the peptide backbone, the use of non-proteinogenic amino acids enables the introduction of chemical moieties that are not presented by the proteinogenic amino acids, and which can be used to dissect the binding mode of peptides. A prominent example is the substitution of aromatic side chains of phenylalanine or tryptophan with larger aromatic groups such as naphthyl or biphenyl (Figure 2B), which

increases the size and hydrophobicity of the side chain and affects π -stacking with the respective protein ligand (Muraki et al., 2000; Bachmann et al., 2011). Functionalized and orthogonally protected amino acids are often used for chemo selective ligation strategies (Tornøe et al., 2002; Kimmerlin and Seebach, 2005). In addition, lysine, among other amino acids, can be used for the synthesis of branched peptides (Franke et al., 2007). Furthermore, a range of scaffold molecules, such as trimesic acid derivatives (Berthelmann et al., 2014), triazacyclophane derivatives (Opatz and Liskamp, 2001; Chamorro et al., 2009), bis-, tris-, and tetrakis(bromomethyl)-benzene (CLIPS technology) (Timmerman et al., 2005) (Figure 2C), as well as cyclic β -tripeptide derivatives (Seebach and Gardiner, 2008) have been introduced for the generation of multivalent peptides.

PROTEIN SECONDARY STRUCTURE MIMICS

The three-dimensional (3D) arrangement of proteins contains unstructured, as well as structured regions, in which peptide chains are organized into secondary structures, such as α -helices and β -sheets. As α -helices and β -sheets mediate protein folding and protein-protein interactions, they are related to various biochemical phenomena and diseases (Fairlie et al., 1998). These secondary structures are stabilized by hydrogen bonds between amide nitrogen and carbonyl oxygen atoms. Bullock et al. have

analyzed the full set of helical protein interfaces in the Protein Data Bank (Berman et al., 2000) and found that about 62% of the helical interfaces contribute to protein–protein interactions (Bullock et al., 2011). Although natural proteins contain less β -sheet structure than α -helical structure, β -sheets contribute to protein aggregation, as well as to protein–protein interactions. Thus, peptides that mimic α -helices and β -sheets of proteins are attractive targets for drug development and tools to explore protein binding mechanism. A range of α -helix and β -sheet mimics have been developed, which will be discussed below. The various strategies of mimicking protein-binding sites through secondary structure mimics have also been extensively reviewed recently (Pelay-Gimeno et al., 2015).

α -Helix Mimics

The α -helical conformation of a peptide can be stabilized, and even induced, by introducing covalent links between amino acid side chains at selected positions. These links can be formed by lactam (Ösapay and Taylor, 1992; Yu and Taylor, 1999; Sia et al., 2002; Yang et al., 2004; Mills et al., 2006) and disulfide bridges (Jackson et al., 1991; Leduc et al., 2003), triazole-based linkages (Scrima et al., 2010; Kawamoto et al., 2011; Madden et al., 2011), and hydrocarbon staples (Blackwell and Grubbs, 1998; Schafmeister et al., 2000; LaBelle et al., 2012; Verdine and Hilinski, 2012; Brown et al., 2013; Chang et al., 2013; Nomura et al., 2013; Walensky and Bird, 2014; Chu et al., 2015). Replacing hydrogen bonds by salt bridges has been reported by Otaka et al. as an alternative means of stabilizing α -helices (Otaka et al., 2002). Further examples for hydrogen bond surrogates include cation– π interaction (Olson et al., 2001; Shi et al., 2002; Tsou et al., 2002) and π – π interaction (Albert and Hamilton, 1995).

Foldamers are a very prominent class of α -helix mimetic peptides. They are composed of β -amino acid (Seebach and Matthews, 1997; Gellman, 1998; Cheng et al., 2001; Martinek and Fulop, 2003), α/β -amino acid oligomers (Johnson and Gellman, 2013), or *N*-substituted glycine residues (peptoids) (Sun and Zuckermann, 2013). Such foldamers have been shown to inhibit the proteolytic activity of γ -secretase (Imamura et al., 2009), an enzyme that is involved in the processing of amyloid- β (A β) in Alzheimer's disease, by blocking the initial substrate binding site of γ -secretase (Lichtenthaler et al., 1999). For these foldamers, the conformationally constrained β -amino acid *trans*-2-aminocyclopentanecarboxylic acid (ACPC) was used as a building block. As such α -helix mimics can increase α -helicity, stability, and cell-permeability, they are increasingly attracting the attention both in academia and the pharmaceutical industry as candidates for novel therapeutics. Apart from biomedical use, α -helical peptide mimics are also of interest as biomaterials, such as self-assembling nanotubes (Burgess et al., 2015) and hydrogels (Mehrban et al., 2015).

β -Sheet Mimics

In β -sheets, two or more β -strands are connected via loops or turns, and the parallel or antiparallel orientation of β -strands is stabilized by hydrogen bonds between carbonyl oxygen atoms in one strand and amide nitrogen atoms of the opposite strand. Methods to mimic turn structures include macrocyclization

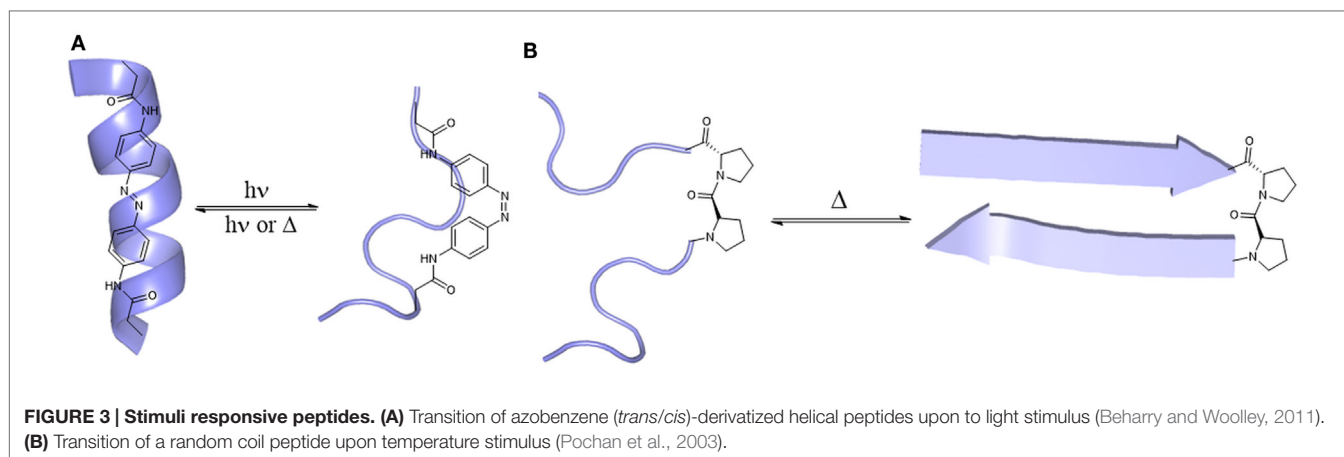
as well as the use of turn-inducing building blocks, such as a dipeptide of D-proline and L-proline (Robinson, 2008), or α -aminoisobutyric acid in combination with either a D- α -amino acid or an achiral α -amino acid (Aravinda et al., 2002; Masterson et al., 2007). One noteworthy example for macrocyclization used cyclic cysteine ladders of θ -defensin as a scaffold to stabilize a turn structure (Conibear et al., 2014). The cyclic cysteine ladder of θ -defensin comprises two antiparallel β -strands connected via two β -turns, and has a high thermal and serum stability. Grafting of the integrin-binding peptide Arg–Gly–Asp (RGD) onto this molecule resulted in 10-fold increase in affinity to integrin, illustrating the utility of θ -defensin as a molecular scaffold.

It has been difficult to develop robust chemical models of β -sheets, which tolerate a wide range of amino acid sequences because amyloidogenic sequences vary enormously and folding of β -sheet mimics depends on their amino acid sequences. Woods et al. overcame this problem by using 42-membered rings, which contain two strands connected via two δ -linked ornithine turns (Woods et al., 2007). Forty-two-membered ring macrocyclic β -sheets present a pentapeptide β -strand on one side (recognition strand), while the other β -strand contains the unnatural amino acid Hao (5-hydrazino-2-methoxybenzoic acid) and two α -amino acids. The relatively rigid structure of Hao-containing peptides preserves the structure of the recognition strand, and at the time serves as a template for the recognition strand. Furthermore, Hao is useful for the intermolecular β -sheet interaction to form fibril-like assembled oligomers (Pham et al., 2014). Similar to α -helical peptides, β -sheet mimics have also been used for biomaterials, such as nanotubes (Hamley, 2014).

STIMULI RESPONSIVE PEPTIDES IN BIOMATERIAL ENGINEERING

Some peptides are able to be structurally rearranged in response to external stimuli, such as temperature, pH, ionic strength, and presence of special ions and light. In 2006, Mart et al. (2006) reviewed different responsive systems based on peptides and their applications, including switchable surfaces, nanoparticle (dis)-assembly, hydrogel-formation, metal ion sensing, and electron transfer. In addition, special applications in medicine, such as drug delivery, tissue engineering, tissue regeneration, wound healing, and nerve cell regrowth rely upon stimuli-responsive peptides. Several conformational transitions of peptides have been reported, ranging from α -helix to random coil and *vice versa* or β -sheet to random coil and *vice versa*, among others. In this review, two selected examples are presented.

One example is the use of an azobenzene moiety as light-sensitive switch (Woolley, 2005; Renner and Moroder, 2006). As a photoswitchable device, azobenzene, which is more stable in the *trans*-conformation, can switch into *cis*-conformation upon irradiation with light at 340 nm, leading to a 3.5 Å shortening of the C–C-distance of azobenzene (Fliegl et al., 2003; Beharry and Woolley, 2011). Incorporation of the reactive azobenzene derivative 3,3'-bis(sulfonate)-4,4'-bis(chloroacetamido)azobenzene at defined positions of the sequence can result either in a loss of helical conformation (positions *i*, *i* + 11, **Figure 3A**) or in helix



stabilization (positions i , $i + 7$), upon light stimulus (Woolley, 2005). To make this approach more feasible for *in vivo* application, longer wavelengths should be used for azobenzene isomerization, considering UV-light scattering through cells and tissues. Samanta et al. (2013) recently reported an azobenzene derivative that can be switched using red light (630–660 nm), enabling the development of photo-switchable compounds for *in vivo* use.

Another example of stimuli-responsive peptides is the temperature-dependent formation of hydrogels by β -sheet peptides. Pochan et al. (2003) designed a β -hairpin mimic called Max3 (**Figure 3B**) that undergoes gelation upon heating ($T_{\text{gel}} = 60^\circ\text{C}$), which was completely reversible while cooling. This peptide is composed of alternating non-polar and polar amino acids bridged via a type II' β -turn. Other peptides undergo non-reversible hydrogelation when heated (Max1, Max2) (Pochan et al., 2003). These β -hairpin peptides were the starting point for the design of peptides whose folding can be triggered by UV light (Haines et al., 2005), changes in pH (Rajagopal et al., 2009), or recognition of electronegative cancer cell membranes (Sinthuvanich et al., 2012). Because of their biocompatibility, biodegradability, weak immunogenicity and selectivity, peptidic hydrogels can serve as potential cancer drugs and antimicrobials, as well as for wound healing (Mart et al., 2006; Branco et al., 2011).

PROTEIN MIMICS IN BIOMEDICAL RESEARCH

Current drug discovery and development approaches are focused on three different types of molecules (Craik et al., 2013; Fosgerau and Hoffmann, 2015). The traditional approach of using small molecules as drugs is still widely used. While small molecules have been shown to be excellent tools to block the catalytic site of enzymes, as well as the ligand binding sites of numerous receptors, they are less promising for the inhibition of protein–protein interactions, which often involve larger interfaces, which typically cannot be adequately addressed by small molecules. Therefore, protein-based drugs, so-called Biologics, are increasingly used as inhibitors of protein–protein interactions. Many proteins, however, have additional effector functions or binding sites for other ligands, causing problems in

in vivo applications. Furthermore, proteins can be immunogenic, resulting in immunological clearance before reaching their target site. As an alternative to both small molecule and protein-based drugs, peptides are becoming more relevant as drug candidates, as documented by an increasing number of peptide drugs approved for clinical use (Fosgerau and Hoffmann, 2015). Due to their potential for highly specific binding, combined with low immunogenicity, peptides are promising candidates as inhibitors of protein–protein interactions.

Specific protein–protein interactions are involved in the pathogenesis of numerous diseases. The design and generation of peptides that mimic the respective protein-binding site, as potential inhibitors of the interactions, is therefore a promising therapeutic strategy. Such mimetic molecules are typically designed based on the 3D structure of the protein–protein complex, which yields information on the location of the binding sites within the proteins, as well as the hot spot amino acids directly involved in the intermolecular interaction (Eichler, 2008). This general strategy will be illustrated here using examples of the various protein–protein interactions, which are involved in the entry of the human immunodeficiency virus type 1 (HIV-1) into cells. Furthermore, a range of protein-mimicking peptides used in the treatment of cancer and as antibiotics or anti-inflammatory compounds, will be reviewed.

Peptides as Mimics of the Viral Spike of HIV-1

The highly active antiretroviral therapy (HAART) has been a breakthrough in the treatment of HIV-1 infection, leading to an effective reduction of morbidity and mortality through drastic suppression of viral replication and, hence, reduction of plasma HIV-1 viral load. HAART consists of a mixture of at least three different drugs with at least two different molecular targets [for details see Arts and Hazuda (2012)]. Almost all of these drugs are small molecules that address intracellular targets. Due to the high genetic variability of HIV-1, the virus is able to rapidly become resistant against drugs. Therefore, there is an ongoing need for new therapeutic strategies against HIV-1. One of these strategies is the prevention of HIV-1 entry into its host cell by blocking the interactions between viral and host proteins that are involved in

the entry process. This can be achieved by using peptides, which mimic the binding sites of the involved proteins.

Entry of HIV-1 into its host cells is initiated by a cascade of protein–protein interactions between the viral and host cell proteins. These interactions involve the trimeric viral spike, composed of glycoproteins gp120 and gp41, as well as the primary receptor CD4 and coreceptors CCR5 and CXCR4 on the host cell (Wilén et al., 2012).

The initial event of HIV-1 entry is an interaction of viral gp120 with the host receptor CD4. In contrast to the generally high genetic variability of HIV-1, the CD4-binding site of gp120 is highly conserved. Peptides mimicking the CD4-binding site are therefore promising candidates as HIV-1 entry inhibitors. Furthermore, as the epitopes of various broadly neutralizing anti-HIV-1 antibodies have been shown to overlap the CD4-binding site, this part of gp120 is an immunogen candidate for the generation of HIV-1 neutralizing antibodies. Based on the X-ray structure of gp120 in complex with CD4 (Kwong et al., 1998) (**Figure 4A**), novel peptides that mimic the CD4-binding site have been developed (**Figure 4**) (Franke et al., 2007; Chamorro et al., 2009). A special characteristic of these peptides is the fact that they present three sequentially discontinuous fragments of the gp120 sequence, either in linear form, or as cyclic loops, on molecular scaffolds, such as a branched peptide composed of spacer amino acids, CD4bs-M (**Figure 4A**), and a triazacyclophane scaffold (**Figure 4B**). While the triazacyclophane scaffold peptide did not affect HIV-1 infection (Chamorro et al., 2009),

CD4bs-M was surprisingly found to strongly enhance HIV-1 infection of both CD4 positive and CD4 negative cells, and this effect could be linked to a strong tendency of the peptide to assemble into amyloid fibrils (Groß et al., 2015b).

Understanding the molecular and structural details of the interaction of antibodies with their viral antigens is an important step in the quest for a still elusive HIV-1 vaccine (Burton et al., 2012). A prominent class of anti HIV-1 antibodies recognizes the V3-loop of the gp120 protein (Zolla-Pazner and Cardozo, 2010), which forms a β -hairpin structure when in the antibody-bound state (**Figures 5A,B**). Robinson et al. were able to stabilize this β -hairpin structure in V3-loop peptides by grafting them on to a D-Pro-L-Pro scaffold (Riedel et al., 2011; Robinson, 2013) (**Figure 5C**). Coupling of such a stabilized V3-loop mimic to a lipopeptide carrier, which self-assembles into virus-like particles (Ghasparian et al., 2011), resulted in increased immunogenicity, enabling an alternative, carrier-independent immunization. Phage display peptide libraries (Smith, 1985) have often been used to identify peptides that bind to antibodies and thus mimic their epitopes (mimotopes). Mimotopes of the broadly neutralizing HIV-1 antibody b12 have been found (Boots et al., 1997) this way. As the viral spike proteins gp120 and gp41 are presented as trimers, Schellinger et al. (2011) generated a potential immunogen based on a trimer of the b12 mimotope in conjunction with a T-helper cell epitope peptide (**Figure 6**). This trimeric peptide bound to b12 substantially better than the monomeric mimotope, illustrating the importance of trimeric presentation,

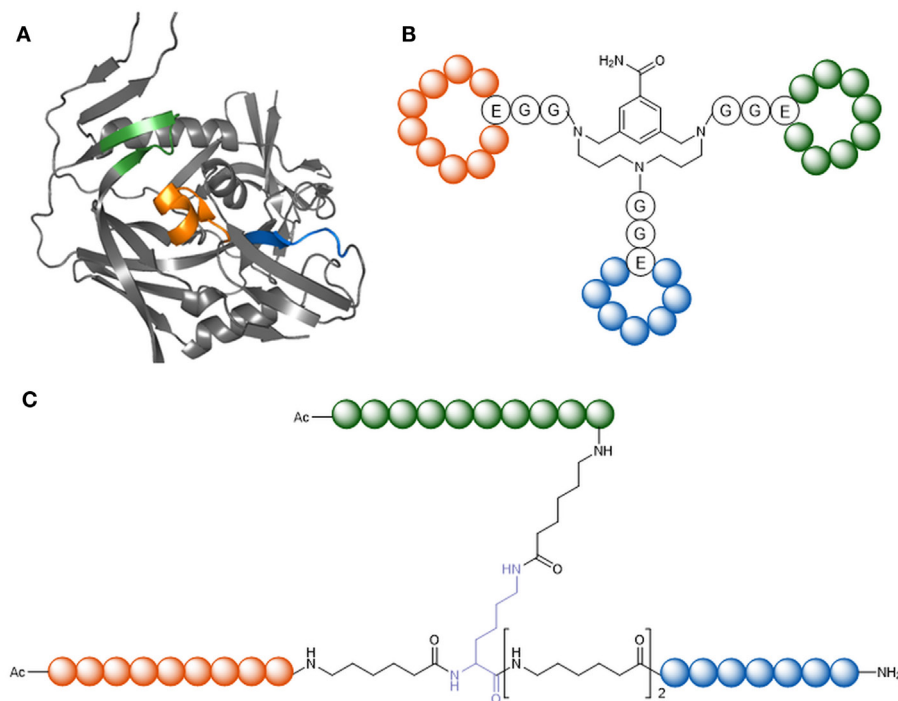


FIGURE 4 | Peptide mimics of the CD4-binding site of gp120. (A) X-ray structure of gp120 in the CD4-bound conformation [1GC1 (Kwong et al., 1998)].

Highlighted in orange, blue, and green are the fragments forming the discontinuous CD4 binding site. **(B)** CD4bs-M (Franke et al., 2007). **(C)** CD4-binding site mimic with triazacyclophane scaffold (Chamorro et al., 2009).

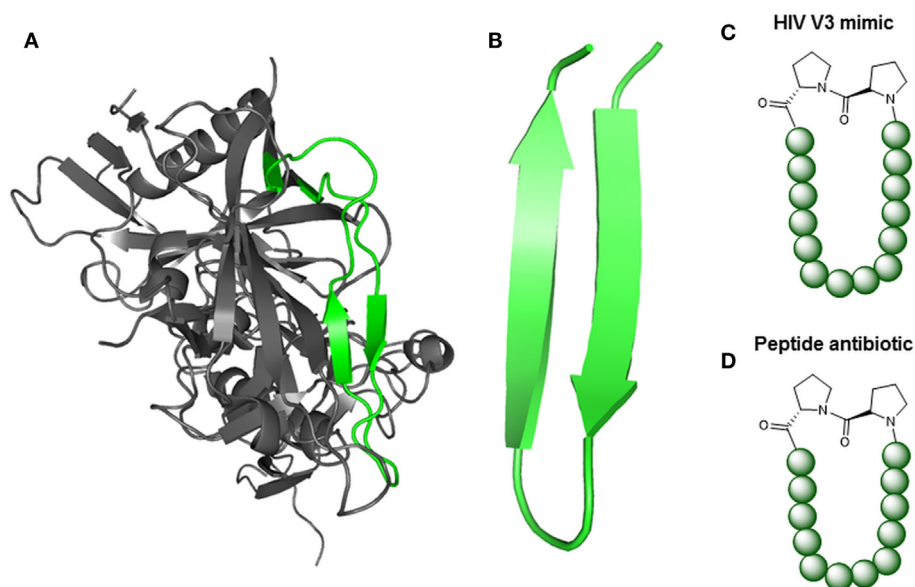


FIGURE 5 | Peptide mimics of turn structures. (A) X-ray structure of gp120 [pdb 4TVP (Pancera et al., 2014)] with highlighted V3-loop (green). **(B)** NMR structure of protegrin 1 from porcine leukocytes [pdb 1PG1 (Fahrner et al., 1996)]. **(C)** V3-loop mimic, stabilized via D-Proline and L-Proline (Riedel et al., 2011). **(D)** Protegrin 1 mimic (L27-11), stabilized via D-Proline and L-Proline (Srinivas et al., 2010).

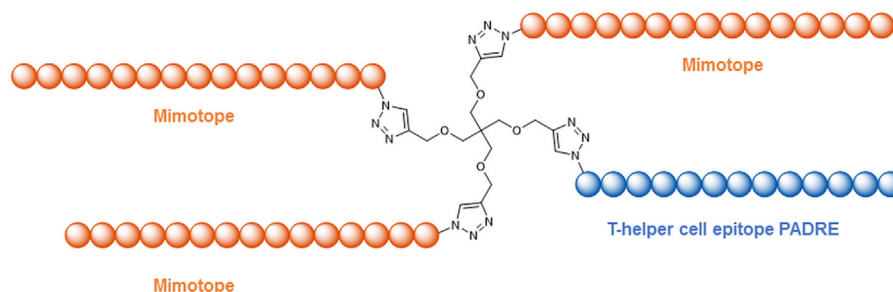


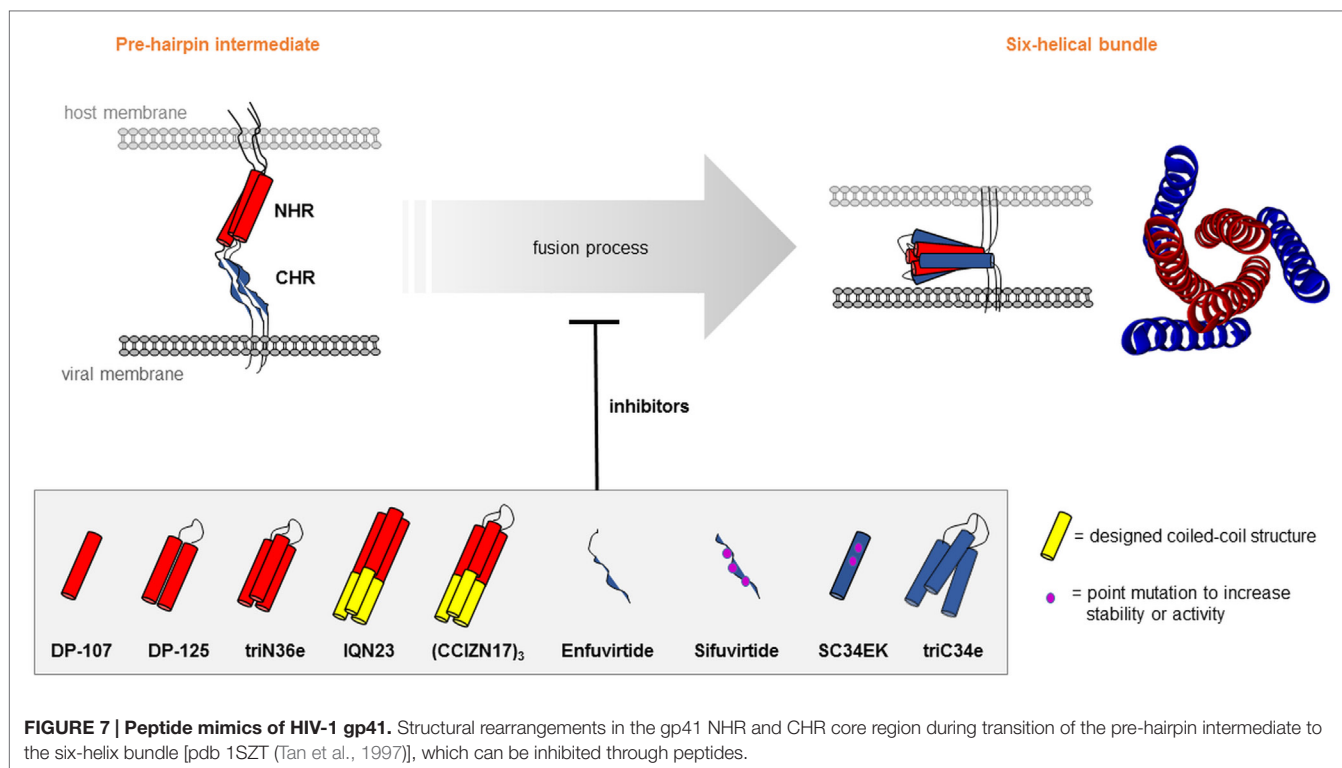
FIGURE 6 | Trimeric presentation of a b12 mimotope in conjunction with a T-helper cell epitope (Schellinger et al., 2011).

which was achieved using the so-called click reaction (Rostovtsev et al., 2002) as a chemoselective ligation reaction.

