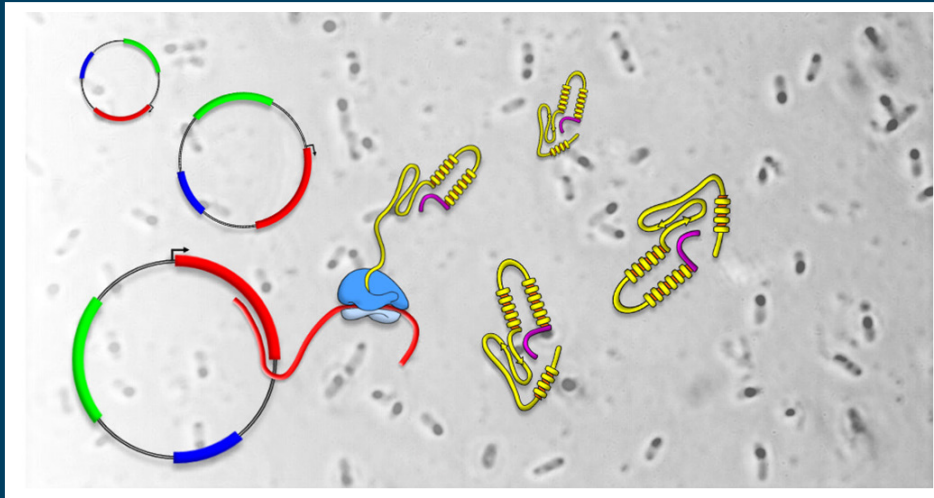


frontiers

RESEARCH TOPICS



RECOMBINANT PROTEIN EXPRESSION IN MICROBIAL SYSTEMS

Topic Editors

Germán L. Rosano and Eduardo A. Ceccarelli



frontiers in
MICROBIOLOGY



frontiers

FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2014
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 978-2-88919-294-6

DOI 10.3389/978-2-88919-294-6

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!

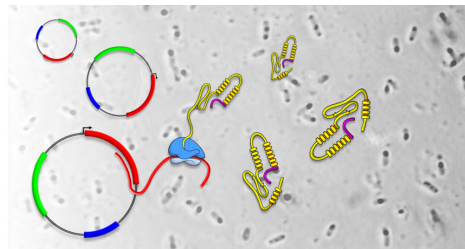
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

RECOMBINANT PROTEIN EXPRESSION IN MICROBIAL SYSTEMS

Topic Editors:

Germán L. Rosano, Universidad Nacional de Rosario, Argentina

Eduardo A. Ceccarelli, Universidad Nacional de Rosario, Argentina



The synthesis of an N-terminally tagged recombinant protein from expression vectors. In the background, *Escherichia coli* cells producing a recombinant protein as seen under a microscope. In this picture, inclusion bodies were detected due to their birefringence under polarized light (dark spots).

With the advent of recombinant DNA technology, expressing heterologous proteins in microorganisms rapidly became the method of choice for their production at laboratory and industrial scale. Bacteria, yeasts and other hosts can be grown to high biomass levels efficiently and inexpensively. Obtaining high yields of recombinant proteins from this material was only feasible thanks to constant research on microbial genetics and physiology that led to novel strains, plasmids and cultivation strategies.

Despite the spectacular expansion of the field, there is still much room for progress. Improving the levels of expression and the

solubility of a recombinant protein can be quite challenging. Accumulation of the product in the cell can lead to stress responses which affect cell growth. Buildup of insoluble and biologically inactive aggregates (inclusion bodies) lowers the yield of production. This is particularly true for obtaining membrane proteins or high-molecular weight and multi-domain proteins. Also, obtaining eukaryotic proteins in a prokaryotic background (for example, plant or animal proteins in bacteria) results in a product that lack post-translational modifications, often required for functionality. Changing to a eukaryotic host (yeasts or filamentous fungi) may not be a proper solution since the pattern of sugar modifications is different than in higher eukaryotes.

Still, many advances in the last couple of decades have provided to researchers a wide variety of strategies to maximize the production of their recombinant protein of choice. Everything starts with the careful selection of the host. Be it bacteria or yeast, a broad list of strains is available for overcoming codon use bias, incorrect disulfide bond formation, protein toxicity and lack of post-translational modifications. Also, a huge catalog of plasmids allows choosing for different fusion partners for improving solubility, protein secretion, chaperone co-expression, antibiotic resistance and promoter strength. Next, controlling culture

conditions like temperature, inducer and media composition can bolster recombinant protein production.

With this Research Topic, we aim to provide an encyclopedic account of the existing approaches to the expression of recombinant proteins in microorganisms, highlight recent discoveries and analyze the future prospects of this exciting and ever-growing field.

Table of Contents

- 05 *Recombinant Protein Expression in Microbial Systems***
Germán L. Rosano and Eduardo A. Ceccarelli
- 07 *Recombinant Protein Expression in Escherichia Coli: Advances and Challenges***
Germán Leandro Rosano and Eduardo A. Ceccarelli
- 24 *Fusion Tags for Protein Solubility, Purification and Immunogenicity in Escherichia Coli: The Novel Fh8 System***
Sofia Costa, André Almeida, António Castro and Lucília Domingues
- 44 *Generation of a Vector Suite for Protein Solubility Screening***
Agustin Correa, Claudia Ortega, Gonzalo Obal, Pedro Alzari, Renaud Vincentelli and Pablo Oppezzo
- 53 *Inclusion Bodies: Not that Bad...***
Ana Ramón, Mario Señorale-Pose and Mónica Marín
- 59 *Expression of Codon Optimized Genes in Microbial Systems: Current Industrial Applications and Perspectives***
Claudia Elena, Pablo Ravasi, María E. Castelli, Salvador Peirú and Hugo G. Menzella
- 67 *Flavoprotein Monooxygenases for Oxidative Biocatalysis: Recombinant Expression in Microbial Hosts and Applications***
Romina D. Ceccoli, Dario A. Bianchi and Daniela V. Rial
- 81 *Playing Catch-Up With Escherichia Coli: Using Yeast to Increase Success Rates in Recombinant Protein Production Experiments***
Roslyn M. Bill
- 86 *Making Recombinant Proteins in Filamentous Fungi- Are We Expecting Too Much?***
Helena Nevalainen and Robyn Peterson
- 96 *Algae-Based Oral Recombinant Vaccines***
Elizabeth A. Specht and Stephen P. Mayfield



Recombinant protein expression in microbial systems

Germán L. Rosano* and Eduardo A. Ceccarelli

Instituto de Biología Molecular y Celular de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

*Correspondence: rosano@ibr-conicet.gov.ar

Edited and reviewed by:

William James Hickey, University of Wisconsin-Madison, USA

Keywords: recombinant proteins, microorganism, inclusion bodies, fusion tags, *Escherichia coli*, yeast, filamentous fungi, microalgae

INTRODUCTION

The emergence of recombinant DNA technology during the early 70's set a revolution in molecular biology. This set of techniques was strengthened even further later on with the introduction of the polymerase chain reaction and allowed scientists to explore and understand essential life processes in an easy and straightforward way. It also marked the birth of the modern biotech industry. At that time, it was shown that eukaryotic DNA could be propagated in *Escherichia coli* (Morrow et al., 1974) and functional products could be synthesized from heterologous genes cloned in bacterial plasmids (Ratzkin and Carbon, 1977; Vapnek et al., 1977). After these successful cases, it was soon realized that the potential applications of these techniques were almost limitless. In fact, US patent 4,237,224 granted to Cohen and Boyer (1980) claimed to commercial ownership of the methodology for cloning virtually all possible DNAs in all possible vectors. While cloning any gene in any given vector is feasible, obtaining a functional product from its expression is not that simple.

In this series of articles, the authors describe the methods and technologies available for producing recombinant proteins in different microbes. They also introduce and discuss recent advances that attempt to tackle common pitfalls in the process. Taken together, this E-book will be of great importance for those entering the field as well as for experienced researchers that are looking for an update in the state of the art.

Before proceeding any further, it is necessary to clarify an important aspect of this topic. In biology, the universal accepted definition of “expression” is “production of an observable phenotype by a gene—usually by directing the synthesis of a protein” (Alberts et al., 2002). By this definition, the term “gene expression” is correct while “protein expression” is basically lab jargon. We do think that correct usage of scientific language is of great importance, yet in this particular case, the usage of “protein expression” in the scientific community is so pervasive that readers will immediately understand what we are talking about. So, considering that “protein expression” found its way into journal names, book names and high-impact reviews (Sørensen and Mortensen, 2005) and research papers (Ghaemmaghami et al., 2003) (>1800 citations in Scopus) we and other authors have used it interchangeably with more correct terms like protein production or protein synthesis.

CURRENT STATUS IN RECOMBINANT PROTEIN EXPRESSION IN MICROBIAL SYSTEMS

Without a doubt, *E. coli* is the most widely used host for heterologous gene expression. It has been used for this purpose for more than 40 years, so there is much accumulated knowledge about its advantages and disadvantages as an expression platform. Rosano and Ceccarelli review the tools at hand (expression vectors, strains, media composition, etc.) when using *E. coli* as a host (Rosano and Ceccarelli, 2014). Different approaches for solving common problems, such as inclusion body (IB) formation or low yield, are also presented.

Other authors delve a little deeper into these issues. Costa and coauthors give a thorough description of different fusion tags that can be appended to the target protein in order to increase its solubility and/or ease its purification from the cellular milieu (Costa et al., 2014). Along this line, Correa and coauthors present a very promising approach to straightforwardly assess the solubility of a recombinant protein by cloning the corresponding gene in 12 different expression vectors in parallel (Correa et al., 2014). Even though IB formation is mainly regarded as a nuance in the production of recombinant proteins, Ramon and coworkers make the case that this is not always true and focus on the positive side of IBs, highlighting the advantages of producing recombinant proteins as IBs for basic and applied research (Ramon et al., 2014). No or low yield can be the result of codon bias and Elena and coauthors look into one of the strategies used for solving this problem, the expression of codon optimized genes (Elena et al., 2014). Of great importance is their account of the application of this technology in the industrial setting. Not always the desired product is a protein, sometimes metabolites and fine chemicals are the goal. The review of Ceccoli and coauthors details with numerous examples how *E. coli* and other microorganisms can be turned into biocatalysts by strain engineering (Ceccoli et al., 2014). High-value products can thus be obtained from pure recombinant enzyme or from whole-cell systems, turning the host into an exquisitely designed and environmentally friendly chemical factory.

ADVANCES IN NEW TECHNOLOGIES

Microbes other than *E. coli* can be used for heterologous protein production. The impact of yeasts on the biotech industry is paramount, as 20% of biopharmaceutical proteins are synthesized

in yeasts. But still, they are not the first-choice microorganism for recombinant protein production. In her Perspective article, Bill argues that the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* should be considered alongside *E. coli* in any project in need of a recombinant protein (Bill, 2014). Filamentous fungi are excellent protein secretors, a property that makes them ideal as a host at an industrial scale. Nevalainen and Peterson present an authoritative review describing the body of research carried out into cellular mechanisms of fungi that will ultimately lead to the optimization of the system (Nevalainen and Peterson, 2014). Specht and Mayfield explain the usefulness of microalgae as expression systems, focusing on the production of recombinant vaccines (Specht and Mayfield, 2014). The quest for edible vaccines is an active area of research and microalgae will definitely play a major role as they possess greater advantages over plants (which can also be used for this purpose) in terms of cost, safety, and logistics.

AFTERWORD

After getting into all the articles in this E-book, it should be clear to the reader that progress in the field of recombinant protein expression in microbial systems shows no sign of deceleration. Their use and importance in research and in industry cannot be disputed. Their establishment in the bio(techno)logy toolkit was due to the ongoing efforts of researchers that continuously optimize well-known systems such as *E. coli* and those who break new ground in the use of alternative microorganisms. We want to take this opportunity to thank all the experts of this series for their excellent contributions and the reviewers for their insightful comments and remarks.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*. New York, NY: Garland Science.
- Bill, R. M. (2014). Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.* 5:85. doi: 10.3389/fmicb.2014.00085
- Ceccoli, R. D., Bianchi, D. A., and Rial, D. V. (2014). Flavoprotein monooxygenases for oxidative biocatalysis: recombinant expression in microbial hosts and applications. *Front. Microbiol.* 5:25. doi: 10.3389/fmicb.2014.00025
- Cohen, S. N., and Boyer, H. W. (1980). *Process for Producing Biologically Functional Molecular Chimeras*. US Patent 4,237,224, filed 4 January, 1979, and issued 2 December, 1980.
- Correa, A., Ortega, C., Obal, G., Alzari, P., Vincentelli, R., and Oppizzo, P. (2014). Generation of a vector suite for protein solubility screening. *Front. Microbiol.* 5:67. doi: 10.3389/fmicb.2014.00067
- Costa, S., Almeida, A., Castro, A., and Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in: the novel Fh8 system. *Front. Microbiol.* 5:63. doi: 10.3389/fmicb.2014.00063
- Elena, C., Ravasi, P., Castelli, M. E., Peiru, S., and Menzella, H. G. (2014). Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Front. Microbiol.* 5:21. doi: 10.3389/fmicb.2014.00021
- Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., et al. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741. doi: 10.1038/nature02046
- Morrow, J. E., Cohen, S. N., Chang, A. C., Boyer, H. W., Goodman, H. M., and Helling, R. B. (1974). Replication and transcription of eukaryotic DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 71, 1743–1747. doi: 10.1073/pnas.71.5.1743
- Nevalainen, H., and Peterson, R. (2014). Making recombinant proteins in filamentous fungi- are we expecting too much? *Front. Microbiol.* 5:75. doi: 10.3389/fmicb.2014.00075
- Ramon, A., Senorale-Pose, M., and Marin, M. (2014). Inclusion bodies: not that bad. *Front. Microbiol.* 5:56. doi: 10.3389/fmicb.2014.00056
- Ratzkin, B., and Carbon, J. (1977). Functional expression of cloned yeast DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 74, 487–491. doi: 10.1073/pnas.74.2.487
- Rosano, G. L., and Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5:172. doi: 10.3389/fmicb.2014.00172
- Sørensen, H. P., and Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115, 113–128. doi: 10.1016/j.jbiotec.2004.08.004
- Specht, E. A., and Mayfield, S. P. (2014). Algae-based oral recombinant vaccines. *Front. Microbiol.* 5:60. doi: 10.3389/fmicb.2014.00060
- Vapnek, D., Haultala, J. A., Jacobson, J. W., Giles, N. H., and Kushner, S. R. (1977). Expression in *Escherichia coli* K-12 of the structural gene for catabolic dehydroquinase of *Neurospora crassa*. *Proc. Natl. Acad. Sci. U.S.A.* 74, 3508–3512. doi: 10.1073/pnas.74.8.3508

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

Received: 22 May 2014; accepted: 19 June 2014; published online: 08 July 2014.

Citation: Rosano GL and Ceccarelli EA (2014) Recombinant protein expression in microbial systems. *Front. Microbiol.* 5:341. doi: 10.3389/fmicb.2014.00341

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Rosano and Ceccarelli. 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.



Recombinant protein expression in *Escherichia coli*: advances and challenges

Germán L. Rosano^{1,2} * and Eduardo A. Ceccarelli^{1,2}

¹ Instituto de Biología Molecular y Celular de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, Rosario, Argentina

² Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

Edited by:

Peter Neubauer, Technische Universität Berlin, Germany

Reviewed by:

Jose M. Bruno-Barcena, North Carolina State University, USA

Thomas Schweder, Ernst-Moritz-Arndt-Universität Greifswald, Germany

*Correspondence:

Germán L. Rosano, Instituto de Biología Molecular y Celular de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, Esmeralda y Ocampo, Rosario 2000, Argentina
e-mail: rosano@ibr-conicet.gov.ar

Escherichia coli is one of the organisms of choice for the production of recombinant proteins. Its use as a cell factory is well-established and it has become the most popular expression platform. For this reason, there are many molecular tools and protocols at hand for the high-level production of heterologous proteins, such as a vast catalog of expression plasmids, a great number of engineered strains and many cultivation strategies. We review the different approaches for the synthesis of recombinant proteins in *E. coli* and discuss recent progress in this ever-growing field.

Keywords: recombinant protein expression, *Escherichia coli*, expression plasmid, inclusion bodies, affinity tags, *E. coli* expression strains

INTRODUCTION

There is no doubt that the production of recombinant proteins in microbial systems has revolutionized biochemistry. The days where kilograms of animal and plant tissues or large volumes of biological fluids were needed for the purification of small amounts of a given protein are almost gone. Every researcher that embarks on a new project that will need a purified protein immediately thinks of how to obtain it in a recombinant form. The ability to express and purify the desired recombinant protein in a large quantity allows for its biochemical characterization, its use in industrial processes and the development of commercial goods.

At the theoretical level, the steps needed for obtaining a recombinant protein are pretty straightforward. You take your gene of interest, clone it in whatever expression vector you have at your disposal, transform it into the host of choice, induce and then, the protein is ready for purification and characterization. In practice, however, dozens of things can go wrong. Poor growth of the host, inclusion body (IB) formation, protein inactivity, and even not obtaining any protein at all are some of the problems often found down the pipeline.

In the past, many reviews have covered this topic with great detail (Makrides, 1996; Baneyx, 1999; Stevens, 2000; Jana and Deb, 2005; Sorensen and Mortensen, 2005). Collectively, these papers gather more than 2000 citations. Yet, in the field of recombinant protein expression and purification, progress is continuously being made. For this reason, in this review, we comment on the most recent advances in the topic. But also, for those with modest experience in the production of heterologous proteins, we describe the many options and approaches that have been successful for expressing a great number of proteins over the last couple of decades, by answering the questions needed to be

addressed at the beginning of the project. Finally, we provide a troubleshooting guide that will come in handy when dealing with difficult-to-express proteins.

FIRST QUESTION: WHICH ORGANISM TO USE?

The choice of the host cell whose protein synthesis machinery will produce the precious protein will initiate the outline of the whole process. It defines the technology needed for the project, be it a variety of molecular tools, equipment, or reagents. Among microorganisms, host systems that are available include bacteria, yeast, filamentous fungi, and unicellular algae. All have strengths and weaknesses and their choice may be subject to the protein of interest (Demain and Vaishnav, 2009; Adrio and Demain, 2010). For example, if eukaryotic post-translational modifications (like protein glycosylation) are needed, a prokaryotic expression system may not be suitable (Sahdev et al., 2008). In this review, we will focus specifically on *Escherichia coli*. Other systems are described in excellent detail in accompanying articles of this series.

The advantages of using *E. coli* as the host organism are well known. (i) It has unparalleled fast growth kinetics. In glucose-salts media and given the optimal environmental conditions, its doubling time is about 20 min (Sezonov et al., 2007). This means that a culture inoculated with a 1/100 dilution of a saturated starter culture may reach stationary phase in a few hours. However, it should be noted that the expression of a recombinant protein may impart a metabolic burden on the microorganism, causing a considerable decrease in generation time (Bentley et al., 1990). (ii) High cell density cultures are easily achieved. The theoretical density limit of an *E. coli* liquid culture is estimated to be about 200 g dry cell weight/l or roughly 1×10^{13} viable bacteria/ml (Lee, 1996; Shiloach and Fass, 2005). However, exponential growth in

complex media leads to densities nowhere near that number. In the simplest laboratory setup (i.e., batch cultivation of *E. coli* at 37°C, using LB media), $<1 \times 10^{10}$ cells/ml may be the upper limit (Sezonov et al., 2007), which is less than 0.1% of the theoretical limit. For this reason, high cell-density culture methods were designed to boost *E. coli* growth, even when producing a recombinant protein (Choi et al., 2006). Being a workhorse organism, these strategies arose thanks to the wealth of knowledge about its physiology. (iii) Rich complex media can be made from readily available and inexpensive components. (iv) Transformation with exogenous DNA is fast and easy. Plasmid transformation of *E. coli* can be performed in as little as 5 min (Pope and Kent, 1996).

SECOND QUESTION: WHICH PLASMID SHOULD BE CHOSEN?

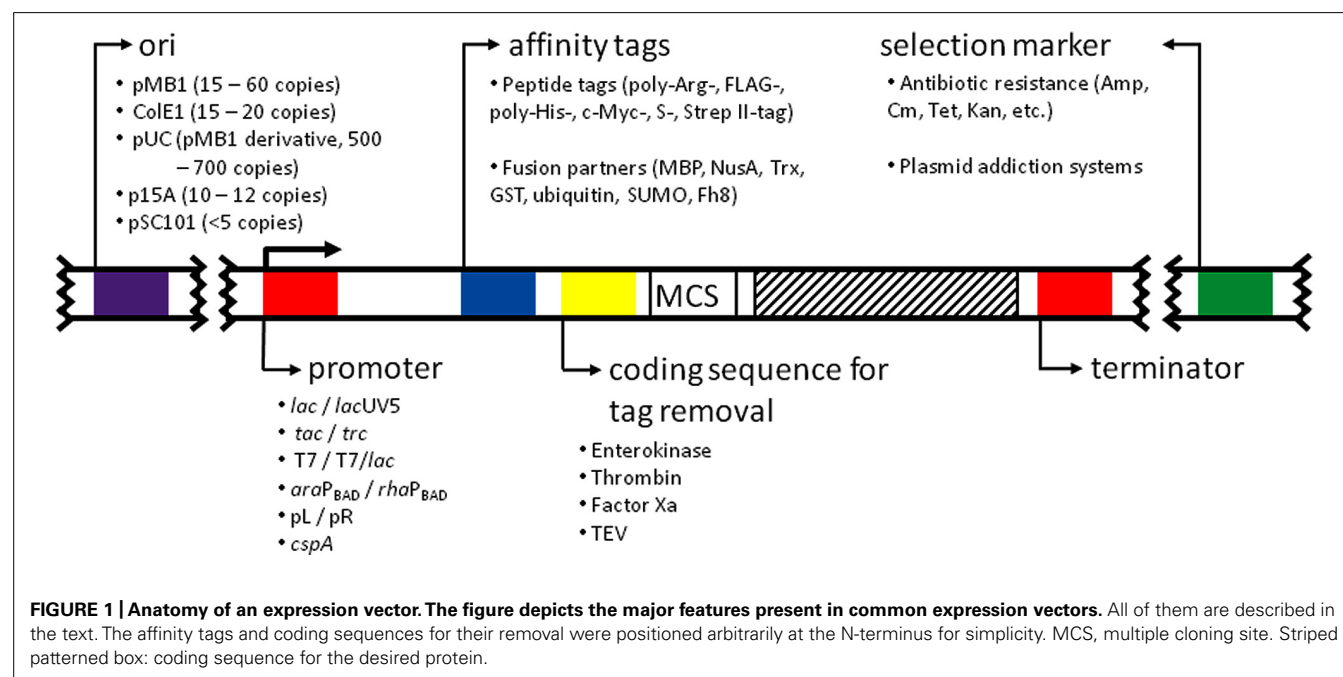
The most common expression plasmids in use today are the result of multiple combinations of replicons, promoters, selection markers, multiple cloning sites, and fusion protein/fusion protein removal strategies (Figure 1). For this reason, the catalog of available expression vectors is huge and it is easy to get lost when choosing a suitable one. To make an informed decision, these features have to be carefully evaluated according to the individual needs.

REPLICON

Genetic elements that undergo replication as autonomous units, such as plasmids, contain a replicon. It consists of one origin of replication together with its associated *cis*-acting control elements. An important parameter to have in mind when choosing a suitable vector is copy number. The control of copy number resides in the replicon (del Solar and Espinosa, 2000). It is logical to think that high plasmid dosage equals more recombinant protein yield as many expression units reside in the cell. However, a high plasmid

number may impose a metabolic burden that decreases the bacterial growth rate and may produce plasmid instability, and so the number of healthy organisms for protein synthesis falls (Bentley et al., 1990; Birnbaum and Bailey, 1991). For this reason, the use of high copy number plasmids for protein expression by no means implies an increase in production yields.

Commonly used vectors, such as the pET series, possess the pMB1 origin (ColE1-derivative, 15–60 copies per cell; Bolivar et al., 1977) while a mutated version of the pMB1 origin is present in the pUC series (500–700 copies per cell; Minton, 1984). The wild-type ColE1 origin (15–20 copies per cell; Lin-Chao and Bremer, 1986; Lee et al., 2006) can be found in the pQE vectors (Qiagen). They all belong to the same incompatibility group meaning that they cannot be propagated together in the same cell as they compete with each other for the replication machinery (del Solar et al., 1998; Camps, 2010). For the dual expression of recombinant proteins using two plasmids, systems with the p15A ori are available (pACYC and pBAD series of plasmids, 10–12 copies per cell; Chang and Cohen, 1978; Guzman et al., 1995). Though rare, triple expression can be achieved by the use of the pSC101 plasmid. This plasmid is under a stringent control of replication, thus it is present in a low copy number (<5 copies per cell; Nordstrom, 2006). The use of plasmids bearing this replicon can be an advantage in cases where the presence of a high dose of a cloned gene or its product produces a deleterious effect to the cell (Stoker et al., 1982; Wang and Kushner, 1991). Alternatively, the use of the Duet vectors (Novagen) simplifies dual expression by allowing cloning of two genes in the same plasmid. The Duet plasmids possess two multiple cloning sites, each preceded by a T7 promoter, a *lac* operator and a ribosome binding site. By combining different compatible Duet vectors, up to eight recombinant proteins can be produced from four expression plasmids.



PROMOTER

The staple in prokaryotic promoter research is undoubtedly the *lac* promoter, key component of the *lac* operon (Müller-Hill, 1996). The accumulated knowledge in the functioning of the system allowed for its extended use in expression vectors. Lactose causes induction of the system and this sugar can be used for protein production. However, induction is difficult in the presence of readily metabolizable carbon sources (such as glucose present in rich media). If lactose and glucose are present, expression from the *lac* promoter is not fully induced until all the glucose has been utilized. At this point (low glucose), cyclic adenosine monophosphate (cAMP) is produced, which is necessary for complete activation of the *lac* operon (Wanner et al., 1978; Postma and Lengeler, 1985). This positive control of expression is known as catabolite repression. In accordance, cAMP levels are low in cells growing in *lac* operon-repressing sugars, and this correlates with lower rates of expression of the *lac* operon (Epstein et al., 1975). Also, glucose abolishes lactose uptake because lactose permease is inactive in the presence of glucose (Winkler and Wilson, 1967). To achieve expression in the presence of glucose, a mutant that reduces (but does not eliminate) sensitivity to catabolite regulation was introduced, the *lacUV5* promoter (Silverstone et al., 1970; Lanzer and Bujard, 1988). However, when present in multicopy plasmids, both promoters suffer from the disadvantage of sometimes having unacceptably high levels of expression in the absence of inducer (a.k.a. “leakiness”) due to titration of the low levels of the *lac* promoter repressor protein LacI from the single chromosomal copy of its gene (about 10 molecules per cell; Müller-Hill et al., 1968). Basal expression control can be achieved by the introduction of a mutated promoter of the *lacI* gene, called *lacI^Q*, that leads to higher levels of expression (almost 10-fold) of LacI (Calos, 1978). The *lac* promoter and its derivative *lacUV5* are rather weak and thus not very useful for recombinant protein production (Deuschle et al., 1986; Makoff and Oser, 1991). Synthetic hybrids that combine the strength of other promoters and the advantages of the *lac* promoter are available. For example, the *tac* promoter consists of the −35 region of the *trp* (tryptophan) promoter and the −10 region of the *lac* promoter. This promoter is approximately 10 times stronger than *lacUV5* (de Boer et al., 1983). Notable examples of commercial plasmids that use the *lac* or *tac* promoters to drive protein expression are the pUC series (*lacUV5* promoter, Thermo Scientific) and the pMAL series of vectors (*tac* promoter, NEB).

The T7 promoter system present in the pET vectors (pMB1 ori, medium copy number, Novagen) is extremely popular for recombinant protein expression. This is not surprising as the target protein can represent 50% of the total cell protein in successful cases (Baneyx, 1999; Graumann and Premstaller, 2006). In this system, the gene of interest is cloned behind a promoter recognized by the phage T7 RNA polymerase (T7 RNAP). This highly active polymerase should be provided in another plasmid or, most commonly, it is placed in the bacterial genome in a prophage (λ DE3) encoding for the T7 RNAP under the transcriptional control of a *lacUV5* promoter (Studier and Moffatt, 1986). Thus, the system can be induced by lactose or its non-hydrolyzable analog isopropyl β -D-1-thiogalactopyranoside

(IPTG). Basal expression can be controlled by *lacI^Q* but also by T7 lysozyme co-expression (Moffatt and Studier, 1987). T7 lysozyme binds to T7 RNAP and inhibits transcription initiation from the T7 promoter (Stano and Patel, 2004). In this way, if small amounts of T7 RNAP are produced because of leaky expression of its gene, T7 lysozyme will effectively control unintended expression of heterologous genes placed under the T7 promoter. T7 lysozyme is provided by a compatible plasmid (pLysS or pLysE). After induction, the amount of T7 RNAP produced surpasses the level of polymerase that T7 lysozyme can inhibit. The “free” T7 RNAP can thus engage in transcription of the recombinant gene. Yet another level of control lies in the insertion of a *lacO* operator downstream of the T7 promoter, making a hybrid T7/*lac* promoter (Dubendorff and Studier, 1991). All three mechanisms (tight repression of the *lac*-inducible T7 RNAP gene by *lacI^Q*, T7 RNAP inhibition by T7 lysozyme and presence of a *lacO* operator after the T7 promoter) make the system ideal for avoiding basal expression.

The problem of leaky expression is a reflection of the negative control of the *lac* promoter. Promoters that rely on positive control should have lower background expression levels (Siegele and Hu, 1997). This is the case of the *araP_{BAD}* promoter present in the pBAD vectors (Guzman et al., 1995). The AraC protein has the dual role of repressor/activator. In the absence of arabinose inducer, AraC represses translation by binding to two sites in the bacterial DNA. The protein–DNA complex forms a loop, effectively preventing RNA polymerase from binding to the promoter. Upon addition of the inducer, AraC switches into “activation mode” and promotes transcription from the *ara* promoter (Schleif, 2000, 2010). In this way, arabinose is absolutely needed for induction.

Another widely used approach is to place a gene under the control of a regulated phage promoter. The strong leftward promoter (pL) of phage lambda directs expression of early lytic genes (Dodd et al., 2005). The promoter is tightly repressed by the λ CI repressor protein, which sits on the operator sequences during lysogenic growth. When the host SOS response is triggered by DNA damage, the expression of the protein RecA is stimulated, which in turn catalyzes the self-cleavage of λ CI, allowing transcription of pL-controlled genes (Johnson et al., 1981; Galkin et al., 2009). This mechanism is used in expression vectors containing the pL promoter. The SOS response (and recombinant protein expression) can be elicited by adding nalidixic acid, a DNA gyrase inhibitor (Lewin et al., 1989; Shatzman et al., 2001). Another way of activating the promoter is to control λ CI production by placing its gene under the influence of another promoter. This two-stage control system has already been described for T7 promoter/T7 RNAP-based vectors. In the pLEX series of vectors (Life Technologies), the λ CI repressor gene was integrated into the bacterial chromosome under the control of the *trp* promoter. In the absence of tryptophan, this promoter is always “on” and λ CI is continuously produced. Upon addition of tryptophan, a tryptophan-TrpR repressor complex is formed that tightly binds to the *trp* operator, thereby blocking λ CI repressor synthesis. Subsequently, the expression of the desired gene under the pL promoter ensues (Mieschendahl et al., 1986).

Transcription from all promoters discussed so far is initiated by chemical cues. Systems that respond to physical signals (e.g., temperature or pH) are also available (Goldstein and Doi, 1995). The pL promoter is one example. A mutant λ CI repressor protein (λ CI⁸⁵⁷) is temperature-sensitive and is unstable at temperatures higher than 37°C. *E. coli* host strains containing the λ CI⁸⁵⁷ protein (either integrated in the chromosome or into a vector) are first grown at 28–30°C to the desired density, and then protein expression is induced by a temperature shift to 40–42°C (Menart et al., 2003; Valdez-Cruz et al., 2010). The industrial advantage of this system lies in part in the fact that during fermentation, heat is usually produced and increasing the temperature in high density cultures is easy. On the other hand, genes under the control of the cold-inducible promoter *cspA* are induced by a downshift in temperature to 15°C (Vasina et al., 1998). This temperature is ideal for expressing difficult proteins as will be explained in another section. The pCold series of plasmids have a pUC118 backbone (a pUC18 derivative; Vieira and Messing, 1987) with the *cspA* promoter (Qing et al., 2004; Hayashi and Kojima, 2008). In the original paper, successful expression was achieved for more than 30 recombinant proteins from different sources, reaching levels as high as 20–40% of the total expressed proteins (Qing et al., 2004). However, it should be noted that in various cases the target proteins were obtained in an insoluble form.

SELECTION MARKER

To deter the growth of plasmid-free cells, a resistance marker is added to the plasmid backbone. In the *E. coli* system, antibiotic resistance genes are habitually used for this purpose. Resistance to ampicillin is conferred by the *bla* gene whose product is a periplasmic enzyme that inactivates the β -lactam ring of β -lactam antibiotics. However, as the β -lactamase is continuously secreted, degradation of the antibiotic ensues and in a couple of hours, ampicillin is almost depleted (Korpimäki et al., 2003). Under this situation, cells not carrying the plasmid are allowed to increase in number during cultivation. Although not experimentally verified, selective agents in which resistance is based on degradation, like chloramphenicol (Shaw, 1983) and kanamycin (Umezawa, 1979), could also have this problem. For this reason, tetracycline has been shown to be highly stable during cultivation (Korpimäki et al., 2003), because resistance is based on active efflux of the antibiotic from resistant cells (Roberts, 1996).

The cost of antibiotics and the dissemination of antibiotic resistance are major concerns in projects dealing with large-scale cultures. Much effort has been put in the development of antibiotics-free plasmid systems. These systems are based on the concept of plasmid addiction, a phenomenon that occurs when plasmid-free cells are not able to grow or live (Zielenkiewicz and Ceglowski, 2001; Peubez et al., 2010). For example, an essential gene can be deleted from the bacterial genome and then placed on a plasmid. Thus, after cell division, plasmid-free bacteria die. Different subtypes of plasmid-addiction systems exist according to their principle of function: (i) toxin/antitoxin-based systems, (ii) metabolism-based systems, and (iii) operator repressor titration systems (Kroll et al., 2010). While this promising technology has been proved successful in large-scale

fermentors (Voss and Steinbüchel, 2006; Peubez et al., 2010), expression systems based on plasmid addiction are still not widely distributed.

AFFINITY TAGS

When devising a project where a purified soluble active recombinant protein is needed (as is often the case), it is invaluable to have means to (i) detect it along the expression and purification scheme, (ii) attain maximal solubility, and (iii) easily purify it from the *E. coli* cellular milieu. The expression of a stretch of amino acids (peptide tag) or a large polypeptide (fusion partner) *in tandem* with the desired protein to form a chimeric protein may allow these three goals to be straightforwardly reached (Nilsson et al., 1997).

Being small, peptide tags are less likely to interfere when fused to the protein. However, in some cases they may provoke negative effects on the tertiary structure or biological activity of the fused chimeric protein (Bucher et al., 2002; Klose et al., 2004; Chant et al., 2005; Khan et al., 2012). Vectors are available that allow positioning of the tag on either the N-terminal or the C-terminal end (the latter option being advantageous when a signal peptide is positioned at the N-terminal end for secretion of the recombinant protein, see below). If the three-dimensional structure of the desired protein is available, it is wise to check which end is buried inside the fold and place the tag in the solvent-accessible end. Common examples of small peptide tags are the poly-Arg-, FLAG-, poly-His-, c-Myc-, S-, and Strep II-tags (Terpe, 2003). Since commercial antibodies are available for all of them, the tagged recombinant protein can be detected by Western blot along expression trials, which is extremely helpful when the levels of the desired proteins are not high enough to be detected by SDS-PAGE. Also, tags allow for one-step affinity purification, as resins that tightly and specifically bind the tags are available. For example, His-tagged proteins can be recovered by immobilized metal ion affinity chromatography using Ni²⁺ or Co²⁺-loaded nitrilotriacetic acid-agarose resins (Porath and Olin, 1983; Bornhorst and Falke, 2000), while anti-FLAG affinity gels (Sigma-Aldrich) are used for capturing FLAG fusion proteins (Hopp et al., 1988).