In addition to gp120 mimetic peptides, peptides that present parts of gp41 are also intensively researched (Cai et al., 2011). In particular this applies to peptides that mimic a six-helix bundle, consisting of a three-stranded coiled-coil structures formed by an N-terminal (NHR) and a C-terminal (CHR) heptad repeat of gp41 (Chan et al., 1997). This region of gp41 plays a key role in the process of fusion of the viral and cellular membranes (Figure 7). Peptides presenting parts of the six-helical bundle are thought to be able to interfere with its correct formation and, consequently, inhibit virus-cell fusion. Already in 1992, Wild et al. (1992) described an approach to mimic the secondary structure of NHR, which was predicted to be α -helical. Using CD spectroscopy, it could be shown that the NHR-mimetic peptide forms a stable α -helix under physiological conditions. Furthermore, the peptide

exhibited a strong anti-HIV-1 activity, which could be further enhanced through dimerization. Trimers of the NHR-mimetic peptide were later found to be better HIV-1 entry inhibitors than the respective monomeric peptide (Nakahara et al., 2010). Covalent stabilization of such peptide trimers through inter-chain disulfide bridges dramatically increased the antiviral potency (Bianchi et al., 2005), as well as the HIV-1 neutralizing capacity of anti-peptide antisera (Bianchi et al., 2010).

Similar to the NHR mimics, peptides mimicking the CHR region of gp41 were developed to inhibit the formation of the six-helical bundle. In 1994, Wild et al. (1994) demonstrated a strong anti-HIV-1 activity of a peptide that overlaps the CHR. Later on, the first and so far only HIV-1 fusion inhibitor approved for clinical use (Enfuvirtide) was developed based on this peptide (Kilby and Eron, 2003; Lalezari et al., 2003). Another fusion inhibitor, called Sifuvirtide, was developed based on the 3D structure of HIV-1



gp41 and computer modeling (He et al., 2008; Wang et al., 2009). Sifuvirtide could effectively block six-helical bundle formation and was active even against Enfuvirtide-resistant HIV-1 strains. Otake et al. increased the α -helicity of a CHR mimetic peptide by introducing Glu–Lys pairs at the i and $i + 4$ positions of the helix (Otake et al., 2002), which greatly enhanced the solubility and stability of the peptide. Trimeric presentation of a CHR mimetic peptide on a C_3 -symmetric scaffold dramatically increased the antiviral activity of the peptide (Nomura et al., 2012).

Peptides as Mimics of Cellular Receptors

Cellular receptors play important roles in signal transduction pathways, as well as in viral entry. As discussed in the previous chapter, HIV-1 contacts two receptors on the host cell surface prior to fusion with the cell membrane. Peptides that mimic these receptors are useful tools to explore the details of virus infection mechanism, as well as to develop new drugs against HIV-1. In 1998, Drakopoulou et al. (1998) developed a peptidic CD4 mimic, called CD4M, based on the analysis of site-directed mutagenesis studies, antibody-blocking experiments and the structure of the extracellular fragment of CD4, which identified the CDR H2-like loop of CD4 as the binding site for gp120 of HIV-1. To retain the native structure of the CDR H2-like loop, the peptide was transferred onto a scorpion toxin, which served as a structural scaffold. Optimizing CD4M led to a variant with 100-fold increased affinity to gp120, as well as infection-inhibitory activity (Vita et al., 1999). Based on the X-ray structure of CD4 in complex with gp120 (Kwong et al., 1998), Martin et al. (2003) further optimized the CD4 mimic, resulting in a 27-mer peptide mimicking the CD4 binding site for gp120. This peptide was

able to bind to gp120 at low nanomolar concentrations, inhibit binding of CD4 to gp120, as well as to induce conformational changes in gp120 similar to those triggered by CD4, from which it was derived. The importance of conformational stability of CD4 mimetic peptides could be further confirmed by Meier et al. (2012). Peptides that present the binding site of CD4 for gp120 were covalently stabilized in their loop structure by cyclization through a disulfide bond between the N- and C-terminus. Using alanine and D-phenylalanine substitution analogs, the importance of the hot spot amino acid phenylalanine 43 could be confirmed at the peptide level. These results were further confirmed by molecular dynamics simulations.

The concept of mimicking protein-binding sites through complex synthetic peptides has recently been extended to peptides that mimic the extracellular domains of seven transmembrane G protein-coupled receptors (GPCRs), which is composed of the N-terminus (NT) and the three extracellular loops (ECLs). GPCRs make up the largest class of drug targets, in fact, 27% of all clinically used drugs target a GPCR.

In the context of HIV-1 infection, two GPCRs are important, i.e., the chemokine coreceptors CCR5 and CXCR4. Although the 3D structures of both receptors are available (Wu et al., 2010; Tan et al., 2013), our knowledge of the structural details of their interaction with HIV-1 gp120 remains limited. Therefore, peptides that mimic the binding site of these receptors for gp120 could be useful tools for the exploration of HIV-1-coreceptor interaction at the molecular level. We have generated a peptide that mimics the three ECLs of CXCR4 (Möbius et al., 2012) (Figure 8A). This peptide, named CX4-M1, is able to discriminate between CXCR4- and CCR5-recognizing gp120 (Möbius

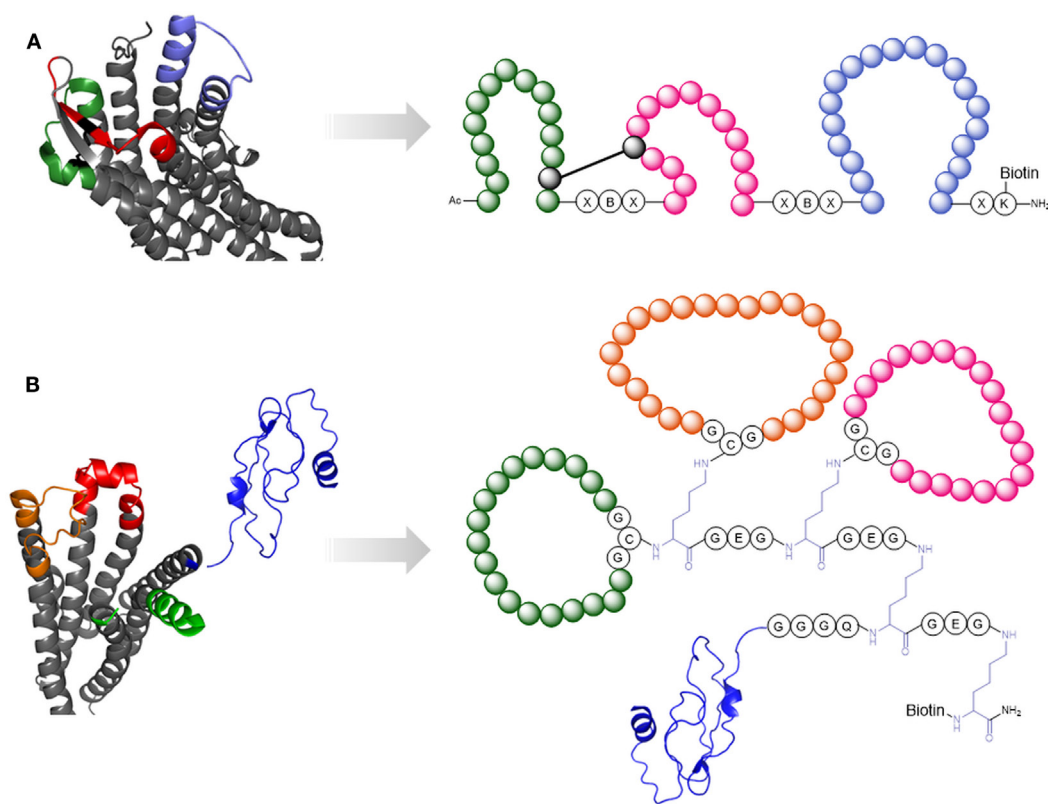


FIGURE 8 | Peptide mimics of cellular receptors. (A) CXCR4 [pdb 3ODU (Wu et al., 2010)]. The extracellular loops (highlighted in green, red, and purple) are presented by the mimetic peptide CX4-M1 (Möbius et al., 2012). **(B)** GPCR CRF1 [pdb 4K5Y (Hollenstein et al., 2013)] and NMR structure of its N-terminus [pdb 2L27 (Grace et al., 2010)]. The extracellular loops are highlighted in orange, red, and green. The N-terminus is depicted in blue. These sequence stretches are presented in the CRF1 mimetic peptide (Pritz et al., 2008).

et al., 2012) and V3-loop peptides mimicking the corresponding binding site on gp120 (Groß et al., 2013), and also inhibits HIV-1 infection of susceptible target cells in a CXCR4-specific manner (Möbius et al., 2012; Groß et al., 2015a). Furthermore, CX4-M1 is recognized by the natural CXCR4-ligand, i.e., the chemokine CXCL12 (also called SDF-1 α), as well as anti-CXCR4-antibodies (Groß et al., 2015a).

In a similar approach, Pritz et al. (2008) generated, via a combination of recombinant, enzymatic and chemical synthesis, a molecule that mimics the extracellular domain of the corticotropin-releasing factor receptor type 1 (CRF1) (**Figure 8B**). Improving the scaffold for the presentation of the ECLs and N-terminus, as well as increasing the overall yields through synthesis optimization, enabled structural analysis of the receptor mimic – ligand interaction through NMR spectroscopy (Abel et al., 2014).

The epidermal growth factor receptor (EGFR), which is a key protein of cell proliferation and differentiation (Yarden and Sliwkowski, 2001), has also been subject to structure-based design of receptor mimetic peptides. As the receptor forms dimers or even oligomers, Hanold et al. (2015) generated a peptide mimic of the EGFR dimerization arm, which forms a β -hairpin in the native conformation. This peptide was stabilized *via* a triazole

crosslink to increase proteolytic stability, while retaining the native structure, resulting in inhibition of EGFR dimerization and, consequently, a reduction of cell viability. Sequence and functional optimization of EGFR mimetic peptides may be useful for the development of novel cancer drugs addressing EGFR overexpression in tumors.

Peptides in Cancer Research

The uncontrolled growth and spread of cells into tumor tissue (Vogelstein and Kinzler, 2004) defines cancer as one of the main fatal diseases worldwide. Therefore, a major focus in peptide drug development is on oncology (Kaspar and Reichert, 2013; Fosgerau and Hoffmann, 2015). Apart from using peptides directly as anticancer drugs (Thundimadathil, 2012), they can also serve as targeting agents to direct highly toxic chemotherapeutics to their respective targets, reducing the systemic toxicity of these drugs [for details see Kaspar and Reichert (2013)].

Structure-based approaches are often used in the design of anticancer peptides, such as the inhibitor of cell migration and invasion published by Bifulco et al. (2008). The urokinase-type plasminogen activator receptor (uPAR), which plays a critical role in cancer cell growth, survival, invasion and metastasis, contains a five amino acid sequence (SRSRY) between two of its

three domains, which is exposed through ligand binding, and mediates chemotactic properties of uPAR. Using the pentapeptide SRSRY as a template, glutamine scanning and insertion of a pyroglutamine (pE) resulted in the identification of the peptide pERERY-NH₂ as a highly active uPAR inhibitor. Further optimization through structure-based design leads to the tetrapeptide Ac-RERF-NH₂, which is 500- to 1000-fold more active than pERERY-NH₂ (Carriero et al., 2009). Ac-RERF-NH₂, which has a high propensity to adopt an α -turn structure, represents a promising drug candidate against cancer.

An important target for the therapy of pancreatic, gastric, and colorectal tumors is gastrin, a peptide hormone, whose activity can be blocked by antibodies that recognize gastrin as their epitope, delaying tumor growth (Watson et al., 1996; Barderas et al., 2008b). Detailed analysis of the antibody epitopes through alanine scanning of gastrin (Barderas et al., 2008b) and docking of the epitope into the antibody binding site, followed by affinity maturation through phage display and *in silico* methods (Barderas et al., 2008a) resulted in the development of antibody fragments with enhanced potency to inhibit gastrin-induced tumor growth. With the aim to shrink these antibodies to the size of peptidomimetics, Timmerman et al. (2005, 2009) used a strategy, in which up to three peptides derived from the complementary-determining regions (CDRs) of an antibody are presented in one molecule using the CLIPS strategy (see Toolbox for peptide synthesis: non-proteinogenic amino acids and site-selective ligation). In most cases, the activity of the obtained peptides was much lower compared to the parent antibodies. Nevertheless, neutralization of gastrin in cell-based assays by the mimetic peptides could be demonstrated (Timmerman et al., 2009). The mode of action of the peptides, however, may be different from that of the parent antibodies (Timmerman et al., 2010), leading to the conclusion that further efforts in peptide design have to be made.

Small GTPases, such as Ras, Rab, and Rho, are key proteins in many cancers, as malfunction of these proteins results in abnormal cell growth and differentiation, prolonged cell survival, membrane trafficking, and vesicular transport (Bourne et al., 1990; Cherfils and Zeghouf, 2013). Inhibiting the activity of these small GTPases could lead to new chemotherapeutic drugs for cancer treatment. One strategy to achieve this is to address the GDP-GTP exchange of Ras, which is the rate-limiting step and requires interaction with the Ras-specific guanine nucleotide exchange factor Sos (Konstantinopoulos et al., 2007). In 2011, Patgiri et al. (2011) published the structure-based design of an α -helical peptide derived from the Sos-Protein, which is able to inhibit Sos-mediated Ras activation through interference with the Sos-Ras interaction, providing a promising lead compound for anti-cancer drugs. Likewise, peptide mimics of the Rab ligands R6IP, LidA, REP1, and Rabin8 have been reported (Spiegel et al., 2014). Using the hydrocarbon-peptide stapling approach, α -helical peptides were stabilized at positions *i* and *i* + 4, resulting in up to 200-fold increased affinity of the peptide to Rab proteins. In addition, one of the peptides, being a pioneer inhibitory compound for Rab GTPase-protein interactions, was found to inhibit the Rab8a-effector interaction.

A challenge in cancer drug delivery is the discrimination between self and non-self, i.e., clearance of drug-loaded nanoparticles

before they reach their target. To overcome this problem, synthetic polymers such as polyethylene glycol are used, but these can hamper uptake by cancer cells (Hong et al., 1999a). As an alternative strategy, Rodriguez et al. (2013) generated, based on the crystal structure of the hCD47-hSIRP α complex, and in combination with computational simulations, a minimal “self” 21-mer peptide. This peptide, which originates from CD47, an established marker of “self” (Rodriguez et al., 2013), was able to prolong the circulation of nanobeads in mice by preventing phagocytosis, providing a new opportunity for enhanced delivery of drugs or imaging agents. As an example for its utility as a marker of “self,” the anti-cancer drug paclitaxel was loaded onto nanoparticles, which also presented the marker-peptide on their surface. Due to delayed clearance, treatment with peptide-coated nanoparticles induced a more efficient size-reduction of lung adenocarcinoma epithelial tumors in mice than beads without the peptide. Although this peptide is not the bio-active compound, it provides an excellent tool for the delivery of drugs to tumor tissues.

Peptides are also promising candidates for cancer immunotherapy, where they are used as vaccines that present tumor-associated antigens, which trigger an immune response against the tumor in the patient. It can be expected that peptides presenting tumor-associated antigens will increasingly gain significance for cancer immunotherapy in the future (Miller et al., 2013).

Peptides as Antibiotics and Anti-Inflammatory Compounds

The growing multi-resistance of bacteria to clinically used antibiotics is one of the current challenges in biomedical research (Dennesen et al., 1998). The development of new antibacterial drugs is therefore an urgent necessity, and peptides have proven beneficial in this area of drug development as well. Robinson et al. (2005) could demonstrate improved antimicrobial activity, as well as plasma half-life of β -hairpin mimics of the naturally occurring membranolytic host-defense peptide protegrin 1 (Shankaramma et al., 2002; Srinivas et al., 2010) (Figure 5D). These peptides were cyclized via a D-proline-L-proline template, reducing flexibility and stabilizing the conformation of the peptide (Shankaramma et al., 2002; Robinson et al., 2005; Srinivas et al., 2010). Furthermore, these peptides were shown to directly interact with the bacterial β -barrel protein LptD, which sets them apart from other antimicrobial peptides, whose effect is mainly based on a membranolytic activity.

An anti-inflammatory peptide, named CHOPS (Bunschoten et al., 2011) (Figure 9), was designed based on the structure of the chemotaxis inhibitory protein of *Staphylococcus aureus* (CHIPS) (Veldkamp et al., 2000; Haas et al., 2005; Ippel et al., 2009). CHIPS is known to bind to the C5a-receptor and to inhibit the C5a-C5a-receptor interaction (Postma et al., 2004), thus addressing an important element in the complement cascade of the innate immunity. As full-length CHIPS is highly immunogenic (Gustafsson et al., 2009), its peptide mimic CHOPS, whose conformation is similar to the respective CHIPS fragment, and which binds to the N-terminus of the C5a-receptor (Bunschoten et al., 2011), may become a promising alternative for the treatment of inflammatory and autoimmune diseases.

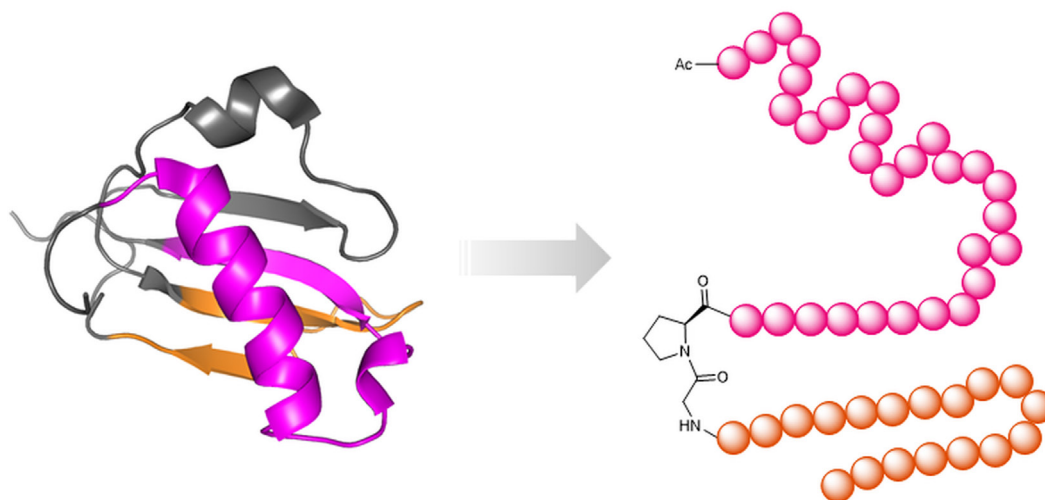


FIGURE 9 | Peptide mimics of an anti-inflammatory protein. Left: NMR structure of the chemotaxis inhibitory protein of *S. aureus* [CHIPS31-121 pdb 1XEE (Haas et al., 2005)] with highlighted discontinuous binding site for the C5a receptor, which was mimicked through the peptide CHOPS (Bunschoten et al., 2011) (right).

Proteins in the outer membrane of Gram-negative bacteria often have β -barrel structures. The proper assembly of these proteins is provided for by the β -barrel assembly machine (Bam) (Hagan et al., 2011). One important component of the Bam multiprotein complex is BamD, which interacts with unfolded protein substrates, like BamA, and facilitates their assembly in the outer membrane (Hagan et al., 2013). Using a peptide scanning approach of the C-terminal region of BamA, a 15-mer peptide was identified as an inhibitor of outer membrane protein assembly (Hagan et al., 2015). *In vivo* expression of this peptide resulted in bacterial growth defects, and sensitized resistant *Escherichia coli* to antibiotics, marking a starting point for the development of new antibiotic compounds for gram-negative bacteria (Hagan et al., 2015).

It should also be noted that a plethora of antimicrobial peptides are found in numerous organisms, including insects, mammals, plants, and bacteria (Mojsoska et al., 2015), which are not subject of this review. Furthermore, computer-based design strategies are aimed at the design of antimicrobial peptides with improved activity and reduced mammalian cell toxicity (Fjell et al., 2012).

CHALLENGES AND FUTURE DIRECTIONS

Due to their intrinsic properties, such as their potential for highly specific interactions with target molecules, generally low toxicity and immunogenicity, and rapid clearance, peptides are increasingly appreciated as candidates for novel drugs. This is particularly true for the development of protein–protein interaction inhibitors, where peptides are often better able than small molecules to cover large protein interface areas.

On the other hand, peptides also present severe bottlenecks that need to be considered and, if necessary, addressed in the development of peptide drugs. The biggest challenge clearly is the limited

metabolic stability of peptides, since they are rapidly degraded by proteolytic enzymes, precluding oral administration of peptide drugs. This challenge can be addressed by different means. First, unlike recombinant protein synthesis, chemical peptide synthesis is not limited to the proteinogenic amino acids as building blocks. A plethora of additional amino acids are currently available for chemical peptide synthesis. Apart from dramatically increasing the metabolic stability of peptides, incorporation of these amino acids also increases the chemical diversity presented by synthetic peptides, as these additional amino acids introduce chemical moieties that are not presented by the proteinogenic amino acids. Furthermore, conformational stabilization through cyclization, or through introduction of defined secondary structures, has been shown to shield peptides from proteolytic enzymes. Such shielding effects can also be achieved by coupling the peptide to larger inert molecules, such as polyethylene glycol (Swierczewska et al., 2015).

Due to their molecular size, peptides are rarely able to passively pass cell membranes, limiting their utility to address intracellular target molecules. This drawback, however can be counteracted by attaching the drug peptide to one of a large group of available cell-penetrating peptides (Kurrikof et al., 2015), which are able to transport a variety of molecular cargo into cells.

In general, the chemical synthesis of peptides through solid-phase synthesis is fairly straightforward and has been optimized over the past decades, so that virtually all peptide sequences are accessible synthetically today. In our experience, however, the synthesis of specific peptides may require the use of specific protected amino acids and other building blocks, solid supports, linkers and other reagents, which significantly increases the cost of synthesis. These considerations may become relevant for the large-scale synthesis of peptide drugs, as well as peptide biomaterials.

The design of peptides as protein–protein interaction inhibitors is typically based on the resolved 3D structure of the respective

protein–protein complex. While such structures are increasingly becoming available through powerful x-ray crystallography technology, their generation is not trivial and contingent on the availability of suitable crystals of the protein complexes.

Overall, taking into account the tremendous technical and scientific progress in the field of using peptides as protein mimics, we strongly believe that the significance of synthetic peptides in biomedical research, as well as in biomaterial engineering, will continue to grow in the future.

CONCLUSION

The design of peptides as protein mimics has evolved as a promising strategy for the exploration of, as well as the controlled interference with, protein–protein interactions. Due to their chemical nature, peptides are an appropriate type of molecules for the mimicry of protein-binding sites, including those involving large protein–protein interfaces. The possibility to use non-proteinogenic amino acids, as well as various methods of chemical modification, greatly enhances the scope of chemical and structural versatility, as well as stability, of synthetic peptides. Apart from their significance as molecular tools to explore

protein–protein interactions, such protein mimetic peptides are also candidates for the inhibition of protein–protein interactions involved in disease processes. Furthermore, peptides play an important role in biomaterial engineering, as they are biocompatible, biodegradable, and functionally selective. Photo-switchable peptides can be used to temporally and/or spatially control processes in organisms, such as drug release at specific organs or tissues. These applications illustrate the utility and versatility of synthetic peptides as molecular tools in biomedical research, as well as in synthetic biology.

AUTHOR CONTRIBUTIONS

AG and CH have written individual chapters and prepared figures. HS and JE have written and edited the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge support through the Emerging Fields Initiative, Project Synthetic Biology, by the University of Erlangen-Nuremberg. CH was supported by a Research Fellowship from the Alexander von Humboldt Foundation.

REFERENCES

- Abel, S., Geltinger, B., Heinrich, N., Michl, D., Klose, A., Beyermann, M., et al. (2014). Semisynthesis and optimization of G protein-coupled receptor mimics. *J. Pept. Sci.* 20, 831–836. doi:10.1002/psc.2680
- Albert, J. S., and Hamilton, A. D. (1995). Stabilization of helical domains in short peptides using hydrophobic interactions. *Biochemistry* 34, 984–990. doi:10.1021/bi00003a033
- Aravinda, S., Shamala, N., Rajkishore, R., Gopi, H. N., and Balaram, P. (2002). A crystalline beta-hairpin peptide nucleated by a type I' Aib-D-Ala beta-turn: evidence for cross-strand aromatic interactions. *Angew. Chem. Int. Ed. Engl.* 41, 3863–3865. doi:10.1002/1521-3773(20021018)41:20<3863::AID-ANIE3863>3.0.CO;2-A
- Arts, E. J., and Hazuda, D. J. (2012). HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Prospect. Med.* 2, a007161. doi:10.1101/cshperspect.a007161
- Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G., and Kiefhaber, T. (2011). Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3952–3957. doi:10.1073/pnas.1012668108
- Barderas, R., Desmet, J., Timmerman, P., Meloen, R., and Casal, J. I. (2008a). Affinity maturation of antibodies assisted by in silico modeling. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9029–9034. doi:10.1073/pnas.0801221105
- Barderas, R., Shochat, S., Timmerman, P., Hollestelle, M. J., Martinez-Torrecuadrada, J. L., Hoppener, J. W., et al. (2008b). Designing antibodies for the inhibition of gastrin activity in tumoral cell lines. *Int. J. Cancer* 122, 2351–2359. doi:10.1002/ijc.23395
- Beharry, A. A., and Woolley, G. A. (2011). Azobenzene photoswitches for biomolecules. *Chem. Soc. Rev.* 40, 4422–4437. doi:10.1039/c1cs15023e
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Berthelmann, A., Lach, J., Grawert, M. A., Groll, M., and Eichler, J. (2014). Versatile C(3)-symmetric scaffolds and their use for covalent stabilization of the foldon trimer. *Org. Biomol. Chem.* 12, 2606–2614. doi:10.1039/c3ob42251h
- Bianchi, E., Finotto, M., Ingallinella, P., Hrin, R., Carella, A. V., Hou, X. S., et al. (2005). Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12903–12908. doi:10.1073/pnas.0502449102
- Bianchi, E., Joyce, J. G., Miller, M. D., Finnefrock, A. C., Liang, X., Finotto, M., et al. (2010). Vaccination with peptide mimetics of the gp41 prehairpin fusion intermediate yields neutralizing antisera against HIV-1 isolates. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10655–10660. doi:10.1073/pnas.1004261107
- Bifulco, K., Longanesi-Cattani, I., Gargiulo, L., Maglio, O., Cataldi, M., De Rosa, M., et al. (2008). An urokinase receptor antagonist that inhibits cell migration by blocking the formyl peptide receptor. *FEBS Lett.* 582, 1141–1146. doi:10.1016/j.febslet.2008.03.001
- Blackwell, H. E., and Grubbs, R. H. (1998). Highly efficient synthesis of covalently cross-linked peptide helices by ring-closing metathesis. *Angew. Chem. Int. Ed.* 37, 3281–3284. doi:10.1002/(SICI)1521-3773(19981217)37:23<3281::AID-ANIE3281>3.0.CO;2-V
- Boots, L. J., McKenna, P. M., Arnold, B. A., Keller, P. M., Gorny, M. K., Zolla-Pazner, S., et al. (1997). Anti-human immunodeficiency virus type 1 human monoclonal antibodies that bind discontinuous epitopes in the viral glycoproteins can identify mimotopes from recombinant phage peptide display libraries. *AIDS Res. Hum. Retroviruses* 13, 1549–1559. doi:10.1089/aid.1997.13.1549
- Bourne, H. R., Sanders, D. A., and McCormick, F. (1990). The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* 348, 125–132. doi:10.1038/348125a0
- Branco, M. C., Sigano, D. M., and Schneider, J. P. (2011). Materials from peptide assembly: towards the treatment of cancer and transmissible disease. *Curr. Opin. Chem. Biol.* 15, 427–434. doi:10.1016/j.cbpa.2011.03.021
- Brown, C. J., Quah, S. T., Jong, J., Goh, A. M., Chiam, P. C., Khoo, K. H., et al. (2013). Stapled peptides with improved potency and specificity that activate p53. *ACS Chem. Biol.* 8, 506–512. doi:10.1021/cb3005148
- Bullock, B. N., Jochim, A. L., and Arora, P. S. (2011). Assessing helical protein interfaces for inhibitor design. *J. Am. Chem. Soc.* 133, 14220–14223. doi:10.1021/ja206074j
- Bunschoten, A., Ippel, J. H., Kruijtz, J. A., Feitsma, L., De Haas, C. J., Liskamp, R. M., et al. (2011). A peptide mimic of the chemotaxis inhibitory protein of *Staphylococcus aureus*: towards the development of novel anti-inflammatory compounds. *Amino Acids* 40, 731–740. doi:10.1007/s00726-010-0711-3
- Burgess, N. C., Sharp, T. H., Thomas, F., Wood, C. W., Thomson, A. R., Zaccari, N. R., et al. (2015). Modular design of self-assembling peptide-based nanotubes. *J. Am. Chem. Soc.* 137, 10554–10562. doi:10.1021/jacs.5b03973
- Burton, D. R., Poignard, P., Stanfield, R. L., and Wilson, I. A. (2012). Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* 337, 183–186. doi:10.1126/science.1225416
- Cai, L., Gochin, M., and Liu, K. (2011). Biochemistry and biophysics of HIV-1 gp41 – membrane interactions and implications for HIV-1 envelope protein

- mediated viral-cell fusion and fusion inhibitor design. *Curr. Top. Med. Chem.* 11, 2959–2984. doi:10.2174/156802611798808497
- Carriero, M. V., Longanesi-Cattani, I., Bifulco, K., Maglio, O., Lista, L., Barbieri, A., et al. (2009). Structure-based design of an urokinase-type plasminogen activator receptor-derived peptide inhibiting cell migration and lung metastasis. *Mol. Cancer Ther.* 8, 2708–2717. doi:10.1158/1535-7163.MCT-09-0174
- Chamorro, C., Kruijtz, J. A., Farsarakis, M., Balzarini, J., and Liskamp, R. M. (2009). A general approach for the non-stop solid phase synthesis of TAC-scaffolded loops towards protein mimics containing discontinuous epitopes. *Chem. Commun. (Camb)* 7, 821–823. doi:10.1039/B817357E
- Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89, 263–273. doi:10.1016/S0092-8674(00)80205-6
- Chang, Y. S., Graves, B., Guerlavais, V., Tovar, C., Packman, K., To, K. H., et al. (2013). Stapled alpha-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3445–E3454. doi:10.1073/pnas.1303002110
- Cheng, R. P., Gellman, S. H., and Degrad, W. F. (2001). beta-Peptides: from structure to function. *Chem. Rev.* 101, 3219–3232. doi:10.1021/cr000045i
- Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* 93, 269–309. doi:10.1152/physrev.00003.2012
- Chongsirawatana, N. P., Patch, J. A., Czyzewski, A. M., Dohm, M. T., Ivankin, A., Gidalevitz, D., et al. (2008). Peptoids that mimic the structure, function, and mechanism of helical antimicrobial peptides. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2794–2799. doi:10.1073/pnas.0708254105
- Chu, Q., Moellering, R. E., Hilinski, G. J., Kim, Y. W., Grossmann, T. N., Yeh, J. T. H., et al. (2015). Towards understanding cell penetration by stapled peptides. *Medchemcomm* 6, 111–119. doi:10.1039/C4MD00131A
- Conibear, A. C., Bochen, A., Rosengren, K. J., Stupar, P., Wang, C., Kessler, H., et al. (2014). The cyclic cystine ladder of theta-defensins as a stable, bifunctional scaffold: a proof-of-concept study using the integrin-binding RGD motif. *Chembiochem* 15, 451–459. doi:10.1002/cbic.201300568
- Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.* 81, 136–147. doi:10.1111/cbdd.12055
- Dennesen, P. J., Bonten, M. J., and Weinstein, R. A. (1998). Multiresistant bacteria as a hospital epidemic problem. *Ann. Med.* 30, 176–185. doi:10.3109/07853899808999401
- Drakopoulou, E., Vizzavona, J., and Vita, C. (1998). Engineering a CD4 mimetic inhibiting the binding of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120 to human lymphocyte CD4 by the transfer of a CD4 functional site to a small natural scaffold. *Letts. Pep. Sci.* 5, 241–245. doi:10.1023/A:1008837427367
- Eichler, J. (2008). Peptides as protein binding site mimetics. *Curr. Opin. Chem. Biol.* 12, 707–713. doi:10.1016/j.cbpa.2008.09.023
- Fahrner, R. L., Dieckmann, T., Harwig, S. S., Lehrer, R. I., Eisenberg, D., and Feigon, J. (1996). Solution structure of protegrin-1, a broad-spectrum antimicrobial peptide from porcine leukocytes. *Chem. Biol.* 3, 543–550. doi:10.1016/S1074-5521(96)90145-3
- Fairlie, D. P., West, M. L., and Wong, A. K. (1998). Towards protein surface mimetics. *Curr. Med. Chem.* 5, 29–62.
- Fjell, C. D., Hiss, J. A., Hancock, R. E., and Schneider, G. (2012). Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* 11, 37–51. doi:10.1038/nrd3591
- Fliegler, H., Kohn, A., Hattig, C., and Ahlrichs, R. (2003). Ab initio calculation of the vibrational and electronic spectra of trans- and cis-azobenzene. *J. Am. Chem. Soc.* 125, 9821–9827. doi:10.1021/ja034433o
- Fosgerau, K., and Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discov. Today* 20, 122–128. doi:10.1016/j.drudis.2014.10.003
- Frank, R. (2002). The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J. Immunol. Methods* 267, 13–26. doi:10.1016/S0022-1759(02)00137-0
- Franke, R., Hirsch, T., Overwin, H., and Eichler, J. (2007). Synthetic mimetics of the CD4 binding site of HIV-1 gp120 for the design of immunogens. *Angew. Chem. Int. Ed. Engl.* 46, 1253–1255. doi:10.1002/anie.200603274
- Gellman, S. H. (1998). Foldamers: a manifesto. *Acc. Chem. Res.* 31, 173–180. doi:10.1021/ar960298r
- Ghasparian, A., Riedel, T., Koomullil, J., Moehle, K., Gorba, C., Svergun, D. I., et al. (2011). Engineered synthetic virus-like particles and their use in vaccine delivery. *Chembiochem* 12, 100–109. doi:10.1002/cbic.201000536
- Grace, C. R., Perrin, M. H., Gulyas, J., Rivier, J. E., Vale, W. W., and Riek, R. (2010). NMR structure of the first extracellular domain of corticotropin-releasing factor receptor 1 (ECD1-CRF-R1) complexed with a high affinity agonist. *J. Biol. Chem.* 285, 38580–38589. doi:10.1074/jbc.M110.121897
- Groß, A., Brox, R., Damm, D., Tschammer, N., Schmidt, B., and Eichler, J. (2015a). Ligand selectivity of a synthetic CXCR4 mimetic peptide. *Bioorg. Med. Chem.* 23, 4050–4055. doi:10.1016/j.bmc.2015.03.003
- Groß, A., Rodel, K., Kneidl, B., Donhauser, N., Moss, M., Lump, E., et al. (2015b). Enhancement and induction of HIV-1 infection through an assembled peptide derived from the CD4 binding site of gp120. *Chembiochem* 16, 446–454. doi:10.1002/cbic.201402545
- Groß, A., Möbius, K., Haussner, C., Donhauser, N., Schmidt, B., and Eichler, J. (2013). Mimicking protein-protein interactions through peptide-peptide interactions: HIV-1 gp120 and CXCR4. *Front. Immunol.* 4:257. doi:10.3389/fimmu.2013.00257
- Gustafsson, E., Haas, P. J., Walse, B., Hijnen, M., Furebring, C., Ohlin, M., et al. (2009). Identification of conformational epitopes for human IgG on chemotaxis inhibitory protein of *Staphylococcus aureus*. *BMC Immunol.* 10:13. doi:10.1186/1471-2172-10-13
- Haas, P. J., De Haas, C. J., Poppelier, M. J., Van Kessel, K. P., Van Strijp, J. A., Dijkstra, K., et al. (2005). The structure of the C5a receptor-blocking domain of chemotaxis inhibitory protein of *Staphylococcus aureus* is related to a group of immune evasive molecules. *J. Mol. Biol.* 353, 859–872. doi:10.1016/j.jmb.2005.09.014
- Hagan, C. L., Silhavy, T. J., and Kahne, D. (2011). Beta-Barrel membrane protein assembly by the Bam complex. *Annu. Rev. Biochem.* 80, 189–210. doi:10.1146/annurev-biochem-061408-144611
- Hagan, C. L., Westwood, D. B., and Kahne, D. (2013). Bam lipoproteins assemble BamA in vitro. *Biochemistry* 52, 6108–6113. doi:10.1021/bi400865z
- Hagan, C. L., Wzorek, J. S., and Kahne, D. (2015). Inhibition of the beta-barrel assembly machine by a peptide that binds BamD. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2011–2016. doi:10.1073/pnas.1415955112
- Haines, L. A., Rajagopal, K., Ozbas, B., Salick, D. A., Pochan, D. J., and Schneider, J. P. (2005). Light-activated hydrogel formation via the triggered folding and self-assembly of a designed peptide. *J. Am. Chem. Soc.* 127, 17025–17029. doi:10.1021/ja054719o
- Hamley, I. W. (2014). Peptide nanotubes. *Angew. Chem. Int. Ed. Engl.* 53, 6866–6881. doi:10.1002/anie.201310006
- Hanold, L. E., Oruganty, K., Ton, N. T., Beedle, A. M., Kannan, N., and Kennedy, E. J. (2015). Inhibiting EGFR dimerization using triazoly-bridged dimerization arm mimics. *PLoS ONE* 10:e0118796. doi:10.1371/journal.pone.0118796
- He, Y., Xiao, Y., Song, H., Liang, Q., Ju, D., Chen, X., et al. (2008). Design and evaluation of siRNA, a novel HIV-1 fusion inhibitor. *J. Biol. Chem.* 283, 11126–11134. doi:10.1074/jbc.M800200200
- Hollenstein, K., Kean, J., Bortolato, A., Cheng, R. K., Dore, A. S., Jazayeri, A., et al. (2013). Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* 499, 438–443. doi:10.1038/nature12357
- Hong, R.-L., Huang, C.-J., Tseng, Y.-L., Pang, V. F., Chen, S.-T., Liu, J.-J., et al. (1999a). Direct comparison of liposomal doxorubicin with or without polyethylene glycol coating in C-26 tumor-bearing mice: is surface coating with polyethylene glycol beneficial? *Clin. Cancer Res.* 5, 3645–3652.
- Hong, S. Y., Oh, J. E., and Lee, K. H. (1999b). Effect of D-amino acid substitution on the stability, the secondary structure, and the activity of membrane-active peptide. *Biochem. Pharmacol.* 58, 1775–1780. doi:10.1016/S0006-2952(99)00259-2
- Houghten, R. A., Pinilla, C., Appel, J. R., Blondelle, S. E., Dooley, C. T., Eichler, J., et al. (1999). Mixture-based combinatorial libraries. *J. Med. Chem.* 42, 3743–3778.
- Imamura, Y., Watanabe, N., Umezawa, N., Iwatsubo, T., Kato, N., Tomita, T., et al. (2009). Inhibition of gamma-secretase activity by helical beta-peptide foldamers. *J. Am. Chem. Soc.* 131, 7353–7359. doi:10.1021/ja9001458
- Ippel, J. H., De Haas, C. J., Bunschoten, A., Van Strijp, J. A., Kruijtz, J. A., Liskamp, R. M., et al. (2009). Structure of the tyrosine-sulfated C5a receptor N terminus in complex with chemotaxis inhibitory protein of *Staphylococcus aureus*. *J. Biol. Chem.* 284, 12363–12372. doi:10.1074/jbc.M808179200
- Jackson, D. Y., King, D. S., Chmielewski, J., Singh, S., and Schultz, P. G. (1991). General approach to the synthesis of short alpha-helical peptides. *J. Am. Chem. Soc.* 113, 9391–9392. doi:10.1021/ja00024a067
- Johnson, L. M., and Gellman, S. H. (2013). Alpha-helix mimicry with alpha/beta-peptides. *Meth. Enzymol.* 523, 407–429. doi:10.1016/B978-0-12-394292-0.00019-9

- Kaspar, A. A., and Reichert, J. M. (2013). Future directions for peptide therapeutics development. *Drug Discov. Today* 18, 807–817. doi:10.1016/j.drudis.2013.05.011
- Kawamoto, S. A., Coleska, A., Ran, X., Yi, H., Yang, C.-Y., and Wang, S. (2011). Design of triazole-stapled BCL9 alpha-helical peptides to target the beta-catenin/B-cell CLL/lymphoma 9 (BCL9) protein-protein interaction. *J. Med. Chem.* 55, 1137–1146. doi:10.1021/jm201125d
- Kilby, J. M., and Eron, J. J. (2003). Novel therapies based on mechanisms of HIV-1 cell entry. *N. Engl. J. Med.* 348, 2228–2238. doi:10.1056/NEJMra022812
- Kimmerlin, T., and Seebach, D. (2005). '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. *J. Pept. Res.* 65, 229–260. doi:10.1111/j.1399-3011.2005.00214.x
- Konstantinopoulos, P. A., Karamouzis, M. V., and Papavassiliou, A. G. (2007). Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. *Nat. Rev. Drug Discov.* 6, 541–555. doi:10.1038/nrd2221
- Kurrikof, K., Gestin, M., and Langel, Ü. (2015). Recent in vivo advances in cell-penetrating peptide-assisted drug delivery. *Expert Opin. Drug Deliv.* doi:10.1517/17425247.2016.1125879
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659. doi:10.1038/31405
- LaBelle, J. L., Katz, S. G., Bird, G. H., Gavathiotis, E., Stewart, M. L., Lawrence, C., et al. (2012). A stapled BIM peptide overcomes apoptotic resistance in hematologic cancers. *J. Clin. Invest.* 122, 2018–2031. doi:10.1172/JCI46231
- Lalezari, J. P., Henry, K., O'Hearn, M., Montaner, J. S., Piliero, P. J., Trottier, B., et al. (2003). Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. *N. Engl. J. Med.* 348, 2175–2185. doi:10.1056/NEJMoa035026
- Leduc, A. M., Trent, J. O., Wittliff, J. L., Bramlett, K. S., Briggs, S. L., Chirgadze, N. Y., et al. (2003). Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor-coactivator interactions. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11273–11278. doi:10.1073/pnas.1934759100
- Li, W., and Caberoy, N. B. (2010). New perspective for phage display as an efficient and versatile technology of functional proteomics. *Appl. Microbiol. Biotechnol.* 85, 909–919. doi:10.1007/s00253-009-2277-0
- Lichtenthaler, S. F., Wang, R., Grimm, H., Uljon, S. N., Masters, C. L., and Beyreuther, K. (1999). Mechanism of the cleavage specificity of Alzheimer's disease gamma-secretase identified by phenylalanine-scanning mutagenesis of the transmembrane domain of the amyloid precursor protein. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3053–3058. doi:10.1073/pnas.96.6.3053
- Madden, M. M., Muppidi, A., Li, Z., Li, X., Chen, J., and Lin, Q. (2011). Synthesis of cell-permeable stapled peptide dual inhibitors of the p53-Mdm2/Mdmx interactions via photoinduced cycloaddition. *Bioorg. Med. Chem. Lett.* 21, 1472–1475. doi:10.1016/j.bmcl.2011.01.004
- Mart, R. J., Osborne, R. D., Stevens, M. M., and Ulijn, R. V. (2006). Peptide-based stimuli-responsive biomaterials. *Soft Matter* 2, 822–835. doi:10.1039/b607706d
- Martin, L., Stricher, F., Misse, D., Sironi, F., Pugniere, M., Barthe, P., et al. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat. Biotechnol.* 21, 71–76. doi:10.1038/nbt768
- Martinek, T. A., and Fulop, F. (2003). Side-chain control of beta-peptide secondary structures – design principles. *Eur. J. Biochem.* 270, 3657–3666. doi:10.1046/j.1432-1033.2003.03756.x
- Masterson, L. R., Etienne, M. A., Porcelli, F., Barany, G., Hammer, R. P., and Veglia, G. (2007). Nonstereogenic alpha-aminoisobutyryl-glycyl dipeptidyl unit nucleates type I' beta-turn in linear peptides in aqueous solution. *Biopolymers* 88, 746–753. doi:10.1002/bip.20738
- Mehl, R. A., Anderson, J. C., Santoro, S. W., Wang, L., Martin, A. B., King, D. S., et al. (2003). Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* 125, 935–939. doi:10.1021/ja0284153
- Mehrban, N., Zhu, B., Tamagnini, F., Young, F. I., Wasmuth, A., Hudson, K. L., et al. (2015). Functionalized α -helical peptide hydrogels for neural tissue engineering. *ACS Biomater. Sci. Eng.* 1, 431–439. doi:10.1021/acsbiomaterials.5b00051
- Meier, J., Kassler, K., Sticht, H., and Eichler, J. (2012). Peptides presenting the binding site of human CD4 for the HIV-1 envelope glycoprotein gp120. *Beilstein J. Org. Chem.* 8, 1858–1866. doi:10.3762/bjoc.8.214
- Merrifield, R. B. (1963). Solid phase peptide synthesis. I. the synthesis of a tetrapeptide. *J. Am. Chem. Soc.* 85, 2149–2154. doi:10.1021/ja00897a025
- Miller, M. J., Foy, K. C., and Kaumaya, P. T. (2013). Cancer immunotherapy: present status, future perspective, and a new paradigm of peptide immunotherapeutics. *Discov. Med.* 15, 166–176.
- Mills, N. L., Daugherty, M. D., Frankel, A. D., and Guy, R. K. (2006). An alpha-helical peptidomimetic inhibitor of the HIV-1 Rev-RRE interaction. *J. Am. Chem. Soc.* 128, 3496–3497. doi:10.1021/ja0582051
- Möbius, K., Durr, R., Haubner, C., Dietrich, U., and Eichler, J. (2012). A functionally selective synthetic mimic of the HIV-1 co-receptor CXCR4. *Chemistry* 18, 8292–8295. doi:10.1002/chem.201200111
- Mojsoska, B., Zuckermann, R. N., and Jenssen, H. (2015). Structure-activity relationship study of novel peptoids that mimic the structure of antimicrobial peptides. *Antimicrob. Agents Chemother.* 59, 4112–4120. doi:10.1128/AAC.00237-15
- Muraki, M., Morii, H., and Harata, K. (2000). Chemically prepared hevein domains: effect of C-terminal truncation and the mutagenesis of aromatic residues on the affinity for chitin. *Protein Eng.* 13, 385–389. doi:10.1093/protein/13.6.385
- Nakahara, T., Nomura, W., Ohba, K., Ohya, A., Tanaka, T., Hashimoto, C., et al. (2010). Remodeling of dynamic structures of HIV-1 envelope proteins leads to synthetic antigen molecules inducing neutralizing antibodies. *Bioconjug. Chem.* 21, 709–714. doi:10.1021/bc900502z
- Nomura, W., Aikawa, H., Ohashi, N., Urano, E., Metifiot, M., Fujino, M., et al. (2013). Cell-permeable stapled peptides based on HIV-1 integrase inhibitors derived from HIV-1 gene products. *ACS Chem. Biol.* 8, 2235–2244. doi:10.1021/cb400495h
- Nomura, W., Hashimoto, C., Ohya, A., Miyauchi, K., Urano, E., Tanaka, T., et al. (2012). A synthetic C34 trimer of HIV-1 gp41 shows significant increase in inhibition potency. *ChemMedChem* 7, 205–208. doi:10.1002/cmcd.201200114
- Olson, C. A., Shi, Z., and Kallenbach, N. R. (2001). Polar interactions with aromatic side chains in alpha-helical peptides: Ch.O H-bonding and cation-pi interactions. *J. Am. Chem. Soc.* 123, 6451–6452. doi:10.1021/ja015590v
- Opatz, T., and Liskamp, R. M. (2001). A selectively deprotectable triazacyclophane scaffold for the construction of artificial receptors. *Org. Lett.* 3, 3499–3502. doi:10.1021/ol0101741
- Ösapay, G., and Taylor, J. W. (1992). Multicyclic polypeptide model compounds. 2. synthesis and conformational properties of a highly alpha-helical uncusapeptide constrained by three side-chain to side-chain lactam bridges. *J. Am. Chem. Soc.* 114, 6966–6973. doi:10.1021/ja00044a003
- Otaka, A., Nakamura, M., Nameki, D., Kodama, E., Uchiyama, S., Nakamura, S., et al. (2002). Remodeling of gp41-C34 peptide leads to highly effective inhibitors of the fusion of HIV-1 with target cells. *Angew. Chem. Int. Ed. Engl.* 41, 2937–2940. doi:10.1002/1521-3773(20020816)41:16<2937::AID-ANIE2937>3.0.CO;2-J
- Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514, 455–461. doi:10.1038/nature13808
- Patgiri, A., Yadav, K. K., Arora, P. S., and Bar-Sagi, D. (2011). An orthosteric inhibitor of the Ras-Sos interaction. *Nat. Chem. Biol.* 7, 585–587. doi:10.1038/nchembio.612
- Pham, J. D., Spencer, R. K., Chen, K. H., and Nowick, J. S. (2014). A fibril-like assembly of oligomers of a peptide derived from beta-amyloid. *J. Am. Chem. Soc.* 136, 12682–12690. doi:10.1021/ja505713y
- Pelay-Gimeno, M., Glas, A., Koch, O., and Grossmann, T. N. (2015). Structure-based design of inhibitors of protein-protein interactions: mimicking peptide binding epitopes. *Angew. Chem. Int. Ed. Engl.* 54, 8896–8927. doi:10.1002/anie.201412070
- Pochan, D. J., Schneider, J. P., Kretsinger, J., Ozbas, B., Rajagopal, K., and Haines, L. (2003). Thermally reversible hydrogels via intramolecular folding and consequent self-assembly of a de novo designed peptide. *J. Am. Chem. Soc.* 125, 11802–11803. doi:10.1021/ja0353154
- Postma, B., Poppelier, M. J., Van Galen, J. C., Prossnitz, E. R., Van Strijp, J. A., De Haas, C. J., et al. (2004). Chemotaxis inhibitory protein of *Staphylococcus aureus* binds specifically to the C5a and formylated peptide receptor. *J. Immunol.* 172, 6994–7001. doi:10.4049/jimmunol.172.11.6994
- Pritz, S., Kraetke, O., Klose, A., Klose, J., Rothemund, S., Fechner, K., et al. (2008). Synthesis of protein mimics with nonlinear backbone topology by a combined recombinant, enzymatic, and chemical synthesis strategy. *Angew. Chem. Int. Ed. Engl.* 47, 3642–3645. doi:10.1002/anie.200705718
- Rajagopal, K., Lamm, M. S., Haines-Butterick, L. A., Pochan, D. J., and Schneider, J. P. (2009). Tuning the pH responsiveness of beta-hairpin peptide folding,