On the other hand, adding a non-peptide fusion partner has the extra advantage of working as solubility enhancers (Hammarström et al., 2002). The most popular fusion tags are the maltose-binding protein (MBP; Kapust and Waugh, 1999), N-utilization substance protein A (NusA; Davis et al., 1999), thioredoxin (Trx; LaVallie et al., 1993), glutathione S-transferase (GST; Smith and Johnson, 1988), ubiquitin (Baker, 1996) and SUMO (Butt et al., 2005). The reasons why these fusion partners act as solubility enhancers remain unclear and several hypothesis have been proposed (reviewed in Raran-Kurussi and Waugh, 2012). In the case of MBP, it was shown that it possesses an intrinsic chaperone activity (Kapust and Waugh, 1999; Raran-Kurussi and Waugh, 2012). In comparison studies, GST showed the poorest solubility enhancement capabilities (Hammarström et al., 2006; Bird, 2011). NusA, MBP, and Trx display the best solubility enhancing properties but their large size may lead to the erroneous assessment of protein solubility (Costa et al., 2013). Indeed, when these tags are removed, the final solubility of the

desired product is unpredictable (Esposito and Chatterjee, 2006). For these reasons, smaller tags with strong solubility enhancing effects are desirable. Recently, the 8-kDa calcium binding protein Fh8 from the parasite *Fasciola hepatica* was shown to be as good as or better than the large tags in terms of solubility enhancement. Moreover, the recombinant proteins maintained their solubility after tag removal (Costa et al., 2013). MBP and GST can be used to purify the fused protein by affinity chromatography, as MBP binds to amylose–agarose and GST to glutathione–agarose. MBP is present in the pMAL series of vectors from NEB and GST in the pGEX series (GE). A peptide tag must be added to the fusion partner-containing protein if an affinity chromatography step is needed in the purification scheme. MBP and GST bind to their substrates non-covalently. On the contrary, the HaloTag7 (Promega) is based on the covalent capture of the tag to the resin, making the system fast and highly specific (Ohana et al., 2009).

A different group of fusion tags are stimulus-responsive tags, which reversibly precipitate out of solution when subjected to the proper stimulus. The addition of β roll tags to a recombinant protein allows for its selective precipitation in the presence of calcium. The final products presented a high purity and the precipitation protocol only takes a couple of minutes (Shur et al., 2013). Another protein-based stimulus-responsive purification tags are elastin-like polypeptides (ELPs), which consist of tandem repeats of the sequence VPGXG, where X is Val, Ala, or Gly in a 5:2:3 ratio (Meyer and Chilkoti, 1999). These tags undergo an inverse phase transition at a given temperature of transition (T_t). When the T_t is reached, the ELP–protein fusion selectively and reversibly precipitates, allowing for quick enrichment of the recombinant protein by centrifugation (Banki et al., 2005). Precipitation can also be triggered by adjusting the ionic strength of the solution (Ge et al., 2005). These techniques represent an alternative to conventional chromatography-based purification methods and can save production costs, especially in large-scale settings (Fong and Wood, 2010). The main characteristics of the tags mentioned in this section are outlined on Table 1.

TAG REMOVAL

If structural or biochemical studies on the recombinant protein are needed, then the fusion partner must be eliminated from the recombinant protein. Peptide tags should be removed too because they can interfere with protein activity and structure (Wu and Filutowicz, 1999; Perron-Savard et al., 2005), but they can be left in place even for crystallographic studies (Bucher et al., 2002; Carson et al., 2007). Tags can be eliminated by either enzymatic cleavage or chemical cleavage.

In the case of tag removal by enzyme digestion, expression vectors possess sequences that encode for protease cleavage sites downstream of the gene coding for the tag. Enterokinase, thrombin, factor Xa and the tobacco etch virus (TEV) protease have all been successfully used for the removal of peptide tags and fusion partners (Jenny et al., 2003; Blommel and Fox, 2007). Choosing among the different proteases is based on specificity, cost, number of amino acids left in the protein after cleavage and ease of removal after digestion (Waugh, 2011). Enterokinase and thrombin were

popular in the past but the use of His-tagged TEV has become an everyday choice due to its high specificity (Parks et al., 1994), it is easy to produce in large quantities (Tropea et al., 2009) and leaves only a serine or glycine residue (or even the natural N-terminus) after digestion (Kapust et al., 2002).

As the name implies, in chemical cleavage the tag is removed by treatment of the fusion protein with a chemical reagent. The advantages of using chemicals for this purpose are that they are easy to eliminate from the reaction mixture and are cheap in comparison with proteolytic enzymes, which makes them an attractive choice in the large-scale production of recombinant proteins (Rais-Beghdadi et al., 1998). However, the reaction conditions are harsh, so their use is largely restricted to purified recombinant proteins obtained from IBs. They also often cause unwanted protein modifications (Hwang et al., 2014). The most common chemical cleavage reagent is cyanogen bromide (CNBr). CNBr cleaves the peptide bond C-terminal to methionine residues, so this amino acid should be present between the tag and the protein of interest (Rais-Beghdadi et al., 1998). Also, the target protein should not contain internal methionines. CNBr cleavage can be performed in common denaturing conditions (6 M guanidinium chloride) or 70% formic acid or trifluoroacetic acid (Andreev et al., 2010). Other chemical methods for protein cleavage can be found in Hwang et al. (2014).

THIRD QUESTION: WHICH IS THE APPROPRIATE HOST?

A quick search in the literature for a suitable *E. coli* strain to use as a host will yield dozens of possible candidates. All of them have advantages and disadvantages. However, something to keep in mind is that many are specialty strains that are used in specific situations. For a first expression screen, only a couple of *E. coli* strains are necessary: BL21(DE3) and some derivatives of the K-12 lineage.

The history of the BL21 and BL21(DE3) strains was beautifully documented in Daegelen et al. (2009) and we recommend this article to the curious. BL21 was described by Studier in 1986 after various modifications of the B line (Studier and Moffatt, 1986), which in turn Daegelen et al. (2009) traced back to d'Herelle. A couple of genetic characteristics of BL21 are worthy of mention. Like other parental B strains, BL21 cells are deficient in the Lon protease, which degrades many foreign proteins (Gottesman, 1996). Another gene missing from the genome of the ancestors of BL21 is the one coding for the outer membrane protease OmpT, whose function is to degrade extracellular proteins. The liberated amino acids are then taken up by the cell. This is problematic in the expression of a recombinant protein as, after cell lysis, OmpT may digest it (Grodberg and Dunn, 1988). In addition, plasmid loss is prevented thanks to the *hsdSB* mutation already present in the parental strain (B834) that gave rise to BL21. As a result, DNA methylation and degradation is disrupted. When the gene of interest is placed under a T7 promoter, then T7 RNAP should be provided. In the popular BL21(DE3) strain, the λ DE3 prophage was inserted in the chromosome of BL21 and contains the T7 RNAP gene under the *lacUV5* promoter, as was explained earlier.

The BL21(DE3) and its derivatives are by far the most used strains for protein expression. Still, there are reports where the

Table 1 | Main characteristics of protein fusion tags.

	Residues/Size (kDa)	Ligand/Matrix	Purification conditions	
Peptide tags				
Poly-Arg	Usually 5/0.80	Cation-exchange resin	NaCl linear gradient (0–400 mM)	
Poly-His	Usually 6/0.84	Ni ²⁺ -nitrilotriacetic acid-agarose	20–250 mM Imidazole/low pH	
FLAG	8/1.01	Anti-FLAG antibody	2–5 mM EDTA	
		immunodecorated agarose		
Strep-tag II	8/1.06	Specially engineered streptavidin (Strep-Tactin)	2–25 mM desthiobiotin	
c-myc	11/1.20	Anti-myc antibody	Low pH	
		immunodecorated agarose		
S-tag	15/1.75	S-protein (RNase A, residues 21–124) agarose	3 M guanidinium thiocyanate; 0.2 M potassium citrate buffer, pH 2 or 3 M MgCl ₂	
Fusion partners ^a				Solubility enhancement ^b
Fh8	69/8.0	Ca ²⁺ -dependent binding to phenyl-Sepharose	10 mM EDTA	ND
Trx	109/11.7	4-amino phenylarsine oxide agarose (alternatively an affinity tag can be added)	5–1000 mM b-βmercaptoethanol	+++
SUMO	ca. 100/12.0	An affinity tag must be added (usually His-tag)		++++
BRT17 (β roll tag)	153/14.7		Precipitation in the presence of 25–75 mM Ca ²⁺	ND
GST	211/26.0	Glutathione–agarose	10–20 mM reduced glutathione	+
HaloTag7	ca. 300/34.0	Chloroalkane ligand attached to agarose	A protease cleavage site is added between the tag and the protein for in-column cleavage	ND
MBP	396/ca. 42.5	Cross-linked amylose	10 mM maltose	+++
ELPs	550 (for 110 repeats)/ca. 47.0		Precipitation by temperature shifts and/or high concentrations of NaCl (> 1.5 M)	ND
NusA	495/54.8	An affinity tag must be added (usually His-tag)		++

^aNumber of residues and size of fusion partners are approximate in some cases, as many variants exist. ^bThe grading in the solubility enhancement column is based on the results of Bird (Bird, 2011); ND, not determined in that study.

K-12 lineage is used for this purpose. The AD494 and OrigamiTM (Novagen) strains are *trxB* (thioredoxin reductase) mutants, so disulfide bond formation in the cytoplasm is enhanced (the Origami strain also lacks the glutathione reductase gene; Derman et al., 1993). Another widely used strain from the K-12 repertoire is HMS174, a *recA* mutant (Campbell et al., 1978). This mutation has a positive effect on plasmid stability (Marisch et al., 2013). Plasmid multimer formation, an important cause of instability, relies on the recombination system of *E. coli* (Summers et al., 1993). All three strains have their λDE3-containing derivative (available at Novagen) so the T7 RNAP system can be used.

FOURTH QUESTION: WHICH IS THE COMBINATION FOR SUCCESS?

At this point, it should be pretty clear that the number of options when designing an expression system is considerably high. Choosing the perfect combination is not possible *a priori*, so multiple conditions should be tested to obtain the desired protein. If the project demands expressing two protein constructs, cloned in six different expression vectors, each transformed in three different expression strains, then you are in for 36 expression trials. This number may be even higher when other variables are taken into account. This trial-and-error and time consuming pilot study can be made faster if micro-expression trials are performed before

scale-up. Small-scale screens can be performed in 2-ml tubes or 96-well plates (Shih et al., 2002). High throughput protocols adapting automatic liquid handling robots have been described, making it possible for a single person to test more than 1000 culture conditions within a week.

TROUBLESHOOTING RECOMBINANT PROTEIN PRODUCTION

This section of the review covers different strategies for optimizing recombinant protein production in *E. coli*. Even after careful selection of plasmid and host, it cannot be predicted if the protein will be obtained in high amounts and in a soluble active form. Various situations that impede reaching that goal can be encountered, which unfortunately happen very often. Many things to try in each case are discussed in the following paragraphs and, for convenience of the readers; a summary is included in **Table 2**.

NO OR LOW PRODUCTION

This situation may be regarded as the worst case scenario. When the protein of interest cannot be detected through a sensitive technique (e.g., Western blot) or it is detected but at very low levels (less than micrograms per liter of culture), the problem often lies in a harmful effect that the heterologous protein exerts on the cell (Miroux and Walker, 1996; Dumon-Seignovet et al., 2004).

Protein toxicity

The problem of protein toxicity may arise when the recombinant protein performs an unnecessary and detrimental function in the host cell. This function interferes with the normal proliferation and homeostasis of the microorganism and the visible result is slower growth rate, low final cell density, and death (Doherty et al., 1993; Dong et al., 1995).

As a first measure, cell growth should be monitored before induction. If the growth rate of the recombinant strain is slower compared to an empty-vector bearing strain then two causes may explain the phenotype: gene toxicity and basal expression of the toxic mRNA/protein. Gene toxicity will not be discussed here and the review of Saida et al. (2006) is recommended.

The control of basal synthesis was covered in some detail in Section “Promoter.” As stated, the expression of LacI from *lacI* or *lacI^Q* represses transcription of *lac*-based promoters. For high copy number plasmids (>100 copies per cell), *lacI^Q* should be cloned in the expression vector. The pQE vectors from Qiagen utilize two *lac* operator sequences to increase control of the T5 promoter, which is recognized by the *E. coli* RNA polymerase (see The QIAexpressionist™ manual from Qiagen). A tighter control can be achieved by the addition of 0.2–1% w/v glucose in the medium as rich media prepared with tryptone or peptone may contain the inducer lactose (Studier, 2005). Another option could be to prepare defined media using glucose as a source of carbon. In T7-based promoters, leaky expression is avoided by co-expression of T7 lysozyme from the pLysS or pLysE plasmids (see above). Use of lower copy number plasmids containing tightly regulated promoters (like the *araP_{BAD}* promoter) is suggested. An interesting case of copy number control is the one employed in pETcoco vectors (Novagen). These plasmids possess two origins of replication. The *oriS* origin and its control

elements maintain pETcoco at one copy per cell (Wild et al., 2002). However, the TrfA replicator activates the medium-copy origin of replication (*oriV*) and amplification of copy number is achieved (up to 40 copies per cell). The *trfA* gene is on the same vector and is under control of the *araP_{BAD}* promoter, so copy number can be controlled by arabinose (Wild et al., 2002).

After control of basal expression, the culture should grow well until the proper time of induction. At this moment, if the protein is toxic, cell growth will be arrested. In many cases, the level of toxicity of a protein becomes apparent when a certain threshold of host tolerance is reached and exceeded. In such situations, the level of expression should be manipulated at will. Tunable expression can be achieved using the Lemo21(DE3) strain. This strain is similar to the BL21(DE3)pLysS strain, however, T7 lysozyme production from the *lysY* gene is under the tunable promoter *rhaP_{BAD}* (Wagner et al., 2008). At higher concentrations of the sugar L-rhamnose, more T7 lysozyme is produced, less active T7 RNAP is present in the cell and less recombinant protein is expressed. Trials using L-rhamnose concentrations from 0 to 2,000 μ M should be undertaken to find the best conditions for expression. By contrast, dose-dependent expression when using IPTG as inducer is not possible since IPTG can enter the cell by active transport through the Lac permease or by permease-independent pathways (Fernandez-Castane et al., 2012). Since expression of Lac permease is heterogeneous and the number of active permeases in each cell is highly variable, protein expression does not respond predictably to IPTG concentration. The Tuner™ (DE3) strain (Novagen) is a BL21 derivative that possesses a lac permease (*lacY*) mutation that allows uniform entry of IPTG into all *lacY⁻* cells in the population, which produces a concentration-dependent, homogeneous level of induction (Khlebnikov and Keasling, 2002). In the same line of thought, an *E. coli* strain was constructed by exchanging the wild-type operator by the derivative *lacO^c*, thus converting the *lac* operon into a constitutive one. This modification avoids the transient non-genetic *lacY⁻* phenotype of a fraction of the cells, allowing uniform entry of the inducer lactose. A second modification (*gal⁺*) permits the full utilization of lactose as an energy source (Menzella et al., 2003).

A word of caution needs to be said in regard to “tunable promoters” that are inducible by sugars (lactose, arabinose, rhamnose). In the case of the *araP_{BAD}* promoter, the yields of the target protein can be reproducibly increased over a greater than 100-fold range by supplementing the culture with different sub-maximal concentrations of arabinose (Guzman et al., 1995). This led to the erroneous belief that within each cell, the level of recombinant protein synthesis can be manipulated at will. However, it was shown that the range in protein expression arises from the heterogeneity in the amount of active sugar permeases in each cell, as was also explained for LacY (Siegele and Hu, 1997). So, even though the final protein yield can be controlled, the amount of protein per cell is widely variable, with cells producing massive amounts of protein and others not producing any protein at all. This can be a nuance, since in the case of toxic products; the subpopulation of cells with high-level synthesis may perish (Doherty et al., 1993; Dong et al., 1995).

Table 2 | Strategies for overcoming common problems during recombinant protein expression in *E. coli*.

Problem	Possible explanation	Solutions
No or low expression	Protein may be toxic before induction	Control basal induction: <ul style="list-style-type: none"> • add glucose when using expression vectors containing <i>lac</i>-based promoters • use defined media with glucose as source of carbon • use pLysS/pLysE bearing strains in T7-based systems • use promoters with tighter regulation
	Protein may be toxic after induction	Lower plasmid copy number Control level of induction: <ul style="list-style-type: none"> • Tuneable promoters • Use strains that allow control of induction [Lemo21(DE3) strain] or <i>lacY</i>[−] strains (TunerTM)
	Codon bias	Lower plasmid copy number Use strains that are better for the expression of toxic proteins (C41 or C43) Direct protein to the periplasm
		Optimize codon frequency in cDNA to better reflect the codon usage of the host Use codon bias-adjusted strains Increase biomass: <ul style="list-style-type: none"> • Try new media formulations • Provide good aeration and avoid foaming
Inclusion body formation	Incorrect disulfide bond formation	Direct protein to the periplasm Use <i>E. coli</i> strains with oxidative cytoplasmic environment
	Incorrect folding	Co-express molecular chaperones Supplement media with chemical chaperones and cofactors Remove inducer and add fresh media
		Lower production rate: <ul style="list-style-type: none"> • Lower temperature. If possible, use strains with cold-adapted chaperones • Tune inducer concentration
Protein inactivity	Low solubility of the protein	Fuse desired protein to a solubility enhancer (fusion partners)
	An essential post translational modification is needed	Change microorganism
	Incomplete folding	Lower temperature Monitor disulfide bond formation and allow further folding <i>in vitro</i>
	Mutations in cDNA	Sequence plasmid before and after induction. If mutations are detected, the protein may be toxic. Use a <i>recA</i> [−] strain to ensure plasmid stability Transform <i>E. coli</i> before each expression round

Some *E. coli* mutants were specifically selected to withstand the expression of toxic proteins. The strains C41(DE3) and C43(DE3) were found by Miroux and Walker (1996) in a screen designed to isolate derivatives of BL21(DE3) with improved membrane protein overproduction characteristics. It was recently discovered that the previously uncharacterized mutations which prevent cell death during the expression of recombinant proteins in these strains lie on the *lacUV5* promoter. In BL21(DE3) cells, the *lacUV5* promoter drives the expression of the T7 RNAP, but in the Walker strains two mutations in the -10 region revert the *lacUV5* promoter back into the weaker wild-type counterpart. This leads to a lesser (and perhaps more tolerable for the cell) level of synthesis (Wagner et al., 2008).

Another solution could be to remove the protein from the cell. Secretion to the periplasm or to the medium is sometimes the only way to produce a recombinant protein (Mergulhao et al., 2005; de Marco, 2009). The first option for expression in the periplasm is the post-translational Sec-dependent pathway (Georgiou and Segatori, 2005). Routing to the extracytoplasmic space is achieved by fusing the recombinant protein to a proper leader peptide. The signal peptides of the following proteins are widely used for secretion: Lpp, LamB, LTB, MalE, OmpA, OmpC, OmpF, OmpT, PelB, PhoA, PhoE, or SpA (Choi and Lee, 2004). The co-translational translocation machinery based on the SRP (signal recognition particle) pathway can also be used. SRP recognizes its substrates by the presence of a hydrophobic signal sequence located in the N-terminal end. Following interaction with the membrane receptor FtsY, the complex of nascent chain and ribosome is transferred to the SecYEG translocase (Valent et al., 1998). The signal sequence of disulfide isomerase I (DsbA) has been used to target recombinant proteins to the periplasm via the SRP pathway. Notable examples of recombinant proteins secreted through this system include thioredoxin (Schiele et al., 2003) and the human growth hormone (Soares et al., 2003).

Codon bias

Codon bias arises when the frequency of occurrence of synonymous codons in the foreign coding DNA is significantly different from that of the host. At the moment of full synthesis of the recombinant protein, depletion of low-abundance tRNAs occurs. This deficiency may lead to amino acid misincorporation and/or truncation of the polypeptide, thus affecting the heterologous protein expression levels (which will be low at best) and/or its activity (Gustafsson et al., 2004). To check if codon bias could be an issue when expressing a recombinant protein, a large number of free online apps detect the presence of rare codons in a given gene when *E. coli* is used as a host (molbiol.ru/eng/scripts/01_11.html, genscript.com/cgi-bin/tools/rare_codon_analysis, nihserver.mbi.ucla.edu/RACC/, just to name a few). Rare codons were defined as codons used by *E. coli* at a frequency $<1\%$ (Kane, 1995). For example, the AGG codon (Arg) is used in *E. coli* at a frequency of $<0.2\%$, but it is not rare in plant mRNAs where it can reach frequencies $>1.5\%$.

Two strategies for solving codon usage bias have been used: codon optimization of the foreign coding sequence or increasing the availability of underrepresented tRNAs by host modification

(Sorensen and Mortensen, 2005). The rationale behind codon usage optimization is to modify the rare codons in the target gene to mirror the codon usage of the host (Burgess-Brown et al., 2008; Welch et al., 2009; Menzella, 2011). The amino acid sequence of the encoded protein must not be altered in the process. This can be done by site-directed silent mutagenesis or resynthesis of the whole gene or parts of it. Codon optimization by silent mutagenesis is a cumbersome and expensive process, so is not very useful when many recombinant proteins are needed. On the other hand, gene synthesis by design is not a trivial issue since it requires choosing the best sequence from a vast number of possible combinations (Gustafsson et al., 2004). The simplest approach is to replace all instances of a given amino acid in the target gene by the most abundant codon of the host, a strategy called “one amino acid-one codon.” More advanced algorithms, which employ several other optimization parameters such as codon context and codon harmonization, have been described (Gao et al., 2004; Supek and Vlahovicek, 2004; Jayaraj et al., 2005; Angov et al., 2011). Some are freely available as web servers or standalone software. For a comprehensive list, please refer to Puigbo et al. (2007).

Correcting codon usage is a tricky situation. The “one amino acid-one codon” strategy disregards factors other than codon rarity that influence protein expression levels. For example, in bacterial genes enriched in rare codons at the N-terminus, protein expression is actually improved. The cause lies not in codon rarity *per se* but in the reduction of RNA secondary structure (Goodman et al., 2013). In addition, a recent report has shown that high levels of protein production are mainly (but not only) determined by the decoding speed of the open reading frame (i.e., the time it takes for a ribosome to translate an mRNA), especially if “fast” codons are located at the 5'-end of the mRNA (Chu et al., 2014). This causes a fast ribosome clearance at the initiation site, so that new recruited ribosomes encounter a free start codon and can engage in translation. Finally, some codon combinations can create Shine–Dalgarno-like structures that cause translational pausing by hybridization between the target mRNA and the 16S rRNA of the translating ribosome (Li et al., 2012). Translational pausing along the mRNA has a beneficial effect in protein folding, as it allows for the newly synthesized chain to adopt a well-folded intermediate conformation (Thanaraj and Argos, 1996; Oresic and Shalloway, 1998; Tsai et al., 2008; Yona et al., 2013). All of this new evidence in translational control mechanisms poses a challenge in the rational design of synthetic genes. Newer algorithms should account for 5' RNA structure, presence of strategically located Shine–Dalgarno-like motifs, ribosome clearance rates at the initiation site and presence of slowly translated regions that are beneficial in co-translational folding.

On the other hand, when the cell is producing massive amounts of proteins (as in the case of recombinant expression of heterologous genes), charged tRNA availability for rare codons does become the major determinant of the levels of produced protein (Pedersen, 1984; Li et al., 2012). Low-abundance tRNA depletion causes ribosome stalling and its subsequent detachment from the RNA strand and thus, failure to generate a full-length product (Buchan and Stansfield, 2007). Several strains carrying plasmids containing extra copies of problematic tRNAs genes can

be used to circumvent this issue. The BL21(DE3)CodonPlus strain (Stratagene) contains the pRIL plasmid (p15A replicon, which is compatible with the ColE1 and ColE1-like origins contained in most commonly used expression vectors), which provides extra genes for the tRNAs for AGG/AGA (Arg), AUA (Ile), and CUA (Leu). BL21(DE3)CodonPlus-RP (Stratagene) corrects for the use of AGG/AGA (Arg) and CCC (Pro). The Rosetta(DE3) strains (Novagen) are TunerTM derivatives containing the pRARE plasmid (p15A replicon), supplying tRNAs for all the above-mentioned codons plus GGA (Gly). It should be noted that the use of these strains often improves the levels of protein production but sometimes can cause a decrease in protein solubility. We have found that proteins with higher than 5% content of RIL codons (AGG/AGA, AUA, and CUA) are less soluble when expressed in the Codon-Plus strain. In this host, the translational pauses introduced by the RIL codons are probably overridden, increasing translation speed and consequently, protein aggregation (Rosano and Ceccarelli, 2009).

Limiting factors in batch cultivation

When the expression of the recombinant protein is low and cannot be increased by the proposed mechanisms, then the volumetric yield of desired protein can be augmented by growing the culture to higher densities. This can be achieved by changing a few parameters, like medium composition and providing better aeration by vigorous shaking (McDaniel and Bailey, 1969; Cui et al., 2006; Blommel et al., 2007).

LB is the most commonly used medium for culturing *E. coli*. It is easy to make, it has rich nutrient contents and its osmolarity is optimal for growth at early log phase. All these features make it adequate for protein production and compensate for the fact that it is not the best option for achieving high cell density cultures. Despite being a rich broth, cell growth stops at a relatively low density. This happens because LB contains scarce amounts of carbohydrates (and other utilizable carbon sources) and divalent cations (Sezonov et al., 2007). Not surprisingly, increasing the amount of peptone or yeast extract leads to higher cell densities (Studier, 2005). Also, divalent cation supplementation (MgSO₄ in the millimolar range) results in higher cell growth. Adding glucose is of limited help in this regard because acid generation by glucose metabolism overwhelms the limited buffer capacity of LB, at least in shake flasks where pH control can be laborious (Weuster-Botz et al., 2001; Scheidle et al., 2011). If culture acidification poses a problem, the media can be buffered with phosphate salts at 50 mM. 2xYT, TB (Terrific Broth) and SB (Super Broth) media recipes are available elsewhere and have been shown to be superior to LB for reaching higher cell densities (Madurawe et al., 2000; Atlas, 2004; Studier, 2005).

A major breakthrough in media composition came in 2005 by the extensive work of Studier. In that report, the concept of autoinduction was developed (Studier, 2005). In autoinduction media, a mixture of glucose, lactose, and glycerol is used in an optimized blend. Glucose is the preferred carbon source and is metabolized preferentially during growth, which prevents uptake of lactose until glucose is depleted, usually in mid to late log phase. Consumption of glycerol and lactose follows, the latter being also the inducer of *lac*-controlled protein expression. In this

way, biomass monitoring for timely inducer addition is avoided, as well as culture manipulation (Studier, 2014).

As the number of cells per liter increases, oxygen availability becomes an important factor with profound influence on growth (O'Beirne and Hamer, 2000; Losen et al., 2004). Oxygen limitation triggers the expression of more than 200 genes in an attempt to adjust the metabolic capacities of the cell to the availability of oxygen, all of which hinder optimal growth over long culture periods (Unden et al., 1995). The easiest way to increase the amount of available oxygen in shake vessels is to increase shaking speed. For regular flasks, the optimal shaking speed range is 400–450 rpm. More agitation is generated in baffled flasks; under these conditions, 350–400 rpm are enough for good aeration. However, vigorous shaking can induce the formation of foam, which will lower oxygen transfer. For this reason, the addition of an antifoaming agent is recommended, although it was shown that antifoams can affect the growth rate of several microorganisms and the yield of recombinant protein (Routledge et al., 2011; Routledge, 2012). Also, proper aeration depends on the ratio of culture volume to vessel capacity. As a rule of thumb, the culture volume should be less or equal to 10% of the shaking flask capacity, although in our hands, protein production with culture volumes occupying 20% of the flask capacity was possible (Rosano et al., 2011). A strategy that can produce significant increases in cell density is fed-batch fermentation. This approach has a wide availability of tools and methods, but it is beyond the scope of this paper and is addressed elsewhere (Yamanè and Shimizu, 1984; Yee and Blanch, 1992; Moulton, 2013).

Two rarely discussed parameters in the process of recombinant protein production are the preparation of the starting culture and the time of induction. Most protocols call for diluting a saturated overnight preculture (dilution factor 1/100) into the larger culture (Sivashanmugam et al., 2009). However, leaky expression of the chosen system can lead to plasmid instability, which may result in a poor yield of target protein. Also, in the starter culture, cells can be in dissimilar metabolic states. Upon dilution into fresh media, cells will grow at different rates leading to irreproducible induction points (Huber et al., 2009). A proper preculture (cells in an active equalized growing phase) can be prepared by growing the overnight starter culture at 20–25°C or by using a slow-release system for glucose, among other methods (Busso et al., 2008; Huber et al., 2009; Sivashanmugam et al., 2009). After inoculation and further growth, the inducer is often added in mid-log phase because the culture is growing fast and protein translation is maximal. However, induction at early stationary phase is also possible (Ou et al., 2004). In fact, in some cases the target protein was more soluble when inducer was added at this stage (Galloway et al., 2003). Presumably, the reduced rate of protein synthesis may result in less aggregation in IBs, as we describe below.

INCLUSION BODIES FORMATION

When a foreign gene is introduced in *E. coli*, spatio-temporal control of its expression is lost. The newly synthesized recombinant polypeptide is expressed in the microenvironment of *E. coli*, which may differ from that of the original source in terms of pH, osmolarity, redox potential, cofactors, and folding

mechanisms. Also, in high level expression, hydrophobic stretches in the polypeptide are present at high concentrations and available for interaction with similar regions. All of these factors lead to protein instability and aggregation (Hartley and Kane, 1988; Carrio and Villaverde, 2002). These buildups of protein aggregates are known as IBs. IB formation results from an unbalanced equilibrium between protein aggregation and solubilization. So, it is possible to obtain a soluble recombinant protein by strategies that ameliorate the factors leading to IB formation (Carrio and Villaverde, 2001, 2002). One is to fuse the desired protein to a fusion partner that acts as a solubility enhancer. Some examples were already described in Section “Affinity Tags.” In some cases the generation of IB can be an advantage, especially if the protein can be refolded easily *in vitro*. If that is the case, conditions can be adjusted to favor the formation IB, providing a simple method for achieving a significant one-step purification of the expressed protein (Burgess, 2009; Basu et al., 2011).

Disulfide bond formation

For many recombinant proteins, the formation of correct disulfide bonds is vital for attaining their biologically active three-dimensional conformation. The formation of erroneous disulfide bonds can lead to protein misfolding and aggregation into IB. In *E. coli*, cysteine oxidation takes place in the periplasm, where disulfide bonds are formed in disulfide exchange reactions catalyzed by a myriad of enzymes, mainly from the Dsb family (Messens and Collet, 2006). By contrast, disulfide bond formation in the cytoplasm is rare, maybe because cysteine residues are part of catalytic sites in many enzymes. Disulfide bond formation at these sites may lead to protein inactivation, misfolding, and aggregation (Derman et al., 1993). The cytoplasm has a more negative redox potential and is maintained as a reducing environment by the thioredoxin–thioredoxin reductase (trxB) system and the glutaredoxin–glutaredoxin reductase (gor) system (Stewart et al., 1998). This situation has a huge impact in the production of recombinant proteins with disulfide bonds. One option would be to direct the protein to the periplasm, as we have discussed in Section “Protein Toxicity.”

Nevertheless, expression in the cytoplasm is still possible thanks to engineered *E. coli* strains that possess an oxidative cytoplasmic environment that favors disulfide bond formation (Derman et al., 1993). Worthy of mention are the Origami (Novagen) and SHuffle (NEB) strains. We described earlier the OrigamiTM strain, as having a *trxB*[−] *gor*[−] genotype in the K-12 background (as this double mutant is not viable, a suppressor mutation in the *ahpC* gene is necessary to maintain viability; Bessette et al., 1999). OrigamiTM is also available in the BL21(DE3) *lacY* (TunerTM, Novagen) background. Addition of the pRARE plasmid for the extra advantage of correcting codon bias resulted in the construction of the Rosetta-gamiTM B strain (Novagen). The SHuffle[®] T7 Express strain [BL21(DE3) background, NEB] goes a little bit further. Besides the *trxB*[−] and *gor*[−] mutations, it constitutively expresses a chromosomal copy of the disulfide bond isomerase DsbC (Lobstein et al., 2012). DsbC promotes the correction of mis-oxidized proteins into their correct form and is also a chaperone that can

assist in the folding of proteins that do not require disulfide bonds. Due to the action of DsbC, less target protein aggregates into IB.

Chaperone co-expression/chemical chaperones and cofactor supplementation

Molecular chaperones lie at the heart of protein quality control, aiding nascent polypeptides to reach their final structure (Hartl and Hayer-Hartl, 2002). Other specialized types of chaperones, like ClpB, can disassemble unfolded polypeptides present in IB. The high level expression of recombinant proteins results in the molecular crowding of the cytosol and quality control mechanisms may be saturated in this situation (Carrio and Villaverde, 2002). One strategy for solving this problem is to stop protein expression by inducer removal after a centrifugation step and addition of fresh media supplemented with chloramphenicol, an inhibitor of protein synthesis. This allows recruitment of molecular chaperones to aid in the folding of newly synthesized recombinant polypeptides (Carrio and Villaverde, 2001; de Marco and De Marco, 2004).

Given their function, it is not surprising that efforts to inhibit IB formation were directed to the co-expression of individual or sets of molecular chaperones (Caspers et al., 1994; Nishihara et al., 2000; de Marco et al., 2007). Commercially, one of the most used systems is the chaperone plasmid set from Takara (Nishihara et al., 1998, 2000). This set consists of five plasmids (pACYC derivatives) which allow overexpression of different chaperones or combinations of them: (i) GroES–GroEL, (ii) DnaK/DnaJ/GrpE, (iii) (i) + (ii), (iv) trigger factor, (v) (i) + (iv). On the other hand, if such a system is not at hand, the natural network of chaperones can be induced by the addition of benzyl alcohol or heat shock, though the latter is not recommended (de Marco et al., 2005).

When proteins are purified from IB, urea-denatured and then refolded *in vitro*, addition of osmolytes (also called chemical chaperones) in the 0.1–1 M range of concentration increases the yield of soluble protein (Rudolph and Lilie, 1996; Clark, 1998; Tsumoto et al., 2003; Alibolandi and Mirzahoseini, 2011). This situation can be mimicked *in vivo* by supplementing the culture media with osmolytes such as proline, glycine–betaine, and trehalose (de Marco et al., 2005). Also, the folding pathways that lead to the correct final conformation and stabilization of the proper folded protein may require specific cofactors in the growth media, for example, metal ions (such as iron–sulfur and magnesium) and polypeptide cofactors. Addition of these compounds to the batch culture considerably increases the yield as well as the folding rate of soluble proteins (Sorensen and Mortensen, 2005).

Slowing down production rate

Slower rates of protein production give newly transcribed recombinant proteins time to fold properly. This was previously addressed when we discussed the role of translational pauses at rare codons and their impact in the production of recombinant proteins. Moreover, the reduction of cellular protein concentration favors proper folding. By far, the most commonly used way to lower protein synthesis is reducing incubation temperature (Schein and Noteborn, 1988; Vasina and Baneyx, 1997; Vera

et al., 2007). Low temperatures decrease aggregation, which is favored at higher temperatures due to the temperature dependence of hydrophobic interactions (Baldwin, 1986; Makhatadze and Privalov, 1995; Schellman, 1997).

When IB formation is a problem, recombinant protein synthesis should be carried out in the range 15–25°C, though one report described successful expression at 4°C for 72 h (San-Miguel et al., 2013). However, when working at the lower end of the temperature range, slower growth and reduced synthesis rates can result in lower protein yields. Also, protein folding may be affected as the chaperone network may not be as efficient (McCarty and Walker, 1991; Mendoza et al., 2000; Strocchi et al., 2006). The ArticExpress™ (Stratagene) strain (B line) possesses the cold-adapted chaperonin Cpn60 and co-chaperonin Cpn10 from the psychrophilic bacterium *Oleispira antarctica* (Ferrer et al., 2004). The chaperonins display high refolding activities at temperatures of 4–12°C and confer an enhanced ability for *E. coli* to grow at lower temperatures (Ferrer et al., 2003).

PROTEIN INACTIVITY

Obtaining a nice amount of soluble protein is not the end of the road. The protein may still be of bad quality; i.e., it does not have the activity it should. Incomplete folding could be the culprit in this scenario (Gonzalez-Montalban et al., 2007; Martinez-Alonso et al., 2008). In this case, the protein adopts a stable soluble conformation but the exact architecture of the active site is still unsuitable for activity. Some options already addressed can be helpful in these cases. Some proteins require small molecules or prosthetic groups to acquire their final folded conformation. Adding these compounds to the culture media can increase the yield and the quality of the expressed protein significantly (Weickert et al., 1999; Yang et al., 2003). Also, erroneous disulfide bond formation can lead to protein inactivity (Kurokawa et al., 2000). In addition, protein production at lower temperatures has a profound impact on protein quality. Work by the Villaverde lab has shown that conformational quality and functionality of highly soluble recombinant proteins increase when the temperature of the culture is reduced (Vera et al., 2007). This was also the case when the intracellular concentration of the chaperone DnaK was elevated (Martinez-Alonso et al., 2007). This phenomenon calls into question the use of solubility as an indicator of quality. Based on this fact, then it may be wise to express all recombinant proteins at low temperatures or at least, to compare the specific activity of a recombinant protein obtained at different temperatures.