- self-assembly, and hydrogel material formation. *Biomacromolecules* 10, 2619–2625. doi:10.1021/bm900544e
- Renner, C., and Moroder, L. (2006). Azobenzene as conformational switch in model peptides. *Chembiochem* 7, 868–878. doi:10.1002/cbic.200500531
- Riedel, T., Ghasparian, A., Moehle, K., Rusert, P., Trkola, A., and Robinson, J. A. (2011). Synthetic virus-like particles and conformationally constrained peptidomimetics in vaccine design. *Chembiochem* 12, 2829–2836. doi:10.1002/cbic.201100586
- Robinson, J. A. (2008). Beta-hairpin peptidomimetics: design, structures and biological activities. *Acc. Chem. Res.* 41, 1278–1288. doi:10.1021/ar700259k
- Robinson, J. A. (2013). Max Bergmann lecture protein epitope mimetics in the age of structural vaccinology. *J. Pept. Sci.* 19, 127–140. doi:10.1002/psc.2482
- Robinson, J. A., Shankaramma, S. C., Jetter, P., Kienzl, U., Schwendener, R. A., Vrijbloed, J. W., et al. (2005). Properties and structure-activity studies of cyclic beta-hairpin peptidomimetics based on the cationic antimicrobial peptide protegrin I. *Bioorg. Med. Chem.* 13, 2055–2064. doi:10.1016/j.bmc.2005.01.009
- Rodriguez, P. L., Harada, T., Christian, D. A., Pantano, D. A., Tsai, R. K., and Discher, D. E. (2013). Minimal “self” peptides that inhibit phagocytic clearance and enhance delivery of nanoparticles. *Science* 339, 971–975. doi:10.1126/science.1229568
- Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002). A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* 41, 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4
- Samanta, S., Beharry, A. A., Sadovski, O., McCormick, T. M., Babalhavaej, A., Tropepe, V., et al. (2013). Photoswitching azo compounds in vivo with red light. *J. Am. Chem. Soc.* 135, 9777–9784. doi:10.1021/ja402220t
- Schafmeister, C. E., Po, J., and Verdine, G. L. (2000). An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *J. Am. Chem. Soc.* 122, 5891–5892. doi:10.1021/ja000563a
- Schellinger, J. G., Danan-Leon, L. M., Hoch, J. A., Kassa, A., Srivastava, I., Davis, D., et al. (2011). Synthesis of a trimeric gp120 epitope mimic conjugated to a T-helper peptide to improve antigenicity. *J. Am. Chem. Soc.* 133, 3230–3233. doi:10.1021/ja1083915
- Scrima, M., Le Chevalier-Isaad, A., Rovero, P., Papini, A. M., Chorev, M., and D’ursi, A. M. (2010). Cu-I-catalyzed azide-alkyne intramolecular i-to-(i+4) side-chain-to-side-chain cyclization promotes the formation of helix-like secondary structures. *J. Org. Chem.* 2010, 446–457. doi:10.1002/efoc.200901157
- Seebach, D., Beck, A. K., and Bierbaum, D. J. (2004). The world of beta- and gamma-peptides comprised of homologated proteinogenic amino acids and other components. *Chem. Biodivers.* 1, 1111–1239. doi:10.1002/cbdv.200490087
- Seebach, D., and Gardiner, J. (2008). Beta-peptidic peptidomimetics. *Acc. Chem. Res.* 41, 1366–1375. doi:10.1021/ar700263g
- Seebach, D., and Matthews, J. L. (1997). Beta-peptides: a surprise at every turn. *Chem. Commun.* 2015–2022. doi:10.1039/a704933a
- Shankaramma, S. C., Athanassiou, Z., Zerbe, O., Moehle, K., Mouton, C., Bernardini, F., et al. (2002). Macrocyclic hairpin mimetics of the cationic antimicrobial peptide protegrin I: a new family of broad-spectrum antibiotics. *Chembiochem* 3, 1126–1133. doi:10.1002/1439-7633(20021104)3:11<1126::AID-CBIC1126>3.0.CO;2-I
- Shi, Z., Olson, C. A., and Kallenbach, N. R. (2002). Cation-pi interaction in model alpha-helical peptides. *J. Am. Chem. Soc.* 124, 3284–3291. doi:10.1021/ja0174938
- Sia, S. K., Carr, P. A., Cochran, A. G., Malashkevich, V. N., and Kim, P. S. (2002). Short constrained peptides that inhibit HIV-1 entry. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14664–14669. doi:10.1073/pnas.232566599
- Simon, R. J., Kania, R. S., Zuckermann, R. N., Huebner, V. D., Jewell, D. A., Banville, S., et al. (1992). Peptoids: a modular approach to drug discovery. *Proc. Natl. Acad. Sci. U.S.A.* 89, 9367–9371. doi:10.1073/pnas.89.20.9367
- Sinthuvanich, C., Veiga, A. S., Gupta, K., Gaspar, D., Blumenthal, R., and Schneider, J. P. (2012). Anticancer beta-hairpin peptides: membrane-induced folding triggers activity. *J. Am. Chem. Soc.* 134, 6210–6217. doi:10.1021/ja210569f
- Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317. doi:10.1126/science.4001944
- Spiegel, J., Cromm, P. M., Itzen, A., Goody, R. S., Grossmann, T. N., and Waldmann, H. (2014). Direct targeting of Rab-GTPase-effector interactions. *Angew. Chem. Int. Ed. Engl.* 53, 2498–2503. doi:10.1002/anie.201308568
- Srinivas, N., Jetter, P., Ueberbacher, B. J., Werneburg, M., Zerbe, K., Steinmann, J., et al. (2010). Peptidomimetic antibiotics target outer-membrane biogenesis in *Pseudomonas aeruginosa*. *Science* 327, 1010–1013. doi:10.1126/science.1182749
- Sun, J., and Zuckermann, R. N. (2013). Peptoid polymers: a highly designable bioinspired material. *ACS Nano* 7, 4715–4732. doi:10.1021/nn4015714
- Swierczewska, M., Lee, K. C., and Lee, S. (2015). What is the future of PEGylated therapies? *Expert Opin. Emerg. Drugs* 20, 531–536. doi:10.1517/14728214.2015.1113254
- Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997). Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12303–12308. doi:10.1073/pnas.94.23.12303
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G. W., Kufareva, I., et al. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341, 1387–1390. doi:10.1126/science.1241475
- Thundimadathil, J. (2012). Cancer treatment using peptides: current therapies and future prospects. *J. Amino Acids* 2012, 967347. doi:10.1155/2012/967347
- Timmerman, P., Barderas, R., Desmet, J., Altschuh, D., Shochat, S., Hollestelle, M. J., et al. (2009). A combinatorial approach for the design of complementarity-determining region-derived peptidomimetics with in vitro anti-tumoral activity. *J. Biol. Chem.* 284, 34126–34134. doi:10.1074/jbc.M109.041459
- Timmerman, P., Beld, J., Puijk, W. C., and Meloen, R. H. (2005). Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *Chembiochem* 6, 821–824. doi:10.1002/cbic.200400374
- Timmerman, P., Shochat, S. G., Desmet, J., Barderas, R., Casal, J. I., Meloen, R. H., et al. (2010). Binding of CDR-derived peptides is mechanistically different from that of high-affinity parental antibodies. *J. Mol. Recognit.* 23, 559–568. doi:10.1002/jmr.1017
- Tornøe, C. W., Christensen, C., and Meldal, M. (2002). Peptidotriazoles on solid phase: [1,2,3]-triazoles by regioselective copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* 67, 3057–3064. doi:10.1021/jo011148j
- Tsou, L. K., Tatko, C. D., and Waters, M. L. (2002). Simple cation-pi interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J. Am. Chem. Soc.* 124, 14917–14921. doi:10.1021/ja026721a
- Veldkamp, K. E., Heezius, H. C., Verhoef, J., Van Strijp, J. A., and Van Kessel, K. P. (2000). Modulation of neutrophil chemokine receptors by *Staphylococcus aureus* supernate. *Infect. Immun.* 68, 5908–5913. doi:10.1128/IAI.68.10.5908-5913.2000
- Verdine, G. L., and Hilinski, G. J. (2012). Stapled peptides for intracellular drug targets. *Meth. Enzymol.* 503, 3–33. doi:10.1016/B978-0-12-396962-0.00001-X
- Vita, C., Drakopoulou, E., Vizzavona, J., Rochette, S., Martin, L., Menez, A., et al. (1999). Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13091–13096. doi:10.1073/pnas.96.23.13091
- Vogelstein, B., and Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nat. Med.* 10, 789–799. doi:10.1038/nm1087
- Walensky, L. D., and Bird, G. H. (2014). Hydrocarbon-stapled peptides: principles, practice, and progress. *J. Med. Chem.* 57, 6275–6288. doi:10.1021/jm4011675
- Wang, R. R., Yang, L. M., Wang, Y. H., Pang, W., Tam, S. C., Tien, P., et al. (2009). Sifuvirtide, a potent HIV fusion inhibitor peptide. *Biochem. Biophys. Res. Commun.* 382, 540–544. doi:10.1016/j.bbrc.2009.03.057
- Watson, S. A., Michaeli, D., Grimes, S., Morris, T. M., Robinson, G., Varro, A., et al. (1996). Gastrimmune raises antibodies that neutralize amidated and glycine-extended gastrin-17 and inhibit the growth of colon cancer. *Cancer Res.* 56, 880–885.
- Wild, C., Oas, T., Mcdanal, C., Bolognesi, D., and Matthews, T. (1992). A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10537–10541. doi:10.1073/pnas.89.21.10537
- Wild, C. T., Shugars, D. C., Greenwell, T. K., Mcdanal, C. B., and Matthews, T. J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl. Acad. Sci. U.S.A.* 91, 9770–9774. doi:10.1073/pnas.91.21.9770
- Wilen, C. B., Tilton, J. C., and Doms, R. W. (2012). HIV: cell binding and entry. *Cold Spring Harb. Perspect. Med.* 2, a006866. doi:10.1101/cshperspect.a006866

- Woods, R. J., Brower, J. O., Castellanos, E., Hashemzadeh, M., Khakshoor, O., Russu, W. A., et al. (2007). Cyclic modular beta-sheets. *J. Am. Chem. Soc.* 129, 2548–2558. doi:10.1021/ja0667965
- Woolley, G. A. (2005). Photocontrolling peptide alpha helices. *Acc. Chem. Res.* 38, 486–493. doi:10.1021/ar040091v
- Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., et al. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071. doi:10.1126/science.1194396
- Yang, B., Liu, D., and Huang, Z. (2004). Synthesis and helical structure of lactam bridged BH3 peptides derived from pro-apoptotic Bcl-2 family proteins. *Bioorg. Med. Chem. Lett.* 14, 1403–1406. doi:10.1016/j.bmcl.2003.09.101
- Yarden, Y., and Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* 2, 127–137. doi:10.1038/35052073
- Yu, C., and Taylor, J. W. (1999). Synthesis and study of peptides with semirigid i and i + 7 side-chain bridges designed for alpha-helix stabilization. *Bioorg. Med. Chem.* 7, 161–175. doi:10.1016/S0968-0896(98)00232-6
- Zolla-Pazner, S., and Cardozo, T. (2010). Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nat. Rev. Immunol.* 10, 527–535. doi:10.1038/nri2801
- Zuckermann, R. N., and Kodadek, T. (2009). Peptoids as potential therapeutics. *Curr. Opin. Mol. Ther.* 11, 299–307.

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 © 2016 Groß, Hashimoto, Sticht and Eichler. 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.



Synthetic Protein Scaffolds Based on Peptide Motifs and Cognate Adaptor Domains for Improving Metabolic Productivity

Anselm H. C. Horn* and Heinrich Sticht*

Bioinformatik, Institut für Biochemie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

M. Kalim Akhtar,
University of Edinburgh, UK
Mattheos Koffas,
Rensselaer Polytechnic Institute, USA

*Correspondence:

Anselm H. C. Horn
anselm.horn@fau.de;
Heinrich Sticht
heinrich.sticht@fau.de

Specialty section:

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

Received: 26 August 2015

Accepted: 05 November 2015

Published: 23 November 2015

Citation:

Horn AHC and Sticht H (2015)
*Synthetic Protein Scaffolds Based on
Peptide Motifs and Cognate Adaptor
Domains for Improving Metabolic
Productivity.*
Front. Bioeng. Biotechnol. 3:191.
doi: 10.3389/fbioe.2015.00191

The efficiency of many cellular processes relies on the defined interaction among different proteins within the same metabolic or signaling pathway. Consequently, a spatial colocalization of functionally interacting proteins has frequently emerged during evolution. This concept has been adapted within the synthetic biology community for the purpose of creating artificial scaffolds. A recent advancement of this concept is the use of peptide motifs and their cognate adaptor domains. SH2, SH3, GBD, and PDZ domains have been used most often in research studies to date. The approach has been successfully applied to the synthesis of a variety of target molecules including catechin, D-glucaric acid, H₂, hydroquinone, resveratrol, butyrate, gamma-aminobutyric acid, and mevalonate. Increased production levels of up to 77-fold have been observed compared to non-scaffolded systems. A recent extension of this concept is the creation of a covalent linkage between peptide motifs and adaptor domains, which leads to a more stable association of the scaffolded systems and thus bears the potential to further enhance metabolic productivity.

Keywords: adaptor domain, linear peptide motif, protein scaffold, fusion protein, metabolic engineering

INTRODUCTION

Nature has developed highly efficient ways for signal and substrate processing in living cells. Synthetic biology is inspired by the natural archetype and tries to mimic and optimize biological processes for tailor-made applications (Luo et al., 2013). Milestone achievements of this relatively young area include the microbial production of artemisinic acid, a key precursor of the antimalarial drug artemisinin (Martin et al., 2003; Ro et al., 2006), the industrial production of 1,3-propanediol, used in a multitude of further applications (Nakamura and Whited, 2003), and the reconstruction of a complete microbial genome (Gibson et al., 2010).

The rational design in synthetic biology is frequently inspired by the spatial proximity of enzymes observed in nature (Conrado et al., 2008; Luo et al., 2013). Following evolution, such artificial bioreactors implement proximity mostly via modular scaffolds (Carroll, 2005; Bhattacharyya et al., 2006). In this concept, modular building blocks are used for the creation of large custom scaffold systems. Such scaffolds define the spatial organization of enzymes and allow substrate channeling like in natural systems, which has several advantages: it rescues the intermediates from diffusion or competing pathways, decreases their transit times, and avoids unfavorable equilibria and kinetics from metabolite concentrations in the bulk phase (Miles et al., 1999; Spivey and Ovadi, 1999).

In nature, many organisms have developed multifunctional enzyme systems with a pivotal role in both primary metabolism [e.g., amino acid biosynthesis (Welch and Gaertner, 1980) or fatty acid oxidation (Ishikawa et al., 2004)] and secondary metabolism [e.g., multifunctional polyketide synthases in bacteria (Pfeifer and Khosla, 2001) and flavonoid or alkaloid biosynthesis in plants (Jorgensen et al., 2005)].

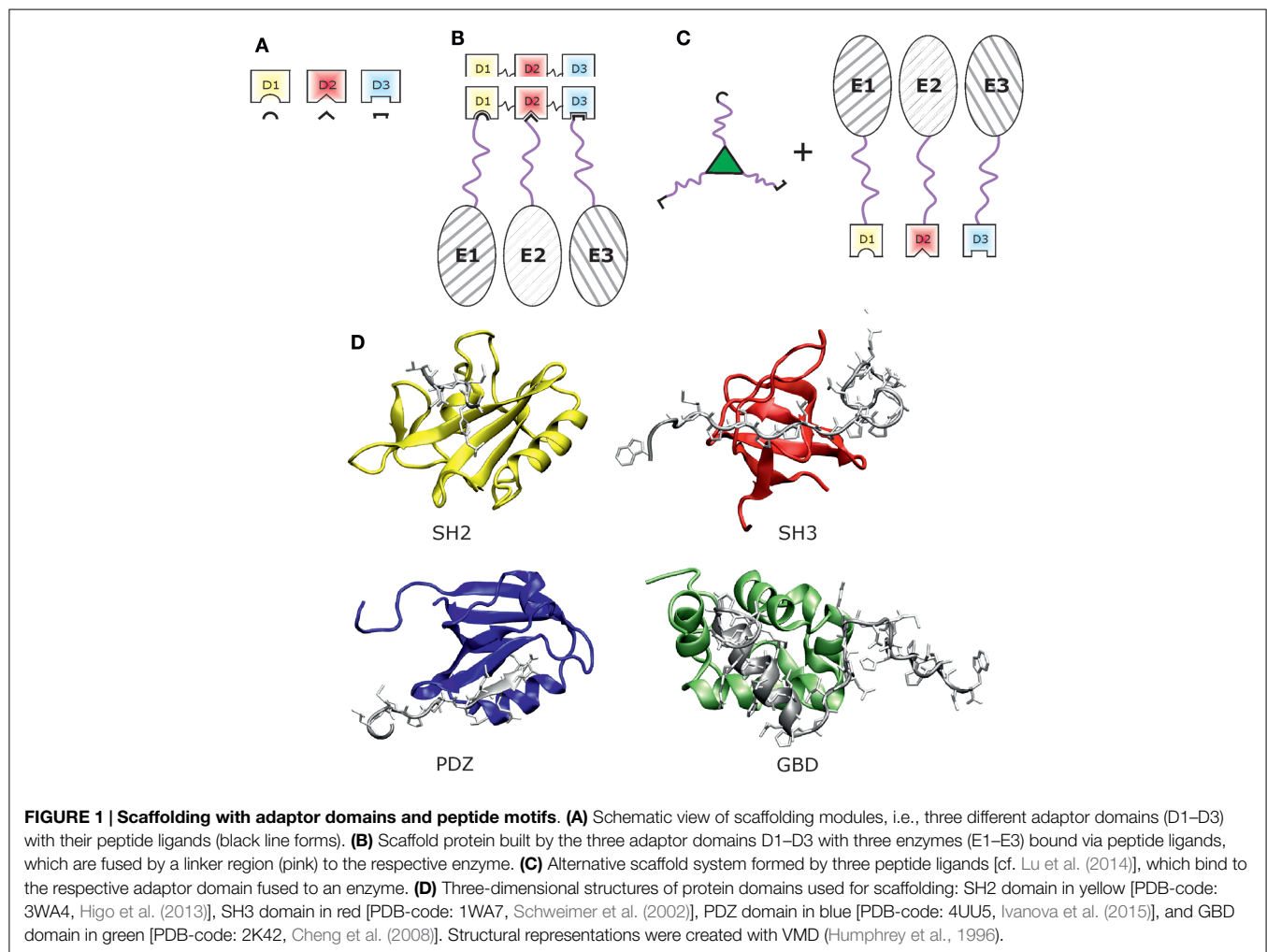
Multifunctional enzyme systems that mimic natural systems can be artificially constructed via at least four general strategies: (I) *colocalization* or *immobilization* of enzymes has been the first approach to be of practical use; (II) *compartmentalization* generates an enclosed reaction area that can be defined similar to biological systems (e.g., in cell organelles); (III) *DNA/RNA building blocks* can be utilized for spatial organization of reactive centers; and (IV) *protein scaffolding* presents a versatile approach in synthetic biology. This last scaffolding principle can be divided further into several approaches: fusion proteins are constructed by linking two or more enzymes into a single protein sequence. Non-covalent protein–protein interactions via the mutual recognition of folded domains or coiled-coil pairs as well as amyloid assemblies can be used to construct scaffolds with defined stoichiometry. A brief overview over all these scaffolding strategies is provided as supplementary information. Another concept in

protein-based scaffolding uses protein adaptor domains and peptide ligands for bringing enzymes in spatial proximity, the focus of this review.

SCAFFOLDING BASED ON ADAPTOR DOMAINS: STRUCTURAL PRINCIPLES

Many important physiological protein interactions are mediated by relatively small protein domains, which bind to peptides exhibiting specific sequence motifs (**Figure 1A**) (Dinkel and Sticht, 2010). In this type of interaction, only the adaptor domain adopts a globular three-dimensional structure while the interaction motif is mostly linear and has, therefore, been termed short linear interaction motif (SLiM). This type of protein–ligand interaction presents a promising concept in protein scaffolding (**Figure 1B**) that has gained a lot of attention in synthetic biology applications.

Protein scaffolding based on cognate adaptor domains and peptide motifs requires a careful selection of candidate domains and SLiMs as well as the choice of proper linkers to interconnect these moieties and to attach them to the enzymes of interest (**Figure 1B**). The properties of these three building blocks, i.e., domain, linker, and peptide ligand, critically affect the shape of



the resulting scaffold (cf. **Figure 1**) and will be described in the following in more detail.

Adaptor Domains

Adaptor domains used in protein-peptide scaffolding need to fulfill two basic requirements. First, they should have a strong affinity toward their peptide ligands to allow for effective coupling. Second, they should provide a distinct specificity for their ligands to allow for defined coupling, when several domain-ligand pairs are used simultaneously. The most often used domains are SH3, SH2, PDZ, and GTPase-binding domain (GBD) (cf. **Figure 1D**).

The first protein modules that were reported to mediate interaction with SLiMs are the “Src homology 2” (SH2) and “Src homology 3” (SH3) domains (Koch et al., 1991). SH3 domains are small modules of ca. 60 residues. They recruit proline-rich ligands, which bind to the domain surface at three shallow grooves formed by conserved aromatic residues (Mayer, 2001) and exhibit two different binding orientations. Over the last few years, an increasing number of SH3 domains with different ligand binding specificity have been described (Saksela and Permi, 2012).

SH2 domains are highly conserved structures of ca. 100 residues comprising two α -helices and seven β -strands (Pawson et al., 2001). In nature, this domain possesses an either promiscuous or strict specificity for a 3–5 residues motif flanking a phosphorylated tyrosine; like for the SH3 domain, additional SH2-binding modes were discovered, underscoring the plasticity of this recognition type in physiological context (Machida and Mayer, 2005).

PDZ domains are also widely used for scaffolding. They are of similar size as SH2 domains and target specific motifs at the C-terminus of the binding partner. The peptide ligand adopts a β -strand and extends an existing β -sheet within the PDZ domain upon binding (Schultz et al., 1998; Harris and Lim, 2001). At least four different classes of ligands are known for PDZ domains exhibiting a distinct binding specificity (Songyang et al., 1997).

The last example of an established domain-ligand pair in synthetic biology originates from GBDs. In contrast to the other domains discussed above, isolated GBD domains do not adopt a single, discrete structure under physiological conditions but rather sample multiple, loosely packed conformations in solution (Abdul-Manan et al., 1999; Kim et al., 2000). The corresponding peptide ligand has been deduced from the autoinhibited form of the GBD (Dueber et al., 2009). **Figure 1D** shows three-dimensional structures of the SH2, SH3, PDZ, and GBD domain. Beyond the examples presented above, other domain/ligand pairs may also be utilized for synthetic scaffolds if they exhibit a sufficiently high affinity and specificity for their ligand.

Linear Motif Peptides

Short linear interaction motifs are the complementary binding partner to protein adaptor domains. These peptide motifs occur in disordered protein regions and are present in 20–50% of all eukaryotic proteins, while up to 17% of the proteins are completely disordered in eukaryotic cells. To date, ~300 known motif patterns

are listed in electronic databases, e.g., ELM database (Dinkel et al., 2014), PROSITE (Hulo et al., 2004), and Minimotif-Miner (Balla et al., 2006). Interestingly, there are estimates that in the proteome the SLiM-mediated instances in signaling pathway modulation outnumber those mediated by globular domains (McEntyre and Gibson, 2004).

Linear motif peptides possess a number of properties, which make them well suited as ligands in synthetic biology. The interaction motifs normally comprise only 3–10 amino acids and are thus rather short and intrinsically disordered. Furthermore, SLiMs may constitute the sites of post-translational modification (e.g., phosphorylation), which enables them to function as inducible switches.

In addition to the key residues necessary for binding, SLiMs also frequently contain variable residues (denoted as “X”) to ensure proper spacing between the binding residues. Due to its lack of a defined structure prior to binding, this peptide-domain interaction differs from the well-known domain-domain interactions in protein complexes. Prominent examples for SLiM sequence patterns include the classical P-x-x-P motif for binding to SH3 domains or a phosphorylated tyrosine with specific sequence neighbors for binding to SH2 domains.

Linkers

The last part necessary for modular protein scaffolding is the linker region connecting the engineered enzymes and the attached peptide ligands (**Figure 1B**). The importance of linker design is well known from fusion proteins (Chen et al., 2013), as length and amino acid composition may influence the activity and folding properties of the protein construct (Robinson and Sauer, 1998; Bai and Shen, 2006; Zhao et al., 2008).

As a guide for the rational design of artificial linkers, an inspection of natural linkers is helpful. Two independent studies with different data sets gave similar results: while Argos found a preferred mean linker length of 6.5 residues (Argos, 1990), George and Heringa obtained a value of 10.0 ± 5.8 residues (George and Heringa, 2002). Generally, polar or charged residues were enriched in the natural linkers, with a secondary structure preference for coil (Argos, 1990) or helix (George and Heringa, 2002), respectively. Natural linkers lack interaction with neighboring protein domains and adopt mainly non-globular conformations (Chen et al., 2013).

Designed linkers may be classified according to their structure, which defines their functionality. Flexible linkers are normally rich in small or hydrophilic amino acids and allow for an increased spatial separation and reorientation of the fused parts. A prominent and very early example for a flexible linker is (GGGGS)₃ that connected the heavy and light chain domains (V_H and V_L) of an engineered antibody fragment (Huston et al., 1988). Rigid linkers with the sequence (EAAAK)_n exhibit a stable helical structure and thus pertain a certain distance between the fused parts. This linker type has been successfully used to increase the enzymatic efficiency of bifunctional fusions of β -glucanase and xylanase (Lu and Feng, 2008).

It should also be noted that linker regions may also have additional benefits. They potentially improve folding and stability (Huston et al., 1988; Takamatsu et al., 1990; Werner et al., 2006;

Hagemeyer et al., 2009), expression (Amet et al., 2009), or even bioactivity (Bai and Shen, 2006).

SCAFFOLDING BASED ON ADAPTOR DOMAINS: APPLICATION TO METABOLIC ENGINEERING

In this section, several applications of scaffolding using adaptor domains and peptide ligands are presented. The key features of the engineered systems are summarized in **Table 1**.

As one prominent example for this approach, Dueber et al. (2009) engineered a model scaffold for the three-step synthesis of mevalonate (Martin et al., 2003), which is an important precursor for the large field of isoprenoids, starting from acetyl-CoA. The enzymatic system comprised three modules, acetoacetyl-CoA thiolase (AtoB), hydroxy-methylglutaryl-CoA synthase (HMGS) and hydroxymethylglutaryl-CoA reductase (HMGR). From these modules, only AtoB is native to the host system *Escherichia coli*, whereas the two others were imported from *Saccharomyces cerevisiae*. To avoid flux imbalances with high metabolic load and to increase the overall production, scaffold constructs of three domains, GBD, SH3, and PDZ connected via flexible linkers were created and AtoB, HMGS, and HMGR were extended by corresponding peptide ligands, respectively. As this first simple scaffold design yielded only slightly increased product titers compared to the scaffold-free system, the authors designed scaffolding proteins with a varying number of SH3 and PDZ domains. This systematic search revealed the best synthetic scaffold GBD₁-SH3₂-PDZ₂ for this system, i.e., one GBD domain linked to two SH3 and PDZ domains, and exhibited a remarkable 77-fold increase of the

product. Furthermore, the authors also investigated the influence of the spatial orientation of the domains toward each other by changing the order of the two SH3 and PDZ domains. A further increase, however, was not observed in these additional systems (Dueber et al., 2009).

In order to demonstrate the generality of this approach, the same group strived to increase the production of D-glucaric acid from D-glucose via scaffolding. The synthetic pathway had originally been constructed by Moon et al. (2009): myo-inositol-1-phosphate synthase (Ino1) from *S. cerevisiae*, myo-inositol oxygenase (MIOX) from mouse, and uronate dehydrogenase (Udh) from *Pseudomonas syringae* were coexpressed in *E. coli*. A domain-based scaffold for the two enzymes Ino1 and MIOX, which were equipped with the respective peptide ligand sequences, tripled the product titers compared to the original system (Dueber et al., 2009). Additional optimization of the system by including Udh into the scaffold and also varying the number of cognate domains within the scaffold allowed for an additional product increase of ~50% (Moon et al., 2010).

The first artificially scaffolded redox pathway was presented by Agapakis et al. (2010). They engineered a hydrogen-producing electron transfer circuit in *E. coli* composed of the heterologously expressed enzymes [Fe-Fe]-hydrogenase, ferredoxin, and pyruvate-ferredoxin oxidoreductase. A major issue was the risk of side reactions caused by high energy electrons stored in iron-sulfur cluster proteins. They, thus, applied several methods to insulate the synthetic pathway, one of which was to utilize a protein scaffold constructed from the three domains GBD, SH3, and PDZ. This approach yielded a threefold increase of H₂ production. Furthermore, the authors investigated the influence of scaffold protein composition and peptide ligand

TABLE 1 | Examples of engineered scaffolds comprising adaptor domains and peptide ligands.

Product	Enzyme pathway ^a	Domains ^b	Host	Fold increase ^c	Reference
Mevalonate	Acetoacetyl-CoA thiolase, hydroxy-methylglutaryl-CoA synthase, hydroxymethylglutaryl-CoA reductase	GBD, SH3, PDZ	<i>E. coli</i>	77	Dueber et al. (2009)
D-glucaric acid	Myo-inositol-1-phosphate synthase, myo-inositol oxygenase, (uronate dehydrogenase)	(GBD), SH3, PDZ	<i>E. coli</i>	3	Dueber et al. (2009)
D-glucaric acid	Myo-inositol-1-phosphate synthase, myo-inositol oxygenase, uronate dehydrogenase	GBD, SH3, PDZ	<i>E. coli</i>	5	Moon et al. (2010)
H ₂	[Fe-Fe]-hydrogenase, ferredoxin, (pyruvate-ferredoxin oxidoreductase)	(GBD), SH3, PDZ	<i>E. coli</i>	3–5	Agapakis et al. (2010)
Hydroquinone	Cutinase	SH2	Self-assembled monolayer	30	Li et al. (2010)
Resveratrol	4-Coumarate:CoA ligase, stilbene synthase	(GBD), SH3, PDZ	<i>S. cerevisiae</i>	5	Wang and Yu (2012)
Butyrate	(Acetoacetyl-CoA thiolase), 3-hydroxybutyryl-CoA dehydrogenase, 3-hydroxybutyryl-CoA dehydratase, trans-enoyl-coenzyme A reductase, (acyl-CoA thioesterase II)	GBD, SH3, PDZ	<i>E. coli</i>	3	Baek et al. (2013)
Gamma-aminobutyric acid	Glutamate decarboxylase, glutamate/GABA antiporter	SH3	<i>E. coli</i>	2.5	Vo et al. (2013)
Catechin	Flavanone 3-hydroxylase, dihydroflavonol 4-reductase, leucoanthocyanidin reductase	GBD, SH3, PDZ	<i>E. coli</i>	1.3	Zhao et al. (2015)

^aUn scaffolded enzymes in parentheses.

^bScaffold domains without a ligand/enzyme counterpart in parentheses.

^cCompared to the unscaffolded system.

linker length on the yield and found both to be a significant factor.

Scaffolds consisting of the same domains, GBD, SH3, and PDZ, were used to increase the production of butyrate in *E. coli* (Baek et al., 2013). For the complete biosynthetic pathway, the five enzymes acetoacetyl-CoA thiolase, 3-hydroxybutyryl-CoA dehydrogenase, 3-hydroxybutyryl-CoA dehydratase, trans-enoyl-coenzyme A reductase, acyl-CoA thioesterase II were overexpressed in the host. For the three enzymes amidst the pathway, a domain scaffold was created to provide a better spatial proximity of the reaction centers. After additional variation of the domain frequency within the scaffold, the production increased to threefold.

Wang and Yu (2012) used the set of scaffold proteins composed of GBD, SH3, and PDZ domains established by Dueber et al. (2009) for another biotechnological application. Their work aimed to recruit two enzymes, 4-coumarate:CoA ligase and stilbene synthase, via covalently attached SH3 and PDZ peptide ligands for the biosynthesis of resveratrol, a naturally occurring defense molecule from plants with significant physiological effects on human and animals. In contrast to the experimental settings discussed above, *S. cerevisiae* was used as host system. The product yield increased fivefold via the scaffolding approach compared to the unscaffolded enzymes and 2.7-fold compared to a direct fusion protein approach.

The biosynthesis pathway of catechins from flavanone was the target of metabolic engineering efforts of Koffas and coworkers (Zhao et al., 2015). In their pathway optimization, they focused on three enzymes: flavanone-3-hydroxylase, dihydroflavonol 4-reductase, and leucoanthocyanidin reductase. Application of scaffolds composed of GBD, SH3, and PDZ domains yielded only marginal metabolic improvement in some cases, whereas most constructs tested exhibited a decreased productivity.

Besides the creation of a scaffold protein containing multiple adaptor domains, domain–ligand interactions can also be exploited in a different fashion as exemplified by the work of Vo et al. (2013). They enhanced the productivity of *E. coli* producing gamma-aminobutyric acid (GABA) by coupling glutamate decarboxylase (GadA/GadB) to the membrane protein glutamate/GABA antiporter (GadC). For that purpose, they attached an SH3 domain to GadA/GadB and three peptide ligand sequences to GadC each separated with flexible linkers. In that way, they could increase the GABA productivity by 2.5-fold.

A further and different approach used the localization of substrate and enzyme on a self-assembled monolayer for a 30-fold product increase (Li et al., 2010); 4-hydroxyphenyl 2-methylvalerate, which is converted by cutinase to a hydroquinone product, and a SH2-ligand were presented on the surface to the enzyme fused to a SH2 domain. As the SH2 domain only recognizes its ligand in phosphorylated form, the system contains a potential switch, which might be exploited in future applications.

OUTLOOK

Inspection of Table 1 reveals that the increase in metabolic productivity is highly dependent on the system investigated, and for some of the systems, there is little benefit from scaffolding. As

suggested by Zhao et al. (2015), a further increase in catechin biosynthesis might be achieved from an optimization of linkers. An additional factor for optimization might be the use of alternative adaptor domains or the rational design of covalent bonds between the two binding partners in order to increase the stability of the scaffolded complex.

Recently, Lu et al. (2014) constructed two domain–ligand pairs for both SH3 and PDZ domains, in which the ligand–domain interaction was reinforced by an engineered thioether bond. For that purpose, a residue within the domain was mutated to cysteine, while the peptide ligand was equipped with an unnatural amino acid carrying a reactive α -chloroacetyl group. Binding of the ligand to the domain brought the two reactants in close proximity and established the covalent bond. Using this approach, the authors constructed several Y-shaped ligand structures via triazole bonds branched from a lysine site as mini-scaffolds (Figure 1C).

Similarly, Guan et al. (2013) created a disulfide bond between a PDZ domain and its ligand by mutating one residue to cysteine in each of the binding partners to reinforce the domain–ligand interaction. Fusing these modified moieties to the trimeric protein CutA, they were able to build stable hydrogels. By adding a second peptide ligand sequence to one CutA species, the hydrogel could be functionalized by an enzyme and formed an enzymatic biocathode for direct electron transfer.

A new versatile approach for covalent protein linkage is based on CnaB domains from bacteria, which autocatalytically establish isopeptide bonds between the sidechains of a lysine and an asparagine/aspartate residue (Veggiani et al., 2014). By a structure-based splitting of the CnaB domain into two parts, it was possible to create a domain–ligand pair that enables spontaneous formation of intermolecular isopeptide linkages (Zakeri et al., 2012; Li et al., 2014). A modification of this approach even allows that two peptides become covalently joined by an artificial ligase (Fierer et al., 2014). A recently described ester bond that forms autocatalytically in a bacterial cell surface adhesion protein (Kwon et al., 2014) also bears the potential for the construction of orthogonal covalent domain/ligand pairs. The enhanced stability due to the isopeptide or ester bonds may present a promising strategy to design more efficient scaffolds for artificial bioreactors in the future.

Synthetic biology is an emerging field with tremendous biotechnological potential. The efforts reviewed above clearly demonstrate that promising steps in this field have been made, though individual system design will require a tailor-made approach for achieving optimization. More complex metabolic pathways or large-scale industrial applications, however, would clearly benefit from an extended and well characterized tool-box of scaffolding components (Kwok, 2010).

FUNDING

This work was supported by the Emerging Fields Initiative from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), project “Synthetic Biology.”

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00191>

REFERENCES

- Abdul-Manan, N., Aghazadeh, B., Liu, G. A., Majumdar, A., Ouerfelli, O., Siminovich, K. A., et al. (1999). Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. *Nature* 399, 379–383. doi:10.1038/20726
- Agapakis, C. M., Ducat, D. C., Boyle, P. M., Wintermute, E. H., Way, J. C., and Silver, P. A. (2010). Insulation of a synthetic hydrogen metabolism circuit in bacteria. *J. Biol. Eng.* 4, 3. doi:10.1186/1754-1611-4-3
- Amet, N., Lee, H. F., and Shen, W. C. (2009). Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharm. Res.* 26, 523–528. doi:10.1007/s11095-008-9767-0
- Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol.* 211, 943–958. doi:10.1016/0022-2836(90)90085-Z
- Baek, J. M., Mazumdar, S., Lee, S. W., Jung, M. Y., Lim, J. H., Seo, S. W., et al. (2013). Butyrate production in engineered *Escherichia coli* with synthetic scaffolds. *Biotechnol. Bioeng.* 110, 2790–2794. doi:10.1002/bit.24925
- Bai, Y., and Shen, W. C. (2006). Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharm. Res.* 23, 2116–2121. doi:10.1007/s11095-006-9059-5
- Balla, S., Thapar, V., Verma, S., Luong, T., Faghri, T., Huang, C. H., et al. (2006). Minimoto Miner: a tool for investigating protein function. *Nat. Methods* 3, 175–177. doi:10.1038/nmeth856
- Bhattacharyya, R. P., Reményi, A., Yeh, B. J., and Lim, W. A. (2006). Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* 75, 655–680. doi:10.1146/annurev.biochem.75.103004.142710
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245. doi:10.1371/journal.pbio.0030245
- Chen, X., Zaro, J. L., and Shen, W. C. (2013). Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* 65, 1357–1369. doi:10.1016/j.addr.2012.09.039
- Cheng, H. C., Skehan, B. M., Campellone, K. G., Leong, J. M., and Rosen, M. K. (2008). Structural mechanism of WASP activation by the enterohaemorrhagic *E. coli* effector EspF(U). *Nature* 454, 1009–1013. doi:10.1038/nature07160
- Conrado, R. J., Varner, J. D., and Delisa, M. P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* 19, 492–499. doi:10.1016/j.copbio.2008.07.006
- Dinkel, H., and Sticht, H. (2010). "Motif-mediated protein interactions and their role in disease," in *Protein-Protein Complexes: Analysis, Modeling and Drug Design*, ed. M. Zacharias (London: Imperial College Press), 273–295.
- Dinkel, H., Van Roey, K., Michael, S., Davey, N. E., Weatheritt, R. J., Born, D., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42, D259–D266. doi:10.1093/nar/gkt1047
- 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
- Fierer, J. O., Veggiani, G., and Howarth, M. (2014). SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1176–E1181. doi:10.1073/pnas.1315776111
- George, R. A., and Heringa, J. (2002). An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.* 15, 871–879. doi:10.1093/protein/15.11.871
- Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi:10.1126/science.1190719
- Guan, D., Ramirez, M., Shao, L., Jacobsen, D., Barrera, I., Lutkenhaus, J., et al. (2013). Two-component protein hydrogels assembled using an engineered disulfide-forming protein-ligand pair. *Biomacromolecules* 14, 2909–2916. doi:10.1021/bm400814u
- Hagemeyer, C. E., von Zur Muhlen, C., von Elverfeldt, D., and Peter, K. (2009). Single-chain antibodies as diagnostic tools and therapeutic agents. *Thromb. Haemost.* 101, 1012–1019. doi:10.1160/TH08-12-0816
- Harris, B., and Lim, W. A. (2001). Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell. Sci.* 114, 3219–3231.
- Higo, K., Ikura, T., Oda, M., Morii, H., Takahashi, J., Abe, R., et al. (2013). High resolution crystal structure of the Grb2 SH2 domain with a phosphopeptide derived from CD28. *PLoS ONE* 8:e74482. doi:10.1371/journal.pone.0074482
- Hulo, N., Sigrist, C. J., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., et al. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Res.* 32, D134–D137. doi:10.1093/nar/gkh044
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph. Model.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotny, J., Margolies, M. N., et al. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 85, 5879–5883. doi:10.1073/pnas.85.16.5879
- Ishikawa, M., Tsuchiya, D., Oyama, T., Tsunaka, Y., and Morikawa, K. (2004). Structural basis for channelling mechanism of a fatty acid beta-oxidation multienzyme complex. *EMBO J.* 23, 2745–2754. doi:10.1038/sj.emboj.7600298
- Ivanova, M. E., Fletcher, G. C., O'Reilly, N., Purkiss, A. G., Thompson, B. J., and McDonald, N. Q. (2015). Structures of the human Pals1 PDZ domain with and without ligand suggest gated access of Crb to the PDZ peptide-binding groove. *Acta Crystallogr. D Biol. Crystallogr.* 71, 555–564. doi:10.1107/S139900471402776X
- Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., et al. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291. doi:10.1016/j.pbi.2005.03.014
- Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A., and Rosen, M. K. (2000). Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* 404, 151–158. doi:10.1038/35004513
- Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* 252, 668–674. doi:10.1126/science.1708916
- Kwok, R. (2010). Five hard truths for synthetic biology. *Nature* 463, 288–290. doi:10.1038/463288a
- Kwon, H., Squire, C. J., Young, P. G., and Baker, E. N. (2014). Autocatalytically generated Thr-Gln ester bond cross-links stabilize the repetitive Ig-domain shaft of a bacterial cell surface adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1367–1372. doi:10.1073/pnas.1316855111
- Li, J., Nayak, S., and Mrksich, M. (2010). Rate enhancement of an interfacial biochemical reaction through localization of substrate and enzyme by an adaptor domain. *J. Phys. Chem. B* 114, 15113–15118. doi:10.1021/jp102820e
- Li, L., Fierer, J. O., Rapoport, T. A., and Howarth, M. (2014). Structural analysis and optimization of the covalent association between SpyCatcher and a peptide Tag. *J. Mol. Biol.* 426, 309–317. doi:10.1016/j.jmb.2013.10.021
- Lu, P., and Feng, M. G. (2008). Bifunctional enhancement of a beta-glucanase-xylanase fusion enzyme by optimization of peptide linkers. *Appl. Microbiol. Biotechnol.* 79, 579–587. doi:10.1007/s00253-008-1468-4
- Lu, Y., Huang, F., Wang, J., and Xia, J. (2014). Affinity-guided covalent conjugation reactions based on PDZ-peptide and SH3-peptide interactions. *Bioconjug. Chem.* 25, 989–999. doi:10.1021/bc500134w
- Luo, Y., Lee, J. K., and Zhao, H. (2013). Challenges and opportunities in synthetic biology for chemical engineers. *Chem. Eng. Sci.* 103:115–9. doi:10.1016/j.ces.2012.06.013
- Machida, K., and Mayer, B. J. (2005). The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta* 1747, 1–25. doi:10.1016/j.bbapap.2004.10.005
- 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
- Mayer, B. J. (2001). SH3 domains: complexity in moderation. *J. Cell. Sci.* 114, 1253–1263.
- McEntyre, J. R., and Gibson, T. J. (2004). Patterns and clusters within the PSM column in TIBS, 1992–2004. *Trends Biochem. Sci.* 29, 627–633. doi:10.1016/j.tibs.2004.10.006
- Miles, E. W., Rhee, S., and Davies, D. R. (1999). The molecular basis of substrate channeling. *J. Biol. Chem.* 274, 12193–12196. doi:10.1074/jbc.274.18.12193
- Moon, T. S., Dueber, J. E., Shiue, E., and Prather, K. L. (2010). Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* 12, 298–305. doi:10.1016/j.ymben.2010.01.003
- Moon, T. S., Yoon, S. H., Lanza, A. M., Roy-Mayhew, J. D., and Prather, K. L. (2009). Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 589–595. doi:10.1128/AEM.00973-08

- Nakamura, C. E., and Whited, G. M. (2003). Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* 14, 454–459. doi:10.1016/j.copbio.2003.08.005
- Pawson, T., Gish, G. D., and Nash, P. (2001). SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* 11, 504–511. doi:10.1016/S0962-8924(01)02154-7
- Pfeifer, B. A., and Khosla, C. (2001). Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* 65, 106–118. doi:10.1128/MMBR.65.1.106-118.2001
- Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943. doi:10.1038/nature04640
- Robinson, C. R., and Sauer, R. T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5929–5934. doi:10.1073/pnas.95.11.5929
- Saksela, K., and Permi, P. (2012). SH3 domain ligand binding: what's the consensus and where's the specificity? *FEBS Lett.* 586, 2609–2614. doi:10.1016/j.febslet.2012.04.042
- Schultz, J., Hoffmüller, U., Krause, G., Ashurst, J., Macias, M. J., Schmieder, P., et al. (1998). Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat. Struct. Biol.* 5, 19–24. doi:10.1038/nsb0198-19
- Schweimer, K., Hoffmann, S., Bauer, F., Friedrich, U., Kardinal, C., Feller, S. M., et al. (2002). Structural investigation of the binding of a herpesviral protein to the SH3 domain of tyrosine kinase Lck. *Biochemistry* 41, 5120–5130. doi:10.1021/bi015986j
- Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., et al. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* 275, 73–77. doi:10.1126/science.275.5296.73
- Spivey, H. O., and Ovadi, J. (1999). Substrate channeling. *Methods* 19, 306–321. doi:10.1006/meth.1999.0858
- Takamatsu, N., Watanabe, Y., Yanagi, H., Meshi, T., Shiba, T., and Okada, Y. (1990). Production of enkephalin in tobacco protoplasts using tobacco mosaic virus RNA vector. *FEBS Lett.* 269, 73–76. doi:10.1016/0014-5793(90)81121-4
- Veggiani, G., Zakeri, B., and Howarth, M. (2014). Superglue from bacteria: unbreakable bridges for protein nanotechnology. *Trends Biotechnol.* 32, 506–512. doi:10.1016/j.tibtech.2014.08.001
- Vo, T. D. L., Ko, J. S., Park, S. J., Lee, S. H., and Hong, S. H. (2013). Efficient gamma-aminobutyric acid bioconversion by employing synthetic complex between glutamate decarboxylase and glutamate/GABA antiporter in engineered *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 40, 927–933. doi:10.1007/s10295-013-1289-z
- Wang, Y., and Yu, O. (2012). Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. *J. Biotechnol.* 157, 258–260. doi:10.1016/j.jbiotec.2011.11.003
- Welch, G. R., and Gaertner, F. H. (1980). Enzyme organization in the polyaromatic-biosynthetic pathway: the arom conjugate and other multienzyme systems. *Curr. Top. Cell. Regul.* 16, 113–162. doi:10.1016/B978-0-12-152816-4.50009-1
- Werner, S., Marillonnet, S., Hause, G., Klimyuk, V., and Gleba, Y. (2006). Immunoabsorbent nanoparticles based on a *Tobamovirus* displaying protein A. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17678–17683. doi:10.1073/pnas.0608869103
- Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., et al. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 109, E690–E697. doi:10.1073/pnas.1115485109
- Zhao, H. L., Yao, X. Q., Xue, C., Wang, Y., Xiong, X. H., and Liu, Z. M. (2008). Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Expr. Purif.* 61, 73–77. doi:10.1016/j.pep.2008.04.013
- Zhao, S., Jones, J. A., Lachance, D. M., Bhan, N., Khalidi, O., Venkataraman, S., et al. (2015). Improvement of catechin production in *Escherichia coli* through combinatorial metabolic engineering. *Metab. Eng.* 28, 43–53. doi:10.1016/j.ymben.2014.12.002

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 Horn and Sticht. 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 of metabolic pathways by artificial enzyme channels

Marlene Pröschel¹, Rainer Detsch², Aldo R. Boccaccini² and Uwe Sonnewald^{1*}

¹ Department of Biology, Biochemistry Division, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany,

² Department of Materials Science and Engineering, Institute of Biomaterials, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

OPEN ACCESS

Edited by:

Zoran Nikoloski,
Max-Planck Institute of Molecular
Plant Physiology, Germany

Reviewed by:

Daehee Lee,
Korea Research Institute of
Bioscience and Biotechnology,
South Korea
Lee Sweetlove,
University of Oxford, UK

*Correspondence:

Uwe Sonnewald
uwe.sonnewald@fau.de

Specialty section:

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

Received: 29 July 2015

Accepted: 06 October 2015

Published: 21 October 2015

Citation:

Pröschel M, Detsch R, Boccaccini AR
and Sonnewald U (2015) Engineering
of metabolic pathways by artificial
enzyme channels.
Front. Bioeng. Biotechnol. 3:168.
doi: 10.3389/fbioe.2015.00168

Application of industrial enzymes for production of valuable chemical compounds has greatly benefited from recent developments in Systems and Synthetic Biology. Both, *in vivo* and *in vitro* systems have been established, allowing conversion of simple into complex compounds. Metabolic engineering in living cells needs to be balanced which is achieved by controlling gene expression levels, translation, scaffolding, compartmentation, and flux control. *In vitro* applications are often hampered by limited protein stability/half-life and insufficient rates of substrate conversion. To improve stability and catalytic activity, proteins are post-translationally modified and arranged in artificial metabolic channels. Within the review article, we will first discuss the supramolecular organization of enzymes in living systems and second summarize current and future approaches to design artificial metabolic channels by additive manufacturing for the efficient production of desired products.