If the activity of the heterologous protein is toxic to the cell, genetic reorganization of the expression vector leading to loss of activity may occur, allowing the host to survive and eventually take over the culture (Corchero and Villaverde, 1998). This structural instability of the plasmid can be detected by DNA sequencing after purification of the plasmid at the end of process. Any point mutation, deletion, insertion, or rearrangement may explain the low activity of a purified recombinant protein (Palomares et al., 2004).

CONCLUDING REMARKS

In terms of recombinant expression, *E. coli* has always been the preferred microbial cell factory. *E. coli* is a suitable host for

expressing stably folded, globular proteins from prokaryotes and eukaryotes. Even though membrane proteins and proteins with molecular weights above 60 kDa are difficult to express, several reports have had success in this regard (our laboratory has produced proteins from plants in the 90–95 kDa range; Rosano et al., 2011). Large-scale protein expression trials have shown that <50% of bacterial proteins and <15% of non-bacterial proteins can be expressed in *E. coli* in a soluble form, which demonstrates the versatility of the system (Braun and LaBaer, 2003). However, when coming across a difficult-to-express protein, things can get complicated. We hope to have given a thorough list of possible solutions when facing the challenge of expressing a new protein in *E. coli*. Nevertheless, a word of caution is needed. Many of the approaches described in this review will fail miserably in a lot of cases. This can be explained by the fact that strategies aiming at troubleshooting recombinant protein expression are sometimes protein specific and suffer from positive bias; i.e., things that work get published, all the others, do not. That being said, thanks to the efforts of the scientific community, the general methods available in the literature are no longer anecdotal and can be used systematically. Moreover, the field is always expanding and even after almost 40 years from the first human protein obtained in *E. coli* (Itakura et al., 1977), there is still much room for improvement.

AUTHOR CONTRIBUTIONS

Germán L. Rosano and Eduardo A. Ceccarelli wrote the manuscript and approved its final version.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their insightful comments on the manuscript, as their remarks led to an improvement of the work. Germán L. Rosano and Eduardo A. Ceccarelli are staff members of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). Also, Germán L. Rosano is a Teaching Assistant and Eduardo A. Ceccarelli is a Professor of the Facultad de Ciencias Bioquímicas y Farmacéuticas, UNR, Argentina. This study was supported by grants from CONICET and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, Argentina).

REFERENCES

- Adrio, J. L., and Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioeng. Bugs* 1, 116–131. doi: 10.4161/bbug.1.2.10484
- Alibolandi, M., and Mirzahoseini, H. (2011). Chemical assistance in refolding of bacterial inclusion bodies. *Biochem. Res. Int.* 2011:631607. doi: 10.1155/2011/631607
- Andreev, Y. A., Kozlov, S. A., Vassilevski, A. A., and Grishin, E. V. (2010). Cyanogen bromide cleavage of proteins in salt and buffer solutions. *Anal. Biochem.* 407, 144–146. doi: 10.1016/j.ab.2010.07.023
- Angov, E., Legler, P. M., and Mease, R. M. (2011). Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol. Biol.* 705, 1–13. doi: 10.1007/978-1-61737-967-3_1
- Atlas, R. M. (2004). *Handbook of Microbiological Media*, 3rd Edn. Boca Raton, FL: Taylor & Francis.
- Baker, R. T. (1996). Protein expression using ubiquitin fusion and cleavage. *Curr. Opin. Biotechnol.* 7, 541–546. doi: 10.1016/S0958-1669(96)80059-0
- Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8069–8072. doi: 10.1073/pnas.83.21.8069
- Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* 10, 411–421. doi: 10.1016/S0958-1669(99)00003-8

- Banki, M. R., Feng, L., and Wood, D. W. (2005). Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nat. Methods* 2, 659–662. doi: 10.1038/nmeth787
- Basu, A., Li, X., and Leong, S. S. (2011). Refolding of proteins from inclusion bodies: rational design and recipes. *Appl. Microbiol. Biotechnol.* 92, 241–251. doi: 10.1007/s00253-011-3513-y
- Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990). Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnol. Bioeng.* 35, 668–681. doi: 10.1002/bit.260350704
- Besette, P. H., Aslund, F., Beckwith, J., and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13703–13708. doi: 10.1073/pnas.96.24.13703
- Bird, L. E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods* 55, 29–37. doi: 10.1016/j.ymeth.2011.08.002
- Birnbaum, S., and Bailey, J. E. (1991). Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* 37, 736–745. doi: 10.1002/bit.260370808
- Blommel, P. G., Becker, K. J., Duvnjak, P., and Fox, B. G. (2007). Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol. Prog.* 23, 585–598. doi: 10.1021/bp070011x
- Blommel, P. G., and Fox, B. G. (2007). A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr. Purif.* 55, 53–68. doi: 10.1016/j.pep.2007.04.013
- Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., et al. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2, 95–113. doi: 10.1016/0378-1119(77)90000-2
- Bornhorst, J. A., and Falke, J. J. (2000). Purification of proteins using poly-histidine affinity tags. *Methods Enzymol.* 326, 245–254. doi: 10.1016/S0076-6879(00)26058-8
- Braun, P., and LaBaer, J. (2003). High throughput protein production for functional proteomics. *Trends Biotechnol.* 21, 383–388. doi: 10.1016/S0167-7799(03)00189-6
- Buchan, J. R., and Stansfield, I. (2007). Halting a cellular production line: responses to ribosomal pausing during translation. *Biol. Cell* 99, 475–487. doi: 10.1042/BC20070037
- Bucher, M. H., Evdokimov, A. G., and Waugh, D. S. (2002). Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallogr. D. Biol. Crystallogr.* 58, 392–397. doi: 10.1107/S0907444901021187
- Burgess, R. R. (2009). Refolding solubilized inclusion body proteins. *Methods Enzymol.* 463, 259–282. doi: 10.1016/S0076-6879(09)63017-2
- Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. (2008). Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. *Protein Expr. Purif.* 59, 94–102. doi: 10.1016/j.pep.2008.01.008
- Busso, D., Stierle, M., Thierry, J. C., and Moras, D. (2008). A comparison of inoculation methods to simplify recombinant protein expression screening in *Escherichia coli*. *Biotechniques* 44, 101–106. doi: 10.2144/000112632
- Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* 43, 1–9. doi: 10.1016/j.pep.2005.03.016
- Calos, M. P. (1978). DNA sequence for a low-level promoter of the lac repressor gene and an ‘up’ promoter mutation. *Nature* 274, 762–765. doi: 10.1038/274762a0
- Campbell, J. L., Richardson, C. C., and Studier, F. W. (1978). Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. *Proc. Natl. Acad. Sci. U.S.A.* 75, 2276–2280. doi: 10.1073/pnas.75.5.2276
- Camps, M. (2010). Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Pat. DNA Gene Seq.* 4, 58–73. doi: 10.2174/187221510790410822
- Carrio, M. M., and Villaverde, A. (2001). Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett.* 489, 29–33. doi: 10.1016/S0014-5793(01)02073-7
- Carrio, M. M., and Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol.* 96, 3–12. doi: 10.1016/S0168-1656(02)00032-9
- Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and Delucas, L. J. (2007). His-tag impact on structure. *Acta Crystallogr. D. Biol. Crystallogr.* 63, 295–301. doi: 10.1107/S0907444906052024
- Caspers, P., Stieger, M., and Burn, P. (1994). Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*. *Cell. Mol. Biol. (Noisy-le-grand)* 40, 635–644.
- Chang, A. C., and Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134, 1141–1156.
- Chant, A., Kraemer-Pecore, C. M., Watkin, R., and Kneale, G. G. (2005). Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expr. Purif.* 39, 152–159. doi: 10.1016/j.pep.2004.10.017
- Choi, J. H., Keum, K. C., and Lee, S. Y. (2006). Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chem. Eng. Sci.* 61, 9. doi: 10.1016/j.ces.2005.03.031
- Choi, J. H., and Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 64, 625–635. doi: 10.1007/s00253-004-1559-9
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., and Von Der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 33, 21–34. doi: 10.1002/emboj.201385651
- Clark, E. D. B. (1998). Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* 9, 157–163. doi: 10.1016/S0958-1669(98)80109-2
- Corchero, J. L., and Villaverde, A. (1998). Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng.* 58, 625–632. doi: 10.1002/(SICI)1097-0290(19980620)58:6<625::AID-BIT8>3.0.CO;2-K
- Costa, S. J., Almeida, A., Castro, A., Domingues, L., and Besir, H. (2013). The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol.* 97, 6779–6791. doi: 10.1007/s00253-012-4559-1
- Cui, F. J., Li, Y., Xu, Z. H., Xu, H. Y., Sun, K., and Tao, W. Y. (2006). Optimization of the medium composition for production of mycelial biomass and exo-polymer by *Grifola frondosa* GF9801 using response surface methodology. *Bioresour. Technol.* 97, 1209–1216. doi: 10.1016/j.biortech.2005.05.005
- Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., and Kim, J. F. (2009). Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21 (DE3). *J. Mol. Biol.* 394, 634–643. doi: 10.1016/j.jmb.2009.09.022
- Davis, G. D., Elisee, C., Newham, D. M., and Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* 65, 382–388. doi: 10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
- de Boer, H. A., Comstock, L. J., and Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. U.S.A.* 80, 21–25. doi: 10.1073/pnas.80.1.21
- del Solar, G., and Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* 37, 492–500. doi: 10.1046/j.1365-2958.2000.02005.x
- del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434–464.
- Demain, A. L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27, 297–306. doi: 10.1016/j.biotechadv.2009.01.008
- de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* 8, 26. doi: 10.1186/1475-2859-8-26
- de Marco, A., and De Marco, V. (2004). Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. *J. Biotechnol.* 109, 45–52. doi: 10.1016/j.jbiotec.2003.10.025
- de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., and Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* 7:32. doi: 10.1186/1472-6750-7-32
- de Marco, A., Vigh, L., Diamant, S., and Goloubinoff, P. (2005). Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones* 10, 329–339. doi: 10.1379/CSC-139R.1

- Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* 262, 1744–1747. doi: 10.1126/science.8259521
- Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986). Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* 5, 2987–2994.
- Dodd, I. B., Shearwin, K. E., and Egan, J. B. (2005). Revisited gene regulation in bacteriophage lambda. *Curr. Opin. Genet. Dev.* 15, 145–152. doi: 10.1016/j.gde.2005.02.001
- Doherty, A. J., Connolly, B. A., and Worrall, A. F. (1993). Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector. *Gene* 136, 337–340. doi: 10.1016/0378-1119(93)90491-K
- Dong, H., Nilsson, L., and Kurland, C. G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* 177, 1497–1504.
- Dubendorff, J. W., and Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219, 45–59. doi: 10.1016/0022-2836(91)90856-2
- Dumon-Seignover, L., Cariot, G., and Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* 37, 203–206. doi: 10.1016/j.pep.2004.04.025
- Epstein, W., Rothman-Denes, L. B., and Hesse, J. (1975). Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 72, 2300–2304. doi: 10.1073/pnas.72.6.2300
- Esposito, D., and Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* 17, 353–358. doi: 10.1016/j.copbio.2006.06.003
- Fernandez-Castane, A., Vine, C. E., Caminal, G., and Lopez-Santin, J. (2012). Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. *J. Biotechnol.* 157, 391–398. doi: 10.1016/j.jbiotec.2011.12.007
- Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N., and Timmis, K. N. (2003). Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat. Biotechnol.* 21, 1266–1267. doi: 10.1038/nbt1103-1266
- Ferrer, M., Lunsdorf, H., Chernikova, T. N., Yakimov, M., Timmis, K. N., and Golyshin, P. N. (2004). Functional consequences of single: double ring transitions in chaperonins: life in the cold. *Mol. Microbiol.* 53, 167–182. doi: 10.1111/j.1365-2958.2004.04077.x
- Fong, B. A., and Wood, D. W. (2010). Expression and purification of ELP-integrated target proteins in high cell density *E. coli* fermentation. *Microb. Cell Fact.* 9, 77. doi: 10.1186/1475-2859-9-77
- Galkin, V. E., Yu, X., Bielnicki, J., Ndjona, D., Bell, C. E., and Egelman, E. H. (2009). Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *J. Mol. Biol.* 385, 779–787. doi: 10.1016/j.jmb.2008.10.081
- Galloway, C. A., Sowden, M. P., and Smith, H. C. (2003). Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. *Biotechniques* 34, 524–526, 528, 530.
- Gao, W., Rzewski, A., Sun, H., Robbins, P. D., and Gambotto, A. (2004). UpGene: application of a web-based DNA codon optimization algorithm. *Biotechnol. Prog.* 20, 443–448. doi: 10.1021/bp0300467
- Ge, X., Yang, D. S. C., Trabbic-Carlson, K., Kim, B., Chilkoti, A., and Filipe, C. D. M. (2005). Self-cleavable stimulus responsive tags for protein purification without chromatography. *J. Am. Chem. Soc.* 127, 11228–11229. doi: 10.1021/ja0531125
- Georgiou, G., and Segatori, L. (2005). Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr. Opin. Biotechnol.* 16, 538–545. doi: 10.1016/j.copbio.2005.07.008
- Goldstein, M. A., and Doi, R. H. (1995). Prokaryotic promoters in biotechnology. *Biotechnol. Annu. Rev.* 1, 105–128. doi: 10.1016/S1387-2656(08)70049-8
- Gonzalez-Montalban, N., Garcia-Fruitos, E., and Villaverde, A. (2007). Recombinant protein solubility - does more mean better? *Nat. Biotechnol.* 25, 718–720. doi: 10.1038/nbt0707-718
- Goodman, D. B., Church, G. M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479. doi: 10.1126/science.1241934
- Gottesman, S. (1996). Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30, 465–506. doi: 10.1146/annurev.genet.30.1.465
- Graumann, K., and Premstaller, A. (2006). Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnol. J.* 1, 164–186. doi: 10.1002/biot.200500051
- Grodberg, J., and Dunn, J. J. (1988). ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* 170, 1245–1253.
- Gustafsson, C., Govindarajan, S., and Minshall, J. (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22, 346–353. doi: 10.1016/j.tibtech.2004.04.006
- Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130.
- Hammarstrom, M., Hellgren, N., Van Den Berg, S., Berglund, H., and Hard, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* 11, 313–321. doi: 10.1110/ps.22102
- Hammarstrom, M., Woestenenk, E. A., Hellgren, N., Hard, T., and Berglund, H. (2006). Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J. Struct. Funct. Genomics* 7, 1–14. doi: 10.1007/s10969-005-9003-7
- Hartl, F. U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852–1858. doi: 10.1126/science.1068408
- Hartley, D. L., and Kane, J. F. (1988). Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochem. Soc. Trans.* 16, 101–102.
- Hayashi, K., and Kojima, C. (2008). pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. *Protein Expr. Purif.* 62, 120–127. doi: 10.1016/j.pep.2008.07.007
- Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., et al. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat. Biotechnol.* 6, 1204–1210. doi: 10.1038/nbt1088-1204
- Huber, R., Scheidle, M., Dittrich, B., Klee, D., and Buchs, J. (2009). Equalizing growth in high-throughput small scale cultivations via precultures operated in fed-batch mode. *Biotechnol. Bioeng.* 103, 1095–1102. doi: 10.1002/bit.22349
- Hwang, P. M., Pan, J. S., and Sykes, B. D. (2014). Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. *FEBS Lett.* 588, 247–252. doi: 10.1016/j.febslet.2013.09.028
- Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., et al. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056–1063. doi: 10.1126/science.412251
- Jana, S., and Deb, J. K. (2005). Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 67, 289–298. doi: 10.1007/s00253-004-1814-0
- Jayaraj, S., Reid, R., and Santi, D. V. (2005). GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res.* 33, 3011–3016. doi: 10.1093/nar/gki614
- Jenny, R. J., Mann, K. G., and Lundblad, R. L. (2003). A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* 31, 1–11. doi: 10.1016/S1046-5928(03)00168-2
- Johnson, A. D., Potete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. (1981). lambda Repressor and cro – components of an efficient molecular switch. *Nature* 294, 217–223. doi: 10.1038/294217a0
- Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 6, 494–500. doi: 10.1016/0958-1669(95)80082-4
- Kapust, R. B., Tozser, J., Copeland, T. D., and Waugh, D. S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* 294, 949–955. doi: 10.1016/S0006-291X(02)00574-0
- Kapust, R. B., and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668–1674. doi: 10.1110/ps.8.8.1668
- Khan, F., Legler, P. M., Mease, R. M., Duncan, E. H., Bergmann-Leitner, E. S., and Angov, E. (2012). Histidine affinity tags affect MSP1(42) structural stability and immunodominance in mice. *Biotechnol. J.* 7, 133–147. doi: 10.1002/biot.201100331

- Khlebnikov, A., and Keasling, J. D. (2002). Effect of lacY expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers. *Biotechnol. Prog.* 18, 672–674. doi: 10.1021/bp010141k
- Klose, J., Wendt, N., Kubald, S., Krause, E., Fechner, K., Beyermann, M., et al. (2004). Hexa-histidin tag position influences disulfide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a. *Protein Sci.* 13, 2470–2475. doi: 10.1110/ps.04835904
- Korpimäki, T., Kurittu, J., and Karp, M. (2003). Surprisingly fast disappearance of beta-lactam selection pressure in cultivation as detected with novel biosensing approaches. *J. Microbiol. Methods* 53, 37–42. doi: 10.1016/S0167-7012(02)00213-0
- Kroll, J., Klintner, S., Schneider, C., Voss, I., and Steinbüchel, A. (2010). Plasmid addition systems: perspectives and applications in biotechnology. *Microb. Biotechnol.* 3, 634–657. doi: 10.1111/j.1751-7915.2010.00170.x
- Kurokawa, Y., Yanagi, H., and Yura, T. (2000). Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl. Environ. Microbiol.* 66, 3960–3965. doi: 10.1128/AEM.66.9.3960-3965.2000
- Lanzer, M., and Bujard, H. (1988). Promoters largely determine the efficiency of repressor action. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8973–8977. doi: 10.1073/pnas.85.23.8973
- LaVallie, E. R., Diblasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (N. Y.)* 11, 187–193. doi: 10.1038/nbt0293-187
- Lee, C., Kim, J., Shin, S. G., and Hwang, S. (2006). Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J. Biotechnol.* 123, 273–280. doi: 10.1016/j.jbiotec.2005.11.014
- Lee, S. Y. (1996). High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 14, 98–105. doi: 10.1016/0167-7799(96)80930-9
- Lewin, C. S., Howard, B. M., Ratcliffe, N. T., and Smith, J. T. (1989). 4-quinolones and the SOS response. *J. Med. Microbiol.* 29, 139–144. doi: 10.1099/00222615-29-2-139
- Li, G. W., Oh, E., and Weissman, J. S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541. doi: 10.1038/nature10965
- Lin-Chao, S., and Bremer, H. (1986). Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. *Mol. Gen. Genet.* 203, 143–149. doi: 10.1007/BF00330395
- Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* 11, 56. doi: 10.1186/1475-2859-11-56
- Losen, M., Frolich, B., Pohl, M., and Buchs, J. (2004). Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnol. Prog.* 20, 1062–1068. doi: 10.1021/bp034282t
- Madurawe, R. D., Chase, T. E., Tsao, E. I., and Bentley, W. E. (2000). A recombinant lipoprotein antigen against Lyme disease expressed in *E. coli*: fermentor operating strategies for improved yield. *Biotechnol. Prog.* 16, 571–576. doi: 10.1021/bp0000555
- Makhatadze, G. I., and Privalov, P. L. (1995). Energetics of protein structure. *Adv. Protein Chem.* 47, 307–425. doi: 10.1016/S0065-3233(08)60548-3
- Makoff, A. J., and Oxer, M. D. (1991). High level heterologous expression in *E. coli* using mutant forms of the lac promoter. *Nucleic Acids Res.* 19, 2417–2421. doi: 10.1093/nar/19.9.2417
- Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 60, 512–538.
- Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M., and Striedner, G. (2013). Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. *Microb. Cell Fact.* 12, 58. doi: 10.1186/1475-2859-12-58
- Martinez-Alonso, M., Gonzalez-Montalban, N., Garcia-Fruitos, E., and Villaverde, A. (2008). The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum. *Appl. Environ. Microbiol.* 74, 7431–7433. doi: 10.1128/AEM.01446-08
- Martinez-Alonso, M., Vera, A., and Villaverde, A. (2007). Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming *Escherichia coli* cells. *FEMS Microbiol. Lett.* 273, 187–195. doi: 10.1111/j.1574-6968.2007.00788.x
- McCarty, J. S., and Walker, G. C. (1991). DnaK as a thermometer: threonine-199 is site of autophosphorylation and is critical for ATPase activity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9513–9517. doi: 10.1073/pnas.88.21.9513
- McDaniel, L. E., and Bailey, E. G. (1969). Effect of shaking speed and type of closure on shake flask cultures. *Appl. Microbiol.* 17, 286–290.
- Menart, V., Jevsevar, S., Vilar, M., Trobis, A., and Pavko, A. (2003). Constitutive versus thermoinducible expression of heterologous proteins in *Escherichia coli* based on strong PR, PL promoters from phage lambda. *Biotechnol. Bioeng.* 83, 181–190. doi: 10.1002/bit.10660
- Mendoza, J. A., Dulin, P., and Warren, T. (2000). The lower hydrolysis of ATP by the stress protein GroEL is a major factor responsible for the diminished chaperonin activity at low temperature. *Cryobiology* 41, 319–323. doi: 10.1006/cryo.2000.2287
- Menzella, H. G. (2011). Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli*. *Microb. Cell Fact.* 10, 15. doi: 10.1186/1475-2859-10-15
- Menzella, H. G., Ceccarelli, E. A., and Gramajo, H. C. (2003). Novel *Escherichia coli* strain allows efficient recombinant protein production using lactose as inducer. *Biotechnol. Bioeng.* 82, 809–817. doi: 10.1002/bit.10630
- Mergulhao, F. J., Summers, D. K., and Monteiro, G. A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* 23, 177–202. doi: 10.1016/j.biotechadv.2004.11.003
- Messens, J., and Collet, J. F. (2006). Pathways of disulfide bond formation in *Escherichia coli*. *Int. J. Biochem. Cell Biol.* 38, 1050–1062. doi: 10.1016/j.biocel.2005.12.011
- Meyer, D. E., and Chilkoti, A. (1999). Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* 17, 1112–1115. doi: 10.1038/15100
- Mieschendahl, M., Petri, T., and Hanggi, U. (1986). A novel prophage independent trp regulated lambda PL expression system. *Nat. Biotechnol.* 4, 6. doi: 10.1038/nbt0986-802
- Minton, N. P. (1984). Improved plasmid vectors for the isolation of translational lac gene fusions. *Gene* 31, 269–273. doi: 10.1016/0378-1119(84)90220-8
- Miroux, B., and Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298. doi: 10.1006/jmbi.1996.0399
- Moffatt, B. A., and Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* 49, 221–227. doi: 10.1016/0092-8674(87)90563-0
- Moulton, G. G. (2013). *Fed-Batch Fermentation: A Practical Guide to Scalable Recombinant Protein Production in Escherichia coli*. Cambridge: Woodhead Publishing Limited.
- Müller-Hill, B. (1996). *The Lac Operon: A Short History of a Genetic Paradigm*. Berlin: Walter de Gruyter. doi: 10.1515/9783110879476
- Müller-Hill, B., Crapo, L., and Gilbert, W. (1968). Mutants that make more lac repressor. *Proc. Natl. Acad. Sci. U.S.A.* 59, 1259–1264. doi: 10.1073/pnas.59.4.1259
- Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M., and Nygren, P. A. (1997). Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif.* 11, 1–16. doi: 10.1006/prep.1997.0767
- Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998). Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 1694–1699.
- Nishihara, K., Kanemori, M., Yanagi, H., and Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* 66, 884–889. doi: 10.1128/AEM.66.3.884-889.2000
- Nordstrom, K. (2006). Plasmid R1 – replication and its control. *Plasmid* 55, 1–26. doi: 10.1016/j.plasmid.2005.07.002
- O’Beirne, D., and Hamer, G. (2000). Oxygen availability and the growth of *Escherichia coli* W3110: a problem exacerbated by scale-up. *Bioprocess Eng.* 23, 487–494. doi: 10.1007/s004499900185
- Ohana, R. F., Encell, L. P., Zhao, K., Simpson, D., Slater, M. R., Urh, M., et al. (2009). HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr. Purif.* 68, 110–120. doi: 10.1016/j.pep.2009.05.010

- Oresic, M., and Shalloway, D. (1998). Specific correlations between relative synonymous codon usage and protein secondary structure. *J. Mol. Biol.* 281, 31–48. doi: 10.1006/jmbi.1998.1921
- Ou, J., Wang, L., Ding, X., Du, J., Zhang, Y., Chen, H., et al. (2004). Stationary phase protein overproduction is a fundamental capability of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 314, 174–180. doi: 10.1016/j.bbrc.2003.12.077
- Palomares, L. A., Estrada-Mondaca, S., and Ramirez, O. T. (2004). Production of recombinant proteins: challenges and solutions. *Methods Mol. Biol.* 267, 15–52. doi: 10.1385/1-59259-774-2:015
- Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994). Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* 216, 413–417. doi: 10.1006/abio.1994.1060
- Pedersen, S. (1984). *Escherichia coli* ribosomes translate in vivo with variable rate. *EMBO J.* 3, 2895–2898.
- Perron-Savard, P., De Crescenzo, G., and Le Moual, H. (2005). Dimerization and DNA binding of the *Salmonella enterica* PhoP response regulator are phosphorylation independent. *Microbiology* 151, 3979–3987. doi: 10.1099/mic.0.28236-0
- Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., et al. (2010). Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microb. Cell Fact.* 9, 65. doi: 10.1186/1475-2859-9-65
- Pope, B., and Kent, H. M. (1996). High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Res.* 24, 536–537. doi: 10.1093/nar/24.3.536
- Porath, J., and Olin, B. (1983). Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron and nickel ions. *Biochemistry* 22, 1621–1630. doi: 10.1021/bi00276a015
- Postma, P. W., and Lengeler, J. W. (1985). Phosphoenolpyruvate: carbohydrate phosphotransferase system of bacteria. *Microbiol. Rev.* 49, 232–269.
- Puigbo, P., Guzman, E., Romeu, A., and Garcia-Vallve, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* 35, W126–W131. doi: 10.1093/nar/gkm219
- Qing, G., Ma, L. C., Khorchid, A., Swapna, G. V., Mal, T. K., Takayama, M. M., et al. (2004). Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat. Biotechnol.* 22, 877–882. doi: 10.1038/nbt984
- Rais-Beghdadi, C., Roggero, M. A., Fasel, N., and Reymond, C. D. (1998). Purification of recombinant proteins by chemical removal of the affinity tag. *Appl. Biochem. Biotechnol.* 74, 95–103. doi: 10.1007/BF02787176
- Raran-Kurussi, S., and Waugh, D. S. (2012). The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS ONE* 7:e49589. doi: 10.1371/journal.pone.0049589
- Roberts, M. C. (1996). Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiol. Rev.* 19, 1–24. doi: 10.1111/j.1574-6976.1996.tb00251.x
- Rosano, G. L., Bruch, E. M., and Ceccarelli, E. A. (2011). Insights into the Clp/HSP100 chaperone system from chloroplasts of *Arabidopsis thaliana*. *J. Biol. Chem.* 286, 29671–29680. doi: 10.1074/jbc.M110.211946
- Rosano, G. L., and Ceccarelli, E. A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb. Cell Fact.* 8, 41. doi: 10.1186/1475-2859-8-41
- Routledge, S. (2012). Beyond de-foaming: the effects of antifoams on bioprocess productivity. *Comput. Struct. Biotechnol. J.* 3, 1–7. doi: 10.5936/csbj.2012.10014
- Routledge, S. J., Hewitt, C. J., Bora, N., and Bill, R. M. (2011). Antifoam addition to shake flask cultures of recombinant *Pichia pastoris* increases yield. *Microb. Cell Fact.* 10, 17. doi: 10.1186/1475-2859-10-17
- Rudolph, R., and Lilie, H. (1996). In vitro folding of inclusion body proteins. *FASEB J.* 10, 49–56.
- Sahdev, S., Khattar, S. K., and Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol. Cell. Biochem.* 307, 249–264. doi: 10.1007/s11010-007-9603-6
- Saida, F., Uzan, M., Odaert, B., and Bontems, F. (2006). Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr. Protein Pept. Sci.* 7, 47–56. doi: 10.2174/138920306775474095
- San-Miguel, T., Perez-Bermudez, P., and Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in is favoured in early log-phase cultures induced at low temperature. *Springerplus* 2, 89. doi: 10.1186/2193-1801-2-89
- Scheidle, M., Dittrich, B., Klinger, J., Ikeda, H., Klee, D., and Buchs, J. (2011). Controlling pH in shake flasks using polymer-based controlled-release discs with pre-determined release kinetics. *BMC Biotechnol.* 11:25. doi: 10.1186/1472-6750-11-25
- Schein, C., and Noteborn, M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Nat. Biotechnol.* 6, 3. doi: 10.1038/nbt0388-291
- Schellman, J. A. (1997). Temperature, stability, and the hydrophobic interaction. *Biophys. J.* 73, 2960–2964. doi: 10.1016/S0006-3495(97)78324-3
- Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D., and Beckwith, J. (2003). The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. *J. Bacteriol.* 185, 5706–5713. doi: 10.1128/JB.185.19.5706-5713.2003
- Schleif, R. (2000). Regulation of the l-arabinose operon of *Escherichia coli*. *Trends Genet.* 16, 559–565. doi: 10.1016/S0168-9525(00)02153-3
- Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* 34, 779–796. doi: 10.1111/j.1574-6976.2010.00226.x
- Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* 189, 8746–8749. doi: 10.1128/JB.01368-07
- Shatzman, A. R., Gross, M. S., and Rosenberg, M. (2001). Expression using vectors with phage lambda regulatory sequences. *Curr. Protoc. Mol. Biol.* Chapter 16, Unit 16.13. doi: 10.1002/0471142727.mb1603s11
- Shaw, W. V. (1983). Chloramphenicol acetyltransferase: enzymology and molecular biology. *CRC Crit. Rev. Biochem.* 14, 1–46. doi: 10.3109/10409238309102789
- Shih, Y. P., Kung, W. M., Chen, J. C., Yeh, C. H., Wang, A. H., and Wang, T. F. (2002). High-throughput screening of soluble recombinant proteins. *Protein Sci.* 11, 1714–1719. doi: 10.1110/ps.0205202
- Shiloach, J., and Fass, R. (2005). Growing *E. coli* to high cell density – a historical perspective on method development. *Biotechnol. Adv.* 23, 345–357. doi: 10.1016/j.biotechadv.2005.04.004
- Shur, O., Dooley, K., Blenner, M., Baltimore, M., and Banta, S. (2013). A designed, phase changing RTX-based peptide for efficient bioseparations. *Biotechniques* 54, 197–198, 200, 202, 204, 206. doi: 10.2144/000114010
- Siegele, D. A., and Hu, J. C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8168–8172. doi: 10.1073/pnas.94.15.8168
- Silverstone, A. E., Arditti, R. R., and Magasanik, B. (1970). Catabolite-insensitive revertants of lac promoter mutants. *Proc. Natl. Acad. Sci. U.S.A.* 66, 773–779. doi: 10.1073/pnas.66.3.773
- Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* 18, 936–948. doi: 10.1002/pro.102
- Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67, 31–40. doi: 10.1016/0378-1119(88)90005-4
- Soares, C. R., Gomide, F. I., Ueda, E. K., and Bartolini, P. (2003). Periplasmic expression of human growth hormone via plasmid vectors containing the lambdaPL promoter: use of HPLC for product quantification. *Protein Eng.* 16, 1131–1138. doi: 10.1093/protein/gzq114
- Sorensen, H. P., and Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115, 113–128. doi: 10.1016/j.biotech.2004.08.004
- Stano, N. M., and Patel, S. S. (2004). T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *J. Biol. Chem.* 279, 16136–16143. doi: 10.1074/jbc.M400139200
- Stevens, R. C. (2000). Design of high-throughput methods of protein production for structural biology. *Structure* 8, R177–R185. doi: 10.1016/S0969-2126(00)00193-3

- Stewart, E. J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *EMBO J.* 17, 5543–5550. doi: 10.1093/emboj/17.19.5543
- Stoker, N. G., Fairweather, N. F., and Spratt, B. G. (1982). Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene* 18, 335–341. doi: 10.1016/0378-1119(82)90172-X
- Strocchi, M., Ferrer, M., Timmis, K. N., and Golyshin, P. N. (2006). Low temperature-induced systems failure in *Escherichia coli*: insights from rescue by cold-adapted chaperones. *Proteomics* 6, 193–206. doi: 10.1002/pmic.200500031
- Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234. doi: 10.1016/j.pep.2005.01.016
- Studier, F. W. (2014). Stable expression clones and auto-induction for protein production in *E. coli*. *Methods Mol. Biol.* 1091, 17–32. doi: 10.1007/978-1-62703-691-7_2
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130. doi: 10.1016/0022-2836(86)90385-2
- Summers, D. K., Beton, C. W., and Withers, H. L. (1993). Multicopy plasmid instability: the dimer catastrophe hypothesis. *Mol. Microbiol.* 8, 1031–1038. doi: 10.1111/j.1365-2958.1993.tb01648.x
- Supek, F., and Vlahovick, K. (2004). INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics* 20, 2329–2330. doi: 10.1093/bioinformatics/bth238
- Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533. doi: 10.1007/s00253-002-1158-6
- Thanaraj, T. A., and Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* 5, 1594–1612. doi: 10.1002/pro.5560050814
- Tropea, J. E., Cherry, S., and Waugh, D. S. (2009). Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol.* 498, 297–307. doi: 10.1007/978-1-59745-196-3_19
- Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., and Nussinov, R. (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.* 383, 281–291. doi: 10.1016/j.jmb.2008.08.012
- Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* 28, 1–8. doi: 10.1016/S1046-5928(02)00641-1
- Umezawa, H. (1979). Studies on aminoglycoside antibiotics: enzymic mechanism of resistance and genetics. *Jpn. J. Antibiot.* 32(Suppl.), S1–S14.
- Unden, G., Becker, S., Bongaerts, J., Holighaus, G., Schirawski, J., and Six, S. (1995). O₂-sensing and O₂-dependent gene regulation in facultatively anaerobic bacteria. *Arch. Microbiol.* 164, 81–90. doi: 10.1007/s002030050238
- Valdez-Cruz, N. A., Caspeta, L., Perez, N. O., Ramirez, O. T., and Trujillo-Roldan, M. A. (2010). Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb. Cell Fact.* 9, 18. doi: 10.1186/1475-2859-9-18
- Valent, Q. A., Scotti, P. A., High, S., De Gier, J. W., Von Heijne, G., Lentzen, G., et al. (1998). The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J.* 17, 2504–2512. doi: 10.1093/emboj/17.9.2504
- Vasina, J. A., and Baneyx, F. (1997). Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* cspA and tac promoter systems. *Protein Expr. Purif.* 9, 211–218. doi: 10.1006/prep.1996.0678
- Vasina, J. A., Peterson, M. S., and Baneyx, F. (1998). Scale-up and optimization of the low-temperature inducible cspA promoter system. *Biotechnol. Prog.* 14, 714–721. doi: 10.1021/bp980061p
- Vera, A., Gonzalez-Montalban, N., Aris, A., and Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.* 96, 1101–1106. doi: 10.1002/bit.21218
- Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol.* 153, 3–11. doi: 10.1016/0076-6879(87)53044-0
- Voss, I., and Steinbuchel, A. (2006). Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16. *Metab. Eng.* 8, 66–78. doi: 10.1016/j.ymben.2005.09.003
- Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., et al. (2008). Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14371–14376. doi: 10.1073/pnas.0804090105
- Wang, R. F., and Kushner, S. R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* 100, 195–199. doi: 10.1016/0378-1119(91)90366-J
- Wanner, B. L., Kodaira, R., and Neidhardt, F. C. (1978). Regulation of lac operon expression: reappraisal of the theory of catabolite repression. *J. Bacteriol.* 136, 947–954.
- Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* 80, 283–293. doi: 10.1016/j.pep.2011.08.005
- Weickert, M. J., Pagratis, M., Glascock, C. B., and Blackmore, R. (1999). A mutation that improves soluble recombinant hemoglobin accumulation in *Escherichia coli* in heme excess. *Appl. Environ. Microbiol.* 65, 640–647.
- Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshall, J., et al. (2009). Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* 4:e7002. doi: 10.1371/journal.pone.0007002
- Weuster-Botz, D., Altenbach-Rehm, J., and Arnold, M. (2001). Parallel substrate feeding and pH-control in shaking-flasks. *Biochem. Eng. J.* 7, 163–170. doi: 10.1016/S1369-703X(00)00117-0
- Wild, J., Hradecna, Z., and Szybalski, W. (2002). Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* 12, 1434–1444. doi: 10.1101/gr.130502
- Winkler, H. H., and Wilson, T. H. (1967). Inhibition of beta-galactoside transport by substrates of the glucose transport system in *Escherichia coli*. *Biochim. Biophys. Acta* 135, 1030–1051. doi: 10.1016/0005-2736(67)90073-9
- Wu, J., and Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochim. Pol.* 46, 591–599.
- Yamanè, T., and Shimizu, S. (1984). “Fed-batch techniques in microbial processes,” in *Bioprocess Parameter Control*, ed. P. Agrawal (Berlin: Springer), 147–194. doi: 10.1007/BFb0006382
- Yang, Q., Xu, J., Li, M., Lei, X., and An, L. (2003). High-level expression of a soluble snake venom enzyme, glosheobin, in *E. coli* in the presence of metal ions. *Biotechnol. Lett.* 25, 607–610. doi: 10.1023/A:1023067626846
- Yee, L., and Blanch, H. W. (1992). Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Biotechnology (N. Y.)* 10, 1550–1556. doi: 10.1038/nbt1292-1550
- Yona, A. H., Bloom-Ackermann, Z., Frumkin, I., Hanson-Smith, V., Chrapak-Amikam, Y., Feng, Q., et al. (2013). tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* 2:e01339. doi: 10.7554/eLife.01339
- Zielenkiewicz, U., and Ceglowski, P. (2001). Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. *Acta Biochim. Pol.* 48, 1003–1023.

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

Received: 20 December 2013; accepted: 29 March 2014; published online: 17 April 2014.

Citation: Rosano GL and Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5:172. doi: 10.3389/fmicb.2014.00172

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Rosano and Ceccarelli. 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.



Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system

Sofia Costa^{1,2}, André Almeida³, António Castro² and Lucília Domingues^{1*}

¹ Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Braga, Portugal

² Instituto Nacional de Saúde Dr. Ricardo Jorge, Porto, Portugal

³ Hitag Biotechnology, Lda., Biocant, Parque Tecnológico de Cantanhede, Cantanhede, Portugal

Edited by:

Germán Leandro Rosano, Instituto de Biología Molecular y Celular de Rosario, Argentina

Reviewed by:

Grzegorz Węgrzyn, University of Gdansk, Poland

Helena Berglund, Karolinska Institutet, Sweden

*Correspondence:

Lucília Domingues, Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
e-mail: luciliad@deb.uminho.pt

Proteins are now widely produced in diverse microbial cell factories. The *Escherichia coli* is still the dominant host for recombinant protein production but, as a bacterial cell, it also has its issues: the aggregation of foreign proteins into insoluble inclusion bodies is perhaps the main limiting factor of the *E. coli* expression system. Conversely, *E. coli* benefits of cost, ease of use and scale make it essential to design new approaches directed for improved recombinant protein production in this host cell. With the aid of genetic and protein engineering novel tailored-made strategies can be designed to suit user or process requirements. Gene fusion technology has been widely used for the improvement of soluble protein production and/or purification in *E. coli*, and for increasing peptide's immunogenicity as well. New fusion partners are constantly emerging and complementing the traditional solutions, as for instance, the Fh8 fusion tag that has been recently studied and ranked among the best solubility enhancer partners. In this review, we provide an overview of current strategies to improve recombinant protein production in *E. coli*, including the key factors for successful protein production, highlighting soluble protein production, and a comprehensive summary of the latest available and traditionally used gene fusion technologies. A special emphasis is given to the recently discovered Fh8 fusion system that can be used for soluble protein production, purification, and immunogenicity in *E. coli*. The number of existing fusion tags will probably increase in the next few years, and efforts should be taken to better understand how fusion tags act in *E. coli*. This knowledge will undoubtedly drive the development of new tailored-made tools for protein production in this bacterial system.

Keywords: *Escherichia coli*, fusion tags, soluble production, protein purification, tag removal, Fh8 tag, H tag, protein immunogenicity

OUTLINE

Proteins are key elements of life, constituting the major part of the living cell. They play important roles in a variety of cell processes, including cell signaling, immune responses, cell adhesion, and the cell cycle, and their failure is consequently correlated with several diseases.

With the introduction of the DNA recombinant technology in the 1970s, proteins started to be expressed in several host organisms resulting in a faster and easier process compared to their natural sources (Demain and Vaishnav, 2009). *Escherichia coli* remains the dominant host for producing recombinant proteins, owing to its advantageous fast and inexpensive, and high yield protein production, together with the well-characterized genetics and variety of available molecular tools (Demain and Vaishnav, 2009).

The recombinant protein production in *E. coli* has greatly contributed for several structural studies; for instance, about 90% of the structures available in the Protein Data Bank were determined on proteins produced in *E. coli*. (Nettlehip et al., 2010; Bird, 2011). The *E. coli* recombinant production has also boosted the biopharmaceutical industry: 30% of the recombinant biopharmaceuticals

licensed up to 2011 by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA) were obtained using this host cell (Ferrer-Miralles et al., 2009; Walsh, 2010; Berlec and Strukelj, 2013).

Escherichia coli recombinant protein-based products can also be found in major sectors of the enzyme industry and the agricultural industry with applications ranging from catalysis (e.g., washing detergents) and therapeutic use (e.g., vaccine development) to functional analysis and structure determination (e.g., crystallography; Demain and Vaishnav, 2009).

As a bacterial system, the *E. coli* has, however, limitations at expressing more complex proteins due to the lack of sophisticated machinery to perform posttranslational modifications, resulting in poor solubility of the protein of interest that are produced as inclusion bodies (Demain and Vaishnav, 2009; Kamionka, 2011). Previous studies (Bussow et al., 2005; Pacheco et al., 2012) reported that up to 75% of human proteins are successfully expressed in *E. coli* but only 25% are produced in an active soluble form using this host system. Other problems found within this host system include proper formation of disulfide bonds, absence of chaperones for the correct folding, and the miss-match between the

codon usage of the host cell and the protein of interest (Terpe, 2006; Demain and Vaishnav, 2009; Pacheco et al., 2012). Moreover, the industrial culture of *E. coli* leads cells to grow in harsh conditions, resulting in cell physiology deterioration (Chou, 2007; Pacheco et al., 2012).

Despite the above-mentioned issues of *E. coli* recombinant protein production, the benefits of cost and ease of use and scale make it essential to design new strategies directed for recombinant soluble protein production in this host cell. Several strategies have been made for efficient production of proteins in *E. coli*, namely, the use of different mutated host strains, co-production of chaperones and foldases, lowering cultivation temperatures, and addition of a fusion partner (Terpe, 2006; Demain and Vaishnav, 2009). The combination of some of these strategies has improved the soluble production of recombinant proteins in *E. coli*, but the prediction of robust soluble protein production processes is still a “a challenge and a necessity” (Jana and Deb, 2005).

Nowadays, with the aid of genetic and protein engineering, novel tailor-made strategies can be designed to suit user or process requirements.

This review describes the key solubility factors that correlate with successful protein production in *E. coli*, and it presents a comprehensive summary of the available fusion partners for protein production and purification in the bacterial host. A main focus is given to the novel Fh8 fusion system (Hitag[®]) for soluble protein production, purification and immunogenicity in *E. coli* (Costa, 2013).

SOLUBLE PROTEIN PRODUCTION IN *ESCHERICHIA COLI*

The production of recombinant proteins requires a successful correlation between the gene's expression, protein solubility, and its purification (Esposito and Chatterjee, 2006). The production levels of recombinant proteins synthesized in *E. coli* are no longer pointed as a limitation for the success of the overall process, but care should be taken with the protein solubility, which is still a major bottleneck in the field. The downstream processing is deeply associated with an efficient protein production strategy, and thus it must be tailor-designed to maximize the recovery of pure recombinant proteins.

All these three properties – expression, solubility, and purification – shall always be considered together as determinants for the effective protein production in *E. coli*. Several aspects are though essential for each individual success, as resumed in **Figure 1** and described.

STRATEGIES FOR THE SUCCESSFUL AND EFFICIENT SOLUBLE PROTEIN PRODUCTION IN *E. COLI* – PREVENTION OF PROTEIN AGGREGATION

Escherichia coli recombinant protein production systems are designed to achieve a high accumulation of soluble protein product in the bacterial cell. However, a strong and rapid protein production can lead to stressful situations for the host cell, resulting in protein misfolding *in vivo*, and consequent aggregation into inclusion bodies (Schumann and Ferreira, 2004; Sorensen and Mortensen, 2005a,b; Sevastyanovich et al., 2010). For instance, macromolecular crowding of proteins at high concentrations in the *E. coli* cytoplasm often impairs the correct folding of proteins, leading to the formation of folding intermediates that,

when inefficiently processed by molecular chaperones, promote inclusion body formation (Sorensen and Mortensen, 2005a,b).

Strategies that direct the soluble production of proteins in *E. coli* are, thus, envisaged, and become more attractive than protein refolding procedures from inclusion bodies.

Several methods have been shown to prevent or decrease protein aggregation during protein production in *E. coli* on a trial-and-error basis, including:

- (i) **Lower expression temperatures:** bacteria cultivation at reduced temperatures is often used to reduce protein aggregation, since it slows down the rate of protein synthesis and folding kinetics, decreasing the hydrophobic interactions that are involved in protein self-aggregation (Schumann and Ferreira, 2004; Sorensen and Mortensen, 2005b). Low cultivation temperatures can also reduce or impair protein degradation due to a poor activity of heat shock proteases that are usually induced during protein overproduction in *E. coli* (Cheshyre and Hipkiss, 1989). This strategy has, however, some drawbacks as the reduction of temperature can also affect replication, transcription, and translation rates, besides decreasing the bacterial growth and protein production yields. Nevertheless, these limitations can be circumvented by the use of cold-inducible promoters that maximize protein production under low temperature conditions (Mujacic et al., 1999).
- (ii) ***E. coli*-engineered host strains:** *E. coli* mutant strains are a significant advance toward the soluble production of difficult recombinant proteins. Several targeted strain strategies have been developed through the introduction of DNA mutations that affect protein synthesis, degradation, secretion, or folding (reviewed in Makino et al., 2011), including: engineered strains for improved protein processing at low temperatures, such as the Arctic Express strain (Agilent Technologies); mutated strains that increase mRNA stability by attenuation of RNases activity, which is responsible for the shorter half-life of mRNA in *E. coli* cells (Lopez et al., 1999); engineered strains that supply extra copies of rare tRNAs, such as the Rosetta strains (Invitrogen) and the BL21 Codon Plus strains (Novagen; Baca and Hol, 2000; Sorensen et al., 2003b); mutant strains that facilitate disulfide bond formation and protein folding in the *E. coli* cytoplasm by render it oxidizing due to mutations in glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) genes, and/or by co-production of Dsb proteins (Bessette et al., 1999; Lobstein et al., 2012), such as the Origami strains (Novagen) or the new SHuffle strain (New England Biolabs; Lobstein et al., 2012); and C41 and C43 (Avidis) BL21 (DE3) mutant strains that improve the synthesis of membrane proteins (Miroux and Walker, 1996).
- (iii) **Cultivation conditions:** protein production can be efficiently improved by the use of high cell-density culture systems like batch, which offers a limited control of the cell growth, and fed-batch, which allows the real time optimization of growth conditions (Sorensen and Mortensen, 2005b). The composition of the cell growth medium and the fermentation variables such as temperature, pH, induction time, and inducer concentration are also essential for the prevention of protein aggregation, whereby a careful optimization improves the

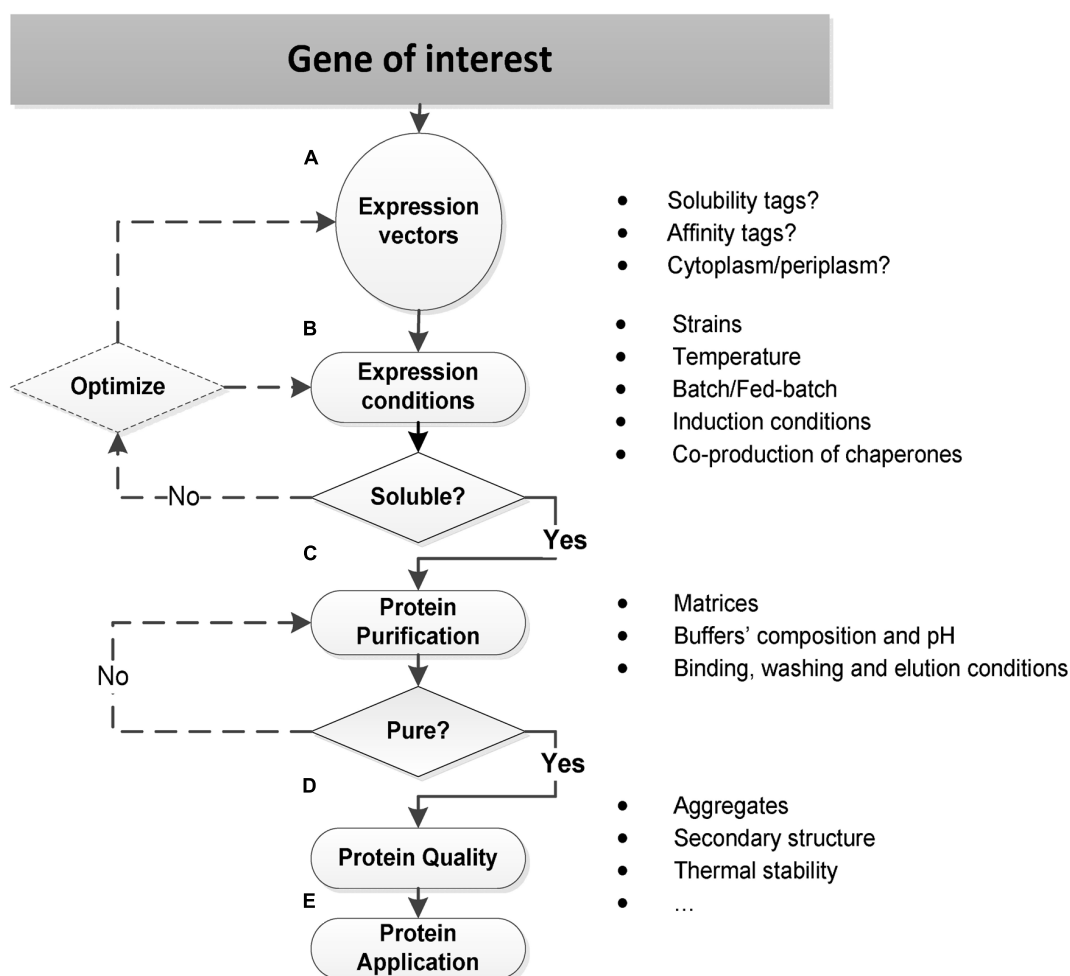


FIGURE 1 | Strategies for soluble protein production in *E. coli*.

(A) Expression vectors should be carefully selected in order to incorporate specific features that affect the protein production in *E. coli* such as solubility and/or affinity fusion tags, and to direct the protein synthesis to the *E. coli* cytoplasm or periplasm. Other important features include: the replicon, antibiotic-resistance markers, and transcriptional promoters (Jana and Deb, 2005; Sorensen and Mortensen, 2005a). (B) The optimization of expression conditions often directs the soluble protein production in *E. coli*, and it relies on a trial-and-error basis: to get a soluble TP, it may require the selection and testing of several engineered *E. coli* strains and cultivation conditions, and sometimes the initial expression vector has also to be re-designed. (C) The protein purification strategy should already be

defined at the beginning when selecting the expression vector: if an affinity tag is incorporated, then a first affinity chromatography step should be conducted. On the other hand, if an affinity tag is prohibitive, other strategies, namely, ion exchange, size exclusion, or hydrophobic interaction chromatography should be tested. After the first purification step, the TP may or may not be sufficiently pure. When it is not pure, further purification steps with other chromatographic strategies need to be conducted. (D–E) The protein quality is an essential requirement for many structural and functional application studies: a purified soluble protein may be aggregated, without a defined secondary structure, and it may also present a low thermal stability. Therefore, a biophysical characterization is often required before proceeding to the final protein's application.

yield and quality of soluble protein production (Jana and Deb, 2005).

- (iv) *Co-production of molecular chaperones and folding modulators*: the initial folding of proteins can be assisted by molecular chaperones that prevent protein aggregation through binding exposed hydrophobic patches on unfolded, partially folded or misfolded polypeptides, and traffic molecules to their sub-cellular destination. Protein aggregation is also prevented by folding catalysts that catalyze important events in protein folding such as the disulfide bond formation (Kolaj et al., 2009). A low concentration of these folding modulators in the cell often

results in protein folding failures; thereby their co-production together with the target protein becomes a suitable strategy for the improvement of soluble protein production in *E. coli* (reviewed in Thomas et al., 1997; Schlieker et al., 2002; Baneyx and Palumbo, 2003; Hoffmann and Rinas, 2004; Betiku, 2006; Gasser et al., 2008; Kolaj et al., 2009). Chaperones like trigger factor, DnaK, GroEL, members of the heat shock protein Hsp70 and Hsp60 families (hsHsp proteins), and ClpB assist protein folding in the *E. coli* cytoplasm, and their individual or cooperative activities presents different contributions for target protein solubility (Nishihara et al., 1998;

Kuczynska-Wisnik et al., 2002; Schlieker et al., 2002; Deuring et al., 2003; de Marco and De Marco, 2004; de Marco et al., 2007).

- (v) *Fusion partner proteins*: in contrast to the above-mentioned strategies, the use of fusion partners involves the target protein engineering. Fusion partners are very stable peptide or protein molecules soluble expressed in *E. coli* that are genetically linked with target proteins to mediate their solubility and purification.

CHROMATOGRAPHIC STRATEGIES FOR RECOMBINANT PROTEIN PURIFICATION

The protein purification accounts for most of the expenses in recombinant protein production. Hence, the design of a straightforward and cost-effective protein isolation and purification is one of the first steps to be considered in the production strategy.

There is no single or simple way to purify all kinds of proteins because of their diversity and different properties. Therefore, several strategies have been developed in the past decades to address a broad range of samples. With the introduction of recombinant DNA technology in the seventies, novel affinity tagging methodologies have revolutionized protein purification processes and several easy-to-use affinity tags have emerged since then. Besides the isolation of recombinant proteins, the purification process is also used to concentrate the desired protein. The target protein is usually first designed to be affinity tagged, thus facilitating the purification process and allowing the target protein to maintain its properties without interacting directly with a matrix. However, if the target protein cannot be affinity tagged or if further purification is needed, other purification strategies are added to the process.

When designing a purification strategy, one must consider the final goal of the target protein to be purified. For instance, recombinant proteins for therapeutic and biomedical applications require a high-level of protein purity and they probably should undergo several subsequent purification steps.

The available protein purification methodologies separate the target proteins according to differences between the properties of the protein to be purified and properties of the rest of the protein mixture. Recombinant proteins are nowadays purified using column chromatography in scales from micrograms or milligrams in research laboratories to kilograms in industrial settings. The purification of a target protein from a crude cell extract is, however, not always easy and even with all the progresses achieved so far, additional physicochemical-based chromatography methods such as size exclusion (SEC), ion exchange (IEX), and hydrophobic interaction (HIC) are often used to complement the affinity tagging. These methods rely on minor differences between various proteins properties such as size, charge, and hydrophobicity, respectively (GE Healthcare, 2010).

In a traditional purification pipeline, the chromatography starts with a capturing step, where the target protein binds to the absorbent while the impurities do not. Then, weakly bound proteins are washed out of the column, and conditions are changed so that the target protein is eluted from the column.

Size exclusion chromatography

This technique is a non-binding method that separates protein samples with different molecular sizes under mild conditions. Size exclusion chromatography (SEC) can be used for protein purification, in which it usually dilutes the sample, or for group separation, which is mainly used for desalting and buffer exchange of samples. This technique is ideal for the final polishing in a multiple-step purification strategy. Analytical SEC allows the determination of the hydrodynamic radius of protein molecules and the corresponding molecular weight (GE Healthcare, 2010).

Ion exchange chromatography

This technique separates proteins with different surface charge and it offers a high-resolution separation combined with high sample loading capacity. The purification relies on a reversible interaction between a charged protein and an oppositely charged chromatography medium. Proteins purified by ion exchange chromatography (IEX) are usually obtained in a concentrated form. The net surface charge of proteins is influenced by the surrounding pH: when the pH is above the protein isoelectric point (pI), the target protein has a negatively charged shield that is used for binding to a positively charged anion exchanger; when the pH is below its pI, the target protein has a positively charged shield that is used for binding to a negatively charged cation exchanger. The IEX purification protocol is initiated under low ionic strength, and the conditions are then changed so that the bound substances can be eluted differentially by increasing salt concentration or changing pH using a gradient or stepwise strategy. In general, the IEX is used to bind the target protein, but it can also be used to bind impurities when required. The IEX is the most common technique used for the capture step in a multiple-step purification strategy, but it can be used in the intermediate step as well (GE Healthcare, 2010).

Hydrophobic interaction chromatography

Hydrophobic interaction chromatography (HIC) separates proteins according to differences in their surface hydrophobicity by using a reversible interaction between non-polar regions on the surface of these proteins and the immobilized hydrophobic ligands of a HIC medium (Queiroz et al., 2001). The proteins are separated according to differences in the amount of exposed hydrophobic amino acids. This technique is ideal for capture and intermediate steps in a multiple-step purification strategy.

The interaction between hydrophobic proteins and a HIC medium is influenced significantly by several parameters (reviewed in Queiroz et al., 2001; Lienqueo et al., 2007), including:

- (i) *The type of the ligand and degree of substitution*: the type of immobilized ligand (alkyl or aryl) determines the protein adsorption selectivity of the HIC adsorbent. In general, alkyl ligands show more pure hydrophobic character than aryl ligands. The protein binding capacities of HIC adsorbents increase with increased degree of substitution of immobilized ligand. The degree of substitution is the average number of substituent groups attached per milliliter of gel, and it correlates with the protein binding capacities of HIC adsorbents as follows: higher binding capacities are obtained

with an increased degree of substitution of immobilized ligand. At a reasonably high degree of ligand substitution, the apparent binding capacity of the adsorbent remains constant (the plateau is reached) but the strength of the interaction increases. Solutes bound under such circumstances are difficult to elute due to multi-point attachment (GE Healthcare, 2006).

- (ii) *The type of base matrix*: the matrix should be neutral to avoid ionic interactions between the protein and the matrix, and it should also be hydrophilic. The two most widely used matrices are strongly hydrophilic carbohydrates, such as cross-linked agarose, or synthetic copolymer materials (GE Healthcare, 2006).
- (iii) *The type and concentration of salt*: a high salt concentration enhances the interaction, while lowering the salt concentration weakens the interaction. The effect of the salt type on protein retention follows the Hofmeister series for the precipitation of proteins from aqueous solutions (Collins and Washabaugh, 1985; Zhang and Cremer, 2006). In Hofmeister series, the chaotropic salts (magnesium sulfate and magnesium chloride) randomize the structure of the liquid water and thus tend to decrease the strength of hydrophobic interactions. In contrast, the kosmotropic salts (sodium, potassium, or ammonium sulfates) promote hydrophobic interactions and protein precipitation, due to higher “salting-out” or molar surface tension increment effects.
- (iv) *pH*: when pH is close to a protein’s pI, net charge is zero and hydrophobic interactions are maximum, due to the minimum electrostatic repulsion between the protein molecules allowing them to get closer. In general, an increase in the pH weakens the hydrophobic interaction probably due to an increased titration of charged groups, thereby leading to an increase of protein hydrophilicity. A decrease of the pH may result in an increase of hydrophobic interactions. However, the effect of pH in HIC is not always straightforward (GE Healthcare, 2006).
- (v) *Temperature*: the role of temperature in HIC is complex, but in general, increased temperatures enhance the protein retention. Careful should be taken when conducting protein purifications at room temperature as the protein performance in the HIC will probably not be reproducible in a cold room, and vice versa.
- (vi) *Additives*: low concentrations of water-miscible alcohols, detergents, and aqueous solutions of chaotropic (“salting-in”) salts result in a weakening of the protein–ligand interactions in HIC leading to the desorption of the bound solutes. The non-polar parts of alcohols and detergents compete with the bound proteins for the adsorption sites on the HIC media resulting in the displacement of the latter. Chaotropic salts affect the ordered structure of water and/or that of the bound proteins. Both types of additives also decrease the surface tension of water thus weakening the hydrophobic interactions to give a subsequent dissociation of the ligand–solute complex. The use of additives should be carefully considered as they might compromise the target protein structure and activity (GE Healthcare, 2006, 2010).

Proteins bound to HIC media can be eluted using some of the above-mentioned conditions such as reduced salt concentration, increased pH, or addition of alcohols or detergents (Lienqueo et al., 2007), but trial-and-error experiments should be conducted to select the best option for each specific target protein.

Besides protein purification, the HIC methodology offers several potentialities in protein production, being described as one of the most used strategies for endotoxin clearance (Wilson et al., 2001; Magalhães et al., 2007; Ongkudon et al., 2012). It can also be used for protein refolding (Hwang et al., 2010).

The HIC methodology has been applied for the purification of calcium-binding proteins (CaBPs; Rozanas, 1998; Shimizu et al., 2003; McCluskey et al., 2007). These proteins expose a large hydrophobic surface in the presence of calcium that can absorb to hydrophobic matrices such as phenyl sepharose, even in the presence of low salt concentration. Most of the contaminant proteins will not bind under these conditions, which benefits the recovery of a pure CaBP. The elution step is often achieved by removal of the bound calcium through the use of chelating agents like EDTA (Rozanas, 1998).

Affinity chromatography

This technique separates proteins through a reversible interaction between the target protein and a specific ligand attached to a chromatographic matrix. The interaction can be performed via an antibody (biospecific interaction), or via an immobilized metal ion (non-biospecific interaction) or dye substance. The affinity chromatography usually offers high selectivity and resolution together with an intermediate-high capacity. The sample is first bound to the ligand using favorable conditions for that binding. Then, the unbound material is washed out of the column and the elution of pure protein is achieved using a competitive ligand or by changing the pH, ionic strength or polarity (GE Healthcare, 2010). This purification strategy can profit from the use of recombinant DNA technology as the affinity tag can be fused to the protein of interest during cloning and it is further presented in the next section.

FUSION PROTEIN TECHNOLOGY

Fusion partners or tags are used in *E. coli* to improve protein production yields, solubility and folding, and to facilitate protein purification. They can also confer specific properties for target proteins characterization and study, such as protein immunodetection, quantification, and structural and interaction studies (Malhotra, 2009). Fusion partners can also be of use when producing toxic proteins. An example is the production of antimicrobial peptides (AMPs) by *E. coli* using cellulose binding modules as fusion partner (Guerreiro et al., 2008; Ramos et al., 2010, 2013). The use of carbohydrate-binding modules (CBMs) as fusion partner has also been applied for targeting peptides and/or functionalizing specific supports/biomaterials for biomedical applications (Moreira et al., 2008; Andrade et al., 2010a,b; Pértile et al., 2012). Besides the fusion(s) partner(s) coding gene, *E. coli* expression vectors can contain a protease recognition sequence between the fusion partner coding gene and the passenger protein coding gene that allows the tag removal when the latter protein is for using in protein therapies, vaccine development and structural analyses.

Some fusion partners also protect target proteins from degradation by promoting the translocation of the passenger protein to different cellular locations, where less protease content exists (Butt et al., 2005). Both maltose-binding protein (MBP) and small ubiquitin related modifier (SUMO) fusion partners present this feature, passing target proteins from the *E. coli* cytosol for cell membrane and nucleus, respectively (Nikaido, 1994; Kishi et al., 2003).

When designing a fusion strategy, the choice of the fusion partner depends on several aspects (Young et al., 2012), including:

- (i) *Purpose of the fusion*: is it for solubility improvement or for affinity purification? Nowadays, a variety of fusion tags that render different purposes are available, and systems containing both solubility and affinity tags like, for instance, the dual hexahistidine (His₆)-MBP tag, can be designed in order to get a rapid “in one step” protein production. Some protein tags can also function in both affinity and solubility roles, as for instance, the MBP or glutathione-S-transferase (GST; Esposito and Chatterjee, 2006). If the fusion tag is to be used in protein purification, the cost and buffer conditions are often the criteria for selection. For instance, proteins that require chelating agents as EDTA are not suitable for immobilized metal affinity chromatography (IMAC) via the His₆ tag as nickel ions in the affinity matrix are chelated by EDTA (Malhotra, 2009).
- (ii) *Amino acid composition and size*: these two factors should be considered when selecting a fusion partner because target proteins may require larger or smaller tags depending on their application. Larger tags can present a major diversity in the amino acid content, and will impose a metabolic burden in the host cell different from that imposed by small tags (Malhotra, 2009).
- (iii) *Required production levels*: structural studies require higher protein production levels that can be rapidly achieved with a larger fusion tag, which has strong translational initiation signals, whereas the study of physiological interactions demands for lower production levels and small tags (Malhotra, 2009).
- (iv) *Tag location*: fusion partners can promote different effects when located at the N-terminus or C-terminus of the passenger protein. Usually, N-terminal tags are advantageous over C-terminal tags because: (1) they provide a reliable context for efficient translation initiation, in which fusion proteins take advantage of efficient translation initiation sites on the tag; (2) they can be removed leaving none or few additional residues at the native N-terminal sequence of the target protein, since most of endoproteases cleave at or near the C-terminus of their recognition sites (Waugh, 2005; Malhotra, 2009).

Fusion tags can be incorporated using different strategies: affinity and solubility tags are set individually or together, and sites for protease cleavage are designed between the fusion tags and target proteins.

Solubility enhancer partners

In spite of all the approaches conducted so far, the choice of a fusion partner is still a trial-and-error experience. Fusion partners do not perform equally with all target proteins, and each

target protein can be differentially affected by several fusion tags (Esposito and Chatterjee, 2006). In the past decade, parallel high throughput (HTP) screenings using different fusion partners have developed soluble protein production, and facilitated a rapid, tailored, and cost-effective choice of the best fusion partner for each target protein (Hammarstrom et al., 2002; Shih et al., 2002; Dyson et al., 2004; Dummmler et al., 2005; Cabrita et al., 2006; Hammarstrom, 2006; Marblestone et al., 2006; Kim and Lee, 2008; Kohl et al., 2008; Ohana et al., 2009; Bird, 2011).

The mechanisms by which fusion tags enhance the solubility of their partner proteins remain unclear, but several hypotheses have been suggested (Butt et al., 2005; Nallamsetty and Waugh, 2007):

- (i) Fusion proteins form *micelle-like structures*: misfolded or unfolded proteins are sequestered and protected from the solvent and the soluble protein domains face outward;
- (ii) Fusion partners *attract chaperones*: the fusion tag drives its partner protein into a chaperone-mediated folding pathway. MBP and N-utilization substance (NusA) are two fusion tags that present this mechanism, being previously reported to interact with GroEL in *E. coli* (Huang and Chuang, 1999; Douette et al., 2005);
- (iii) Fusion partners have an *intrinsic chaperone-like activity*: hydrophobic patches of the fusion tag interact with partially folded passenger proteins, preventing their self-aggregation, and promoting their proper folding. MBP was previously reported to act also as a chaperone in the fusion context (Kapust and Waugh, 1999; Fox et al., 2001). Solubility enhancer partners may thus play a passive role in the folding of their target proteins, reducing the chances for protein aggregation (Waugh, 2005; Nallamsetty and Waugh, 2006);
- (iv) Fusion partners *net charges*: highly acidic fusion partners were suggested to inhibit protein aggregation by electrostatic repulsion (Zhang et al., 2004; Su et al., 2007).

A large variety of solubility enhancer tags are available (**Table 1**), including the well-known MBP, NusA, thioredoxin (TrxA), GST, and SUMO, and several other novel moieties recently discovered, for instance, the Fh8 tag.

MBP is a large (43 kDa) periplasmic and highly soluble protein of *E. coli* that acts as a solubility enhancer tag (Kapust and Waugh, 1999; Fox et al., 2001), and it has a native affinity property to function as a purification handle.

MBP plays an important role in the translocation of maltose and maltodextrins (Nikaido, 1994): it has a natural protein-binding site that it uses to interact with other proteins involved in maltose signaling and chemotaxis, and it has a large hydrophobic cleft close to this site that undergoes conformational changes upon maltose binding (Fox et al., 2001).

When used in the fusion context, MBP promotes target protein solubility by showing chaperone intrinsic activity (Kapust and Waugh, 1999; Bach et al., 2001; Fox et al., 2001), and it is more efficient at the N-terminus of the target proteins rather than at the C-terminus (Sachdev and Chirgwin, 2000). In fact, MBP promotes the proper folding of the target protein by interacting with the latter, and occluding its self-association. This passive role of MBP in protein folding is correlated with the large hydrophobic area exposed on its surface, which is responsible for the contact with

Table 1 | Solubility enhancer tags [adapted from Esposito and Chatterjee (2006), Malhotra (2009)].

Tag	Protein	Size (aa)	Organism	Reference
Fh8	<i>Fasciola hepatica</i> 8-kDa antigen	69	<i>F. hepatica</i>	Costa (2013), Costa et al. (2013a,b)
MBP	Maltose-binding protein	396	<i>Escherichia coli</i>	di Guan et al. (1988), Kapust and Waugh (1999)
NusA	N-utilization substance	495	<i>E. coli</i>	Davis et al. (1999)
Trx	Thioredoxin	109	<i>E. coli</i>	Lavallie et al. (1993)
SUMO	Small ubiquitin modified	~100	<i>Homo sapiens</i>	Butt et al. (2005), Marblestone et al. (2006)
GST	Glutathione-S-transferase	211	<i>Schistosoma japonicum</i>	Smith and Johnson (1988)
SET	Solubility-enhancer peptide sequences	<20	Synthetic	Zhang et al. (2004)
GB1	IgG domain B1 of Protein G	56	<i>Streptococcus sp.</i>	Zhou et al. (2001), Cheng and Patel (2004)
ZZ	IgG repeat domain ZZ of Protein A	116	<i>Staphylococcus aureus</i>	Rondahl et al. (1992), Inouye and Sahara (2009)
HaloTag	Mutated dehalogenase	~300	<i>Rhodococcus sp.</i>	Ohana et al. (2009)
SNUT	Solubility eNhancing Ubiquitous Tag	147	<i>Staphylococcus aureus</i>	Caswell et al. (2010)
Skp	Seventeen kilodalton protein	161	<i>E. coli</i>	Esposito and Chatterjee (2006)
T7PK	Phage T7 protein kinase	~240	<i>Bacteriophage T7</i>	Esposito and Chatterjee (2006)
EspA	<i>E. coli</i> secreted protein A	192	<i>E. coli</i>	Cheng et al. (2010)
Mocr	Monomeric bacteriophage T7 0.3 protein (Orc protein)	117	<i>Bacteriophage T7</i>	DelProposto et al. (2009)
Ecotin	<i>E. coli</i> trypsin inhibitor	162	<i>E. coli</i>	Malik et al. (2006, 2007)
CaBP	Calcium-binding protein	134	<i>Entamoeba histolytica</i>	Reddi et al. (2002)
ArsC	Stress-responsive arsenate reductase	141	<i>E. coli</i>	Song et al. (2011)
IF2-domain I	N-terminal fragment of translation initiation factor IF2	158	<i>E. coli</i>	Sorensen et al. (2003a)
Expressivity tag (part of IF2-domain I)	N-terminal fragment of translation initiation factor IF2	7 (21 nt)	<i>E. coli</i>	Hansted et al. (2011)
RpoA, SlyD, Tsf,	Stress-responsive proteins	329, 196, 283,	<i>E. coli</i>	Ahn et al. (2007), Han et al. (2007a,b,c),
RpoS, PotD, Crr		330, 348, 169		Park et al. (2008)
msyB, yjgD, rpoD	<i>E. coli</i> acidic proteins	124, 138, 613	<i>E. coli</i>	Su et al. (2007), Zou et al. (2008)

aa— amino acids; nt— nucleotides.

other proteins in the maltose transport apparatus (Kapust and Waugh, 1999; Fox et al., 2001). Hence, the MBP hydrophobic cleft is pointed as the site where fused polypeptides interact with the fusion partner (Kapust and Waugh, 1999; Fox et al., 2001; Nallamsetty and Waugh, 2007), similar to what it is reported for GroEL and DnaK molecular chaperones (Buckle et al., 1997; Chatellier et al., 1999; Tanaka and Fersht, 1999). The presence of this cleft can explain why only certain soluble proteins like MBP act as solubilizing agents. Moreover, MBP presents certain conformational flexibility associated with the cleft; thereby it can adjust its shape to accommodate several different polypeptides.