Keywords: metabolic engineering, matrix-bound enzymes, protein scaffolding, enzyme arrays, metabolic channels, isopeptide-bonding, SpyCatcher/SpyTag, additive manufacturing

INTRODUCTION

Living cells are highly dynamic and complex metabolic systems in which most enzymes do not function in isolation but form supramolecular complexes (Jørgensen et al., 2005). By providing spatial and temporal organization of molecules within the cell, these complexes allow optimized substrate channeling and thereby prevent loss of intermediates and improve control and efficiency of catalysis. To mimic supramolecular complexes, several approaches to co-localize functionally related enzymes have been followed. These include scaffolding of enzymes to generate artificial substrate channels (Dueber et al., 2009; Lee et al., 2012). *In vitro* scaffolding of enzymes can be achieved by, i.e., cross-linking, encapsulation, and binding to nucleic acid or protein scaffolds. The latter two options allow the sequential arrangement of enzymes in a correct, programmable, and defined spatial order. Protein-based scaffolding requires specific binding domains for interaction. This bears some problems: only a limited number of high-affinity interaction domains are available, binding efficiency of different domains may not be comparable and interactions are reversible that may result in a short half-life of the artificial channel. To circumvent these problems, covalent linkages between the synthetic scaffold platform and the enzymes to be arranged would be advantageous. In nature, inter- and intramolecular isopeptide bonds are formed to stabilize proteins or to label proteins for proteolysis by ubiquitylation (Kang and Baker, 2011). By dissecting the mechanism of spontaneous intramolecular isopeptide formation within the CnaB2 domain of the fibronectin-binding protein FbaB from *Streptococcus Pyogenes* (Spy), Howarth and co-workers developed a versatile tool to allow covalent binding of tagged-enzymes to modified macromolecules (Zakeri and Howarth, 2010). This

approach can be applied to cell free and possibly even to cellular systems. Besides designing covalent/irreversible or reversible synthetic protein complexes for metabolic engineering, three-dimensional (3-D) printing of enzyme arrays may enable the design of *in vitro* protein channels. These channels do not rely on protein–protein interactions but are based on the sequential printing of individual enzymes. Within the review article, we will describe examples of supramolecular organization in cells, attempts to immobilize and stabilize enzymes for industrial use, and finally summarize current approaches to design artificial metabolic channels by additive manufacturing (AM) for efficient production of valuable chemical products.

CELLULAR PROTEINS ARE ORGANIZED IN SUPRAMOLECULAR STRUCTURES

Cellular systems are highly complex and contain high concentrations of macromolecules (Long et al., 2005; Conrado et al., 2008; Good, 2011; Chen and Silver, 2012). Within the cell, these molecules are organized in a temporal and spatial manner allowing the cell to fulfill its many distinct reactions that take place simultaneously (Good, 2011). Coordination and organization of cellular processes is achieved through compartmentation (Chen and Silver, 2012). The need for spatial and temporal organization of proteins in signaling pathways and metabolism is evident when looking at the crowded milieu of macromolecules inside cells and the many complex and competing reactions running concurrently (Sweetlove and Fernie, 2013). In signaling pathways the question arises, how correct interaction partners find each other while avoiding interaction and cross-talk with the wrong ones (Good, 2011). This is important since the correct communication of functionally interacting proteins is a prerequisite for the coordination and regulation of many cellular processes required for appropriate cellular responses to external and internal stimuli (Chen et al., 2014). Strict control and tight regulation of flux through metabolic pathways is of equal importance (Dueber et al., 2009). Metabolic regulation faces many challenges, including avoidance of flux imbalances, slow turnover rates of enzymes, toxic pathway intermediates, and competing metabolic reactions (Figure 1; Conrado et al., 2008; Chen and Silver, 2012; Lee et al., 2012). Consequently, engineering of artificial metabolic pathways in living cells often suffers from low productivity and yield if spatial organization/compartmentation strategies are not included in the engineering concepts (Conrado et al., 2008). To increase the overall cellular efficiency, accuracy, and specificity, nature has evolved compartmentation strategies to control and regulate flux through metabolic and signaling pathways (Chen and Silver, 2012; Conrado et al., 2012).

Intracellular compartmentation can be divided into macro-compartmentation and micro-compartmentation (Sweetlove and Fernie, 2013). Macro-compartmentation refers to the separation of reaction compartments, organelles, by biological membranes. Organelles, which are a hallmark of eukaryotic cells, contain a certain subset of metabolic enzymes that carry out distinct biological reactions. This physical separation of biological reactions increases the overall metabolic efficiency and allows even incompatible

or contradictory reactions such as synthesis (anabolism) and degradation (catabolism) or oxidation and reduction to take place within the same cell at the same time. Compartmentation also allows detoxification of toxic pathway intermediates without harming the cell. In peroxisomes, for example, the cytotoxic reactive oxygen species H_2O_2 is efficiently degraded by catalase that is specifically present in these organelles. As the amount of catalase in peroxisomes is so high, released H_2O_2 is degraded instantly.

In nature, most cellular multi-cascaded reactions are not catalyzed by free-floating, isolated enzymes but by multienzyme complexes. In these microcompartments, several enzymes form so-called metabolons or metabolic channels overcoming flux imbalances, diffusion and loss of intermediates, and release of toxic intermediates (Figure 2; Conrado et al., 2008; Lee et al., 2012; Jia et al., 2014). The assembly of sequential pathway enzymes into metabolons offers several advantages (Jørgensen et al., 2005; Conrado et al., 2008) when compared to isolated, soluble enzymes. The overall catalytic efficiency is increased because active centers of sequential pathway enzymes are brought into close proximity (enforced proximity) allowing direct transfer of intermediates from one enzyme to the other while avoiding metabolic interference. This phenomenon is called substrate or metabolic channeling and is based on the ordered cascading of subsequent enzymatic steps in which one enzyme produces the substrate of the following enzyme. Thanks to the reduction in diffusion distance and transit time, the effective local concentration of pathway intermediates is higher around the enzyme complex compared to the rest of the cell (Lee et al., 2012). This increase in local concentrations of metabolites prevents unspecific side reactions, favors reaction kinetics, and directs the intracellular flux of metabolites toward the synthesis of the desired product. However, it should be noted that this model is only true for diffusion-limited enzymes (Lee et al., 2012; Sweetlove and Fernie, 2013). In fact, the majority of enzymes are not diffusion-limited which means that the chemistry/conversion process is slower than the diffusion rate. For those non-diffusion-limited enzymes, the most prominent benefits of metabolic channeling are a reduced time to steady state and a better control of reaction specificity and regulation of metabolic branch points (Lee et al., 2012; Sweetlove and Fernie, 2013). Moreover, the release of toxic and/or unstable intermediates into the bulk phase is restricted by such multienzyme complexes that function as pipelines that strictly control the metabolic flux.

There are many examples in a wide range of organisms, including eukaryotes and prokaryotes, where the assembly of different individual sequential enzymes of a metabolic pathway into functional multiprotein complexes increases the overall catalytic efficiency. In the case of primary metabolism, multienzyme complexes are involved in central carbon metabolism (e.g., glycolysis, citric acid cycle), fatty acid oxidation, the Calvin cycle, amino acid biosynthesis (Tryptophan synthesis), the carboxysome, and the proteasome (Conrado et al., 2008). Al-Habori (1995) stated for example that glycolytic enzymes are co-localized on actin filaments to form an active complex. Depending on the energy demand, other authors observed that many of the glycolytic enzymes can be functionally associated with the outer mitochondrial membrane when there is a need for pyruvate to fuel

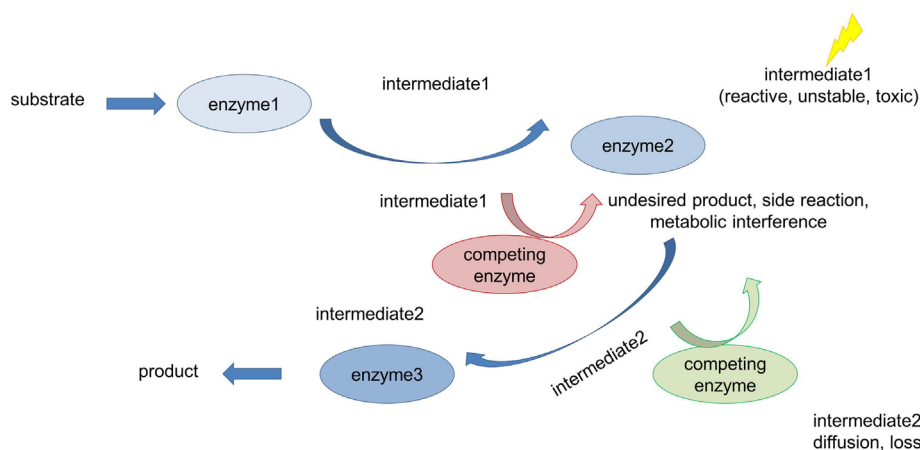


FIGURE 1 | Metabolic challenges cells have to deal with. Many metabolic reactions are characterized by low productivity and product yield for the following reasons: accumulation of toxic and/or unstable intermediates, loss of intermediates due to diffusion into the bulk phase (long distances between interacting proteins), and competing pathways (metabolic interference) also leading to undesirable side reactions.

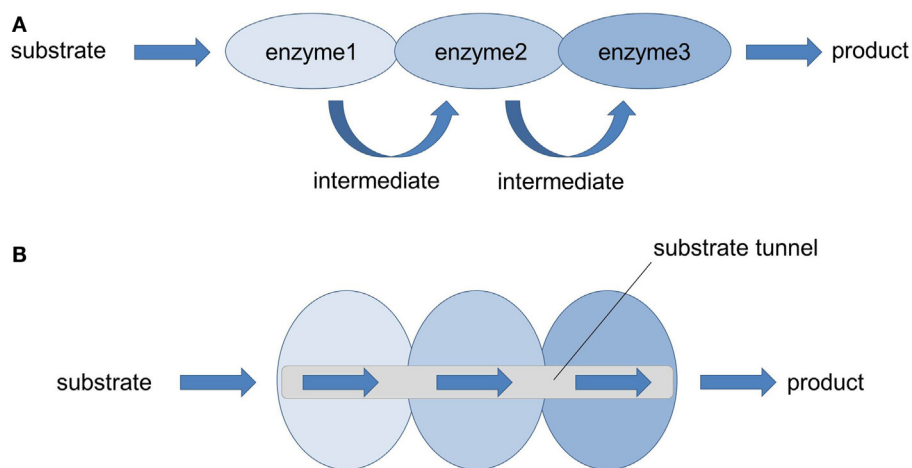


FIGURE 2 | Multienzyme complexes (= metabolons): nature's strategy to solve metabolic issues. (A) In nature, many enzymes performing sequential multistep transformation of a substrate are co-localized into multienzyme complexes. Thereby the local concentration of pathway metabolites is increased, the accumulation of pathway intermediates (toxic, reactive, unstable) is limited and the probability of metabolic interference and cross-talk with other cellular components is avoided (Lee et al., 2012). **(B)** Thanks to the enforced proximity of functionally related enzymes, efficient enzyme-to-enzyme channeling through a substrate tunnel can take place, leading to an increased catalytic performance and product yield.

respiration (Giegé et al., 2003; Lunn, 2007; Møller, 2010). This defined intracellular spatial localization of the glycolytic enzymes makes sense as the product pyruvate can directly be transferred into the mitochondria where the next reactions of the central carbon catabolism (citric acid cycle) take place. Reversibility of this spatial organization is important, since pyruvate is an important and central intermediate of several biosynthetic pathways. Several enzymes performing sequential conversion steps in the citric acid cycle are also thought to be associated in a multienzyme complex within the mitochondrial matrix (Barnes and Weitzman, 1986; Lunn, 2007; Jia et al., 2014). The pyruvate dehydrogenase complex that catalyzes the conversion of pyruvate into acetyl-CoA also

exhibits an efficient multienzyme structure allowing for substrate channeling and active-site coupling (Smolle and Lindsay, 2006; Jia et al., 2014).

In plants secondary metabolism, biosynthesis of isoprenoids, alkaloids, flavonoids, cyanogenic glucosides (e.g., dhurrin) and phenylpropanoids are examples of the presence/involvement of multienzyme complexes (Winkel, 2004; Jørgensen et al., 2005; Conrado et al., 2008). The phenylpropanoid pathway demands a substantial portion of carbon and energy fixed during photosynthesis. It is organized in a metabolic grid giving rise to a large number of different metabolites (Laursen et al., 2015). Synthesis of the cyanogenic glucoside dhurrin requires seven enzymatic

steps starting from tyrosine. In *Sorghum bicolor*, these reactions are catalyzed by two multifunctional enzymes and one monofunctional enzyme that form a metabolic channel. The importance of this channel became evident in transgenic *Arabidopsis* plants expressing only the two first enzymes of the pathway. As a consequence, the resulting transgenic plants showed significant stunting, most likely caused by accumulation of the toxic pathway intermediate, p-hydroxymandelonitrile. Co-expression of the third enzyme restored normal plant growth and eliminated accumulation of the intermediate.

In fungi, supramolecular enzyme organization can be found in the polyaromatic/shikimate pathway. This pathway includes a multifunctional enzyme known as the AROM complex that has evolved to link five distinct enzymatic activities into a single pentafunctional polypeptide (Conrado et al., 2008). The fungal AROM complex is encoded by a single gene cluster that is expressed in a coordinate manner and also remains associated after synthesis. Interestingly, in *Escherichia coli* and other bacteria, the enzymes are encoded by individual genes distributed throughout the genome (Bachmann, 1983). This precise organization of multiple enzymatic activities on one single peptide, as found in the fungal AROM complex, allows very efficient substrate tunneling between adjacent active sites. Another example of supramolecular enzyme organization in the polyaromatic pathway is the tryptophan synthase. The enzyme is composed of two subunits α and β that assemble as a stable $\alpha\beta\alpha$ multienzyme complex. Close vicinity of both subunits allows the α subunit to channel the reactive indole intermediate to the β subunit via a hydrophobic, physical tunnel exactly matching indole (Conrado et al., 2008; Dueber et al., 2009). In the case of the tryptophan synthase complex, the active sites are only 25 Å apart from each other leading to the prevention of diffusion of the reactive indole intermediate (Dueber et al., 2009). Thereby the cell is protected and the enzymatic conversion is significantly increased because of the high effective, local concentration of indole (Winkel, 2004; Conrado et al., 2008; Dueber et al., 2009; Chen and Silver, 2012; Lee et al., 2012; Jia et al., 2014).

Bacteria also spatially organize their interior milieu for specialized functions (Chen and Silver, 2012) although lacking membrane-bound organelles as isolated reaction compartments. The interior of prokaryotes can contain various protein-based compartments to physically separate distinct, often critical, enzymatic reactions (Boyle and Silver, 2012). These protein-based compartments are composed of multiple proteins that co-assemble to form thin protein shells and typically encapsulate sequential pathway enzymes (encapsulation). These shells are named bacterial microcompartments (BMCs). The number or amount of encapsulated enzymes is defined by the size of the protein-based compartments. The incorporation of the enzymes into the complex is thought to take place during the assembly of the shell proteins. Small pores (~0.5 nm) on the shells allow the exchange of metabolites according to the size exclusion limit of the pores. Therefore, unrestricted metabolite diffusion across the shell is prevented. Additionally, some of these pores have been shown to be selective (Lee et al., 2012). One example of BMCs found in cyanobacteria and other autotrophic prokaryotes (Chen

and Silver, 2012) are carboxysomes (approximately 100 nm in diameter) that are estimated to contain 270 molecules of the key carbon fixation enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Lee et al., 2012; Chen et al., 2014). In these proteinaceous microcompartments, the local CO₂ concentration is dramatically increased in the vicinity of RuBisCO preventing the oxygenase reaction of the enzyme and hence increasing its catalytic activity (Lee et al., 2012; Chen et al., 2014). Another example of defined spatial organization to increase metabolic efficiency in prokaryotes is the multienzyme cellulosome complex found in the cellulolytic *Clostridium thermocellum* (Chen and Silver, 2012). Due to the specific interaction of the complementary protein domains Dockerin and Cohesin, many different enzymes required for the degradation of plant cell walls are organized as extracellular nanomachinery on a scaffold on the cell surface (Figure 3; Lytle and Wu, 1998; Bayer et al., 2004; Pinheiro et al., 2009; Mazzoli et al., 2012). The organization and co-localization of the different hydrolytic enzymes and the close proximity to the substrate provides an efficient synergistic strategy to degrade cellulose and hemicellulose (Gefen et al., 2012).

In mammalian cells, the purinosome catalyzes the conversion of phosphoribosyl pyrophosphate (PRPP) to inosine monophosphate (IMP). This conversion requires six enzymes that are co-localized on microtubules to form an efficient metabolon. Assembly and disassembly are highly regulated and link the rate of *de novo* purine synthesis to the cellular purine nucleotide pool (DeLisa and Conrado, 2009). The purinosome complex is formed upon depletion of purines. The complex is associated with microtubules and stabilized by molecular chaperones, including HSP70 and HSP90. Accumulation of purines induces the disassembly of the complex. This process involves reversible protein phosphorylation by protein kinase CK2 (for review, see Laursen et al., 2015). Assembly and disassembly of purinosomes are remarkable examples of the dynamic regulation of multienzyme complexes which is of utmost importance to tightly coordinate metabolic flux through metabolic channels (Conrado et al., 2008; DeLisa and Conrado, 2009; Möller, 2010).

If not encoded by a single gene cluster (i.e., the AROM complex in fungi), the described supramolecular enzyme complexes assemble post-translationally into rather stable or transient complexes. Transient complexes allow dynamic responses to intra- (e.g., metabolic demands) and extracellular stimuli or signals (e.g., abiotic and biotic challenges) which is important for fine-tuning metabolism according to the state/demand of the cell or organism (Conrado et al., 2008; DeLisa and Conrado, 2009; Möller, 2010; Sweetlove and Fernie, 2013). Transient or dynamic metabolons are also called “functioning-dependent structures” (Möller, 2010). This micro-compartmentation can be achieved in different ways. One strategy of living cells is to co-localize interacting proteins/enzymes by anchoring them on membranes or on cytoskeleton structures (actin, microtubules). Alternatively, multienzyme complexes are organized by scaffold proteins (Good, 2011). The latter strategy is found in some metabolic pathways such as the cellulosomes but occurs frequently in signal cascades. In yeast, for example, the Ste5 scaffold protein organizes the interaction between Fus3 (MAPK), Ste7

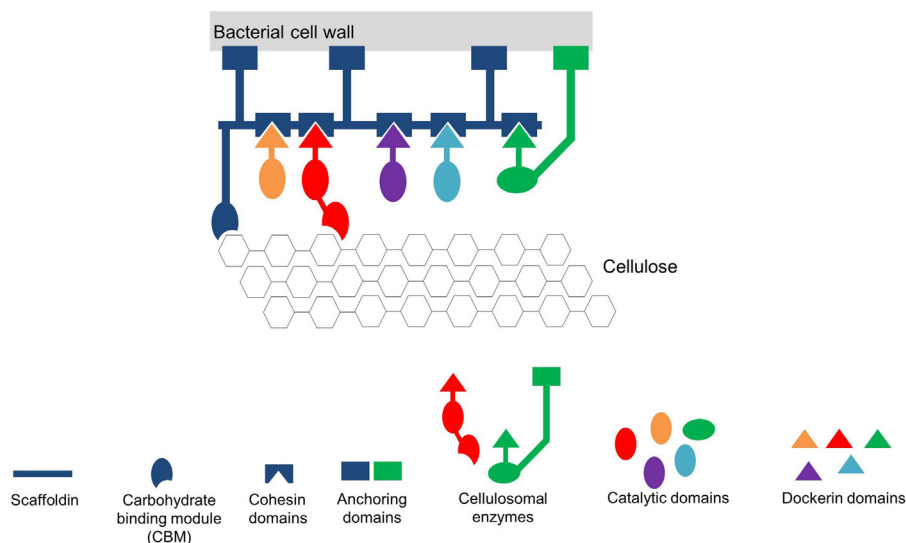


FIGURE 3 | Simplified scheme of the cellulosome from *Clostridium thermocellum* (adapted from Mazzoli et al., 2012). The Cellulosome in this scheme consists of a scaffold protein (Scaffoldin) that is composed of five Cohesin domains. The whole complex is connected to the bacterial cell wall via protein anchors. Each enzyme has its own specific Dockerin domain. By specifically interacting, the Cohesin- and the Dockerin domains form the multienzyme complex in close vicinity to the substrate cellulose. To further increase the affinity of the multienzyme complex toward the substrate the whole scaffold is linked to the substrate via carbohydrate binding modules (CBM).

(MAPKK), and Ste11 (MAPKKK) and is essential for mating. In mammalian cells, the Kinase suppressor of Ras (KSR) functions as a scaffold in the Ras–Raf–MEK–MAPK pathway (Roy et al., 2002) and the PSD-95 synaptic scaffold is crucial for the organization of neuronal synapses controlling the neurotransmitter receptor density (Good, 2011). Scaffold proteins are defined as extremely diverse proteins that coordinate the physical assembly of individual partner molecules (Good, 2011). Scaffold proteins, composed of multiple modular interaction domains (for example, protein–protein interaction domains) or motifs (Good, 2011), form flexible platforms where other proteins or relevant molecular components of a specific pathway can bind to. Thereby, the interaction partners are co-localized in a modular manner. Signaling pathways, where cascaded enzyme reactions often take place, especially benefit from this subcellular, spatial organization. Scaffolding strategy supports specificity, accuracy, and efficiency of signal transduction pathways, by enforcing proximity of the correct interaction partners whereas incorrect interaction partners are excluded (Good, 2011).