MBP fusion proteins bind to immobilized amylose resins, but this binding is highly dependent on the nature of the passenger protein as it can block or reduce the amylose interaction (Pryor and Leiting, 1997). Difficulties found in the binding of MBP fusion proteins to amylose resins corroborate the hypothesis that target proteins interact with MBP via its binding site (Fox et al., 2001).

Other affinity tags, specific proteases and protein cultivation strategies are being employed together with MBP to improve protein soluble production, purification and native protein recovery, as for instance, His₆-MBP fusions (Nallamsetty et al., 2005), His₆-MBP-TEV fusions (Rocco et al., 2008), MBP-His₆-Smt3 fusions in which the *Saccharomyces cerevisiae* Smt3 protein is used for protein processing by proteolytic cleavage between the MBP-His₆ tags and the protein of interest (Motejadded and Altenbuchner, 2009), and secretion of MBP fusion protein into the culture medium (Sommer et al., 2009).

Several commercial expression vectors containing the MBP tag are available for cytoplasmic and periplasmic production of target proteins, including the pMAL series (New England Biolabs) and pIVEX (Roche).

NusA is a transcription termination/anti-termination protein that promotes/prevents RNA polymerase pausing when acting alone or when included in the anti-termination complex, respectively. NusA (55 kDa) is used as a fusion partner to confer

stability and high solubility to its target proteins (De Marco et al., 2004; Dümmler et al., 2005; Turner et al., 2005). The NusA ability to improve the soluble production of fusion proteins may be correlated with its intrinsically solubility and biological activity in *E. coli*. NusA slows down translation at the transcriptional pauses, offering more time for protein folding (Davis et al., 1999; De Marco et al., 2004). In contrast to MBP, NusA does not present an intrinsic affinity property, therefore requiring the addition of an affinity tag for efficient protein production, as for instance, the His₆ tag (Davis et al., 1999). As for MBP, several strategies have been exploited to use the NusA solubility enhancer fusion partner with purification tags and specific proteases like the pETM60 vector (EMBL; De Marco et al., 2004) that render the production of a NusA–His₆–TEV fusion protein, or the pET43 (Novagen), that offers the same NusA–His₆ fusion protein but with a thrombin and enterokinase cleavage sites between the fusion tags and target proteins.

In spite of the different physiochemical and structural properties, as well as different biological functions, MBP and NusA are often reported to promote similar solubility improvements in their target proteins, being ranked as two of the best tags for making soluble proteins (Shih et al., 2002; Kohl et al., 2008; Bird, 2011). Both fusion partners were reported to probably work by similar mechanisms, in which NusA, like MBP, plays a passive role on the target protein folding (Nallamsetty and Waugh, 2006).

TrxA, or Trx, is a 12-kDa intracellular thermostable protein of *E. coli* that is highly soluble expressed in its cytoplasm (Young et al., 2012). The *E. coli* Trx can be used for co-production with a target protein, improving the solubility of the latter (Yasukawa et al., 1995). Trx is also commonly employed as a fusion tag to avoid inclusion body's formation in recombinant protein production by taking advantage of its intrinsic oxido-reductase activity responsible for the reduction of disulfide bonds through thio-disulfide exchange (Stewart et al., 1998; LaVallie et al., 2000; Young et al., 2012). The fusion partner Trx can be placed both at the N- or C-terminal of target proteins (LaVallie et al., 2000) but this fusion partner is more effective at the N-terminal of the target protein (Terpe, 2003; Dyson et al., 2004). In some HTP screenings (Hammarstrom et al., 2002; Dyson et al., 2004; Kim and Lee, 2008), the Trx fusion partner improves target protein solubility similar to MBP tag, being considered one of the best choices for protein production in *E. coli*.

Unlike MBP, Trx does not have intrinsic affinity properties, thus requiring an additional fusion tag for protein purification such as the His₆ tag. The pET32 (Novagen), one of the commercially available vectors for Trx tagging, carries this dual-fusion partners for protein production and purification (Austin, 2003).

Trx fusion partner can also be useful in protein crystallization of certain target proteins because it readily forms several crystals itself, and it offers a rigid connection to the target protein, which is an essential feature for blocking conformational heterogeneity usually found in various attempts of fusion proteins crystallization (Smyth et al., 2003; Corsini et al., 2008).

Small ubiquitin related modifier is a small protein (~11 kDa) found in yeast (one single gene coding for Smt3) and vertebrates (three genes coding for SUMO-1, SUMO-2, and SUMO-3; Kawabe et al., 2000) that has recently been used as an effective N-terminal

solubility enhancer fusion partner, offering advantages over other fusion systems (Marblestone et al., 2006; Bird, 2011).

The robust SUMO protease (catalytic domains of Ulp1) offers significant advantages over other endoproteases because it recognizes the tertiary structure of SUMO, and consequently it does not present unspecific cleavage of the protein linear amino acid sequence. Moreover, when used for tag removal, SUMO protease generates a cleaved target protein with its native N-terminal amino acid composition (Malakhov et al., 2004; Marblestone et al., 2006).

Small ubiquitin related modifier promotes the proper folding and solubility of its target proteins possibly by exerting chaperoning effects in a similar mechanism to the described for its structural homolog Ubiquitin (Ub; Khorasanizadeh et al., 1996). Ub was reported to be the nature's fastest folding protein, and SUMO also presents a tight, rapidly folding soluble structure (Marblestone et al., 2006). In addition, Ub and Ub-like proteins (Ulp) have a highly hydrophobic inner core and a hydrophilic surface that, together with such a rapid folding, may explain the SUMO's behavior as a nucleation site for the proper folding of target proteins (Malakhov et al., 2004; Marblestone et al., 2006).

Small ubiquitin related modifier fusion proteins or peptides are usually purified by affinity chromatography using the His₆ tag (Lee et al., 2008; Gao et al., 2010; Wang et al., 2010; Satakarni and Curtis, 2011). Due to its unique features, SUMO technology has been constantly explored, and novel strategies for a facile and rapid protein production are now available, as the SUMO–intein system (Wang et al., 2012). The SUMO fusion partner is also available for recombinant protein production in other host cells, namely, insect cells and other eukaryotic cells (Panavas et al., 2009).

Glutathione-S-transferase from *Schistosoma japonicum* (26 kDa) that has been used as an affinity fusion partner for the single-step purification of its target proteins (Smith and Johnson, 1988). GST can also promote protein soluble production in *E. coli*, being more efficient when positioned at the N-terminal rather than at the C-terminal end (Malhotra, 2009). This fusion partner can protect its target protein from the proteolytic degradation, stabilizing it into the soluble fraction (Kaplan et al., 1997; Hu et al., 2008; Young et al., 2012). In spite of performing quite well in some HTP studies (Dümmler et al., 2005; Cabrita et al., 2006; Kim and Lee, 2008), GST is often a poor solubility tag when compared to other commonly fusion partners, rendering the target protein production into inclusion bodies (Hammarstrom et al., 2002; Dyson et al., 2004; Hammarstrom, 2006; Kohl et al., 2008; Ohana et al., 2009).

Glutathione transferases are dimeric enzymes that catalyze the nucleophilic addition of the thiol of glutathione to a wide range of hydrophobic electrophilic molecules (Ketterer, 2001). Taking this feature into account, GST can be useful for monitoring the protein production and purification via its catalytic activity, and the purification of GST fusion proteins can be easily performed by affinity chromatography using glutathione derivatives immobilized into a solid support (Viljanen et al., 2008). GST fusion proteins can be eluted with glutathione under mild conditions (Vinckier et al., 2011).

A major disadvantage for using GST as solubility and affinity tag relies on its oligomerized form: GST has four solvent

exposed cysteines that can provide a significant oxidative aggregation (Kaplan et al., 1997), making it a poor choice for tagging oligomeric target proteins (Malhotra, 2009).

As occurs with MBP, GST can be coupled with other affinity strategies, for instance, the His₆ tag, to improve the protein purification (Scheich et al., 2003; Hayashi and Kojima, 2008; Hu et al., 2008). GST expression vectors like the pGEX (Hakes and Dixon, 1992) or pCold-GST (Hayashi and Kojima, 2008) usually contain a protease recognition site between the fusion tag coding gene and the target protein coding gene for GST tag's removal after or during protein purification.

GST has also been applied as a fusion partner in other expression systems apart from the *E. coli* such as yeast (Mitchell et al., 1993), insect cells (Beekman et al., 1994), and mammalian cells (Rudert et al., 1996). This fusion partner has shown to be useful for protein labeling (Ron and Dressler, 1992; Viljanen et al., 2008), antibody production (Aatsinki and Rajaniemi, 2005), and vaccine development (Mctigue et al., 1995).

In addition to these commonly used fusion partners, new solubility enhancer tags are constantly emerging in literature (see the corresponding references in **Table 1**), as for instance, the *Fh8* tag [see The Novel Fh8 Fusion System (Hitag[®])], *HaloTag*, which uses a modified haloalkane dehalogenase protein that improves protein solubility and can bind to several synthetic ligands, the monomeric mutant of Orc protein of the bacteriophage T7 (*Mocr*), the *E. coli* protein *Skp*, stress-responsive proteins *RpoA*, *SlyD*, *Tsf*, *RpoS*, part of the domain I of IF2 (*expressivity tag*), the *E. coli* secreted protein A (*EspA*), and the *SNUT* tag, which is a protein derived from a portion of the bacterial transpeptidase sortase A of *Staphylococcus aureus*.

Affinity purification handles

Affinity fusion partners have widely contributed for the development of recombinant protein production studies in basic research and in HTP structural biology (Waugh, 2011) by simplifying protein purification procedures, and allowing for protein detection, and characterization (Butt et al., 2005; Malhotra, 2009; Young et al., 2012).

Affinity purification handles can be divided into two groups: (1) peptides or proteins that bind a small ligand immobilized on a solid support, as for instance, the His₆ tag and nickel affinity resins, and (2) tags that bind to an immobilized molecule such as antibodies (Arnau et al., 2006).

The purification of a target protein using an affinity handle offers several *advantages* over the conventional chromatographic methodologies, namely:

- (i) The target protein never interacts directly with the chromatographic resin (Waugh, 2005);
- (ii) Target proteins can be easily obtained pure after a single-step purification (Terpe, 2003);
- (iii) Affinity purification offers a variety of strategies to bind the target protein on an affinity matrix (Malhotra, 2009);
- (iv) Affinity tags are an economically favorable and time-saving strategy, and they allow different proteins to be purified using a common method in contrast to highly customized procedures used in conventional chromatographic purification (Arnau et al., 2006).

An affinity tag is often chosen taking into account the purification costs: different affinity media and elution principles present different expenses during the operation process and should therefore be carefully selected at the beginning of the cloning strategy. The buffer requirements are also essential for the designing of an efficient purification strategy (Malhotra, 2009). In addition, the choice of an affinity can also rely on the size: small tags are useful for protein detection and antibody production, as they are not immunogenic as large tags (Terpe, 2003).

Tandem affinity purification (TAP) or *dual-tagging* strategies are now commonly used in recombinant protein production: they offer a highly specific isolation of target proteins with minimal background and under mild conditions, and they are very useful in the study of protein interactions, allowing the separation of different mixed protein complexes (Arnau et al., 2006; Li, 2010).

Table 2 lists some of the common and novel purification tags used in recombinant protein production.

The *polyhistidine affinity tag* or *His tag* consists of a variable number of consecutive histidine residues (usually six) that coordinate, via the histidine imidazole ring, transition metal ions such as Ni²⁺ or Co²⁺ immobilized on beads or a resin for IMAC (Gaber-Porekar and Menart, 2001; Terpe, 2003; Kimple and Sondek, 2004; Malhotra, 2009). Commonly used IMAC resins such as nitrilotriacetic acid agarose (Ni-NTA, from Qiagen), or carboxymethylaspartate agarose (Talon, from ClonTech) have a high binding capacity, and can be used for purification of fusion proteins directly from crude cell lysates (Terpe, 2003; Kimple and Sondek, 2004; Li, 2010).

The His tag is one of the most widely used purification tags, and it offers several advantages (Kimple and Sondek, 2004; Li, 2010):

- (i) Its small size and charge rarely interferes with protein function and structure;
- (ii) It can be used under native and denaturing conditions
- (iii) Target proteins can be eluted under mild conditions by imidazole competition or low pH.

The His tag has been used in several HTP screenings, placed at the N- or C-terminal end, or even in the middle of the fusion protein (Cabrita et al., 2006; Hammarstrom, 2006; Marblestone et al., 2006; Bird, 2011), and it is also a useful tool in protein crystallization as well as protein detection (Carson et al., 2003; Kimple and Sondek, 2004).

Taking into account the mechanism of protein interaction with the immobilized ions, careful should be taken in IMAC to avoid strong reducing and chelating agents in any of the buffers (as for instance, EDTA), as they will reduce or strip the immobilized metal ions (Carson et al., 2003; Kimple and Sondek, 2004; Li, 2010).

Epitope tags are short sequences of amino acids that serve as the antigen region to which the antibody binds, being suitable for several immunoapplications. These include affinity chromatography on immobilized monoclonal antibodies, and protein trafficking *in vitro* or in cell cultures (Kimple and Sondek, 2004; Young et al., 2012). Epitope tagging engages an expensive purification that often limits its wide application.

The following partners are often used as epitope tags: the FLAG tag (Einhauser and Jungbauer, 2001), the hemagglutinin, and the c-Myc (Fritze and Anderson, 2000). Their short sequences rarely

Table 2 | Affinity purification tags [adapted from Esposito and Chatterjee (2006), Malhotra (2009)].

Tag	Protein	Size (aa)	Affinity matrix	Elution	Reference
His ₆	Hexahistidine tag	6–10	Immobilized metal ion – Ni, Co, Cu, Zn	Competition with imidazole	Gaberc-Porekar and Menart (2001)
Fh8	<i>Fasciola hepatica</i> 8-kDa antigen	69	Hydrophobic (calcium dependent interaction)	Ca ²⁺ -chelating agents such as EDTA or pH manipulation	Costa (2013), Costa et al. (2013b)
GST	Glutathione-S-transferase	211	Glutathione	Competition with free glutathione	Smith and Johnson (1988)
MBP	Maltose-binding protein	396	Amylose	Competition with maltose	di Guan et al. (1988), Pryor and Leiting (1997)
FLAG	FLAg tag peptide	8	Anti-FLAG antibody octapeptide when using anti-FLAG M2 antibody	Competition with FLAG	Einbauer and Jungbauer (2001)
Strep-II	Streptavidin binding peptide	8	Streptavidin	Competition with biotin and derivatives	Schmidt and Skerra (1994)
CBP	Calmodulin-binding protein	26	Immobilized calmodulin	Ca ²⁺ -chelating agents	Vaillancourt et al. (2000)
HaloTag	Mutated dehalogenase	~300	Chloroalkane	Covalent binding and proteolytic release of target protein	Ohana et al. (2009)
Protein A	Staphylococcal Protein A	280	Immobilized IgG	pH manipulation (acidic)	Stahl and Nygren (1997)
IMPACT (CBD)	Intein mediated purification with the chitin-binding domain	51	Chitin	Intein self-cleavage induction with dithiothreitol, β-mercaptoethanol or cysteine	Chong et al. (1997), Sheibani (1999)
CBM	Cellulose-binding module	*	Cellulose	Urea and guanidine-HCl or ethylene glycol	Tomme et al. (1998), Ramos et al. (2010), Ramos et al. (2013)
Dock	Dockerin domain of <i>Clostridium josui</i>	22	Cohesin – Cellulose	Ca ²⁺ -chelating agents	Kamezaki et al. (2010)
Tamavidin	fungal avidin-like protein	~140	Biotin	Free biotin in excess when using the Tamavidin 2-REV tag	Takakura et al. (2010, 2013)

*Several sizes, from 4 to 20 kDa.

interfere with structure or function of target proteins, and are very specific for their respective primary antibodies (Kimple and Sondek, 2004; Malhotra, 2009). The *FLAG* tag is a short hydrophilic eight amino-acid peptide, and it was the first tag to be used in the epitope context. This tag works either for protein detection or purification (Hopp et al., 1988; Knappik and Pluckthun, 1994), and it has an intrinsic enterokinase cleavage site at its C-terminus end, allowing its complete removal from the target protein (Einbauer and Jungbauer, 2001; Young et al., 2012).

Strep II tag is a short tag of only eight amino acid residues that possesses a strong and specific binding to streptavidin via its biotin pocket (Schmidt and Skerra, 1994). This affinity partner can be fused at both N- or C-terminal ends, or within the target protein. Strep II-fused proteins elute from streptavidin columns with biotin derivatives under gentle conditions (Terpe, 2003; Li, 2010).

The *CBP* tag is a calmodulin-binding peptide derived from the C-terminus of skeletal muscle myosin light chain kinase, and it has been used as an N- or C-terminal affinity tag of target protein purification on a calmodulin immobilized matrix (Terpe, 2003;

Malhotra, 2009). The *CBP* interaction with calmodulin is calcium-dependent, and hence, the addition of calcium-chelating allows the single step elution of target proteins under gentle conditions (Terpe, 2003; Malhotra, 2009; Li, 2010). This tag is an affinity system highly specific for protein purification in *E. coli* but not in eukaryotic systems, as *E. coli* does not contain endogenous proteins that interact with calmodulin (Terpe, 2003; Malhotra, 2009).

In addition to the above-mentioned affinity tags, new affinity purification strategies are now described in literature for protein isolation and detection (see the corresponding references in **Table 2**) such as the *Fh8* tag [see The Novel Fh8 Fusion System (Hitag[®])], cellulose-binding domains I, II, and III (*CBD*), the *HaloTag*, the dockerin domain *Dock* tag, and the avidin-like protein, *Tamavidin* tag.

Tag removal

The removal of the fusion partner from the final protein is often necessary because the tag can potentially interfere with the proper

structure and functioning of the target protein (Waugh, 2005; Malhotra, 2009; Young et al., 2012).

Fusion partners are removed from their target proteins either by *enzymatic cleavage*, in which site specific proteases are used under mild conditions, or by *chemical cleavage*, like for instance formic acid (Ramos et al., 2010, 2013), that offers a less expensive tag removal but it is also less specific compared to the enzymatic strategy, besides presenting harsh conditions that can affect the target protein stability and solubility (Malhotra, 2009; Li, 2011). Fusion partners can also be cleaved from the target protein using an *in vivo* cleavage strategy, in which a controlled intracellular processing (CIP) is applied as follows: the fusion protein and protease are produced from separate compatible expression vectors that can be regulated independently of one another. The protease cleaves the fusion protein *in vivo*, offering the advantage of not compromising the target protein's purity level or its production yields like often occurs in *in vitro* cleavage strategies (Kapust and Waugh, 2000).

The efficiency of the *enzymatic* removal of fusion proteins may vary in an unpredicted manner with different proteins (Li, 2011; Vergis and Wiener, 2011; Young et al., 2012), and it often requires the optimization of cleavage conditions through a trial-and-error process (Malhotra, 2009). Two types of proteases can be used for tag removal (reviewed in Waugh, 2011):

- (i) *Endoproteases*: they are divided into *serine proteases* such as the activated blood coagulation factor X (factor Xa), enterokinase, and α -thrombin, and *viral proteases* like the tobacco etch virus (TEV), and the human rhinovirus 3C protease (Table 3). In spite of recognizing a similar number of substrate amino acid residues, viral proteases have usually more stringent sequence specificity than serine proteases, presenting also slower rates than the latter. Endoproteases are useful tools for the removal of N-terminal fusion tags, since they cleave close to the C-terminus end of their recognition sites thus leaving the target protein with its native N-terminal sequence.
- (ii) *Exoproteases*: they are often used together with endoproteases mainly for the removal of C-terminal fusion tags. The available exoproteases include metallocarboxypeptidases, and aminopeptidases.

The removal of a fusion tag is usually accomplished by two purification steps, as follows: after the initial affinity purification step (e.g., via a histidine tag located at the N-terminal of the fusion protein), the purified fusion protein is mixed in solution with the endoprotease (e.g., a his-tagged protease) to cleave off the tag. The cleaved target protein is recovered in the flow-through sample after a second affinity purification step, in which the cleaved fusion tag and the added protease are collected in the eluted sample.

In spite of widely employed, the removal of fusion partners has always been the Achilles' heel of affinity tagging, presenting several *difficulties* such as:

- (i) Unspecific cleavage due to the recognition of a linear amino acid sequence (except for SUMO protease);
- (ii) Inefficient processing due to steric hindrance or the presence of unfavorable residues around the cleavage site (Li, 2011; Waugh, 2011). The inclusion of extra amino acid residues (a spacer or linker) between the cleavage site and target protein (Esposito and Chatterjee, 2006; Malhotra, 2009) can alleviate this problem;
- (iii) Low protein yields after tag removal, and failure in recover active, structurally organized target proteins due to protein precipitation and aggregation (Butt et al., 2005; Waugh, 2011);
- (iv) High costs of proteases and tedious optimization of cleavage conditions (Smyth et al., 2003).

Independently of the cleavage type, additional chromatographic steps are often required to purify the target protein from the cleavage mixture. Although conventional affinity technologies have greatly simplified recombinant protein production, resins, and buffers are still too expensive. Hence, the tag removal adds another layer of complexity and expense to the recombinant protein production process (Mee et al., 2008; Li, 2011).

Self-cleaving tags are a special group of fusion tags that possess inducible proteolytic activity, therefore being considered an attractive alternative to the existent affinity strategies for simple and costless protein purification and tag removal (Chong et al., 1997; Li, 2011).

The protein splicing is a process in which the intervening sequence (intein) removes itself and binds the flanking residues (exteins) to produce two independent protein products (Perler

Table 3 | Common endoproteases for tag removal [adapted from Malhotra (2009)].

Protease	Source	Cleavage site	Reference
TEV	Tobacco etch virus protease	ENLYFQ/G	Parks et al. (1994), Kapust et al. (2001)
EntK	Enterokinase	DDDDK/	Choi et al. (2001)
Xa	Factor Xa	IEGR/	Jenny et al. (2003)
Thr	Thrombin	LVPR/GS	Jenny et al. (2003)
PreScission	Genetically engineered derivative of human rhinovirus 3C protease	LEVLFQ/GP	Cordingley et al. (1990), GE Healthcare (2010)
SUMO protease	Catalytic core of Ulp1	Recognizes SUMO tertiary structure and cleaves at the C-terminal end of the conserved Gly–Gly sequence in SUMO	Malakhov et al. (2004), Butt et al. (2005), Marblestone et al. (2006)

et al., 1994). Self-cleaving tags undergo specific cleavage upon being triggered by low molecular weight compounds or upon a change of conformation. The available technologies include inteins, the *S. aureus* sortase A, the N-terminal protease (N^{pro}), the *Neisseria meningitidis* iron-regulated protein FrpC, and the cysteine protease domain secreted by *Vibrio cholerae*, all of them reviewed in Li (2011).

THE NOVEL Fh8 FUSION SYSTEM (Hitag®)

Fh8 (GenBank ID: AF213970.1) is one of the promising new fusion technologies, advancing the existing tags by acting simultaneously as an effective solubility enhancer partner (Costa et al., 2013a) and robust purification handle (Costa et al., 2013b). Actually, the Fh8 is one of the few existent fusion tags to offer this combined feature of enhancing protein solubility and purification, and its low molecular weight (8 kDa) is also a great advantage over other large fusion partners for recombinant protein production in *E. coli* (Costa, 2013).

The Fh8 is a small antigen (8 kDa) excreted-secreted by the parasite *F. hepatica* in the early stages of infection (Silva et al., 2004). This protein is located on the surface of the parasite, and it was suggested as a useful tool for the diagnosis, vaccine, and drug development against *F. hepatica* infections (Silva et al., 2004). The use of recombinant Fh8 produced in *E. coli* led to the development of a novel, rapid, and simple immunodetection of *F. hepatica* infections (Silva et al., 2004). Moreover, when produced recombinantly in *E. coli*, the Fh8 revealed to be a highly soluble and unusual thermal stable protein (keeping secondary structure integrity up to 74°C; Silva et al., 2004; Fraga et al., 2010).

The Fh8 has high homology with 8-kDa calcium-binding proteins (CaBPs) of *Schistosoma mansoni* (Sm8; Ram et al., 1989), of *Clonorchis sinensis* (Ch8), and of *S. japonicum* (Sj8; Lv et al., 2009), and it belongs to the calmodulin-like EF-hand CaBP family (Fraga et al., 2010).

CaBPs are structurally organized by EF-hand motifs, which are helix–loop–helix structures that participate in Ca^{2+} coordination (Bhattacharya et al., 2004; Zhou et al., 2006; Chazin, 2011). Upon calcium binding, Ca^{2+} sensor proteins, like calmodulin (Nelson and Chazin, 1998; Chin and Means, 2000) and troponin C (Nelson and Chazin, 1998), translate the physiological changes in calcium levels by undergoing a conformational change. This then allows the binding of other proteins downstream the process. In EF-hand proteins, the open of the EF-hand structure exposes a hydrophobic surface, which binds the target sequence (Lewit-Bentley and Rety, 2000; Bhattacharya et al., 2004). Ca^{2+} buffer proteins, such as calbindin D_{9k} and parvalbumin (Schwaller, 2010), are involved in calcium signal modulation, undergoing minimal conformational changes upon calcium binding.

The Fh8 presents two EF-hand motifs, and it was characterized as a Ca^{2+} sensor protein: when calcium binds, the Fh8 switches from a closed (apo-state) to an open (calcium-loaded state) conformation due to the reorientation of the four helices, exposing a large hydrophobic region that acts as a target-binding surface (Fraga et al., 2010).

Previous studies for the prediction of the Fh8 three-dimensional structure (unpublished data) showed that almost all the Fh8's amino acid sequence is involved or affected by the

calcium-binding, with the exception of small residue sequences in the N-terminal (11 amino acid residues) and C-terminal (six amino acid residues). Considering that the N-terminal of a protein is very important for its half-life, the first N-terminal 11 residues of Fh8 were named the “H sequence” and were initially suggested to play a key role in the stability and production of the entire Fh8 protein. This H sequence could also be critical for the immunological response of the Fh8 antigen.

Taking into account the Fh8 high solubility and stability when expressed in *E. coli* together with its calcium-binding properties, and given the potential importance of the H sequence, both Fh8 and H peptides were suggested to function as fusion tags for protein production and solubility in *E. coli*, protein purification, and antibody production.

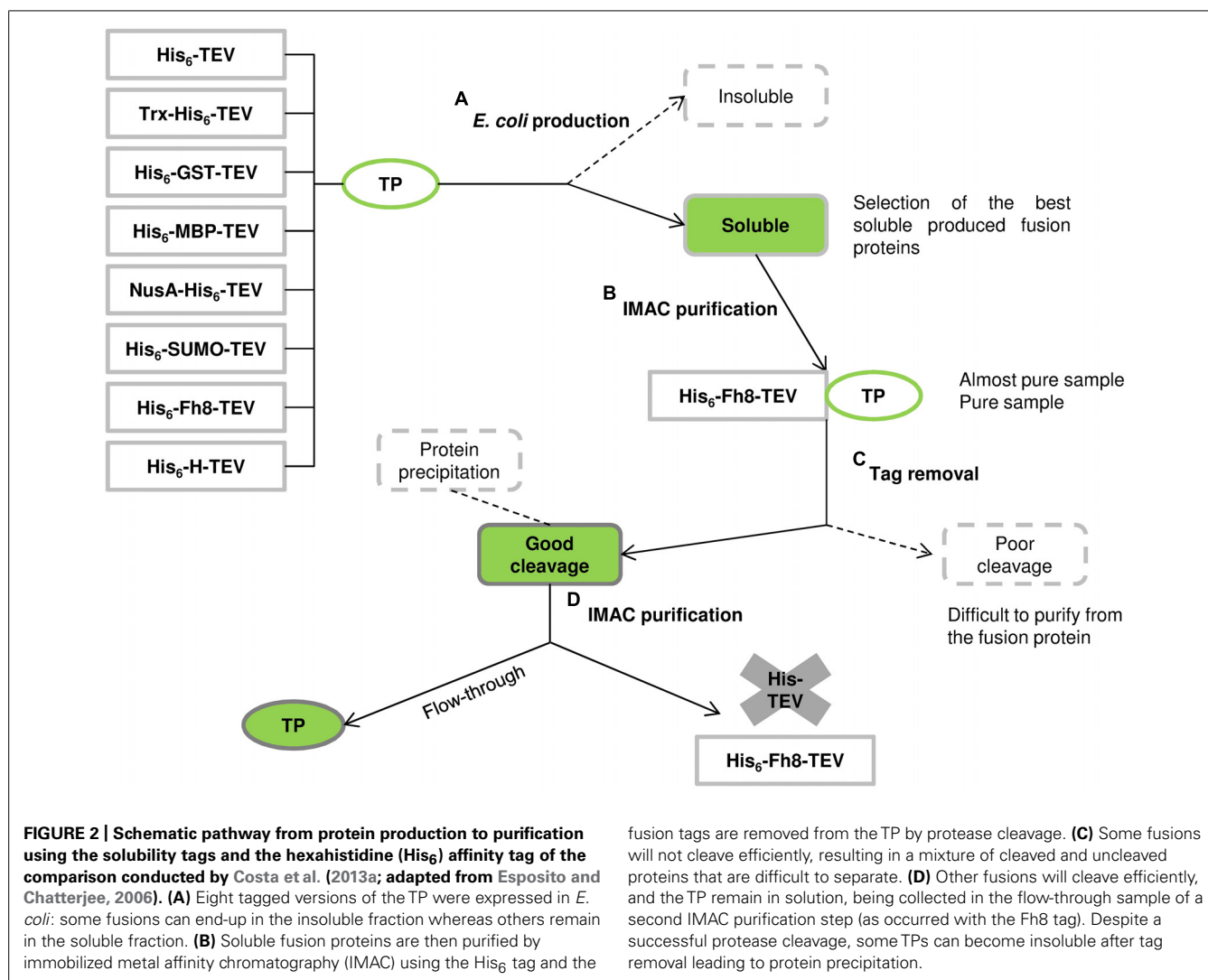
The application of both Fh8 (8 kDa) and H (1 kDa) peptides as fusion tags for protein overproduction in *E. coli* was first reported by Conceição and co-workers, using the following recombinant proteins: a 12-kDa surface protein of *Cryptosporidium parvum* (CP12), the interleukin-5 of human origin (IL-5), and an oocyst wall protein of *Toxoplasma gondii* (TgOWP; Conceição et al., 2010). This initial study showed that both Fh8 and H peptides have indeed a positive effect on the *E. coli* production levels of all target proteins, reaching values three- to 16-fold higher than those obtained with non-fused target proteins.

The Fh8 and H fusion tags were then studied as solubility enhancer tags, and their performance was compared with other commonly used fusion tags available in the Protein Expression and Purification Core Facility of the European Molecular Biology Laboratory (Costa, 2013; Costa et al., 2013a). **Figure 2** illustrates the schematic pathway from protein production to purification with the studied solubility tags (His₆ tag, GST, MBP, NusA, Trx, SUMO, H, and Fh8). Here, the selected target proteins included the 12-kDa surface protein of *C. parvum* (CP12), the lectin frutalin from the *Artocarpus incisa* plant (FTL; Oliveira et al., 2008, 2009a,b, 2011), and four proteins from the yeast *S. cerevisiae*: reduced viability upon starvation protein 167 (RVS167), phospholipase D1 (SPO14), and serine/threonine-protein kinases 1 and 2 (YPK1 and YPK2). These target proteins were all known as difficult-to-express in *E. coli*, and presented different molecular weights, locations, and functions. The evaluation of their solubility and consequent effect of each fusion tag was performed after nickel affinity purification and upon tag removal in 10-mL cultures and in 500-mL cultures.

This comparison study showed that the Fh8 fusion partner stands among the well-described best fusion partners, MBP, NusA, and Trx, for soluble protein production. For the proteins tested, both GST and H fusion tags did not improve target protein solubility in *E. coli*.

The novel Fh8 fusion partner is thus an excellent candidate for testing production and solubility next to the other well-known fusion tags. Its low molecular weight and its solubility enhancing effect make Fh8 an advantageous option compared to larger fusion tags for soluble protein production in *E. coli*.

Apart from its solubility enhancer effect, the Fh8 was also explored by Costa (2013), Costa et al. (2013b) as a purification handle via its calcium-binding behavior combined with HIC. Two different model proteins were used within this study: green fluorescent protein (GFP) and superoxide dismutase (SOD), and the



Fh8-HIC performance was also compared to the one of His tag technology (via IMAC).

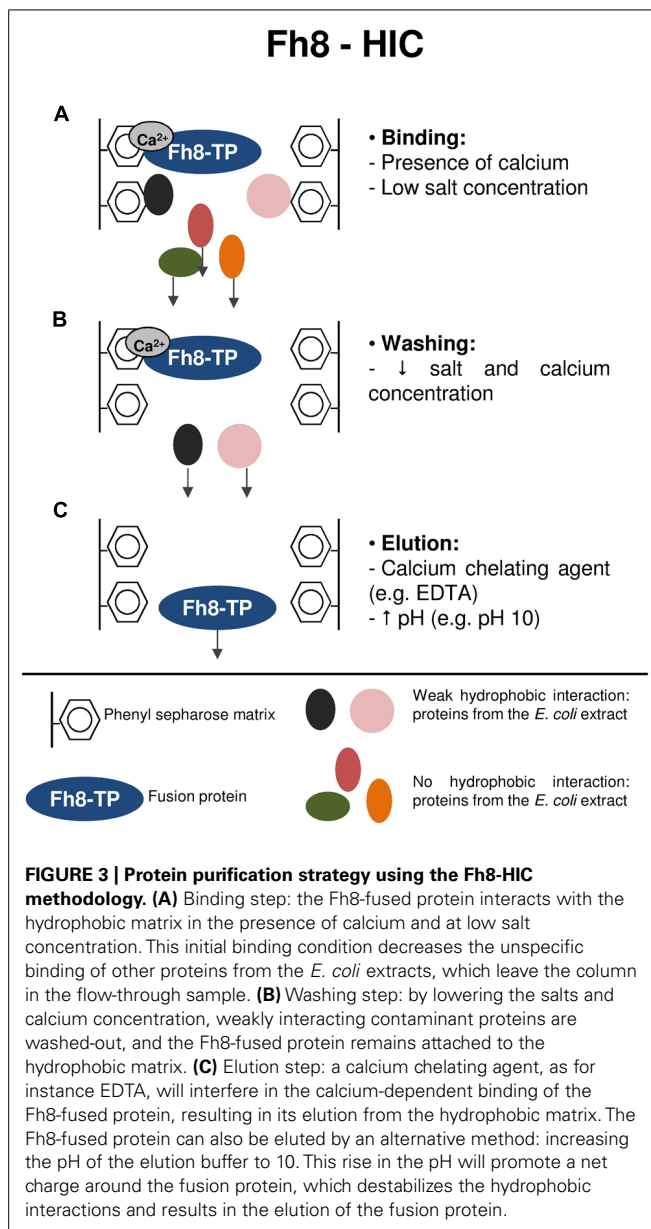
Figure 3 resumes the purification mechanism of target proteins using the Fh8-HIC strategy. As previously mentioned, the Fh8 is a Ca²⁺-sensor protein that opens its structure upon calcium accommodation. The opening of the Fh8's structure exposes a large hydrophobic surface that becomes available for interaction with its targets (Fraga et al., 2010). In this study, the Fh8 tag and Fh8-fused proteins presented a calcium-dependent interaction with a hydrophobic resin, and, as reported for other calcium-binding proteins (Rozanas, 1998; Shimizu et al., 2003), this interaction was still occurring even with low salt concentration in the mobile phase. The low salt concentration decreases the unspecific binding of other proteins from the *E. coli* extracts, thus promoting selectivity toward the purification of the fusion protein of interest (Costa et al., 2013b). Moreover, it was also shown that, as a calcium-binding protein, the Fh8 tag and Fh8-fused proteins can be eluted by using a calcium chelating agent, such as EDTA. One can also use for elution a mobile phase with an increased pH (e.g., pH 10), which creates a net charge

that destabilizes hydrophobic interactions. This elution strategy allows a single-step and rapid elution of all bound proteins (Costa et al., 2013b).