SYNTHETIC MULTIENZYME COMPLEX FORMATION TO MIMIC NATURE'S STRATEGY TO INCREASE METABOLIC AND SIGNALING EFFICIENCY

In metabolic engineering where a natural endogenous biosynthetic pathway is manipulated to increase productivity and yield of a valuable molecule (Sonnewald, 2003; Capell and Christou, 2004; Na et al., 2010; Chen and Silver, 2012), several

challenges have to be overcome. One is the expression level of heterologous enzymes. In bacterial systems, expression levels are often much higher compared to endogenous enzymes. This can lead to cellular stress responses, for example, due to the huge amount of proteins produced that are often even unfolded or misfolded, flux imbalances coupled with unpredictable and non-controllable metabolic changes and high energy and molecule (amino acids, nucleotides) consumption. An additional challenge is to balance expression of consecutive enzymes (Na et al., 2010). If the reactions are not balanced in a manipulated pathway, toxic intermediates are likely to accumulate which can lead to death of the expression host (Dueber et al., 2009). With nature's strategies to increase metabolic efficiency in mind, metabolic engineers are trying to engineer artificial multienzyme complexes, where the enzymes performing consecutive reactions are spatially organized (directed enzyme organization). The idea is to co-localize functional enzymes into complexes. Due to the enforced proximity of the enzyme active sites and the formation of enzyme microdomains built as a consequence of coclustering of multiple enzymes into higher aggregates, catalytic efficiency and metabolic pathway performance are improved (Sweetlove and Fernie, 2013; Castellana et al., 2014). Overall, pathway balancing involves several layers, including DNA copy number, transcriptional and translational regulation, scaffolding and compartmentation, as well as inclusion of metabolic sensors balancing the flux through synthetic pathways (Boyle and Silver, 2012; Jones et al., 2015). In addition to balancing protein amount and organization, enzyme engineering allows to improve activity, selectivity, and stability of enzymes (Otte and Hauer, 2015). Several approaches can

be followed to stabilize enzymes. One promising approach involves cyclization of enzymes which has been achieved by using the split intein (Zhao et al., 2010) or SpyTag/SpyCatcher system (Schoene et al., 2014).

NON-PROGRAMMABLE MATRIX-BOUND ENZYME COMPLEXES

One strategy of engineering efficient multienzyme complexes that mimic those found in nature is to co-immobilize multiple enzymes of a sequential/cascaded pathway on the same carrier or on the same matrix. Early studies on immobilized enzymes clearly demonstrated that tethering of enzymes to particles significantly improves product formation. Comparing the activity of three successive enzymes, β -galactosidase, hexokinase, and glucose-6-phosphate-dehydrogenase, either matrix-bound or soluble, Mattiasson and Mosbach (1971) observed a faster conversion of the substrate when the sequential enzymes are co-localized on the same particle. Coupling was achieved by the CNBr method yielding covalently bound enzymes. Obvious disadvantages of the system include (i) spatial arrangement of enzymes is impossible and (ii) chemical cross-linking bears the risk of losing enzyme activity because chemical cross-linking is random and can also affect amino acid residues in the active center of the enzyme. The group of Mallapragada (Jia et al., 2013) sequentially co-localized the two model enzymes glucose oxidase (GOX) and horseradish peroxidase (HRP) on dual-functionalized polystyrene nanoparticles. The nanoparticles had been functionalized with carboxyl groups which have partially been modified using biotin hydrazide resulting in biotinylated carboxyl-polystyrene nanoparticles. This dual-functionalization allowed the use of different attachment strategies for each enzyme to better control the relative amounts of the enzymes on the nanoparticle. The streptavidin-tagged HRP was attached to the nanoparticles via the high-affinity biotin–streptavidin interaction. The unmodified carboxyl groups on the nanoparticles were used to covalently attach GOX by amide bond formation between the reactive carboxyl groups on the nanoparticle and amino groups of the enzyme. Immobilized enzymes retained their enzymatic activity that was comparable to free enzymes. Interestingly, sequential co-localization of GOX and HRP resulted in a twofold enhancement of the overall product conversion rate compared to the free enzymes and a mixture of individual immobilized enzymes on nanoparticles (Jia et al., 2013).

In several studies, immobilization of enzymes caused positive catalytic/kinetic effects and at the same time resulted in the stabilization of various enzymes (Sheldon, 2007; Garcia-Galan et al., 2011; Homaei et al., 2013; Guzik et al., 2014). For industrial applications, immobilized enzymes have additional advantages. They can be reused over multiple cycles and the enzymes are sequestered from the product stream. These properties improve industrial processes and first applications of immobilized enzymes include glucose isomerase for high fructose corn syrup, lipase for biodiesel production from triacylglycerides or thermolysin for aspartame synthesis (DiCosimo et al., 2013). Compared to the conventional fossil fuel-based chemistry, biomanufacturing

offers many advantages, including biocompatibility and sustainability that leads to massive growth rates of the world market for industrial enzymes.

SYNTHETIC COMPARTMENTATION BY PROTEIN ENCAPSULATION

Besides sequestration or tethering of enzymes using different matrices, encapsulation of proteins into semi-permeable compartments allows the physical separation of different metabolic reactions leading to an increase in pathway efficiency. As discussed above, bacteria encode proteinaceous shells, BMCs, in which functionally related enzymes are sequestered. Up to now BMC-shell encoding genes have been found in over 400 different sequenced bacterial genomes (Choudhary et al., 2012). The proteins form polyhedral structures similar to virus-like particles. Co-expression of recombinant shell proteins and selected proteins of interest (e.g., pathway enzymes), fused to shell-targeting signal peptides, in *E. coli* cells, leads to the functional compartmentation of heterologous enzymes in recombinant shells (Choudhary et al., 2012). In addition to bacterial shell proteins, also viral capsid proteins have been used as a tool for compartmentalizing engineered pathways (Chen et al., 2014). Co-expression of two or three enzymes fused by linker sequence to form one multifunctional protein with the bacteriophage P22 capsid protein allowed the design of a synthetic metabolon (Patterson et al., 2014). To improve the suitability of capsid and cargo proteins, several modifications have been tested. By adding positively charged peptides to capsid proteins and negatively charged peptides to cargo proteins, binding could be improved via enhanced electrostatic forces (Chen et al., 2014).

PROGRAMMABLE AND REVERSIBLE SCAFFOLDING

Besides the co-immobilization strategy of enzymes on carriers to build efficient synthetic multiprotein complexes, various genetic modules (proteins, nucleic acids) have been described to function as building blocks for generating programmable, modular scaffolds on which enzymes can specifically be co-localized in a spatially ordered manner by simple tethering mechanisms. Specific protein–protein, DNA–DNA/RNA–RNA, and DNA–protein/RNA–protein interactions have been used to co-localize metabolic enzymes to improve pathway flux. Dueber et al. (2009) designed synthetic protein scaffolds out of known, well-characterized, and widespread protein–protein interaction domains from metazoan signaling proteins (SH3-, PDZ-, and GBD binding domain). By fusing enzymes of the mevalonate biosynthetic pathway (AtoB, HMGS, and HMGR) to the respective peptide ligands (SH3-, PDZ-, and GBD-ligand), the authors generated a modular genetically encoded scaffold system on which the enzymes can be co-localized in a programmable and defined manner (Figure 4). This scaffold approach is based on simple, specific high-affinity interactions between protein binding domains and their cognate, specific peptide ligands. In addition to the close proximity of the enzymes co-immobilized on the protein scaffold and the substrate

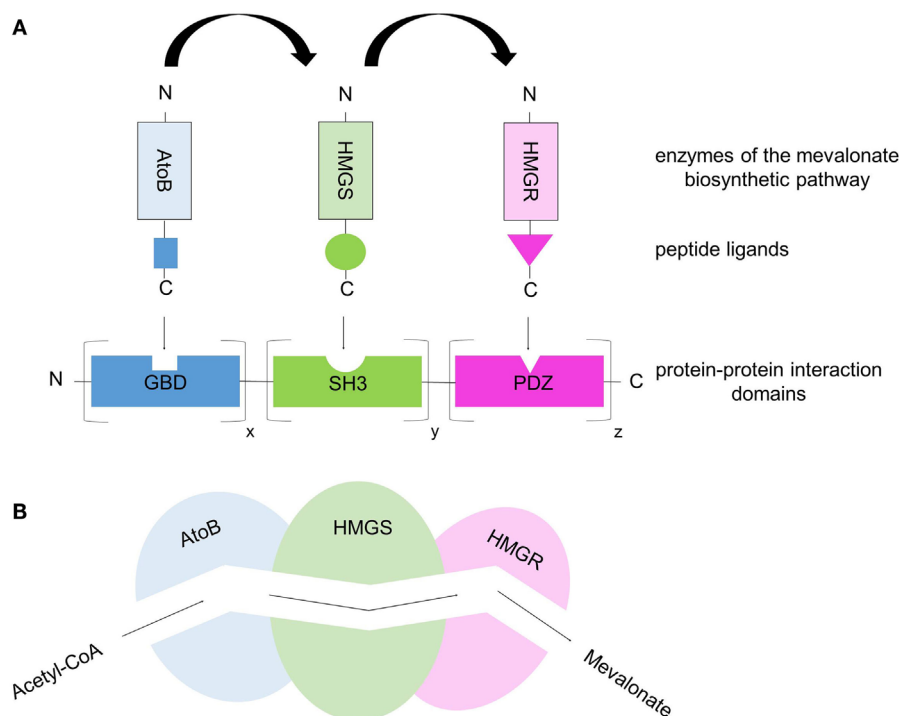


FIGURE 4 | Protein-scaffold-based multienzyme complex offers a synthetic metabolic pipeline that allows substrate channeling. (A) The three enzymes of the mevalonate biosynthetic pathway (acetoacetyl-CoA thiolase, AtoB; hydroxy-methylglutaryl-CoA synthase, HMGS; hydroxy-methylglutaryl-CoA reductase, HMGR) were co-localized on a synthetic protein-scaffold via high-affinity interactions between protein domains and their specific, cognate peptide ligands. The protein scaffold consists of various repeats of well-characterized, metazoan protein domains (SH3, GBD, PDZ) and the corresponding peptide ligands are fused to the enzymes. By co-expressing the synthetic scaffold and the enzyme–peptide ligand fusions in *E. coli* cells, the three enzymes are targeted to the scaffold building a multienzyme complex. Through the variation of the binding domain repeats the ratio/stoichiometry of enzymes can be controlled. **(B)** Simplified scheme of substrate channeling through a tunnel/metabolic pipeline connecting the active sites of the co-localized enzymes.

channeling resulting from this, the reaction efficiency and the overall production rate/product yield can be further improved by varying the number of protein binding domain repeats in the scaffold taking into account the kinetic properties of the individual enzymes. Therefore, potential enzymatic bottlenecks, e.g., caused by low K_{cat} values or binding affinities (K_M values), can be compensated by simply increasing the amount of “weak” enzymes with slow turnover rates. When the optimal scaffold stoichiometry (optimal enzyme ratio) is used, the conversion from acetyl-CoA to mevalonate performed by the three sequential enzymes AtoB, HMGS, and HMGR produced a 77-fold higher level of the product mevalonate compared to that of the un-scaffolded pathway (Dueber et al., 2009; Lee et al., 2012; Chen et al., 2014). This demonstrates once again that the spatial organization of enzymes into functional complexes allowing effective substrate channeling increases the overall metabolic efficiency because the local concentrations of metabolic intermediates are increased while their accumulation to toxic levels is prevented (Dueber et al., 2009). The same modular protein-based scaffold strategy was applied to the three-enzyme glucaric acid pathway, where glucaric acid is produced from glucose. Compared to the free enzymes, glucaric acid levels have been improved fivefold by scaffolding in *E. coli* cells (Moon et al., 2010; Lee et al., 2012). Another example of the use of synthetic protein scaffolds to increase pathway flux is

resveratrol biosynthesis in yeast cells (Wang and Yu, 2012). Here, the authors scaffolded two enzymes, 4-coumarate: CoA ligase and stilbene synthase and achieved a fivefold increase in resveratrol synthesis compared to the un-scaffolded control. While the above given examples demonstrate the power of protein scaffolding using protein–protein interaction domains from metazoan signal transduction pathways, these domains may not be applicable in all systems. They may misfold or aggregate and cross-talk between engineered scaffolds with native signaling molecules may occur. The potential risk of cross-talk is dependent on the organism and may be neglectable for *E. coli* cells in which the described domains are not present (Dueber et al., 2009). To circumvent unintended perturbations of signaling pathways of the expression host, minimized synthetic domains or alternatives are required. One alternative are interaction domains derived from the bacterial cellulosomes. Based on the modular architecture of bacterial cellulosomes (discussed above), complementary protein modules of Cohesin and Dockerin can be used to generate artificial multienzyme complexes. Any enzyme of interest can be genetically fused to Dockerin domains. A scaffold consisting of various Cohesin domains leads to the targeting of the enzymes in a defined spatial orientation due to the specific protein–protein interaction between Cohesin domains and their cognate Dockerin domains. Thereby the so-called “Designer-Cellulosomes” can be generated.

An alternative to protein scaffolds are nucleic acid-based (DNA or RNA) scaffolds. DNA and RNA molecules represent suitable modular tools for the specific programmable spatial organization of pathway enzymes as they provide specific interactions by either hybridization (base pair complementarity, DNA–DNA/RNA–RNA binding) or protein binding sequences specific for engineered zinc-finger or TALE proteins (DNA/RNA–protein binding). Conrado et al. (2012) used a configurable DNA-based scaffold to spatially arrange multiple pathway enzymes in a distinct order in the cytoplasm of *E. coli*. To specifically target pathway enzymes to the plasmid DNA-based scaffold consisting of multiple unique zinc-finger binding sites, the enzymes of interest were genetically fused to corresponding zinc-finger domains that specifically bind the DNA sequences present in the engineered scaffold with high affinity. Increased overall production rates due to the DNA scaffold-mediated enzyme co-assembly could be observed when enzymes of the resveratrol biosynthesis and the mevalonate synthesis were used (Conrado et al., 2012).

Another example of a DNA-assembled artificial multienzyme complex made use of luciferase and oxidoreductase catalyzing two consecutive reactions of flavin mononucleotide reduction and aldehyde oxidation (Niemeyer et al., 2002; Müller and Niemeyer, 2008). DNA hybridization of complementary ssDNA oligonucleotides in combination with the high-affinity biotin–streptavidin protein interaction was used to build the bienzymatic complex. Biotinylated enzymes were linked to covalent ssDNA–streptavidin conjugates resulting in enzyme–DNA conjugates due to the specific and very strong high affine, but non-covalent interaction between biotin and streptavidin. As single-stranded DNA carrier strands, containing complementary regions to the DNA oligomers found in the enzyme–DNA conjugates, were previously immobilized on a surface (microtiterplate) via the biotin–streptavidin interaction, co-immobilization and therefore multienzyme complex formation of the enzymes takes place by specific DNA–DNA hybridization (complementary base pairing) (Niemeyer et al., 2002). To explore the proximity effect caused by the spatial arrangement of the two enzymes, activity assays were performed. Results indicate that the enforced spatial proximity of the enzymes luciferase and oxidoreductase increases the overall enzymatic activity of the bienzyme complex in a scaffold architecture dependent manner (Niemeyer et al., 2002). Almost the same DNA-scaffold approach was used by Müller and Niemeyer (2008). The authors reported on the DNA-directed assembly of GOX and HRP. Both enzymes were used because they are a suitable reporter system for which the kinetic rates and the output can easily be measured as the two-step reaction performed by these enzymes produces a highly fluorescent dye (Resorufin) when providing Amplex Red. Catalytic efficiency could be measured by monitoring the fluorescence emission. In this study, the authors could show that the efficiency of the two DNA–enzyme conjugates was dependent on the position and steric parameters (Müller and Niemeyer, 2008).

RNA molecules also provide a modular tool for scaffold approaches where multiple pathway enzymes can be targeted and thereby spatially organized. Delebecque et al. (2011) used RNA aptamers (= short single-stranded oligonucleotides) to create 1D and 2D scaffolds. These scaffolds were used for the

spatial organization of two bacterial enzymes required for hydrogen production. Similar to protein- and DNA-scaffolds, RNA-based scaffolds increased the rate of substrate conversion and product yield (Delebecque et al., 2011; Conrado et al., 2012). While proteins and especially large protein fusions often tend to misfold or aggregate, nucleic acids (DNA and RNA) have highly predictable local structures, and enzymes can be arranged into a given and programmable order by changing the distance between the protein binding sites (Conrado et al., 2012; Chen et al., 2014). Furthermore, DNA and RNA can easily be produced, fold into various structures (high-order assemblies), can have different lengths, and can consist of flexible numbers of repetitive scaffold units (Chen et al., 2014). For protein-based scaffolds, the addition of only short peptide ligands to each pathway enzyme is recommended to ensure the correct folding and therefore the function of the enzymes (Lee et al., 2012). In addition of course the fusion of, for example, zinc-finger domains to enzymes is critical, therefore an impact on the enzyme stability and activity cannot be excluded. Consequently, all modifications performed on the enzymes as well as on the modules (proteins, nucleic acids) to design the synthetic scaffold have to be optimized.

Besides all the advantages such as modularity and specificity, a big disadvantage of these scaffolds based on high affine protein–protein interaction domains is that the interactions are reversible depending on the different dissociation constants of the binding domain/peptide ligand pairs and they do not resist forces or boiling. Furthermore, not so many protein–protein or protein–peptide interactions are known that have low (nM) dissociation constants which is a measure for the affinity. In addition, each specific pair has its own dissociation constant and often they are not comparable with each other. This influences the targeting of the ligand-fused enzymes to the scaffold out of the protein binding domains. The different affinities complicate the specific and exact targeting of single enzymes to the scaffold in a defined ratio. There is also the risk of uncontrolled dissociation of the binding domain/peptide ligand pairs that are reversible and do not resist forces as the interaction is not covalent. Therefore, it would be desirable to create covalent linkages between proteins that are specific, formed under a wide range of conditions and which require fusion of small peptides only.

PROGRAMMABLE AND IRREVERSIBLE SCAFFOLDING OF ENZYMES

Inspiration for the design of intermolecular covalent linkages between proteins originates from the analysis of pili formation in Gram-positive bacteria. Pili of *Spy* are composed of three subunits, *Spy0125*, *Spy0128*, and *Spy0130* and are required for adhesion to host cells. Besides the well-known disulfide bridges between two cysteine residues, structural analysis of these and other bacterial surface proteins uncovered several additional covalent intramolecular bonds, including isopeptide bonds between lysine and asparagine or lysine and aspartate (Kang et al., 2014), ester bonds between threonine and glutamine (Kwon et al., 2014) or thioesters between cysteine and glutamine (Pointon et al., 2010; Walden et al., 2015).

The covalent SpySystem, pioneered by the group of Howarth, is based on the CnaB2 domain (immunoglobulin-like collagen adhesion domain) of the extracellular surface protein FbaB (fibronectin-binding protein) from the Gram-positive bacteria *Spy*. Within this CnaB2 domain, an autocatalytic, spontaneous, intramolecular reaction between a reactive lysine residue and a reactive aspartic acid residue takes place. This covalent isopeptide bond between Lys³¹ and Asp¹¹⁷ is catalyzed by a third amino acid residue, Glu⁷⁷. These three amino acids form a catalytic triad and are directly involved in the isopeptide bond formation. In nature, a lot of Gram-positive bacteria were discovered to form spontaneous, intramolecular isopeptide bonds within their surface proteins. As a consequence of the covalent bond within a single protein, the protein is connected to itself conferring thermal, proteolytic, and pH stability to the protein. Consequently, bacteria use this covalent intramolecular bond formation to stabilize their extracellular proteins that are, for example, essential for the penetration and invasion of the host cells. By splitting the CnaB2 domain, two protein partners, SpyTag and SpyCatcher, were generated (Zakeri et al., 2012). The SpyTag (13 amino acids) contains the reactive aspartic acid residue whereas the SpyCatcher (138 amino acids) harbors the reactive lysine residue and the catalytic glutamic acid residue. *In vitro* and *in vivo* experiments have shown that the two protein partners find each other, reconstitute and undergo covalent reaction simply upon mixing. The isopeptide bond forms within minutes and is stable under a wide range of conditions (pH, temperature, buffer composition, reducing agents, and detergents) (Zakeri et al., 2012). This fact is important and should be highlighted as other systems such as covalent disulfide bonds easily dissociate under reducing conditions and are therefore reversible (Veggiani et al., 2014). Due to the robustness and covalent character of this SpySystem and the fact that all components are genetically encoded and can be expressed efficiently in *E. coli*, the SpySystem is ideal for irreversibly attaching

proteins to each other. By fusing either SpyTag or SpyCatcher to the enzymes of interest, the catalyst can be covalently linked to corresponding SpyTag- or SpyCatcher-modified carriers (Figure 5). The SpySystem can also be used to create artificial multienzyme complexes that are stable. A synthetic scaffold composed of repeats of the SpyCatcher domain and SpyTag-enzyme conjugates are needed to co-localize sequential pathway enzymes. The final ratio of enzymes bound to SpyCatcher scaffolds is determined by the ratio of soluble SpyTag-modified enzymes mixed with the SpyCatcher scaffold and can only be controlled by a defined input enzyme stoichiometry which might be a disadvantage of this covalent SpySystem. The artificial multienzyme complex forms spontaneously upon mixing SpyTag- and SpyCatcher-modified enzymes and scaffolds. The SpyTag and SpyCatcher domains find each other, reconstitute and form a covalent isopeptide bond. To provide more flexibility so that the SpyCatcher- and SpyTag domains can fold properly and can undergo the isopeptide bonding, specific linker sequences, such as glycine-serine linkers, are inserted between the enzymes and the SpyTag or SpyCatcher domains. As the SpyTag sequence is very small, the fusion to enzymes should be no problem regarding correct folding, structure, and function of the enzyme. The covalent SpyCatcher–SpyTag system also can be used to covalently immobilize sequential pathway enzymes on carriers such as magnetic beads. By co-localizing pathway enzymes on the same magnetic bead, the metabolic efficiency should be increased due to the enforced proximity of the enzymes and the favored enzyme-to-enzyme substrate channeling. In addition to the metabolic/kinetic benefits, the stability of the enzymes can be enhanced thanks to the co-immobilization. It is also feasible to create multifunctional proteins by fusing SpyTag sequences to the N- and C-terminus of a selected enzyme. Mixing this double tagged enzyme with two enzymes carrying one SpyCatcher each, trimeric complexes are possible (Figure 6).

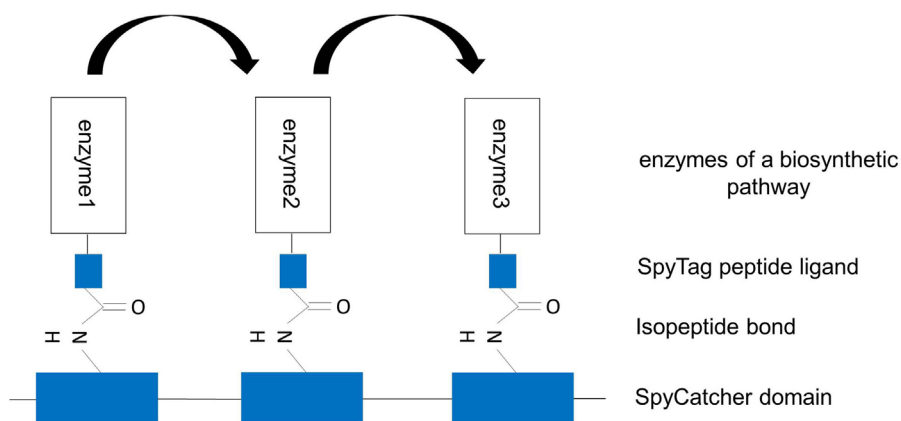


FIGURE 5 | Artificial multienzyme complex by covalently tethering pathway enzymes to the synthetic scaffold via the SpySystem. The protein-based scaffold is based on repeats of SpyCatcher domains. The enzymes to be co-localized are fused to SpyTag peptide ligands. By co-expressing the components, the enzymes were co-immobilized on the scaffold via the covalent, irreversible interaction between SpyCatcher and SpyTag (isopeptide bond). Metabolic channeling between the single enzymes (indicated by the arrows) should be facilitated.

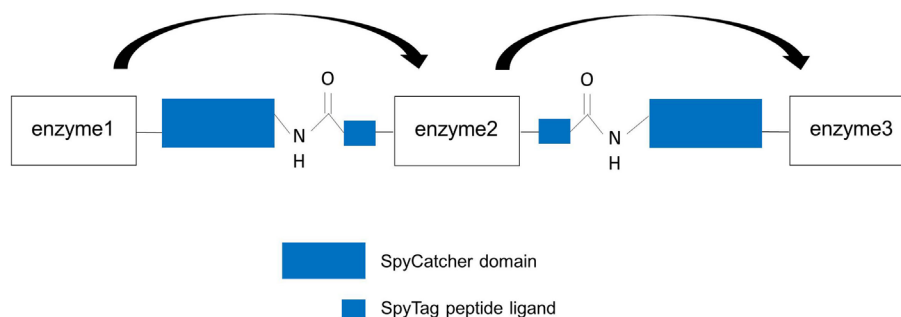


FIGURE 6 | Design of artificial trimeric proteins (functional trimer or even oligomers) by using the covalent SpySystem. By mixing SpyCatcher-, respectively, SpyTag-modified enzymes artificial, covalent enzyme pipes can be generated. Linker sequences between the enzymes and the Spy-domains should ensure correct protein folding and structure. Metabolic channeling between the single, covalently linked enzymes (indicated by the arrows) should be facilitated.