The Fh8-HIC methodology presented also the advantage of being compatible with the IMAC technique, thus, allowing a dual protein purification strategy that can be used sequentially, complementing each other, to obtain an active and more purified protein when desired. In addition, the use of two consecutive purification steps and the distinct nature of HIC and IMAC methodologies is known to help for the efficient removal of contaminating proteins (McCluskey et al., 2007).

Regarding the H tag, it did not function as a solubility enhancer tag, but it improved the production levels of target proteins in *E. coli* similarly to the Fh8 tag (Costa et al., 2013a). Taking that into account, the H tag was further explored for the recombinant production of antigens of interest in *E. coli*, and their subsequent immunization and polyclonal antibody production.

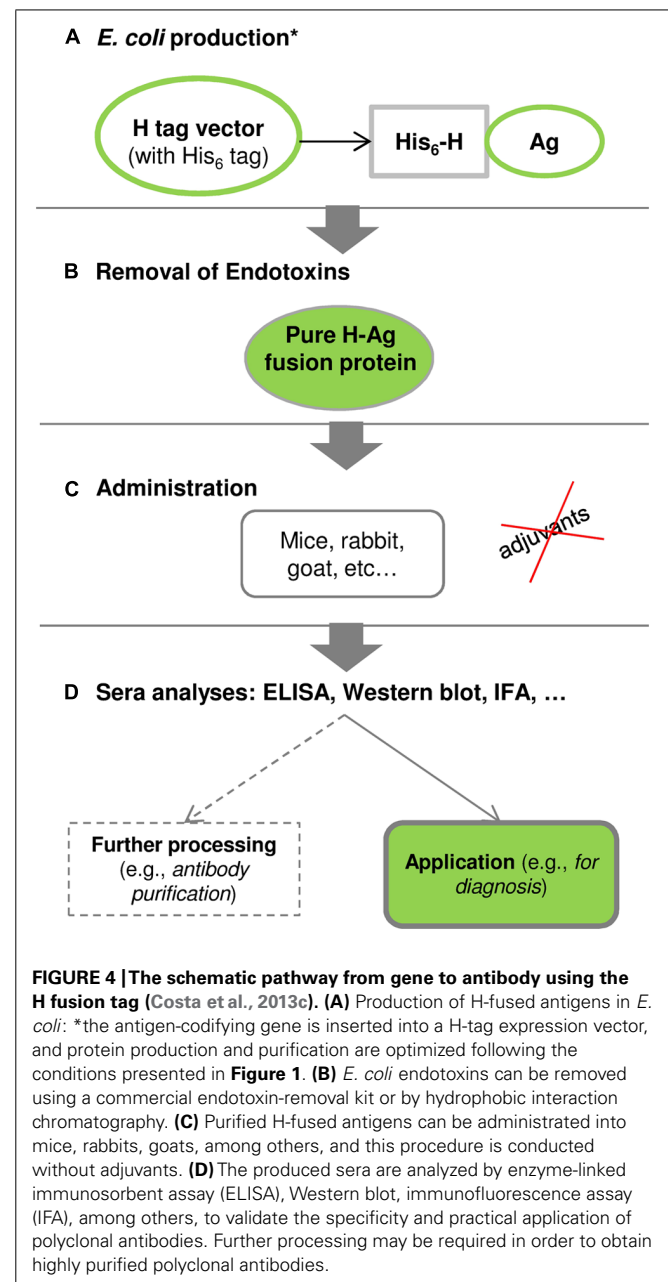
The major novelty of the H tag relies on its small size (1 kDa) combined with the adjuvant-free immunization of antigens (Conceição et al., 2011; Costa, 2013; Costa et al., 2013c). **Figure 4**



shows the schematic pathway of using the H fusion tag from gene to antibody.

Costa et al. (2013c) showed a successful case study with the CP12 antigen, which has a low molecular weight that can hinder the production of polyclonal antibodies. The HCP12 fusion antigen elicited an earlier immune response and higher (approximately 2-fold) polyclonal antibody titers than the non-fused CP12 (Conceição et al., 2011; Costa et al., 2013c). This application study demonstrated that the H partner improves the specific polyclonal antibody production against the CP12 antigen without using adjuvants, and the resulting polyclonal antibodies can be used as a diagnostic tool for immunodetection of *C. parvum* infections in humans or animals (Costa et al., 2013c).

Apart from CP12, several H-fused antigens have already been produced in *E. coli* (Conceição et al., 2010) and immunized



in mice and rabbits, such as, the human interleukin-5 (IL-5), the cyst wall protein-1 from *T. gondii* (TgOWP), the cyst wall protein from *Giardia lamblia* cysts (CWG), the β-giardin cytoskeletal protein of the ventral disk from the *G. lamblia* trophozoite (βG), the cyst wall specific-glycoprotein Jacob from *Entamoeba histolytica* (Ent), and the falcipain-1 trophozoite cysteine proteinase from *Plasmodium falciparum* (Pfsp), among others (Conceição et al., 2011).

CONCLUSIONS AND FUTURE TRENDS

The growing demand for effective health and environmental biotechnology resources has advancing the design of different strategies for the successful protein production in *E. coli*. Its

benefits of cost and ease of use and scale make *E. coli* one of the most widely used host systems for recombinant protein production, but one must be aware that success is not always guaranteed in this prokaryotic host system, mainly when working with recombinant proteins of human origin.

This review highlighted several key factors that contribute to the soluble protein production and purification in *E. coli*, including the use of different mutated host strains, co-production of chaperones and foldases and testing different cultivation conditions, with a main focus in the gene fusion technology.

The use of fusion partners was an important turning point for the *E. coli* host system: fusion tags promote or increase protein solubility, help on protein purification and can also be used to increase protein's immunogenicity. Traditional fusion systems like MBP, GST, NusA, or Trx have constantly been challenged and complemented by novel fusion solutions such as the SUMO tag (Butt et al., 2005; Marblestone et al., 2006), the HaloTag (Ohana et al., 2009), the SNUT tag (Caswell et al., 2010), and the expressivity tag (Hansted et al., 2011), among others.

More recently, a novel and unique fusion system for simple and inexpensive soluble protein overproduction and purification in *E. coli* was developed and studied: the Fh8 tag (Costa, 2013).

The Fh8 is ranked among the best solubility enhancer tags as Trx, MBP, or NusA (Costa et al., 2013a), and it offers a specific and simple purification of the target proteins by using its natural calcium-binding properties and mild conditions for HIC (Costa et al., 2013b). The Fh8 fusion partner is one of the few existing tags to promote simultaneously target protein solubility directly into the *E. coli* cytoplasm and a simple and cost-effective protein purification.

The novel Fh8 fusion system overcomes several issues related with recombinant protein production in *E. coli*: by using a straightforward methodology, this novel system increases protein production levels, promotes protein solubility and low cost purification, and helps for protein immunogenicity, in which the H tag facilitates a simple, rapid, and adjuvant-free production from gene to antibody (Costa et al., 2013c). This novel fusion system offers the great advantage of combining these four abilities into the two lowest molecular weight fusion partners described so far. Hence, the Fh8 fusion system appears as a valuable tool for the efficient and economical recombinant protein production in *E. coli*.

While this review applies to the use of Fh8 and H tags for recombinant protein production in bacterial host systems, it is hoped that the novel fusion system presented here will apply to other hosts, as for instance, eukaryotes and mammalian cells and thus, this must be investigated.

Despite being widely employed to improve soluble protein production in *E. coli*, fusion tags are not yet well comprehended as suggested by the general lack in literature of studies regarding their mechanism of action. Therefore, efforts should be taken to disclose how fusion tags work while promoting such a positive effect in the protein production in *E. coli*. Perhaps, a wide systems biology analysis can help to reveal the different pathways that fusion tags undergo in *E. coli*, leading also to their organization into functional groups.

Taking into account the broad range of applications, the trend is that the number of available fusion tags will increase, and the

understanding of their way of action will, undoubtedly, allow the development of tailored-made tools for protein production.

AUTHOR CONTRIBUTIONS

Sofia Costa drafted the review, participated in the study design of the novel Fh8 and H fusion tags, and in most of its experimental work. André Almeida, António Castro, and Lucília Domingues participated in the study design of the novel Fh8 and fusion tags. Lucília Domingues conceived and helped to draft the review. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Sofia Costa acknowledges support from Fundação para a Ciência e a Tecnologia (FCT), Portugal (by the fellowship SFRH/BD/46482/2008). The authors thank the FCT Strategic Project PEst-OE/EQB/LA0023/2013 and the Project “BioInd – Biotechnology and Bioengineering for improved Industrial and Agro-Food processes, REF. NORTE-07-0124-FEDER-000028” Co-funded by the Programa Operacional Regional do Norte (ON.2 – O Novo Norte), QREN, FEDER. The authors gratefully acknowledge Hüseyin Besir (from the EMBL, Heidelberg, Germany) for his constructive discussions and contribution throughout the study of the novel fusion tags.

REFERENCES

- Aatsinki, J. T., and Rajaniemi, H. J. (2005). An alternative use of basic pGEX vectors for producing both N- and C-terminal fusion proteins for production and affinity purification of antibodies. *Protein Expr. Purif.* 40, 287–291. doi: 10.1016/j.pep.2004.11.012
- Ahn, K. Y., Song, J. A., Hah, K. Y., Park, J. S., Seo, H. S., and Lee, J. (2007). Heterologous protein expression using a novel stress-responsive protein of *E. coli* RpoA as fusion expression partner. *Enzyme Microb. Technol.* 41, 859–866. doi: 10.1016/j.enzmictec.2007.07.009
- Andrade, F. K., Moreira, S. M. G., Domingues, L., and Gama, F. M. (2010a). Improving the affinity of fibroblasts for bacterial cellulose using carbohydrate-binding modules fused to RGD. *J. Biomed. Mater. Res. A* 92, 9–17. doi: 10.1002/jbm.a.32284
- Andrade, F. K., Costa, R., Domingues, L., Soares, R., and Gama, M. (2010b). Improving bacterial cellulose for blood vessel replacement: functionalization with a chimeric protein containing a cellulose-binding module and an adhesion peptide. *Acta Biomater.* 6, 4034–4041. doi: 10.1016/j.actbio.2010.04.023
- Arnau, J., Lauritzen, C., Petersen, G. E., and Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* 48, 1–13. doi: 10.1016/j.pep.2005.12.002
- Austin, C. (2003). Novel approach to obtain biologically active recombinant heterodimeric proteins in *Escherichia coli*. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 786, 93–107. doi: 10.1016/S1570-0232(02)00720-1
- Baca, A. M., and Hol, W. G. J. (2000). Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.* 30, 113–118. doi: 10.1016/S0020-7519(00)00019-9
- Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D. L., et al. (2001). *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J. Mol. Biol.* 312, 79–93. doi: 10.1006/jmbi.2001.4914
- Baneyx, F., and Palumbo, J. L. (2003). Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Methods Mol. Biol.* 205, 171–197. doi: 10.1385/1-59259-301-1:171
- Beekman, J. M., Cooney, A. J., Elliston, J. E., Tsai, S. Y., and Tsai, M. J. (1994). A rapid one-step method to purify baculovirus-expressed human estrogen-receptor to be used in the analysis of the oxytocin promoter. *Gene* 146, 285–289. doi: 10.1016/0378-1119(94)90307-7

- Berlec, A., and Strukelj, B. (2013). Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J. Industr. Microbiol. Biotechnol.* 40, 257–274. doi: 10.1007/s10295-013-1235-0
- Bessette, P. H., Aslund, F., Beckwith, J., and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13703–13708. doi: 10.1073/pnas.96.24.13703
- Betiku, E. (2006). Molecular chaperones involved in heterologous protein folding in *Escherichia coli*. *Biotechnol. Mol. Biol. Rev.* 1, 66–75.
- Bhattacharya, S., Bunick, C. G., and Chazin, W. J. (2004). Target selectivity in EF-hand calcium binding proteins. *Biochim. Biophys. Acta Mol. Cell Res.* 1742, 69–79. doi: 10.1016/j.bbamer.2004.09.002
- Bird, L. E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods* 55, 29–37. doi: 10.1016/j.ymeth.2011.08.002
- Buckle, A. M., Zahn, R., and Fersht, A. R. (1997). A structural model for GroEL-polypeptide recognition. *Proc. Natl. Acad. Sci. U.S.A.* 94, 3571–3575. doi: 10.1073/pnas.94.8.3571
- Bussow, K., Scheich, C., Sievert, V., Harttig, U., Schultz, J., Simon, B., et al. (2005). Structural genomics of human proteins: target selection and generation of a public catalogue of expression clones. *Microb. Cell Fact.* 4. doi: 10.1186/1475-2859-4-21
- Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* 43, 1–9. doi: 10.1016/j.pep.2005.03.016
- Cabrita, L. D., Dai, W. W., and Bottomley, S. P. (2006). A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* 6:12. doi: 10.1186/1472-6750-6-12
- Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and Delucas, L. J. (2003). His-tag impact on structure. *Acta Crystallogr. D Biol. Crystallogr.* 63, 295–301. doi: 10.1107/S0907444906052024
- Caswell, J., Snoddy, P., McMeel, D., Buick, R. J., and Scott, C. J. (2010). Production of recombinant proteins in *Escherichia coli* using an N-terminal tag derived from sortase. *Protein Expr. Purif.* 70, 143–150. doi: 10.1016/j.pep.2009.10.012
- Chatellier, J., Buckle, A. M., and Fersht, A. R. (1999). GroEL recognises sequential and non-sequential linear structural motifs compatible with extended beta-strands and alpha-helices. *J. Mol. Biol.* 292, 163–172. doi: 10.1006/jmbi.1999.3040
- Chazin, W. J. (2011). Relating form and function of EF-hand calcium binding proteins. *Acc. Chem. Res.* 44, 171–179. doi: 10.1021/ar100110d
- Cheng, Y., Gu, J. A., Wang, H. G., Yu, S., Liu, Y. Q., Ning, Y. L., et al. (2010). EspA is a novel fusion partner for expression of foreign proteins in *Escherichia coli*. *J. Biotechnol.* 150, 380–388. doi: 10.1016/j.jbiotec.2010.09.940
- Cheng, Y., and Patel, D. J. (2004). An efficient system for small protein expression and refolding. *Biochem. Biophys. Res. Commun.* 317, 401–405. doi: 10.1016/j.bbrc.2004.03.068
- Cheshyre, J. A., and Hipkiss, A. R. (1989). Low temperatures stabilize interferon α -2 against proteolysis in *Methylophilus methylotrophus* and *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 31, 158–162. doi: 10.1007/BF00262455
- Chin, D., and Means, A. R. (2000). Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* 10, 322–328. doi: 10.1016/S0962-8924(00)01800-6
- Choi, S. I., Song, H. W., Moon, J. W., and Seong, B. L. (2001). Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotechnol. Bioeng.* 75, 718–724. doi: 10.1002/bit.10082
- Chong, S. R., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., et al. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* 192, 271–281. doi: 10.1016/S0378-1119(97)00105-4
- Chou, C. P. (2007). Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 76, 521–532. doi: 10.1007/s00253-007-1039-0
- Collins, K. D., and Washabaugh, M. W. (1985). The Hofmeister effect and the behavior of water at interfaces. *Q. Rev. Biophys.* 18, 323–422. doi: 10.1017/S0033583500005369
- Conceição, M., Costa, S., Castro, A., and Almeida, A. (2010). *Fusion Proteins, Its Preparation Process and Its Application on Recombinant Protein Expression Systems*. WIPO Patent WO/2010/082097.
- Conceição, M., Costa, S., Castro, A., and Almeida, A. (2011). *Immunogens, Process for Preparation and Use in Systems of Polyclonal Antibodies Production*. WIPO Patent WO/2011/071404.
- Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M., and Colonna, R. J. (1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.* 265, 9062–9065.
- Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M., and Stier, G. (2008). Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Sci.* 17, 2070–2079. doi: 10.1110/ps.037564.108
- Costa, S. (2013). *Development of a Novel Fusion System for Recombinant Protein Production and Purification in Escherichia coli*. Ph.D. Thesis, University of Minho, Braga, Portugal.
- Costa, S. J., Almeida, A., Castro, A., Domingues, L., and Besir, H. (2013a). The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol.* 97, 6779–6791. doi: 10.1007/s00253-012-4559-1
- Costa, S. J., Coelho, E., Franco, L., Almeida, A., Castro, A., and Domingues, L. (2013b). The Fh8 tag: a fusion partner for simple and cost-effective protein purification in *Escherichia coli*. *Protein Expr. Purif.* 92, 163–170. doi: 10.1016/j.pep.2013.09.013
- Costa, S. J., Silva, P., Almeida, A., Conceição, M., Domingues, L., and Castro, A. (2013c). A novel adjuvant-free H fusion system for the production of recombinant immunogens in *Escherichia coli*: its application to a 12 kDa antigen from *Cryptosporidium parvum*. *Bioengineered* 4, 1–7. doi: 10.4161/bioe.26003
- Davis, G. D., Elisee, C., Newham, D. M., and Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* 65, 382–388. doi: 10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
- de Marco, A., and De Marco, V. (2004). Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. *J. Biotechnol.* 109, 45–52. doi: 10.1016/j.jbiotec.2003.10.025
- de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., and Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* 7:32. doi: 10.1186/1472-6750-7-32
- De Marco, V., Stier, G., Blandin, S., and De Marco, A. (2004). The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem. Biophys. Res. Commun.* 322, 766–771. doi: 10.1016/j.bbrc.2004.07.189
- DelProposto, J., Majmudar, C. Y., Smith, J. L., and Brown, W. C. (2009). Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* 63, 40–49. doi: 10.1016/j.pep.2008.08.011
- Demain, A. L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27, 297–306. doi: 10.1016/j.biotechadv.2009.01.008
- Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., et al. (2003). Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* 47, 1317–1328. doi: 10.1046/j.1365-2958.2003.03370.x
- di Guan, C., Li, P., Riggs, P. D., and Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* 67, 21–30. doi: 10.1016/0378-1119(88)90004-2
- Douette, P., Navet, R., Gerkens, P., Galleni, M., Levy, D., and Sluse, F. E. (2005). *Escherichia coli* fusion carrier proteins act as solubilizing agents for recombinant uncoupling protein 1 through interactions with GroEL. *Biochem. Biophys. Res. Commun.* 333, 686–693. doi: 10.1016/j.bbrc.2005.05.164
- Dummler, A., Lawrence, A. M., and De Marco, A. (2005). Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* 4, 34. doi: 10.1186/1475-2859-4-34
- Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., and McCafferty, J. (2004). Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol.* 4:32. doi: 10.1186/1472-6750-4-32
- Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* 49, 455–465. doi: 10.1016/S0165-022X(01)00213-5
- Esposito, D., and Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* 17, 353–358. doi: 10.1016/j.copbio.2006.06.003

- Ferrer-Mirallès, N., Domingo-Espin, J., Corchero, J. L., Vazquez, E., and Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* 8, 17. doi: 10.1186/1475-2859-8-17
- Fox, J. D., Kapust, R. B., and Waugh, D. S. (2001). Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* 10, 622–630. doi: 10.1110/ps.45201
- Fraga, H., Faria, T. Q., Pinto, F., Almeida, A., Brito, R. M. M., and Damas, A. M. (2010). FH8-a small EF-hand protein from *Fasciola hepatica*. *FEBS J.* 277, 5072–5085. doi: 10.1111/j.1742-4658.2010.07912.x
- Fritze, C. E., and Anderson, T. R. (2000). Epitope tagging: general method for tracking recombinant proteins. *Methods Enzymol.* 327, 3–16. doi: 10.1016/S0076-6879(00)27263-7
- Gaberc-Porekar, V., and Menart, V. (2001). Perspectives of immobilized-metal affinity chromatography. *J. Biochem. Biophys. Methods* 49, 335–360. doi: 10.1016/S0165-022X(01)00207-X
- Gao, X., Chen, W., Guo, C., Qian, C., Liu, G., Ge, F., et al. (2010). Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. *Appl. Microbiol. Biotechnol.* 85, 1051–1060. doi: 10.1007/s00253-009-2078-5
- Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., et al. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb. Cell Fact.* 7, 11. doi: 10.1186/1475-2859-7-11
- GE Healthcare. (2006). *Hydrophobic Interaction and Reversed Phase Chromatography: Principles and Methods*. Uppsala: GE Healthcare.
- GE Healthcare. (2010). *Strategies for Protein Purification: Handbook*. Uppsala: GE Healthcare.
- Guerreiro, C. I., Fontes, C. M., Gama, M., and Domingues, L. (2008). *Escherichia coli* expression and purification of four antimicrobial peptides fused to a family 3 carbohydrate-binding module (CBM) from *Clostridium thermocellum*. *Protein Expr. Purif.* 59, 161–168. doi: 10.1016/j.pep.2008.01.018
- Hakes, D. J., and Dixon, J. E. (1992). New vectors for high-level expression of recombinant proteins in bacteria. *Anal. Biochem.* 202, 293–298. doi: 10.1016/0003-2697(92)90108-J
- Hammarström, M. (2006). Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J. Struct. Funct. Genomics* 7, 1–14. doi: 10.1007/s10969-005-9003-7
- Hammarström, M., Hellgren, N., Van Den Berg, S., Berglund, H., and Hard, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* 11, 313–321. doi: 10.1110/ps.22102
- Han, K. Y., Seo, H. S., Song, J. A., Ahn, K. Y., Park, J. S., and Lee, J. (2007a). Transport proteins PotD and Crr of *Escherichia coli*, novel fusion partners for heterologous protein expression. *Biochim. Biophys. Acta Proteins Proteomics* 1774, 1536–1543. doi: 10.1016/j.bbapap.2007.09.012
- Han, K. Y., Song, J. A., Ahn, K. Y., Park, J. S., Seo, H. S., and Lee, J. (2007b). Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. *FEMS Microbiol. Lett.* 274, 132–138. doi: 10.1111/j.1574-6968.2007.00824.x
- Han, K. Y., Song, J. A., Ahn, K. Y., Park, J. S., Seo, H. S., and Lee, J. (2007c). Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive *Escherichia coli* protein, SlyD. *Protein Eng. Des. Select.* 20, 543–549. doi: 10.1093/protein/gzm055
- Hansted, J. G., Pietikainen, L., Hog, F., Sperling-Petersen, H. U., and Mortensen, K. K. (2011). Expressivity tag: a novel tool for increased expression in *Escherichia coli*. *J. Biotechnol.* 155, 275–283. doi: 10.1016/j.jbiotec.2011.07.013
- Hayashi, K., and Kojima, C. (2008). pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. *Protein Expr. Purif.* 62, 120–127. doi: 10.1016/j.pep.2008.07.007
- Hoffmann, F., and Rinas, U. (2004). Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.* 89, 143–161. doi: 10.1007/b93996
- Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., et al. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Biotechnology* 6, 1204–1210. doi: 10.1038/nbt1088-1204
- Hu, J., Qin, H. J., Sharma, M., Cross, T. A., and Gao, F. P. (2008). Chemical cleavage of fusion proteins for high-level production of transmembrane peptides and protein domains containing conserved methionines. *Biochim. Biophys. Acta Biomembr.* 1778, 1060–1066. doi: 10.1016/j.bbamem.2007.12.024
- Huang, Y. S., and Chuang, D. T. (1999). Mechanisms for GroEL/GroES-mediated folding of a large 86-kDa fusion polypeptide in vitro. *J. Biol. Chem.* 274, 10405–10412. doi: 10.1074/jbc.274.15.10405
- Hwang, S. M., Kang, H. J., Bae, S. W., Chang, W. J., and Koo, Y. M. (2010). Refolding of lysozyme in hydrophobic interaction chromatography: effects of hydrophobicity of adsorbent and salt concentration in mobile phase. *Biotechnol. Bioproc. Eng.* 15, 213–219. doi: 10.1007/s12257-009-0216-7
- Inouye, S., and Sahara, Y. (2009). Expression and purification of the calcium binding photoprotein mitrocomin using ZZ-domain as a soluble partner in *E. coli* cells. *Protein Expr. Purif.* 66, 52–57. doi: 10.1016/j.pep.2009.01.010
- Jana, S., and Deb, J. K. (2005). Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 67, 289–298. doi: 10.1007/s00253-004-1814-0
- Jenny, R. J., Mann, K. G., and Lundblad, R. L. (2003). A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* 31, 1–11. doi: 10.1016/S1046-5928(03)00168-2
- Kamezaki, Y., Enomoto, C., Ishikawa, Y., Koyama, T., Naya, S., Suzuki, T., et al. (2010). The Dock tag, an affinity tool for the purification of recombinant proteins, based on the interaction between dockerin and cohesin domains from *Clostridium josui* cellulosome. *Protein Expr. Purif.* 70, 23–31. doi: 10.1016/j.pep.2009.09.024
- Kamionka, M. (2011). Engineering of therapeutic proteins production in *Escherichia coli*. *Curr. Pharm. Biotechnol.* 12, 268–274. doi: 10.2174/138920111794295693
- Kaplan, W., Husler, P., Klump, H., Erhardt, J., Sluiscremer, N., and Dirr, H. (1997). Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* 6, 399–406. doi: 10.1002/pro.5560060216
- Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., et al. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* 14, 993–1000. doi: 10.1093/protein/14.12.993
- Kapust, R. B., and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668–1674. doi: 10.1110/ps.8.8.1668
- Kapust, R. B., and Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* 19, 312–318. doi: 10.1006/prep.2000.1251
- Kawabe, Y., Seki, M., Seki, T., Wang, W. S., Imamura, O., Furuichi, Y., et al. (2000). Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. *J. Biol. Chem.* 275, 20963–20966. doi: 10.1074/jbc.C000273200
- Ketterer, B. (2001). A bird's eye view of the glutathione transferase field. *Chem. Biol. Interact.* 138, 27–42. doi: 10.1016/S0009-2797(01)00277-0
- Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* 3, 193–205. doi: 10.1038/nsb0296-193
- Kim, S., and Lee, S. B. (2008). Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. *Protein Expr. Purif.* 62, 116–119. doi: 10.1016/j.pep.2008.06.015
- Kimple, M. E., and Sondek, J. (2004). Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* doi: 10.1002/0471140864.ps0909s36
- Kishi, A., Nakamura, T., Nishio, Y., Maegawa, H., and Kashiwagi, A. (2003). Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation. *Am. J. Physiol. Endocrinol. Metab.* 284, E830–E840.
- Knappik, A., and Pluckthun, A. (1994). An improved affinity tag based on the Flag(R) peptide for the detection and purification of recombinant antibody fragments. *Biotechniques* 17, 754–761.
- Kohl, T., Schmidt, C., Wiemann, S., Poustka, A., and Korf, U. (2008). Automated production of recombinant human proteins as resource for proteome research. *Proteome Sci.* 6. doi: 10.1186/1477-5956-6-4
- Kolaj, O., Spada, S., Robin, S., and Wall, J. G. (2009). Use of folding modulators to improve heterologous protein production in *Escherichia coli*. *Microb. Cell Fact.* 8. Kuczynska-Wisnik, D., Kedzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipinska, B., et al. (2002). The *Escherichia coli* small heat-shock proteins IbpA and IbpB

- prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. *Microbiol. SGM* 148, 1757–1765.
- Lavallie, E. R., Diblasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *Escherichia coli* cytoplasm. *Biotechnology* 11, 187–193. doi: 10.1038/nbt0293-187
- LaVallie, E. R., Lu, Z. J., Diblasio-Smith, E. A., Collins-Racie, L. A., and McCoy, J. M. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Appl. Chimeric Genes Hybrid Proteins Pt A* 326, 322–340. doi: 10.1016/S0076-6879(00)26063-1
- Lee, C. D., Sun, H. C., Hu, S. M., Chiu, C. F., Homhuan, A., Liang, S. M., et al. (2008). An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci.* 17, 1241–1248. doi: 10.1110/ps.035188.108
- Lewit-Bentley, A., and Rety, S. (2000). EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol.* 10, 637–643. doi: 10.1016/S0959-440X(00)00142-1
- Li, Y. F. (2010). Commonly used tag combinations for tandem affinity purification. *Biotechnol. Appl. Biochem.* 55, 73–83. doi: 10.1042/BA20090273
- Li, Y. F. (2011). Self-cleaving fusion tags for recombinant protein production. *Biotechnol. Lett.* 33, 869–881. doi: 10.1007/s10529-011-0533-8
- Lienqueo, M. E., Mahn, A., Salgado, J. C., and Asenjo, J. A. (2007). Current insights on protein behaviour in hydrophobic interaction chromatography. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 849, 53–68. doi: 10.1016/j.jchromb.2006.11.019
- Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* 11, 56. doi: 10.1186/1475-2859-11-56
- Lopez, P. J., Marchand, I., Joyce, S. A., and Dreyfus, M. (1999). The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol. Microbiol.* 33, 188–199. doi: 10.1046/j.1365-2958.1999.01465.x
- Ly, Z. Y., Yang, L. L., Hu, S. M., Sun, X., He, H. J., He, S. J., et al. (2009). Expression profile, localization of an 8-kDa calcium-binding protein from *Schistosoma japonicum* (SjCa8), and vaccine potential of recombinant SjCa8 (rSjCa8) against infections in mice. *Parasitol. Res.* 104, 733–743. doi: 10.1007/s00436-008-1249-0
- Magalhães, P. O., Lopes, A. M., Mazzola, P. G., Rangel-Yagui, C., Penna, T. C. V., et al. (2007). Methods of endotoxin removal from biological preparations: a review. *J. Pharm. Pharm. Sci.* 10, 388–404.
- Makino, T., Skretas, G., and Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microb. Cell Fact.* 10, 32. doi: 10.1186/1475-2859-10-32
- Malakhov, M. P., Mattern, M., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* 5, 75–86. doi: 10.1023/B:JSFG.0000029237.70316.52
- Malhotra, A. (2009). “Tagging for protein expression,” in *Guide to Protein Purification*, 2nd Edn, eds R. R. Burgess and M. P. Deutscher (San Diego, CA: Elsevier), 463, 239–258. doi: 10.1016/S0076-6879(09)63016-0
- Malik, A., Jenzsch, M., Lubbert, A., Rudolph, R., and Sohling, B. (2007). Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin. *Protein Expr. Purif.* 55, 100–111. doi: 10.1016/j.pep.2007.04.006
- Malik, A., Rudolph, R., and Sohling, B. (2006). A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. *Protein Expr. Purif.* 47, 662–671. doi: 10.1016/j.pep.2006.02.018
- Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* 15, 182–189. doi: 10.1110/ps.051812706
- McCluskey, A. J., Poon, G. M. K., and Gariepy, J. (2007). A rapid and universal tandem-purification strategy for recombinant proteins. *Protein Sci.* 16, 2726–2732. doi: 10.1110/ps.072894407
- McTigue, M. A., Williams, D. R., and Tainer, J. A. (1995). Crystal structures of a schistosomal drug and vaccine target: glutathione-S-transferase from *Schistosoma japonica* and its complex with the leading anti-schistosomal drug praziquantel. *J. Mol. Biol.* 246, 21–27. doi: 10.1006/jmbi.1994.0061
- Mee, C., Banki, M. R., and Wood, D. W. (2008). Towards the elimination of chromatography in protein purification: expressing proteins engineered to purify themselves. *Chem. Eng. J.* 135, 56–62. doi: 10.1016/j.cej.2007.04.021
- Miroux, B., and Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298. doi: 10.1006/jmbi.1996.0399
- Mitchell, D. A., Marshall, T. K., and Deschenes, R. J. (1993). Vectors for the inducible overexpression of glutathione-S-transferase fusion proteins in yeast. *Yeast* 9, 715–722. doi: 10.1002/yea.320090705
- Moreira, S. M. G., Andrade, F. K., Domingues, L., and Gama, F. M. (2008). Development of a strategy to functionalize a starch-based hydrogel for animal cell cultures using a starch-binding module fused to RGD sequence. *BMC Biotechnol.* 8:78. doi: 10.1186/1472-6750-8-78
- Motejaded, H., and Altenbuchner, J. (2009). Construction of a dual-tag system for gene expression, protein affinity purification and fusion protein processing. *Biotechnol. Lett.* 31, 543–549. doi: 10.1007/s10529-008-9909-9
- Mujacic, M., Cooper, K. W., and Baneyx, F. (1999). Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein. *Gene* 238, 325–332. doi: 10.1016/S0378-1119(99)00328-5
- Nallamsetty, S., Austin, B. P., Penrose, K. J., and Waugh, D. S. (2005). Gateway vectors for the production of combinatorially-tagged His(6)-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* 14, 2964–2971. doi: 10.1110/ps.051718605
- Nallamsetty, S., and Waugh, D. S. (2006). Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* 45, 175–182. doi: 10.1016/j.pep.2005.06.012
- Nallamsetty, S., and Waugh, D. S. (2007). Mutations that alter the equilibrium between open and closed conformations of *Escherichia coli* maltose-binding protein impede its ability to enhance the solubility of passenger proteins. *Biochem. Biophys. Res. Commun.* 364, 639–644. doi: 10.1016/j.bbrc.2007.10.060
- Nelson, M. R., and Chazin, W. J. (1998). An interaction-based analysis of calcium-induced conformational changes in Ca^{2+} sensor proteins. *Protein Sci.* 7, 270–282. doi: 10.1002/pro.5560070206
- Nettleship, J. E., Assenberg, R., Diprose, J. M., Rahman-Huq, N., and Owens, R. J. (2010). Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J. Struct. Biol.* 172, 55–65. doi: 10.1016/j.jsb.2010.02.006
- Nikaido, H. (1994). Maltose transport-system of *Escherichia coli*: an ABC-type transporter. *FEBS Lett.* 346, 55–58. doi: 10.1016/0014-5793(94)00315-7
- Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998). Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 1694–1699.
- Ohana, R. F., Encell, L. P., Zhao, K., Simpson, D., Slater, M. R., Urh, M., et al. (2009). HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr. Purif.* 68, 110–120. doi: 10.1016/j.pep.2009.05.010
- Oliveira, C., Felix, W., Moreira, R. A., Teixeira, J. A., and Domingues, L. (2008). Expression of frutalin a alpha-D-galactose-binding jacalin-related lectin, in the yeast *Pichia pastoris*. *Protein Expr. Purif.* 60, 188–193. doi: 10.1016/j.pep.2008.04.008
- Oliveira, C., Costa, S., Teixeira, J. A., and Domingues, L. (2009a). cDNA cloning and functional expression of the alpha-D-galactose-binding lectin frutalin in *Escherichia coli*. *Mol. Biotechnol.* 43, 212–220. doi: 10.1007/s12033-009-9191-7
- Oliveira, C., Teixeira, J. A., Schmitt, F., and Domingues, L. (2009b). A comparative study of recombinant and native frutalin binding to human prostate tissues. *BMC Biotechnol.* 9:78. doi: 10.1186/1472-6750-9-78
- Oliveira, C., Nicolau, A., Teixeira, J. A., and Domingues, L. (2011). Cytotoxic effects of native and recombinant frutalin, a plant galactose-binding lectin, on HeLa cervical cancer cells. *J. Biomed. Biotechnol.* 2011, 568932. doi: 10.1155/2011/568932
- Ongkudon, C. M., Chew, J. H., Liu, B., and Danquah, M. K. (2012). Chromatographic removal of endotoxins: a bioprocess engineer's perspective. *ISRN Chromatogr.* 649746. doi: 10.5402/2012/649746
- Pacheco, B., Crombet, L., Loppnau, P., and Cossar, D. (2012). A screening strategy for heterologous protein expression in *Escherichia coli* with the highest