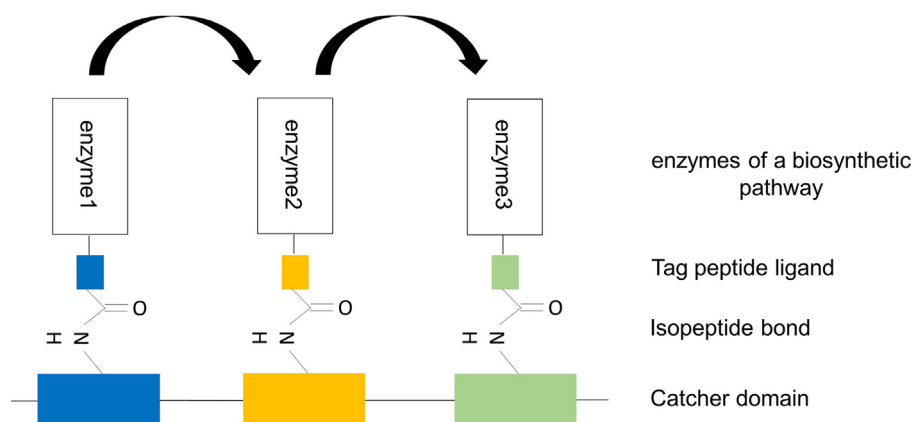
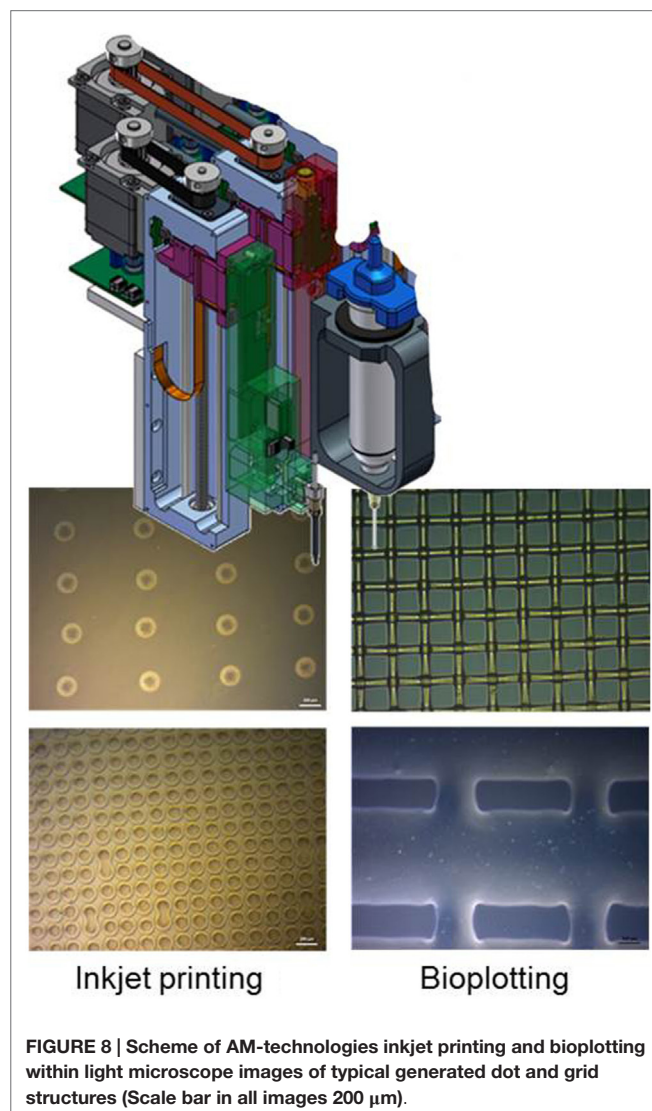


FIGURE 7 | Design of artificial trimeric proteins (functional trimer or even oligomers) by using different, covalent, alternative SpySystems. The scaffold consists of different unique Catcher domains and the enzymes to be co-localized are fused to the corresponding specific Tag domains. Thanks to the specific and covalent interaction between the Catcher domains and their Tag domains, the enzymes are irreversibly targeted and bound to the scaffold. The specificity of this system enables the precise control of the enzyme ratio and the exact position of the enzymes at the scaffold. Metabolic channeling (indicated by the arrows) should be possible because of the enforced proximity of the sequential pathway enzymes.

ARE THERE ADDITIONAL SYSTEMS ALLOWING CONTROLLED FORMATION OF INTERMOLECULAR LINKAGES IN PROTEINS?

To allow the design of precisely ordered enzyme arrays, alternative systems, in addition to the SpyTag-SpyCatcher system, for covalent tagging of proteins are required. Intermolecular isopeptide bonds are long known and form, for instance, between sumo or ubiquitin and lysine residues of target proteins. These modifications are linked to protein function, localization, and degradation. Intramolecular isopeptide bonds have only recently been discovered. They were first found in crystal structures of Spy0128 and have subsequently been detected in *Staphylococcus aureus* adhesin Can, *Enterococcus faecalis* adhesin Ace, *Streptococcus gordonii* antigen I/II adhesin SspB, and others (Kang and Baker, 2011). As intramolecular isopeptide bonds are frequently seen in extracellular proteins of Gram-positive bacteria, there is a

good chance to find alternative covalent SpyCatcher-SpyTag interactions. If this search is successful, one can imagine that the targeting of Tag-enzyme conjugates to their specific Catcher binding domains is more specific and can be precisely controlled (Figure 7). As an alternative to isopeptide bonding, it may also be feasible to utilize the autocatalytic formation of thr-gln ester bonds as has been reported for the *Clostridium perfringens* adhesion protein Cpe0147 (Kwon et al., 2014). In a recent study of Gao et al. (2015) the authors combined non-covalent, high-affinity protein-protein interaction between the PDZ domain and its peptide ligand with covalent disulfide bonding. To this end, the authors substituted one amino acid in the PDZ domain and ligand by cysteine. Fusing these modified interaction domains to target proteins, the authors could demonstrate that depending on the redox potential of the buffer disulfide bridges could be formed between both domains. This disulfide bonding allowed the formation of “disulfide-locked multienzyme supramolecular devices” (Gao et al., 2015).



3-D PRINTING – A NEW TOOL IN DESIGNING ENZYME ARRAYS

Three-dimensional printing is revolutionizing industry and holds great promise in biotechnology/biomedicine. Nowadays, it is possible to process a wide range of biological substances ranging from animal cells, plant cells, bacteria, proteins up to DNA sequences. Different strategies have been employed to transfer biological systems into micropattern, lines, channels, or other 3-D structures. These include photolithography, dip pen nanolithography, microcontact printing, replica molding, and AM eventually coupled with self-assembly techniques (Herzer et al., 2010). For example, photolithographic patterning of proteins on surfaces has been used extensively in the past to analyze cell behavior with micrometer-scale resolution (Kane et al., 1999). However, by photolithographic patterning it is hard depositing several different proteins on the same surface. Furthermore, in order to print microarrays in high-resolution structures down to approx. 100 nm, microcontact printing has been established as the

method of choice. A 3-D shaped stamp transfers its surface layout onto a planar target structure (surface patterning). 3-D printing, as AM technique, adds the desired material in its final shape without subtractive removal. Historically, AM started first at an industrial level in the early 1990s as an alternative to traditional model-making techniques for rapid prototyping (sometimes referred to as 3-D printing). AM techniques are now defined by the American Society for Testing and Materials (ASTM) as processes of joining materials to make objects from 3-D digital data, usually layer upon layer, as opposed to subtractive manufacturing methods such as computer numerical control (CNC) milling (ASTM F2792-10 Standard Terminology for Additive Manufacturing Technologies, ASTM Intern., USA). Engineering of metabolic pathways is a highly complex and demanding approach, including large scale and high-throughput biology, manufacturing techniques with high resolution and enormous flexibility. Inkjet printing and bioplotting could be suitable AM techniques for a successful application in engineering of metabolic pathways.

The inkjet technique is a non-contact digital printing method, creating a functional pattern by delivering ink droplets to the substrate surface. One type is the continuously injecting one, which shoots a permanent ink stream. The second one is the drop on demand (DOD) technique, which expulse only a single drop on a signal. The two main technics used are piezoelectric and electro-thermal, which are fast, cheap, digital controlled and simple methods for structured patterning of biomaterials, cells, and protein molecules (Calvert, 2001). Electro-thermal is the easiest and most used DOD technique, while piezoelectric is a milder printing technique. Inkjet printing as suitable AM method has been applied for the creation of a microenvironment for cells, proteins, and oligonucleotides by printing water-based matrices with spatially defined patterns (Barbulovic-Nad et al., 2006). Generally, inkjet systems enable the fast printing of small droplets or patterns with volumes in the range of 50–250 pL with a high resolution (50–100 μm) as illustrated in **Figure 8**. The bioplotting process is significantly slower and possesses a lower resolution compared to inkjet printing (also **Figure 8**). However, larger volumes of up to 10 nL can be processed so that the process is especially well suited for the handling of immobilized cells (Zehnder et al., 2015). With the aid of the 3-D bioplotter, larger and more complex scaffold geometries without the need for support structures can be realized, in contrast to the inkjet printer. The advantage of this method is the possibility to print cells, growth factors, and enzymes at the same time into precisely defined 3-D structure. By using a multi-nozzle bioplotter, it is possible to create scaffold structures consisting of several different materials and metabolic enzymes (Detsch et al., 2014).

CONCLUSION

Engineering of metabolic pathways holds great promise to pave the way for green and sustainable chemistry. Recent success in Systems and Synthetic Biology opened a large number of possibilities to design more and more complex metabolic pathways *in vitro* and *in vivo*, not foreseen until recently. Thanks to structural analysis and intensive research in protein–protein interactions, it

becomes feasible to organize artificial pathways in ordered and regulated arrays. Synthetic scaffolds offer a modular and highly flexible tool for rationally, post-translationally organizing/co-localizing multiple functionally related enzymes in a defined and controllable manner because the protein binding domain/peptide ligand interactions as well as the DNA–DNA/RNA–RNA, respectively, the DNA–Protein/RNA–Protein interactions are specific, adaptable, engineerable and can be theoretically applied to all metabolic pathways only depending on the developed ligand–enzyme conjugates or DNA/RNA–enzyme conjugates. So the main advantage of the scaffold-based multienzyme complex formation strategy is the modularity. The scaffold architecture and therefore the enzyme stoichiometry and enzyme ratio at the synthetic complex can be easily controlled just by varying single, modular biological parts. Through adding, removing, or replacing distinct domains, the scaffold structure is changed which can lead to a completely new behavior of the immobilized components and thereby of the whole pathway. Strategies for reversible interactions between proteins are complemented by covalent bonding enabling the stable construction of large

multienzyme complexes. In addition to advances in the field of protein chemistry, technical developments such as 3-D printing will enable the design of industrial scale metabolic channels in the near future.

AUTHOR CONTRIBUTIONS

All authors contributed to writing and providing the literature, which has been cited in this review article.

ACKNOWLEDGMENTS

We are grateful to the comments of the reviewers, which improved the quality of our manuscript.

FUNDING

We acknowledge support by the emerging field initiative “Synthetic Biology” of the Friedrich-Alexander University of Erlangen-Nuremberg.

REFERENCES

- Al-Habori, M. (1995). Microcompartmentation, metabolic channelling and carbohydrate metabolism. *Int. J. Biochem. Cell Biol.* 27, 123–132. doi:10.1016/1357-2725(94)00079-Q
- Bachmann, B. J. (1983). Linkage map of *Escherichia coli* K-12, edition 7. *Microbiol. Rev.* 47, 180–230.
- Barbulovic-Nad, I., Lucente, M., Sun, Y., Zhang, M., Wheeler, A. R., and Bussmann, M. (2006). Bio-microarray fabrication techniques – a review. *Crit. Rev. Biotechnol.* 26, 237–259. doi:10.1080/07388550600978358
- Barnes, S. J., and Weitzman, P. D. J. (1986). Organization of citric acid cycle enzymes into a multienzyme cluster. *FEBS Lett.* 201, 267–270. doi:10.1016/0014-5793(86)80621-4
- Bayer, E. A., Belaich, J.-P., Shoham, Y., and Lamed, R. (2004). The celulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* 58, 521–554. doi:10.1146/annurev.micro.57.030502.091022
- Boyle, P. M., and Silver, P. A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metab. Eng.* 14, 223–232. doi:10.1016/j.ymben.2011.10.003
- Calvert, P. (2001). Inkjet printing for materials and devices. *Chem. Mater.* 13, 3299–3305. doi:10.1021/cm0101632
- Capell, T., and Christou, P. (2004). Progress in plant metabolic engineering. *Curr. Opin. Biotechnol.* 15, 148–154. doi:10.1016/j.copbio.2004.01.009
- Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P., Cristea, I. M., Rabinowitz, J. D., et al. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* 32, 1011–1018. doi:10.1038/nbt.3018
- Chen, A. H., and Silver, P. A. (2012). Designing biological compartmentalization. *Trends Cell Biol.* 22, 662–670. doi:10.1016/j.tcb.2012.07.002
- Chen, R., Chen, Q., Kim, H., Siu, K.-H., Sun, Q., Tsai, S.-L., et al. (2014). Biomolecular scaffolds for enhanced signaling and catalytic efficiency. *Curr. Opin. Biotechnol.* 28, 59–68. doi:10.1016/j.copbio.2013.11.007
- Choudhary, S., Quin, M. B., Sanders, M. A., Johnson, E. T., and Schmidt-Dannert, C. (2012). Engineered protein nano-compartments for targeted enzyme localization. *PLoS ONE* 7:e33342. doi:10.1371/journal.pone.0033342
- Conrado, R. J., Varner, J. D., and DeLisa, M. P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* 19, 492–499. doi:10.1016/j.copbio.2008.07.006
- Conrado, R. J., Wu, G. C., Boock, J. T., Xu, H., Chen, S. Y., Lebar, T., et al. (2012). DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* 40, 1879–1889. doi:10.1093/nar/gkr888
- 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
- DeLisa, M. P., and Conrado, R. J. (2009). Synthetic metabolic pipelines. *Nat. Biotechnol.* 27, 728–729. doi:10.1038/nbt0809-728
- Detsch, R., Sarker, B., Zehnder, T., Douglas, T. E. L., and Boccaccini, A. R. (2014). Additive manufacturing of cell-loaded alginate enriched with alkaline phosphatase for bone tissue engineering application. *BioNanoMaterials* 15, 79–87. doi:10.1515/bnm-2014-0007
- DiCosimo, R., McAuliffe, J., Poulou, A. J., and Bohlmann, G. (2013). Industrial use of immobilized enzymes. *Chem. Soc. Rev.* 42, 6437–6474. doi:10.1039/c3cs35506c
- 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
- Gao, X., Zhao, C., Yu, T., Yang, S., Ren, Y., and Wei, D. (2015). Construction of a reusable multi-enzyme supramolecular device via disulfide bond locking. *Chem. Commun.* 51, 10131. doi:10.1039/c5cc02544c
- Garcia-Galan, C., Berenguer-Murcia, Á., Fernandez-Lafuente, R., and Rodrigues, R. C. (2011). Potential of different enzyme immobilization strategies to improve enzyme performance. *Adv. Synth. Catal.* 353, 2885–2904. doi:10.1002/adsc.201100534
- Gefen, G., Anbar, M., Morag, E., Lamed, R., and Bayer, E. A. (2012). Enhanced cellulose degradation by targeted integration of a cohesin-fused β -glucosidase into the *Clostridium thermocellum* cellulosome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10298–10303. doi:10.1073/pnas.1202747109
- Giegé, P., Heazlewood, J. L., Roessner-Tunali, U., Millar, A. H., Fernie, A. R., Leaver, C. J., et al. (2003). Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells. *Plant Cell* 15, 2140–2151. doi:10.1105/tpc.012500
- Good, M. C. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332, 680. doi:10.1126/science.1198701
- Guzik, U., Hupert-Kocurek, K., and Wojcieszynska, D. (2014). Immobilization as a strategy for improving enzyme properties-application to oxidoreductases. *Molecules* 19, 8995–9018. doi:10.3390/molecules19078995
- Herzer, N., Hoeppener, S., and Schubert, U. S. (2010). Fabrication of patterned silane based self-assembled monolayers by photolithography and surface reactions on silicon-oxide substrates. *Chem. Commun. (Camb.)* 46, 5634–5652. doi:10.1039/c0cc00674b
- Homaei, A. A., Sariri, R., Vianello, F., and Stevanato, R. (2013). Enzyme immobilization: an update. *J. Chem. Biol.* 6, 185–205. doi:10.1007/s12154-013-0102-9

- Jia, F., Narasimhan, B., and Mallapragada, S. (2014). Materials-based strategies for multi-enzyme immobilization and co-localization: a review. *Biotechnol. Bioeng.* 111, 209–222. doi:10.1002/bit.25136
- Jia, F., Narasimhan, B., and Mallapragada, S. K. (2013). Biomimetic multienzyme complexes based on nanoscale platforms. *AIChE J.* 59, 355–360. doi:10.1002/aic.13992
- Jones, J. A., Toparlak, ÖD., and Koffas, M. A. G. (2015). Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Curr. Opin. Biotechnol.* 33, 52–59. doi:10.1016/j.copbio.2014.11.013
- Jørgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., et al. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291. doi:10.1016/j.pbi.2005.03.014
- Kane, R. S., Takayama, S., Ostuni, E., Ingber, D. E., and Whitesides, G. M. (1999). Patterning proteins and cells using soft lithography. *Biomaterials* 20, 2363–2376. doi:10.1016/S0142-9612(99)00165-9
- Kang, H. J., and Baker, E. N. (2011). Intramolecular isopeptide bonds: protein crosslinks built for stress? *Trends Biochem. Sci.* 36, 229–237. doi:10.1016/j.tibs.2010.09.007
- Kang, H. J., Paterson, N. G., Kim, C. U., Middleditch, M., Chang, C., Ton-That, H., et al. (2014). A slow-forming isopeptide bond in the structure of the major pilin SpaD from *Corynebacterium diphtheriae* has implications for pilus assembly. *Acta Crystallogr. D Biol. Crystallogr.* 70, 1190–1201. doi:10.1107/S1399004714001400
- Kwon, H., Squire, C. J., Young, P. G., and Baker, E. N. (2014). Autocatalytically generated Thr-Gln ester bond cross-links stabilize the repetitive Ig-domain shaft of a bacterial cell surface adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1367–1372. doi:10.1073/pnas.1316855111
- Laursen, T., Möller, B. L., and Bassard, J.-E. (2015). Plasticity of specialized metabolism as mediated by dynamic metabolons. *Trends Plant Sci.* 20, 20–32. doi:10.1016/j.tplants.2014.11.002
- Lee, H., DeLoache, W. C., and Dueber, J. E. (2012). Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14, 242–251. doi:10.1016/j.ymben.2011.09.003
- Long, M. S., Jones, C. D., Helfrich, M. R., Mangeney-Slavin, L. K., and Keating, C. D. (2005). Dynamic microcompartmentation in synthetic cells. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5920–5925. doi:10.1073/pnas.0409333102
- Lunn, J. E. (2007). Compartmentation in plant metabolism. *J. Exp. Bot.* 58, 35–47. doi:10.1093/jxb/erl134
- Lytle, B., and Wu, J. H. D. (1998). Involvement of both dockerin subdomains in assembly of the *Clostridium thermocellum* cellulosome. *J. Bacteriol.* 180, 6581–6585.
- Mattiasson, B., and Mosbach, K. (1971). Studies on a matrix-bound three-enzyme system. *Biochim. Biophys. Acta* 14, 253–257. doi:10.1016/0005-2744(71)90054-4
- Mazzoli, R., Lamberti, C., and Pessione, E. (2012). Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. *Trends Biotechnol.* 30, 111–119. doi:10.1016/j.tibtech.2011.08.003
- Möller, B. L. (2010). Dynamic metabolons. *Science* 330, 1328. doi:10.1126/science.1194971
- Moon, T. S., Dueber, J. E., Shiue, E., and Prather, K. L. J. (2010). Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* 12, 298–305. doi:10.1016/j.ymben.2010.01.003
- Müller, J., and Niemeyer, C. M. (2008). DNA-directed assembly of artificial multienzyme complexes. *Biochem. Biophys. Res. Commun.* 377, 62–67. doi:10.1016/j.bbrc.2008.09.078
- Na, D., Kim, T. Y., and Lee, S. Y. (2010). Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* 13, 363–370. doi:10.1016/j.mib.2010.02.004
- Niemeyer, C. M., Koehler, J., and Wuerdemann, C. (2002). DNA-directed assembly of bienzymic complexes from in vivo biotinylated NAD(P) H:FMN oxidoreductase and luciferase. *ChemBiochem* 3, 242–245. doi:10.1002/1439-7633(20020301)3:2/3<242::AID-CBIC242>3.0.CO;2-F
- Otte, K. B., and Hauer, B. (2015). Enzyme engineering in the context of novel pathways and products. *Curr. Opin. Biotechnol.* 35, 16–22. doi:10.1016/j.copbio.2014.12.011
- Patterson, D. P., Schwarz, B., Waters, R. S., Gedeon, T., and Douglas, T. (2014). Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle. *ACS Chem. Biol.* 9, 359–365. doi:10.1021/cb4006529
- Pinheiro, B. A., Gilbert, H. J., Sakka, K., Sakka, K., Fernandes, V. A., Prates, J. A. M., et al. (2009). Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem. J.* 424, 375–384. doi:10.1042/BJ20091152
- Pointon, J. A., Smith, W. D., Saalbach, G., Crow, A., Kehoe, M. A., and Banfield, M. J. (2010). A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J. Biol. Chem.* 285, 33858–33866. doi:10.1074/jbc.M110.149385
- Roy, F., Laberge, G., Douziech, M., Ferland-McCollough, D., and Therrien, M. (2002). KSR is a scaffold required for activation of the ERK/MAPK module. *Genes Dev.* 16, 427–438. doi:10.1101/gad.962902
- Schoene, C., Fierer, J. O., Bennett, S. P., and Howarth, M. (2014). SpyTag/SpyCatcher cyclization confers resilience to boiling on a mesophilic enzyme. *Angew. Chem. Int. Ed. Engl.* 53, 6101–6104. doi:10.1002/anie.201402519
- Sheldon, R. A. (2007). Enzyme immobilization: the quest for optimum performance. *Adv. Synth. Catal.* 349, 1289–1307. doi:10.1002/adsc.200700082
- Smolle, M., and Lindsay, J. G. (2006). Molecular architecture of the pyruvate dehydrogenase complex: bridging the gap. *Biochem. Soc. Trans.* 34(Pt 5), 815–818. doi:10.1042/BST0340815
- Sonnenwald, U. (2003). Plant biotechnology: from basic science to industrial applications. *J. Plant Physiol.* 160, 723–725. doi:10.1078/0176-1617-01172
- Sweetlove, L. J., and Fernie, A. R. (2013). The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* 64, 723–746. doi:10.1146/annurev-arplant-050312-120233
- Veggiani, G., Zakeri, B., and Howarth, M. (2014). Superglue from bacteria: unbreakable bridges for protein nanotechnology. *Trends Biotechnol.* 32, 506–512. doi:10.1016/j.tibtech.2014.08.001
- Walden, M., Edwards, J. M., Dziejewska, A. M., Bergmann, R., Saalbach, G., Kan, S.-Y., et al. (2015). An internal thioester in a pathogen surface protein mediates covalent host binding. *eLIFE* 4, e06638. doi:10.7554/eLife.06638
- Wang, Y., and Yu, O. (2012). Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. *J. Biotechnol.* 157, 258–260. doi:10.1016/j.jbiotec.2011.11.003
- Winkel, B. S. J. (2004). Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 55, 85–107. doi:10.1146/annurev.arplant.55.031903.141714
- Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., et al. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 109, E690–E697. doi:10.1073/pnas.1115485109
- Zakeri, B., and Howarth, M. (2010). Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. *J. Am. Chem. Soc.* 132, 4526–4527. doi:10.1021/ja910795a
- Zehnder, T., Sarker, B., Boccaccini, A. R., and Detsch, R. (2015). Evaluation of an alginate–gelatine crosslinked hydrogel for bioplotting. *Biofabrication* 7, 1–12. doi:10.1088/1758-5090/7/2/025001
- Zhao, Z., Ma, X., Li, L., Zhang, W., Ping, S., Xu, M. Q., et al. (2010). Protein cyclization enhanced thermostability and exopeptidase-resistance of green fluorescent protein. *J. Microbiol. Biotechnol.* 20, 460–466. doi:10.4014/jmb.0908.08013

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 Pröschel, Detsch, Boccaccini and Sonnenwald. 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



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

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



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