- return of investment. *Protein Expr. Purif.* 81, 33–41. doi: 10.1016/j.pep.2011.08.030
- Panavas, T., Sanders, C., and Butt, T. R. (2009). SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* 303–317. doi: 10.1007/978-1-59745-566-4_20
- Park, J. S., Han, K. Y., Lee, J. H., Song, J. A., Ahn, K. Y., Seo, H. S., et al. (2008). Solubility enhancement of aggregation-prone heterologous proteins by fusion expression using stress-responsive *Escherichia coli* protein, RpoS. *BMC Biotechnol.* 8:15. doi: 10.1186/1472-6750-8-15
- Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994). Release of proteins and peptides from fusion proteins using a recombinant plant-virus proteinase. *Anal. Biochem.* 216, 413–417. doi: 10.1006/abio.1994.1060
- Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., et al. (1994). Protein splicing elements: inteins and exteins: a definition of terms and recommended nomenclature. *Nucleic Acids Res.* 22, 1125–1127. doi: 10.1093/nar/22.7.1125
- Pértile, R., Moreira, S., Andrade, F., Domingues, L., and Gama, M. (2012). Bacterial cellulose modified using recombinant proteins to improve neuronal and mesenchymal cell adhesion. *Biotechnol. Prog.* 28, 526–532. doi: 10.1002/btpr.1501
- Pryor, K. D., and Leiting, B. (1997). High-level expression of soluble protein in *Escherichia coli* using a His(6)-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* 10, 309–319. doi: 10.1006/prep.1997.0759
- Queiroz, J. A., Tomaz, C. T., and Cabral, J. M. S. (2001). Hydrophobic interaction chromatography of proteins. *J. Biotechnol.* 87, 143–159. doi: 10.1016/S0168-1656(01)00237-1
- Ram, D., Grossman, Z., Markovics, A., Avivi, A., Ziv, E., Lantner, F., et al. (1989). Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 34, 167–175. doi: 10.1016/0166-6851(89)90008-X
- Ramos, R., Domingues, L., and Gama, F. M. (2010). *Escherichia coli* expression and purification of LL37 fused to a family III carbohydrate-binding module from *Clostridium thermocellum*. *Protein Expr. Purif.* 71, 1–7. doi: 10.1016/j.pep.2009.10.016
- Ramos, R., Moreira, S., Rodrigues, A., Gama, M., and Domingues, L. (2013). Recombinant expression and purification of the antimicrobial peptide magainin-2. *Biotechnol. Prog.* 29, 17–22. doi: 10.1002/btpr.1650
- Reddi, H., Bhattacharya, A., and Kumar, V. (2002). The calcium-binding protein of *Entamoeba histolytica* as a fusion partner for expression of peptides in *Escherichia coli*. *Biotechnol. Appl. Biochem.* 36, 213–218. doi: 10.1042/BA20020024
- Rocco, C. J., Dennison, K. L., Klenchin, V. A., Rayment, I., and Escalante-Semerena, J. C. (2008). Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. *Plasmid* 59, 231–237. doi: 10.1016/j.plasmid.2008.01.001
- Ron, D., and Dressler, H. (1992). pGSTag: a versatile bacterial expression plasmid for enzymatic labeling of recombinant proteins. *Biotechniques* 13, 866–869.
- Rondahl, H., Nilsson, B., and Holmgren, E. (1992). Fusions to the 5' end of a gene encoding a 2-domain analog of *Staphylococcal* protein-A. *J. Biotechnol.* 25, 269–287. doi: 10.1016/0168-1656(92)90161-2
- Rozanas, C. (1998). Purification of calcium-binding proteins using hydrophobic interaction chromatography. *Life Science News I, Amersham Biosciences*.
- Rudert, F., Visser, E., Gradl, G., Grandison, P., Shemshadini, L., Wang, Y., et al. (1996). pLEF, a novel vector for expression of glutathione-S-transferase fusion proteins in mammalian cells. *Gene* 169, 281–282. doi: 10.1016/0378-1119(95)00820-9
- Sachdev, D., and Chirgwin, J. M. (2000). Fusions to maltose-binding protein: control of folding and solubility in protein purification. *Appl. Chimeric Genes Hybrid Proteins A*, 326, 312–321. doi: 10.1016/S0076-6879(00)26062-X
- Satakarni, M., and Curtis, R. (2011). Production of recombinant peptides as fusions with SUMO. *Protein Expr. Purif.* 78, 113–119. doi: 10.1016/j.pep.2011.04.015
- Scheich, C., Sievert, V., and Bussow, K. (2003). An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnol.* 3:12. doi: 10.1186/1472-6750-3-12
- Schlieker, C., Bukau, B., and Mogk, A. (2002). Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. *J. Biotechnol.* 96, 13–21. doi: 10.1016/S0168-1656(02)00033-0
- Schmidt, T. G. M., and Skerra, A. (1994). One-step affinity purification of bacterially produced proteins by means of the strep tag and immobilized recombinant core streptavidin. *J. Chromatogr. A* 676, 337–345. doi: 10.1016/0021-9673(94)80434-6
- Schumann, W., and Ferreira, L. C. S. (2004). Production of recombinant proteins in *Escherichia coli*. *Genet. Mol. Biol.* 27, 442–453. doi: 10.1590/S1415-47572004000300022
- Schwaller, B. (2010). Cytosolic Ca²⁺ buffers. *Cold Spring Harb. Perspect. Biol.* 2: a004051. doi: 10.1101/cshperspect.a004051
- Sevastyanovich, Y. R., Alfasi, S. N., and Cole, J. A. (2010). Sense and nonsense from a systems biology approach to microbial recombinant protein production. *Biotechnol. Appl. Biochem.* 55, 9–28. doi: 10.1042/BA20090174
- Sheibani, N. (1999). Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins. *Prep. Biochem. Biotechnol.* 29, 77–90. doi: 10.1080/10826069908544695
- Shih, Y. P., Kung, W. M., Chen, J. C., Yeh, C. H., Wang, A. H. J., and Wang, T. F. (2002). High-throughput screening of soluble recombinant proteins. *Protein Sci.* 11, 1714–1719. doi: 10.1110/ps.0205202
- Shimizu, F., Sanada, K., and Fukada, Y. (2003). Purification and immunohistochemical analysis of calcium-binding proteins expressed in the chick pineal gland. *J. Pineal Res.* 34, 208–216. doi: 10.1034/j.1600-079X.2003.00031.x
- Silva, E., Castro, A., Lopes, A., Rodrigues, A., Dias, C., Conceicao, A., et al. (2004). A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. *J. Parasitol.* 90, 746–751. doi: 10.1645/GE-136R
- Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* 67, 31–40. doi: 10.1016/0378-1119(88)90005-4
- Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* 12, 1313–1322. doi: 10.1110/ps.0243403
- Sommer, B., Friehs, K., Flaschel, E., Reck, M., Stahl, F., and Scheper, T. (2009). Extracellular production and affinity purification of recombinant proteins with *Escherichia coli* using the versatility of the maltose binding protein. *J. Biotechnol.* 140, 194–202. doi: 10.1016/j.jbiotec.2009.01.010
- Song, J. A., Lee, D. S., Park, J. S., Han, K. Y., and Lee, J. (2011). A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. *Enzyme Microb. Technol.* 49, 124–130. doi: 10.1016/j.enzmictec.2011.04.013
- Sorensen, H. P., and Mortensen, K. K. (2005a). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115, 113–128. doi: 10.1016/j.jbiotec.2004.08.004
- Sorensen, H. P., and Mortensen, K. K. (2005b). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* 4.
- Sorensen, H. P., Sperling-Petersen, H. U., and Mortensen, K. K. (2003a). A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr. Purif.* 32, 252–259. doi: 10.1016/j.pep.2003.07.001
- Sorensen, H. P., Sperling-Petersen, H. U., and Mortensen, K. K. (2003b). Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the bottleneck. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 786, 207–214. doi: 10.1016/S1570-0232(02)00689-X
- Stahl, S., and Nygren, P. A. (1997). The use of gene fusions to protein A and protein G in immunology and biotechnology. *Pathol. Biol.* 45, 66–76.
- Stewart, E. J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *EMBO J.* 17, 5543–5550. doi: 10.1093/emboj/17.19.5543
- Su, Y., Zou, Z. R., Feng, S. Y., Zhou, P., and Cao, L. J. (2007). The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *J. Biotechnol.* 129, 373–382. doi: 10.1016/j.jbiotec.2007.01.015
- Takakura, Y., Oka, N., Kajiwar, H., Tsunashima, M., Usami, S., Tsukamoto, H., et al. (2010). Tamavidin, a versatile affinity tag for protein purification and immobilization. *J. Biotechnol.* 145, 317–322. doi: 10.1016/j.jbiotec.2009.12.012
- Takakura, Y., Sofuku, K., and Tsunashima, M. (2013). Tamavidin 2-REV: an engineered tamavidin with reversible biotin-binding capability. *J. Biotechnol.* 164, 19–25. doi: 10.1016/j.jbiotec.2013.01.006
- Tanaka, N., and Fersht, A. R. (1999). Identification of substrate binding site of GroEL minichaperone in solution. *J. Mol. Biol.* 292, 173–180. doi: 10.1006/jmbi.1999.3041

- Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533. doi: 10.1007/s00253-002-1158-6
- Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 72, 211–222. doi: 10.1007/s00253-006-0465-8
- Thomas, J. G., Ayling, A., and Baneyx, F. (1997). Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*: to fold or to refold. *Appl. Biochem. Biotechnol.* 66, 197–238. doi: 10.1007/BF02785589
- Tomme, P., Boraston, A., Mclean, B., Kormos, J., Creagh, A. L., Sturch, K., et al. (1998). Characterization and affinity applications of cellulose-binding domains. *J. Chromatogr. B* 715, 283–296. doi: 10.1016/S0378-4347(98)00053-X
- Turner, P., Holst, O., and Karlsson, E. N. (2005). Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr. Purif.* 39, 54–60. doi: 10.1016/j.pep.2004.09.012
- Vaillancourt, P., Zheng, C. F., Hoang, D. Q., and Briester, L. (2000). Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides. *Appl. Chimic Genes Hybrid Proteins Pt A* 326, 340–362. doi: 10.1016/S0076-6879(00)26064-3
- Vergis, J. M., and Wiener, M. C. (2011). The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expr. Purif.* 78, 139–142. doi: 10.1016/j.pep.2011.04.011
- Viljanen, J., Larsson, J., and Broo, K. S. (2008). Orthogonal protein purification: expanding the repertoire of GST fusion systems. *Protein Expr. Purif.* 57, 17–26. doi: 10.1016/j.pep.2007.09.011
- Vinckier, N. K., Chworos, A., and Parsons, S. M. (2011). Improved isolation of proteins tagged with glutathione-S-transferase. *Protein Expr. Purif.* 75, 161–164. doi: 10.1016/j.pep.2010.09.006
- Walsh, G. (2010). Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* 28, 917–924. doi: 10.1038/nbt0910-917
- Wang, H. Y., Xiao, Y. C., Fu, L. J., Zhao, H. X., Zhang, Y. F., Wan, X. S., et al. (2010). High-level expression and purification of soluble recombinant FGF21 protein by SUMO fusion in *Escherichia coli*. *BMC Biotechnol.* 10:14. doi: 10.1186/1472-6750-10-14
- Wang, Z. Y., Li, N., Wang, Y. Y., Wu, Y. P., Mu, T. Y., Zheng, Y., et al. (2012). Ubiquitin-intein and SUMO2-intein fusion systems for enhanced protein production and purification. *Protein Expr. Purif.* 82, 174–178. doi: 10.1016/j.pep.2011.11.017
- Waugh, D. S. (2005). Making the most of affinity tags. *Trends Biotechnol.* 23, 316–320. doi: 10.1016/j.tibtech.2005.03.012
- Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* 80, 283–293. doi: 10.1016/j.pep.2011.08.005
- Wilson, M. J., Haggart, C. L., Gallagher, S. P., and Walsh, D. (2001). Removal of tightly bound endotoxin from biological products. *J. Biotechnol.* 88, 67–75. doi: 10.1016/S0168-1656(01)00256-5
- Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., and Ishii, S. (1995). Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* 270, 25328–25331. doi: 10.1074/jbc.270.43.25328
- Young, C. L., Britton, Z. T., and Robinson, A. S. (2012). Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* 7, 620–634. doi: 10.1002/biot.201100155
- Zhang, Y. B., Howitt, J., Mccorkle, S., Lawrence, P., Springer, K., and Freimuth, P. (2004). Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* 36, 207–216. doi: 10.1016/j.pep.2004.04.020
- Zhang, Y. J., and Cremer, P. S. (2006). Interactions between macromolecules and ions: the Hofmeister series. *Curr. Opin. Chem. Biol.* 10, 658–663. doi: 10.1016/j.cbpa.2006.09.020
- Zhou, P., Lugovskoy, A. A., and Wagner, G. (2001). A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR* 20, 11–14. doi: 10.1023/A:1011258906244
- Zhou, Y., Yang, W., Kirberger, M., Lee, H. W., Ayalasomayajula, G., and Yang, J. J. (2006). Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins. *Proteins* 65, 643–655. doi: 10.1002/prot.21139
- Zou, Z. R., Cao, L. J., Zhou, P., Su, Y., Sun, Y. H., and Li, W. J. (2008). Hyperacidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J. Biotechnol.* 135, 333–339. doi: 10.1016/j.jbiotec.2008.05.007

Conflict of Interest Statement: The Fh8 tag utilization for the improvement of protein production in *E. coli* and the H tag utilization for the production of immunogens and corresponding polyclonal antibodies are covered by worldwide patents (WO 2010082097 and WO 2011071404, respectively), both licensed to Hitag Biotechnology, Lda. The authors Sofia Costa, André Almeida, and António Castro are co-owners of the patent and are associated with Hitag Biotechnology, Lda.

Received: 29 November 2013; accepted: 30 January 2014; published online: 19 February 2014.

Citation: Costa S, Almeida A, Castro A and Domingues L (2014) Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front. Microbiol.* 5:63. doi: 10.3389/fmicb.2014.00063

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Costa, Almeida, Castro and Domingues. 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.



Generation of a vector suite for protein solubility screening

Agustín Correa¹, Claudia Ortega¹, Gonzalo Obal², Pedro Alzari³, Renaud Vincentelli⁴ and Pablo Oppezzo^{1*}

¹ Recombinant Protein Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay

² Protein Biophysics Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay

³ Unité de Microbiologie Structurale, Institut Pasteur, Paris, France

⁴ Centre National de la Recherche Scientifique, Aix-Marseille Université, CNRS UMR7257, AFMB, Marseille, France

Edited by:

Eduardo A. Ceccarelli, Universidad Nacional de Rosario, Argentina

Reviewed by:

Jun-Jie Zhang, Chinese Academy of Sciences, China

Grzegorz Węgrzyn, University of Gdansk, Poland

*Correspondence:

Pablo Oppezzo, Recombinant Protein Unit, Institut Pasteur de Montevideo, Mataojo 2020, Montevideo 11400, Uruguay
e-mail: poppezzo@pasteur.edu.uy

Recombinant protein expression has become an invaluable tool for academic and biotechnological projects. With the use of high-throughput screening technologies for soluble protein production, uncountable target proteins have been produced in a soluble and homogeneous state enabling the realization of further studies. Evaluation of hundreds of conditions requires the use of high-throughput cloning and screening methods. Here we describe a new versatile vector suite dedicated to the expression improvement of recombinant proteins (RP) with solubility problems. This vector suite allows the parallel cloning of the same PCR product into the 12 different expression vectors evaluating protein expression under different promoter strength, different fusion tags as well as different solubility enhancer proteins. Additionally, we propose the use of a new fusion protein which appears to be a useful solubility enhancer. Above all we propose in this work an economic and useful vector suite to fast track the solubility of different RP. We also propose a new solubility enhancer protein that can be included in the evaluation of the expression of RP that are insoluble in classical expression conditions.

Keywords: recombinant proteins, solubility, expression, vector, cloning, high-throughput

INTRODUCTION

Recombinant protein production has become a routine practice in many laboratories from academic to industrial fields. Several hosts are available for protein production among them, *Escherichia coli* has been by far the most widely used. Some advantages of this host is the low cost, infrastructure of implementation, easy handling, high yield production, and an ever increasing set of tools and genetic information useful for the expression of challenging targets. Despite its importance and utility, recombinant proteins (RP) not always are produced in a soluble and homogeneous state. For these “difficult to express” proteins, several approaches have been developed in order to overcome the problems associated with insolubility. Some parameters that can affect protein expression are: induction temperature, promoter strength, use of specific *E. coli* strains, co-expression of molecular chaperones or biological partners and the use of different solubility enhancer or fusion proteins (Correa and Oppezzo, 2011). In the last decade, the advent of high-throughput screening methods have facilitated the evaluation of hundreds of conditions generated from the combination of the mentioned parameters in order to find one that gives a soluble protein (Vincentelli et al., 2011; Vincentelli and Romier, 2013). However, to exploit all these variables it is necessary to have a method for cloning the target gene in many different vectors in a fast and simple manner. Several techniques were recently generated to facilitate the cloning of target genes in a parallel way, in which the same insert can be introduced into different expression vectors simultaneously. Among these methods are the Gateway technology [Invitrogen, (Esposito et al., 2009)], In-Fusion technology, [Clontech, (Berrow

et al., 2007)], Ligase Independent Cloning, (Aslanidis and de Jong, 1990), and Restriction Free Cloning, [RF cloning, (Unger et al., 2010)]. With these methodologies, the use of restriction endonucleases is avoided, so no special sequence requirements are necessary enabling the development of high-throughput technologies for molecular cloning (Cabrita et al., 2006; Berrow et al., 2007; Curiel et al., 2010; Unger et al., 2010; Luna-Vargas et al., 2011).

In this work, we have modified two commonly used commercial vectors (pET32a and pQE80L, T7 and T5 promoters respectively) for *E. coli* protein expression. We generated 12 different vectors introducing the same sequence at the insertion site, and important features for protein purification like N-terminal (His)₆ tag (Murphy and Doyle, 2005), TEV cleavage site, and C-terminal StrepTag II (Schmidt and Skerra, 2007), in order to set up a high-throughput cloning and purification protocol. The cloning strategy used for the development of the vectors as well as for cloning the target genes on the entire suite is based in the “RF cloning methodology” (Unger et al., 2010). The data reported here, describe the application of an easy methodology to clone any target in 12 different vectors with only two primers. In order to evaluate and find a condition for soluble protein expression, different promoters and solubility enhancer fusion proteins were included in these vectors. Concerning protein solubility enhancers, the target gene can be fused as a C-terminal partner with maltose binding protein (MBP; Kapust and Waugh, 1999), thioredoxin A (Trx; LaVallie et al., 2000), small ubiquitin-like modifier protein (SUMO; Marblestone et al., 2006), disulfide bond isomerase C (DsbC; Nozack et al., 2013), and Histag alone in a T5 or T7 promoter context.

Finally, we propose a new fusion protein which appears to be an efficient solubility enhancer for the RP with previous solubility problems and is included in the vector suite. This solubility enhancer corresponds to a truncated construct of the endoglucanase CelD (CelDnc) from *Clostridium thermocellum*. This is a thermostable protein, highly expressed in *E. coli* system and more interestingly, this molecule maintains a full activity even in the presence of 8M Urea implying a very high stability of its native structure (Chaffotte et al., 1992). All these characteristics make CelDnc a good candidate to study the solubility enhancing properties when fused a target protein. As a proof of concept, we fused to CelDnc the decaprenylphosphoryl- β -D-ribofuranose-2'-epimerase (DprE1) protein from *Micobacterium smegmatis* (Neres et al., 2012) a difficult protein to express in *E. coli* (<0.4 mg/l) and we successfully improved this expression obtaining high yields of soluble and functional monomeric protein.

In summary, here we illustrate how to generate in any laboratory an economic and useful vector suite to fast track the solubility of different RP targets and we propose a new solubility enhancer protein that can be included in the evaluation of the expression of RP that are insoluble in classical expression conditions.

RESULTS

CONSTRUCTION OF A NEW VECTOR SUITE

Aiming to achieve a fast and economical way to evaluate the solubility of RP, we selected two commonly used expression vectors pQE-80L (Qiagen) and pET-32a (Novagen) as the starter plasmids for the suite generation thus giving rise to T5 or T7 based vectors. In order to provide a parallel cloning of the target gene and an easy protein purification method, all the generated vectors contain the same insertion site and antibiotic resistance (ampicillin), an N-terminus His-Tag with the tobacco etch virus (TEV) recognition site and a C-terminus strep-Tag II (Figure 1; Table 1). In addition, we introduced several solubility enhancing proteins including MBP, Trx, DsbC, SUMO, and CelDnc, in

combination with the two promoters (T5 or T7). An extra serine residue was added after the TEV site to decrease steric effects and improve cleavage. This can be avoided by not including it in the forward primer. This extra codon also generates a *Bam*HI site at the beginning of the gene so it can be useful for analysis of clones or to do a restriction based method if preferred (Figure 1).

VALIDATION OF THE NEW VECTOR SUITE

In order to evaluate the expression capabilities and functionality of this new vector suite we selected green fluorescent protein (GFP) as control protein and two "difficult to express" RP such as DprE1 and the MAP kinase 4 from *Leishmania major* (MPK4). All of them were cloned into 12 different vectors and their expression was evaluated. The results showed that all the GFP constructs were produced soluble and at the expected molecular weight. Fractions treated with TEV showed the correct cleavage and release of GFP protein and fusion partner (Figure 2A). The construct DsbC-GFP under the control of T7 promoter was the less productive when working at 37°C. This was over-passed when the expression was done at 17°C over night (ON) where an increment of cleaved proteins was obtained in most of the cases (Figure 2A).

For the case of DprE1 constructs, we can see that despite a correct growth and induction conditions in the culture, it was not possible to obtain any expression of this RP when fused only to a Histag. In contrast, fusion of DprE1 with MBP, Sumo, Trx, and CelDnc give a good soluble production and only low yields account for the DsbC/DprE1 construct (Figure 2B; Table 2). Also, there was an effect of the induction temperature and promoter strength in protein expression where DprE1 was expressed with higher yields at 37°C compared to 17°C and with the T5 promoter compared with T7 for most of the cases. Interestingly, our results suggested that DprE1 fused with CelDnc (in the condition T5-37°C) appear to be one of the most overexpressed fused proteins. For the case of DprE1/CelDnc in T7 at 17°C, there was no cell growth. Finally, the treatment with TEV revealed that DprE1

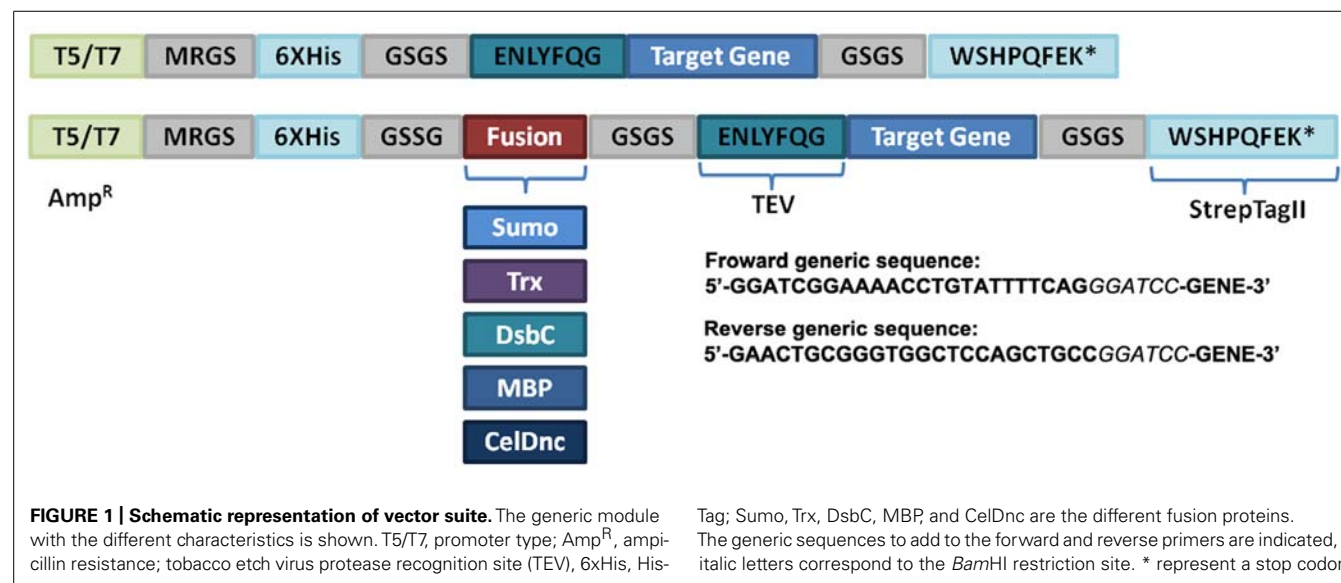


Table 2 | Expression screening of DprE1 protein.

After purification by IMAC, concentration of the entire fusions and yield was determined at 280 nm taking into account the different extinction coefficients. The expected molecular weight as well as construct name and characteristics are indicated.

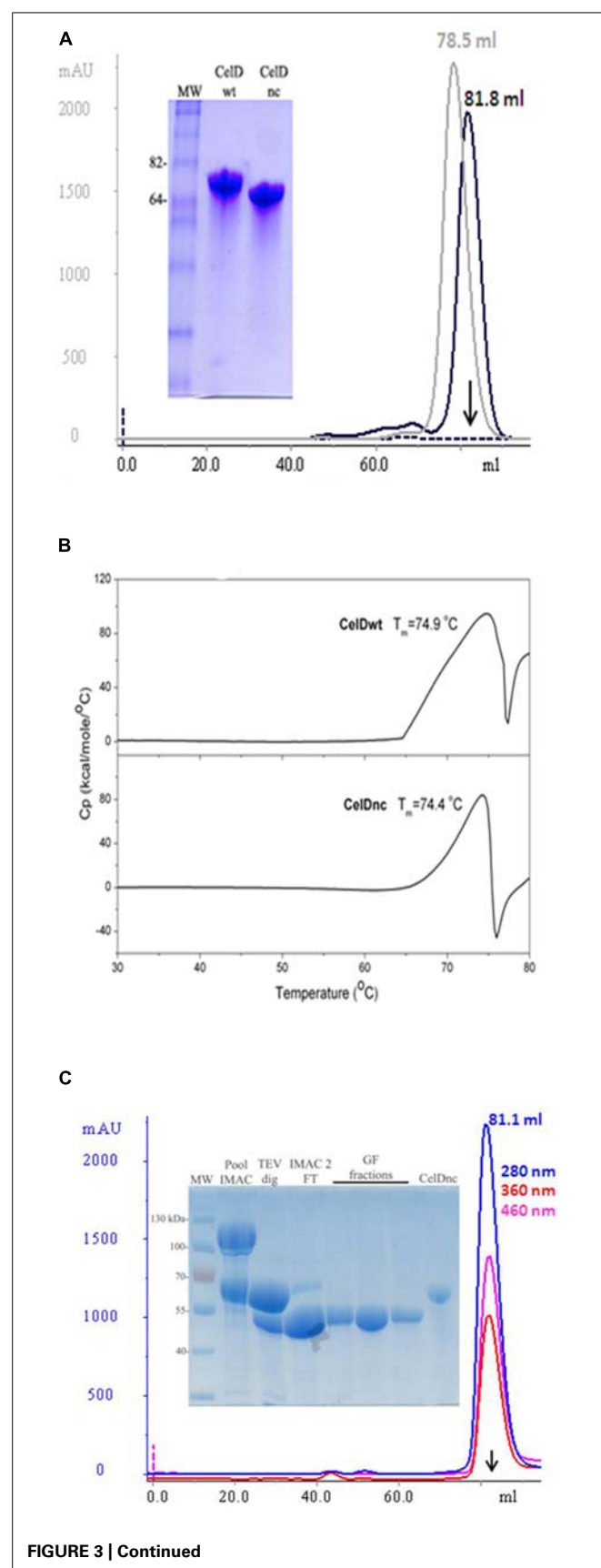


FIGURE 3 | Continued

FIGURE 3 | Continued

(A) Analysis of the purity and monomeric states of CelDwt (gray) and CelDnc (black). SEC was performed in a Superdex 200 16/60 and protein purity evaluated in a 10% SDS-PAGE. **(B)** Differential scanning calorimetry (DSC) curves of CelDwt (top panel) and CelDnc (bottom panel). Determined melting temperature (T_m) is indicated for each case. **(C)** Large scale expression and purification of DprE1. DprE1 was fused to CelD, expressed, and purified by IMAC. After TEV cleavage and second IMAC purification, the monomeric state was confirmed by SEC in a Superdex 200 16/60. FAD binding properties of DprE1 are confirmed by peaks at 360 nm (red) and 460 nm (pink). Purity of DprE1 (53.7 kDa) was evaluated by 12% SDS-PAGE. CelDnc (61.1 kDa) was added as a control. Arrows indicates the retention volume for BSA (66.5 kDa).

as expected for this protein (peaks at 360 and 460 nm; **Figure 3C**). The final yield was of 7 mg/l which corresponds to more than 17 times improvement in soluble protein expression when compared with no fusion (<0.4 mg/l). Moreover, the same experiment done with MBP fusion resulted in a final yield for DprE1 of 2.8 mg/l (data not shown), demonstrating the usefulness of CelDnc as a solubility enhancer of RP.

These results suggest that the construct CelDnc is an interesting new solubility enhancer that could be taken into account for the expression screening of “difficult to express” RP.

DISCUSSION

Purified and soluble proteins are essential tools in academic, industrial and medical areas. The knowledge of the molecular structure of individual proteins allow addressing important questions about the physiological function of these molecules, so as to know the biochemical and regulatory pathways in which they are implicated. However, a common scenario is that the first attempt for obtaining soluble protein often fails, requiring the optimization of many parameters increasing production costs and time. One of the standard procedures to circumvent this problem is to screen a series of constructs to identify the optimal vector and culture conditions able to produce enough soluble protein. This may also include the expression of the full-length protein, mutated and/or truncated variants, as well as specific domains of RP (Dahlroth et al., 2006; Yumerefendi et al., 2010). Series of fusion partners may also be investigated for their effects on driving enhanced expression or their capacity to capture and purify the target protein quickly with minimal impurities (Young et al., 2012).

In this work, we describe the generation of a vector suite composed of 12 different expression vectors using the RF cloning method. This suite engages the expression of the RP with strong promoters such as T7 or T5, with N-terminus His-tag, a TEV specific cleavage site and a C-terminus StrepTag II as well as different fusion proteins such as Sumo, Trx, DsbC, MBP, and CelDnc. All these vectors contain the same site of insertion in order to enable a parallel cloning for solubility screening and the posterior large scale purification in a simple and general manner (IMAC purification, TEV cleavage and dialysis, 2nd IMAC). The suite is based on the commonly used pET and pQE vectors and presents no major changes in expression or sequencing protocols. The cloning strategy occurs in an insert-sequence independent manner, with the additional advantage that no restriction site or extra aminoacids are added to the N-terminus of the expressed protein after TEV cleavage, apart from the last glycine residue. As

purification features we selected the use of the HisTag, because it has demonstrated to be very versatile, cheap and to work well in small and large scale purifications (Schafer et al., 2002; Steen et al., 2006). Additionally, if the stop codon of the target gene is omitted, an additional purification tag, the strepTag II is expressed in the C-terminus of the target protein. This last can be useful if degradation intermediates appear by coupling IMAC purification with StrepTacting purification only a product with an intact N- and C-terminus will be purified. Also the purification via the StrepTag II showed to be very useful for proteins that are expressed in low abundance where usually purification by IMAC gives many contaminants from the host (Magnusdottir et al., 2009). Finally the TEV site was chosen for protein cleavage as it has demonstrated to be very specific, work well at low temperatures and can be produced in the laboratory with high yield reducing production costs (van den Berg et al., 2006). Moreover, it was shown that the last residue of the cleavage site (Gly) can be changed for all the other residues except for proline for an expense in cleavage efficiency, so if a protein with a native N terminus is needed it can be taken into account (Kapust et al., 2002).

The suite was tested with GFP, and we found out that in all cases there were expression and cleavage with TEV demonstrating that all the vectors worked well. By using this suite of vectors the high-throughput screening for soluble expression could be easily achieved manually or automatically as it was demonstrated for the expression of GFP, DprE1 and MPK4.

In order to challenge the vector suite proposed here we selected two “difficult to express” RP like DprE1, and MPK4. For the first protein evaluated (DprE1) the vector suite demonstrated that the expression protein improved when the target protein was fused to Sumo, Trx, DsbC, MBP, and CelDnc solubility enhancer proteins. Among them the best results concerning solubility and quantities of stable protein was achieved when DprE1 was fused to CelDnc and subsequently cleaved by TEV. In the second case, only two out of 12 conditions evaluated were able to express MPK4 in the soluble fraction and only one (pT7-DsbC-MPK4 construct) remains soluble after TEV cleavage. Interestingly, high yield of this fusion construct remained as a decamer before TEV cleavage, so after improving purification protocols (like the use of strepTag II or ion exchange chromatography), the entire fusion can be used for crystallization screenings.

Despite the fact that, many fusion proteins were evaluated, it remains difficult to define a “universal fusion protein.” Different options are commercially available (MBP, GST, Trx, DsbC, NusA, etc), and several groups have found new proteins that can be promising alternatives to obtain a soluble and homogeneous recombinant protein (Chatterjee and Esposito, 2006; DelProposto et al., 2009; Cheng et al., 2010; Song et al., 2011) by fusing the target gene. In this work, we evaluated the use of a novel fusion protein, CelDnc that is thermostable (T_m : 71.4°C) and is expressed in massive amounts in *E. coli* system. CelD is an endo- β -glucanase (EC 3.2.1.4) from *C. thermocellum* and is part of the cellulose degrading complex termed cellulosome composed of a large number of individual enzymes (Kataeva et al., 1997).

When this protein was evaluated as a solubility fusion enhancer for DprE1 the results showed an increasing solubility performance

for this molecule compared with other classical fusion enhancers like MBP. After expression and IMAC purification was done the CelDnc fusion was soluble in large amounts. Moreover, DprE1 was still soluble, monomeric and presented FAD binding properties even after the proteolytical removal of CelDnc demonstrating the utility of this fusion protein that can be taken into account when solubility screening is performed.

In this work we propose a new vector suite and a new fusion enhancer molecule with chances to improve the solubility of different RP. The vector suite proposed here allows the evaluation of five different fusion proteins or only the HisTag in combination with two different promoters, giving rise to 12 different constructs for a single target gene. Altogether, our results suggest that this expression system could be an interesting tool to improve solubility problems of RP.

Moreover, the screening protocol can be further improved. In the present work we used Rosetta cells for the screening of RP production. Different *E. coli* strains can be evaluated in parallel like the use of strains for disulfide bond formation (Shuffle, New England Biolabs), reduced mRNA degradation (BL21 Star, Invitrogen) among others. Also, the co-expression of chaperones or molecular partners can be included if they are in a vector compatible with a ColE1 replication origin. By the complementation of such variables with the vector suite, a great number of conditions can be screened, increasing the chances of finding the optimal context for target protein production.

It was shown that the sequence at the translation initiation region (TIR) can have a detrimental effect in protein production due to the generation of secondary structures in the messenger RNA that can hamper the translation by the ribosome complex. In this regard a predictive method was developed for designing synthetic ribosome binding sites (RBS) that can minimize the formation of secondary structures at RNA level, so increasing the translation rate (Salis et al., 2009; Salis, 2011). Because the nucleotide sequence from +1 to +25 is the same in all vectors, a new RBS can be designed and introduced into the entire suite increasing translation rates.

Finally, despite the cloning of target genes into the suite was very efficient, false positives were found in some cases. This can be improved, for example, if a toxic gene like the toxin CcdB of type II toxin-antitoxin system is added at the insertion site.

Despite the fact that, more proteins should be tested in this vector suite and that there is no magic formula able to ensure the solubility of different proteins, this could be a useful and economic model to fast track the soluble expression of the RP.

MATERIALS AND METHODS

GENERATION OF THE VECTOR SUITE

For the generation of the vector suite we used a modified version of the pQE80L (Qiagen) as the starter plasmid, that contained a TEV cleavage site after the Histag separated by a GSGS linker (pQE80L-TEV). In a first step we cloned the gene DprE1 into this vector and added the different modules for the vector suite (linkers, strepTag and different fusion proteins) thus generating the T5 series. Then the entire constructs were cloned into the vector pET32a in order to generate the T7 series.

All PCR were done using Phusion polymerase (Finnzymes). For the amplification of the fragments (megaprimer generation) conditions were 30 s at 98°C and 28 cycles of 98°C for 10 s, 59°C for 1 min and 72°C for 1 min with a final extension step at 72°C for 5 min and PCR products were purified by agarose gel. The generated megaprimers contained 30 bp in both ends that overlaps with the insertion site in the destination vectors. The integration into the vectors was done by RF cloning (Unger et al., 2010) and the RF reaction was as follows: 30 s at 98°C and 30 cycles of 98°C for 10 s, 60°C for 1 min and 72°C for 5 min with a final extension step at 72°C for 7 min. For RF reactions 120 ng of megaprimers and 30 ng destination vector were used. 20 µl were digested with 2 µl Fast Digest DpnI (Thermo) for 15 min at 37°C in order to remove parental plasmid, and 5 µl were used to transform 50 µl of competent DH5α *E. coli* cells. Positive clones were confirmed by colony PCR by using Taq polymerase (Invitrogen) with the same primers used for megaprimer generation. Colony PCR was as follows, 95°C for 3 min, 25 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 2 min followed by a final extension step at 72°C for 5 min. Positive colonies were selected for plasmid extraction and confirmed by sequencing.

The gene for DprE1 was amplified from *M. smegmatis* genomic DNA using the primers QE3790For and QE3790Rev for the generation of the megaprimer (Table 1). The product was cloned into the vector pQE80L-TEV by RF cloning to generate the construct pDprE1. The genes coding for CelDwt or the truncated version CelDnc (residues 32–577), were amplified from the plasmid pCT603 (Chaffotte et al., 1992) with the primers CelDwtNFor and CelDwtCRev for CelDwt and primers CelDtruncNFor and CelDtruncCRev for CelDnc (Table 1) and cloned by RF in the same vector to generate the constructs pCelD and pCelDnc. The construct pDprE1 was used for the insertion of CelDnc in the 5' of DprE1 (between the HisTag and the GSGS linker, Figure 1). CelDnc was amplified from the pCelDnc construct using primers CelDInsFor and CelDInsRev. The forward primer was designed also to add a GSSG linker to separate the HisTag from the fusion partner generating the construct pCelD-DprE1. The generated constructs (pDprE1 and pCelD-DprE1) were then used to add the last module of the vector, the C-terminal strepTag II. The strepTag II was inserted at the C-terminus separated by a GSGS linker with primers strepCterFor and strepCterRev (Table 1) for the generation of the vector pT5-DprE1 (HisTag alone) and pT5-CelD-DprE1 (CelDnc fusion). The primers anneal each other, so they were used without addition of DNA for the generation of the megaprimer. The generated pT5-CelD-DprE1 vector was then used for the insertion and replacement of CelDnc by other fusion partners. In this regard the primers SumoFor and SumoRev; TrxFor, and TrxRev; MBPFor and MBPRev and DsbCFor and DsbCRev were used for the insertion of Sumo, TrxA, MBP, and DsbC, respectively, (Table 1). The genes were amplified from *Saccharomyces cerevisiae* for Sumo, pET32a (Novagen) for TrxA, pMAL (New England Biolabs) for MBP, and *E. coli* genome for DsbC. By this way, the T5 vector series was completed. All 6 vectors were confirmed by sequencing with the QEFor and QERev plasmid primers. For the case of MBP and CelDnc constructs internal primers were also used in order to cover the entire sequence.

The last step was to transfer the modules into a T7 context. To do this, we selected the pET32a (Novagen) as a destination vector amplifying the entire cassette from T5 series (from MRGS-HisTag up to the strepTag II for the different fusions) with the primers T5T7For and T5T7Rev and replacing the expression cassette of the pET32a vector. The generated megaprimers were used for the RF reactions. By this way the vector suite was completed containing the gene DprE1 in all 12 vectors for expression screening.

CLONING OF GFP AND MPK4 INTO THE SUITE OF VECTORS

Leishmania major MPK4 gene was amplified with primers MPK4For and MPK4Rev from a pGem vector containing the gene. GFP was amplified with primers GFPFor and GFPRev from a pET vector containing a GFP variant that is well expressed in *E. coli* (Waldo et al., 1999).

The 12 vectors were added to 12 different PCR tubes, and the amplified products were used as megaprimers for the RF reaction using the HF buffer from Phusion polymerase. After digestion of 20 µl PCR products with 2 µl DpnI, chemical competent cells were transformed with 5 µl RF reaction in a PCR machine with the following program: 30 min at 4°C, 45 s at 42°C, 3 min at 4°C, addition of 100 µl of LB, 1 h at 37°C, and plating of 100 µl in agar plates containing ampicillin. Four colonies for each construct were selected and confirmed by colony PCR and sequenced. After the analysis we found out that in most cases all were positive (or at least three of four were positive) giving a percentage of success of more than 80%.

EXPRESSION SCREENING OF GFP AND DprE1

Chemocompetent Rosetta-pLysS cells were transformed with 5 µl of purified plasmids as described above and then incubated in a shaker ON at 37°C in 1 ml of LB with chloramphenicol and ampicillin in a 96 × deep-well plate. 100 µl of ON culture were used to inoculate 4 ml of Terrific Broth in 24 × deep-well plates by duplicate. Cultures were incubated at 37°C until D.O.₆₀₀ reached 1.0–1.2. At that moment one plate was induced with 1 mM IPTG and left at 37°C for 4 h. The other 24 deep-well was incubated at 17°C for 15 min to cooling it and then induced with 1 mM IPTG ON at the same temperature. After induction time was reached, cells were harvested, resuspended in 1 ml lysis buffer (50 mM Tris pH 8.0; 300 mM NaCl, 10 mM imidazol, 0.5 mg/ml lysozyme) and frozen at –80°C. After thawing cells, 10 units of DNase I and 10 µl of 2M MgSO₄ were added and incubated with shaking for 20 min at 20°C. Then 200 µl of Nickel beads (Qiagen) equilibrated in binding buffer (50 mM Tris pH 8.0; 300 mM NaCl, 10 mM imidazol) were added to cell extracts and incubated for 15 min at 20°C. Cell extracts were then transferred to a 96×-well filter plate assembled in a vacuum device, and bound protein was washed with 2 ml of binding buffer. An additional wash step was done with 2 ml of binding buffer containing 50 mM imidazol. Elution was done with 160 µl of elution buffer [50 mM Tris pH 8.0; 300 mM NaCl, 500 mM imidazol; for a detailed protocol, see (Saez and Vincentelli, 2013)]. Eluates were divided in two groups for evaluation of uncleaved protein and assessment of TEV cleavage ON at 18°C.

Samples were then loaded into an E-PAGE 96 acrylamide gel (Invitrogen).

EXPRESSION SCREENING OF MPK4

Expression screening and purification of MPK4 constructs was made in a similar way than for GFP and DprE1 but only 17°C of induction was evaluated. Purification steps were the same but the pipeting scheme was done automatically by using a TECAN Freedom EVO®200. Expression analysis was done also automatically by using a Labchip GX II (Caliper, USA) microfluidic detection system.

LARGE SCALE EXPRESSION AND PURIFICATION OF DsbC-MPK4

DsbC-MPK4 was expressed in Terrific Broth (TB) supplemented with ampicillin and chloramphenicol and induction was done at D.O₆₀₀: 1.2 ON at 17°C with 1 mM IPTG. Pellets were resuspended in lysis buffer and frozen at −80°C. After thawing, the pellets were sonicated and centrifugated at 15.000 × g. Soluble fraction was injected in a 1 ml IMAC column (GE Healthcare) equilibrated in binding buffer. Elution was done in a linear gradient of 5–100% B in 10 column volumes (CV) with elution buffer. Purified protein was cleaved with TEV protease in a 1:30 protein:enzyme ratio and dialyzed against cleavage buffer (50 mM Tris pH 8.0; 150 mM NaCl, 1 mM DTT) ON at 8°C. Sample was filtered through 0.22 μm to remove precipitates, and analyzed by SDS-PAGE.

EXPRESSION AND PURIFICATION OF CelD AND CelDnc

Production of CelD and CelDnc was done in M15pREP4 from the constructs pCelD and pCelDnc, respectively, in 1 l 2YT supplemented with ampicillin and kanamycin, and induced with 1 mM IPTG at D.O. 1.0 ON at 37°C. IMAC was done like for the case of DsbC-MPK4 but using a 5 ml column and only half of the soluble fraction was used. TEV cleavage was done as before and desalted in order to remove imidazole. The reaction was injected in a second IMAC under same conditions as above and the flow through containing the cleaved protein was injected in a Superdex 200 16/60 (GE Healthcare) equilibrated with buffer 40 mM Tris pH 7.7.

DSC ANALYSIS OF CelD AND CelDnc

Differential scanning calorimetry (DSC) experiments were carried out in PBS, in a VP-DSC instrument (Microcal, Northampton, MA, USA) and data analyzed with the software supplied with the equipment. The temperature was increased at 1°C per minute from 30 to 80°C, and proteins were added at concentration of 1 mg/ml for CelD and CelDnc.

LARGE SCALE EXPRESSION AND PURIFICATION OF pT5-DprE1, pT5-CelD-DprE1 AND pT5-MBP-DprE1

Induction of p5DprE1, p5CelDnc-DprE1 and p5MBP-DprE1 were done in M15pREP4 with 1 mM IPTG in 1 l 2YT supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml) and 15 μM FAD at D.O.: 1.0–1.2 during 4 h at 37°C. Cells were harvested, resuspended in lysis buffer and frozen at −80°C. After thawing the cells, were lysed and protein purified as before. Purified protein was cleaved with TEV in a 1:30 ratio, and dialysed against cleavage buffer. The product was then purified by a second IMAC and injected in a Superdex 200 16/60 equilibrated with buffer 25 mM Tris pH 8.0; 150 mM NaCl.

ACKNOWLEDGMENTS

This work was partially funded by FOCER (MERCOSUR Structural Convergence Fund), COF 03/11 and CYTED Program. Agustín Correa was supported by a doctoral program of the Agencia Nacional de Investigación e Innovación, Uruguay. We wish to thank Dr. Trajtemberg and Sofía Horjales from the Crystallography Unit (PXF) of the Institut Pasteur de Montevideo for giving the plasmid pQE80L-TEV and pGem-MPK4 and Mrs. Natalia López for helpful secretarial assistance.

REFERENCES

- Aslanidis, C., and de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18, 6069–6074. doi: 10.1093/nar/18.20.6069
- Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., et al. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* 35, e45. doi: 10.1093/nar/gkm047
- Cabrera, L. D., Dai, W., and Bottomley, S. P. (2006). A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* 6:12. doi: 10.1186/1472-6750-6-12
- Chaffotte, A. F., Guillou, Y., and Goldberg, M. E. (1992). Inclusion bodies of the thermophilic endoglucanase D from *Clostridium thermocellum* are made of native enzyme that resists 8 M urea. *Eur. J. Biochem.* 205, 369–373. doi: 10.1111/j.1432-1033.1992.tb16789.x
- Chatterjee, D. K., and Esposito, D. (2006). Enhanced soluble protein expression using two new fusion tags. *Protein Expr. Purif.* 46, 122–129. doi: 10.1016/j.pep.2005.07.028
- Cheng, Y., Gu, J., Wang, H. G., Yu, S., Liu, Y. Q., Ning, Y. L., et al. (2010). EspA is a novel fusion partner for expression of foreign proteins in *Escherichia coli*. *J. Biotechnol.* 150, 380–388. doi: 10.1016/j.jbiotec.2010.09.940
- Correa, A., and Oppezzo, P. (2011). Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening. *Biotechnol. J.* 6, 715–730. doi: 10.1002/biot.201100025
- Curiel, J. A., De Las Rivas, B., Mancheno, J. M., and Munoz, R. (2010). The pURI family of expression vectors: a versatile set of ligation independent cloning plasmids for producing recombinant His-fusion proteins. *Protein Expr. Purif.* 76, 44–53. doi: 10.1016/j.pep.2010.10.013
- Dahlroth, S. L., Nordlund, P., and Cornvik, T. (2006). Colony filtration blotting for screening soluble expression in *Escherichia coli*. *Nat. Protoc.* 1, 253–258. doi: 10.1038/nprot.2006.39
- DelProposto, J., Majmudar, C. Y., Smith, J. L., and Brown, W. C. (2009). Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* 63, 40–49. doi: 10.1016/j.pep.2008.08.011
- Esposito, D., Garvey, L. A., and Chakiath, C. S. (2009). Gateway cloning for protein expression. *Methods Mol. Biol.* 498, 31–54. doi: 10.1007/978-1-59745-196-3_3
- Kapust, R. B., Tozser, J., Copeland, T. D., and Waugh, D. S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* 294, 949–955. doi: 10.1016/S0006-291X(02)00574-0
- Kapust, R. B., and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668–1674. doi: 10.1110/ps.8.8.1668
- Kataeva, I., Guglielmi, G., and Beguin, P. (1997). Interaction between *Clostridium thermocellum* endoglucanase CelD and polypeptides derived from the cellulosome-integrating protein CipA: stoichiometry and cellulolytic activity of the complexes. *Biochem. J.* 326 (Pt 2), 617–624.
- LaVallie, E. R., Lu, Z., Diblasio-Smith, E. A., Collins-Racie, L. A., and McCoy, J. M. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Methods Enzymol.* 326, 322–340. doi: 10.1016/S0076-6879(00)26063-1
- Luna-Vargas, M. P., Christodoulou, E., Alfieri, A., Van Dijk, W. J., Stadnik, M., Hibbert, R. G., et al. (2011). Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. *J. Struct. Biol.* 175, 113–119. doi: 10.1016/j.jsb.2011.03.017

- Magnusdottir, A., Johansson, I., Dahlgren, L. G., Nordlund, P., and Berglund, H. (2009). Enabling IMAC purification of low abundance recombinant proteins from *E. coli* lysates. *Nat. Methods* 6, 477–478. doi: 10.1038/nmeth0709-477
- Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* 15, 182–189. doi: 10.1110/ps.051812706
- Murphy, M. B., and Doyle, S. A. (2005). High-throughput purification of hexahistidine-tagged proteins expressed in *E. coli*. *Methods Mol. Biol.* 310, 123–130. doi: 10.1007/978-1-59259-948-6_9
- Neres, J., Pojer, F., Molteni, E., Chiarelli, L. R., Dhar, N., Boy-Rottger, S., et al. (2012). Structural basis for benzothiazinone-mediated killing of *Mycobacterium tuberculosis*. *Sci. Transl. Med.* 4, 150ra121. doi: 10.1126/scitranslmed.3004395
- Nozach, H., Fruchart-Gaillard, C., Fenaille, F., Beau, F., Ramos, O. H., Douzi, B., et al. (2013). High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. *Microb. Cell Fact.* 12, 37. doi: 10.1186/1475-2859-12-37
- Saez, N. J., and Vincentelli, R. (2013). High-throughput expression screening and purification of recombinant proteins in *E. coli*. *Methods Mol. Biol.* 1091, 33–53. doi: 10.1007/978-1-62703-691-7_3
- Salis, H. M. (2011). The ribosome binding site calculator. *Methods Enzymol.* 498, 19–42. doi: 10.1016/B978-0-12-385120-8.00002-4
- Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950. doi: 10.1038/nbt.1568
- Schafer, F., Romer, U., Emmerlich, M., Blumer, J., Lubenow, H., and Steinert, K. (2002). Automated high-throughput purification of 6xHis-tagged proteins. *J. Biomol. Tech.* 13, 131–142.
- Schmidt, T. G., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* 2, 1528–1535. doi: 10.1038/nprot.2007.209
- Song, J. A., Lee, D. S., Park, J. S., Han, K. Y., and Lee, J. (2011). A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. *Enzyme Microb. Technol.* 49, 124–130. doi: 10.1016/j.enzmictec.2011.04.013
- Steen, J., Uhlen, M., Hober, S., and Ottosson, J. (2006). High-throughput protein purification using an automated set-up for high-yield affinity chromatography. *Protein Expr. Purif.* 46, 173–178. doi: 10.1016/j.pep.2005.12.010
- Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R., and Peleg, Y. (2010). Applications of the restriction free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* 172, 34–44. doi: 10.1016/j.jsb.2010.06.016
- van den Berg, S., Lofdahl, P. A., Hard, T., and Berglund, H. (2006). Improved solubility of TEV protease by directed evolution. *J. Biotechnol.* 121, 291–298. doi: 10.1016/j.jbiotec.2005.08.006
- Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y., and Cambillau, C. (2011). High-throughput protein expression screening and purification in *Escherichia coli*. *Methods* 55, 65–72. doi: 10.1016/j.ymeth.2011.08.010
- Vincentelli, R., and Romier, C. (2013). Expression in *Escherichia coli*: becoming faster and more complex. *Curr. Opin. Struct. Biol.* 23, 326–334. doi: 10.1016/j.sbi.2013.01.006
- Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695. doi: 10.1038/10904
- Young, C. L., Britton, Z. T., and Robinson, A. S. (2012). Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* 7, 620–634. doi: 10.1002/biot.201100155
- Yumerefendi, H., Tarendeau, F., Mas, P. J., and Hart, D. J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J. Struct. Biol.* 172, 66–74. doi: 10.1016/j.jsb.2010.02.021

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.

Received: 21 December 2013; paper pending published: 13 January 2014; accepted: 05 February 2014; published online: 25 February 2014.

Citation: Correa A, Ortega C, Obal G, Alzari P, Vincentelli R and Oppezzo P (2014) Generation of a vector suite for protein solubility screening. *Front. Microbiol.* 5:67. doi: 10.3389/fmicb.2014.00067

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Correa, Ortega, Obal, Alzari, Vincentelli and Oppezzo. 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.



Inclusion bodies: not that bad...

Ana Ramón, Mario Señorale-Pose and Mónica Marín*

Sección Bioquímica, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Edited by:

Germán L. Rosano, Instituto de Biología Molecular y Celular de Rosario, Argentina

Reviewed by:

Mark Sutherland, University of Stellenbosch, South Africa
Marc Blondel, Inserm UMR1078, France

*Correspondence:

Mónica Marín, Sección Bioquímica, Facultad de Ciencias, Universidad de la República, Igua 4225, 11400 Montevideo, Uruguay
e-mail: marin@fcien.edu.uy

The formation of inclusion bodies (IBs) constitute a frequent event during the production of heterologous proteins in bacterial hosts. Although the mechanisms leading to their formation are not completely understood, empirical data have been exploited trying to predict the aggregation propensity of specific proteins while a great number of strategies have been developed to avoid the generation of IBs. However, in many cases, the formation of such aggregates can be considered an advantage for basic research as for protein production. In this review, we focus on this positive side of IBs formation in bacteria. We present a compilation on recent advances on the understanding of IBs formation and their utilization as a model to understand protein aggregation and to explore strategies to control this process. We include recent information about their composition and structure, their use as an attractive approach to produce low cost proteins and other promising applications in Biomedicine.

Keywords: protein aggregation, bacterial inclusion bodies, protein folding, recombinant protein expression, conformational disease model, drug delivery systems, nanoparticles

INTRODUCTION

The deciphering of the genetic code and the availability of the first tools and basic procedures in genetic engineering opened the way for protein production in bacteria. This apparently easy goal encountered an unexpected difficulty: the accumulation of the protein of interest as insoluble form and the generation of inclusion bodies (IBs). Since then the generation of IBs has been traditionally considered an obstacle to avoid and difficult to predict. Notwithstanding in the last decade, IBs have been observed from a different perspective and their study gained considerable interest. From this point of view, the formation of IBs in bacteria is seen as part of a general cellular response related to the presence in the cell of unfolded proteins and as a pathway for the control of aggregation. From this perspective, IBs constitute a valuable model to better understand protein aggregation in eukaryotes, and for the search of specific inhibitors or disaggregation approaches, in relation to relevant conformational diseases. On the other hand, the production of proteins as aggregates in IBs opened new interesting perspectives for diverse applications in Biomedicine. They can be an almost pure source of recombinant protein, and because of their particular structural and functional characteristics they can be potentially exploited as naturally immobilized enzymes or as nanomaterials. In this minireview, we focus on this positive side of IBs formation in bacteria. We include recent information about their composition, formation and structure and their use as an attractive approach to produce proteins at low cost. We also review the role of IBs as a model to understand protein aggregation and to explore strategies to control this process. Finally we describe some promising applications of bacterial IBs as systems for controlled drug delivery and nanotechnology applications.

PROTEIN AGGREGATION AS A CONSERVED CELLULAR RESPONSE

Diverse conditions can lead to an alteration of protein homeostasis and to protein aggregation in all living cells. Among others, the presence of unfolded proteins is increased after a heat shock or other environmental stress conditions. Some mutations can lead to the synthesis of unstable protein structures, as well. Another current event is provided by fast and high expression of recombinant proteins in bacteria, leading to the formation of IBs. Recent evidence indicates that in the cell, the quality control system—regulating chaperones and proteases levels—and protein aggregation and disaggregation are part of a cellular response to altered protein homeostasis. In a recent review Tyedmers et al. (2010) described protein aggregation as a regulated process in bacteria, yeast and mammalian cells. Interestingly, despite differences, similarities of protein aggregation in these cells reflect the universality of the response. In particular, the subcellular localization of aggregates is not random but well-defined in each cellular type. The IBs in bacteria are mainly localized in the cell poles, and also in septation sites, whereas in yeast protein deposits are close to vacuoles or to the nucleus. In mammals deposits in aggresomes are associated to the nucleolus. Another relevant observation is the similarity at the level of protein structure between bacterial IBs and the amyloid aggregates characteristic of several human diseases, the so-called “conformational diseases”, in which protein deposits are observed in specific tissues or cells. This is the case for several neurodegenerative diseases such as Alzheimer or Parkinson, among others.

Under a heat shock stress, about 150–200 different proteins were identified as aggregation-prone in *Escherichia coli* (Winkler et al., 2010). The formation of aggregates can be reversed by the combined effect of chaperones and protease activities and the

disaggregation machinery. As mentioned, aggregates are mainly located at cellular poles by mechanisms not yet completely understood. Recently Winkler et al. proposed that nucleoid occlusion is the main driving force which determines the number and positioning of the protein aggregates in *E. coli*. Also, authors argued against the idea that an active targeting mechanism was involved in polar localization. Interestingly, polar localization allows an asymmetric partitioning of protein aggregates between daughter cells. This asymmetry allows an increased cell division rate in the population devoid of aggregates, beneficial for the ageing of the bacterial cell population (Winkler et al., 2010).

PREDICTION OF A PROTEIN'S AGGREGATION PROPENSITY

The aggregation behavior of a protein is strongly determined by intrinsic properties of its amino acid sequence. This observation supported the development of computational methods to predict protein aggregation propensity. Interestingly, recent algorithms take into account not only the primary sequence of the polypeptide, but also experimental proteome data, including information about cellular localization, cytosolic, periplasmic and membrane proteins. The analysis performed by De Groot and Ventura (2010) using AGGRESCAN—an algorithm developed by their group, (Conchillo-Sole et al., 2007; De Groot et al., 2012)—, indicates that the aggregation propensity of bacterial proteins is associated with their length, conformation, location, function, and abundance. Recently, in AMYLPRED2 11 predictive methods were considered, trying to produce a consensus prediction of amyloidogenic determinants/“aggregation-prone” peptides in proteins, from sequence alone (Tsolis et al., 2013) (<http://biophysics.biol.uoa.gr/AMYLPRED2>). It is worth mentioning that other factors affect the aggregation of recombinant protein expression in bacteria such as temperature and growth rate, fusion to soluble protein tags, specific codon usage, tRNA availability, and general optimization of codons in the heterologous expressed sequence (Cortazzo et al., 2002; Rosano and Ceccarelli, 2009).

IBs STRUCTURE: A SPONGE-LIKE SUPRAMOLECULAR ORGANIZATION

The formation and structure of IBs in *E. coli* expressing heterologous proteins has been extensively analyzed. Several excellent reviews summarize the current knowledge on IBs structure and formation (De Groot et al., 2008, 2009; Wang, 2009; Sabate et al., 2010; Garcia-Fruitos et al., 2011, 2012). Here we present a concise summary and extend a little more on the most recent published data.

IBs are normally observed in the cytoplasm of the producing bacteria as dense, large and apparently spherical or cylindrical particle, ranging from 0.2 to 1.2 μm , composed of 80–95% of the heterologous expressed protein. IBs may also contain other proteins, like small heat shock proteins (IbpA and IbpB) and chaperones (like the DnaK system), phospholipids from membranes and nucleic acids and other background proteins that co-purify with aggregates (Jurgen et al., 2010). Interestingly, the cellular composition of IBs evolves during cell growth so that cellular proteins are predominant during the first steps of formation, while

heterologous proteins become predominant at the end. These changes occur in concert with the evolution of other parameters inherent to cell growth, such as division time and growth rate, leading to the idea that aging of bacterial population could be related to protein aggregation (Lindner et al., 2008).

Different approaches have revealed that IBs have a characteristic cross- β structure, resembling that found in amyloid fibers associated to a wide variety of human degenerative diseases. Notwithstanding, IBs may also contain variable amounts of natively folded proteins or partially folded proteins that can acquire their native conformation even if they are embedded in an aggregate (Gonzalez-Montalban et al., 2008). In fact aggregates have been found to be composed of a wide spectrum of conformations, ranging from native conformation to misfolded aggregates (Schrodel et al., 2005; Rinas et al., 2007). Moreover, aggregation and disaggregation have been shown to occur simultaneously *in vivo* in actively producing recombinant bacteria (Carrio and Villaverde, 2002). The solubilized proteins can then reach their native state or alternatively suffer partial proteolytic degradation (Corchero et al., 1997; Carrio et al., 1999; Cubarsi et al., 2001; Lethanh et al., 2005; Vera et al., 2005; Rinas et al., 2007). The proportion of functional protein is characteristic of the target protein's sequence (Upadhyay et al., 2012), but also depends on cell growth temperature (De Groot and Ventura, 2006; Peternel et al., 2008) and on the genetic traits of the host strain (Garcia-Fruitos et al., 2009).

The cross- β sheet regions have been shown to be refractory to proteinase K (PK) digestion, while native or native-like structures are highly sensitive to PK digestion. Using a GFP reporter model, Cano-Garrido et al. (2013) have shown that when IBs are submitted to mild digestion with the protease their morphology or size is not affected while fluorescence emission and density are notably diminished. They propose that IBs present a sponge-like structure, where the PK resistant fibrils constitute a scaffold which confers mechanical stability to IBs, while the functional, PK sensitive fraction accumulate in the gaps of this scaffold.

In accordance to this, Walther et al. (2014) suggested a mechanism of pore diffusion out of a barrier layer for the solubilization of IBs. They propose that the solubilization process involves different layers in the IBs: a core, consisting of the IBs agglomerates, a reactive and a diffusion layer. The densely packed inner cores of protein shrink as the solubilized protein diffuses to the outer layers and subsequently through a porous barrier layer into free solution. The authors propose that this model correlates well with the IBs structure suggested by Cano-Garrido et al. (2013), the barrier layer corresponding to the amyloid scaffold, which becomes visible only as solubilization progresses.

IBs A MODEL TO STUDY PROTEIN AGGREGATION RELATED TO CONFORMATIONAL DISEASES

The formation and disaggregation of bacterial IBs gained growing attention as models to study insoluble protein deposits observed in some complex human diseases, as in the so-called “conformational diseases”. This approach is strongly supported by the concept that protein aggregation is part of a conserved cellular response. Three examples have been chosen to illustrate how IBs are employed as a model to study aggregation proteins involved

in particular human diseases and as a useful screening approach for the search for aggregation inhibitors.

EXPANDED polyQ IN HUNTINGTON DISEASE

Huntington disease is a neurodegenerative disorder that affects muscle coordination, followed by cognitive and psychiatric problems. The disease is caused by mutations in the Huntingtin gene, in which expansion of the triplet CAG within the first exon of the gene produces a protein carrying stretches of repeated glutamines (polyQ). When polyQ exceeds a critical length, huntingtin protein undergoes amyloid aggregation (Orr and Zoghbi, 2007). *E. coli* has been employed to follow *in vivo* the aggregation process of an artificial protein harboring a polyglutamine (polyQ) tract (Ignatova et al., 2007). *E. coli* growth rate was found to be sensitive to the protein conformational state, and showed that misfolded peptides and soluble aggregates were cytotoxic (Miller et al., 2010).

Related to some pathologies, the relationship between “aggregation” and “toxicity” is often controversial. This question was recently explored in *E. coli* by expressing PolyQ-Containing Ataxin-3 (Invernizzi et al., 2012). For this purpose, the toxicity of three variants expressed in *E. coli* was determined according to reduction of growth rate. The authors showed that toxicity was correlated to the formation of soluble cytosolic oligomers, but not to peptide aggregation. Instead, interestingly, the aggregates appeared to be protective against cell toxicity (Invernizzi et al., 2012).

PRION PROTEIN EXPRESSION IN BACTERIA

Prions are protein aggregates with self-perpetuating ability and thus infectious (reviewed in Villar-Pique and Ventura, 2012). Prions are involved in transmissible spongiform encephalopathies (TSEs), a family of rare progressive neurodegenerative disorders that affect both humans and animals. Bacterial IBs have been exploited as a tool for the study of the structural and functional characteristics of prions. Het-s, from the fungus *Podospora anserina*, was the first prion protein whose bacterial IBs were shown to display amyloid-like properties (Sabate et al., 2009; Wasmer et al., 2009). These *E. coli*-produced aggregates were transfected into prion-free fungal strains, and were shown to promote prionic conversion of Het-s at levels comparable to those induced by homologous amyloid fibrils (Sabate et al., 2009). A similar observation was reported in the case of the yeast prion Sup35. The IBs of this protein were used to induce the [PSI⁺] prion in [psi⁻] prion-free yeast strains. These results highlight the fact that the infectivity rate can be easily modulated by tuning the environmental conditions during the formation of IBs (Radchenko et al., 2011; Sabate et al., 2012).

ALZHEIMER A β 42 AGGREGATION

There is an increasing interest in developing methods to identify cellular factors that trigger the aggregation of proteins inside the organism as well as to discover drugs able to interfere with these factors. Villar-Pique et al. (2012) describe a fast, cost-effective high-throughput approach to study conditions and molecules that affect A β 42 aggregation. The assay is based on the use of IBs formed by an A β 42-GFP fusion protein in bacteria. They

showed the ability of the approach to detect the effect of metal ions on A β 42 aggregation as well as to identify compounds that block metal-induced reaction. The authors further propose that, as many proteins form IBs when expressed in bacteria (De Groot et al., 2009), this approach may have a much larger applicability in the search for aggregation modulators in conformational disorders (Villar-Pique et al., 2012).

METHODS FOR THE RECOVERY OF FUNCTIONAL PROTEINS FROM IBs

Considering IBs as a source of almost pure proteins, one possible way is to attain the dissolution of aggregates in order to obtain native-folded, active protein. The challenge is then to solubilize and refold as much aggregated protein as possible and obtain a stable, functional product. The cost of the whole process must be taken into consideration if the aim is to produce a large-scale manufactured product.

The rate and yield of the solubilization process seem to be influenced by the conditions used, like chaotrope addition, concentration, temperature, pressure, etc. Even if there is no general method to solubilize and refold a protein and the strategy in each case must be “custom-made”, most IBs protein-recovery procedures include the following steps (Burgess, 2009; Basu et al., 2011): (i) overexpression of the selected protein in an appropriate host strain; (ii) isolation of IBs; (iii) solubilization of IBs; (iv) refolding (including disulfide bond formation when necessary); (v) high-resolution chromatography (vi) quality control of obtained material.

Solubilization and refolding are the most critical steps in the procedure and successful conditions still depend mostly on trial-and-error strategies (Burgess, 2009). Notwithstanding, efforts have been made to rationalize the refolding step, the free REFOLD database (<http://refold.med.monash.edu.au>; Chow et al., 2006a,b) can be a useful tool for the design of procedures for the refolding and purification of recombinant proteins.

There are very good recent reviews that compile the different methods employed in the different steps of the recovery process. In accordance to the above mentioned findings on the structure and solubilization mechanisms of IBs, the tendency is to employ milder extraction conditions, avoiding strong denaturation and refolding conditions. In Table 1 we listed the most recent approaches reported in the last 3 years.

IBs NANOPARTICLES AND CONTROLLED DELIVERY OF DRUGS

A recent report showed that peptide hormones of the pituitary gland are stored intracellularly as amyloid aggregates within the secretion granules. The amyloid cross- β structure provides a very stable and highly compacted state from which controlled release of functional monomeric hormone can take place upon signaling (Maji et al., 2009). A somewhat similar situation occurs with bacterial IBs. In fact, IBs are composed by amyloid-like aggregates from which substantial amounts of functional recombinant protein can be released *in vivo* as well as under mild (non-denaturing) conditions *in vitro*. This feature—reminiscent of a drug delivery system—led to the concept of the “nanopills,” i.e., nanoparticles which are able to release proteins with therapeutic effects

Table 1 | Novel approaches employed in the recovery of proteins from IBs.

Recovery step	Strategy	Advantage	Protein	References
Solubilization/Refolding	6M n-propanol	Chaotropic effect at high concentration/kosmotropic effect at dilution helps refolding	r-hGH	Singh et al., 2012
Solubilization/Refolding	Aminoacid-based detergent + arginine assisted refolding	Arg minimizes protein-protein interactions	scFvs	Kudou et al., 2011a,b
Solubilization	Non-reducing buffer + low concentration of urea	Avoids inappropriate disulfide bonds and aggregates. Suitable for proteins that aggregate with native-like structures	PrP	Walsh et al., 2012
Solubilization	Microwave assistance	Shortens solubilization time	various	Datta et al., 2013
Refolding	High hydrostatic pressure	No previous protein denaturation required	LECT2	Zheng et al., 2013
Refolding	Dialysis against PE-PEG (Polyethylene-Polyethyleneglycol)	Simple and cheap method; high yield	FhuA Δ 1-159	Dworeck et al., 2011
Solubilization	Alkaline-shock	No use of chaotropic agents	adiponectin	Heiker et al., 2010
Refolding	Solid-phase refolding in cation-exchange resin with decreasing gradient of urea	Minimize time and number of steps	BMP-2 dimers	Rane et al., 2013

directly from the inside of cells (Vazquez et al., 2012). As a proof of concept, Vazquez et al. demonstrated that the addition of different types of IBs to the culture medium of mammalian cells, for example IBs composed of HSP70, leukemia inhibitory factor or catalase, were able to rescue them from cis-platinum, serum deprivation or oxidative stress, respectively. Also, dihydrofolate reductase IBs complemented the intrinsic cell deficiency of this enzyme. Furthermore, IBs are spontaneously internalized by cultured cells, as was directly demonstrated with green fluorescent protein IBs (Villaverde et al., 2012). Following a similar approach, Liovic et al. (2012) introduced keratin 14 IBs into epithelial cells which do not normally express this protein, and found that intracellular keratin filaments start to form.

CONCLUDING REMARKS

The common view of IBs as undesirable by-products in recombinant protein production have been lastly reconsidered, in view of the potential of these structures for different purposes. On the other hand, IBs provide to constitute an easy to handle model for the study of the molecular basis of conformational diseases. On the other hand, IBs provide a source of almost pure polypeptides and are a potentially useful source of ready-to-use protein. In this sense, the aim is then to obtain IBs containing as much folded, functional protein as possible. The design of strategies to reach this aim requires a deep knowledge of IB's structure and formation, in order to identify possible molecular targets which can be "tuned" to improve the protein recovery yield or to obtain IBs with the desired characteristics to allow their use as enzyme carriers or nanomaterials.

ACKNOWLEDGMENTS

This work was supported by Comisión Sectorial de Investigación Científica (CSIC-Universidad de la República), Uruguay.

REFERENCES

- Basu, A., Li, X., and Leong, S. S. (2011). Refolding of proteins from inclusion bodies: rational design and recipes. *Appl. Microbiol. Biotechnol.* 92, 241–251. doi: 10.1007/s00253-011-3513-y
- Burgess, R. R. (2009). Refolding solubilized inclusion body proteins. *Methods Enzymol.* 463, 259–282. doi: 10.1016/S0076-6879(09)63017-2
- Cano-Garrido, O., Rodríguez-Carmona, E., Díez-Gil, C., Vazquez, E., Elizondo, E., Cubarsi, R., et al. (2013). Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. *Acta Biomater.* 9, 6134–6142. doi: 10.1016/j.actbio.2012.11.033
- Carrio, M. M., Corchero, J. L., and Villaverde, A. (1999). Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms. *Biochim. Biophys. Acta* 1434, 170–176. doi: 10.1016/S0167-4838(99)00177-6
- Carrio, M. M., and Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol.* 96, 3–12. doi: 10.1016/S0168-1656(02)00032-9
- Chow, M. K., Amin, A. A., Fulton, K. F., Fernando, T., Kamau, L., Batty, C., et al. (2006a). The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic Acids Res.* 34, D207–D212. doi: 10.1093/nar/gkj080
- Chow, M. K., Amin, A. A., Fulton, K. F., Whisstock, J. C., Buckle, A. M., and Bottomley, S. P. (2006b). REFOLD: an analytical database of protein refolding methods. *Protein Expr. Purif.* 46, 166–171. doi: 10.1016/j.pep.2005.07.022
- Conchillo-Sole, O., De Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* 8:65. doi: 10.1186/1471-2105-8-65
- Corchero, J. L., Cubarsi, R., Enfors, S., and Villaverde, A. (1997). Limited *in vivo* proteolysis of aggregated proteins. *Biochem. Biophys. Res. Commun.* 237, 325–330. doi: 10.1006/bbrc.1997.7132
- Cortazzo, P., Cervenansky, C., Marin, M., Reiss, C., Ehrlich, R., and Deana, A. (2002). Silent mutations affect *in vivo* protein folding in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 293, 537–541. doi: 10.1016/S0006-291X(02)00226-7
- Cubarsi, R., Carrio, M. M., and Villaverde, A. (2001). *In situ* proteolytic digestion of inclusion body polypeptides occurs as a cascade process. *Biochem. Biophys. Res. Commun.* 282, 436–441. doi: 10.1006/bbrc.2001.4583
- Datta, I., Gautam, S., and Gupta, M. N. (2013). Microwave assisted solubilization of inclusion bodies. *Sustain. Chem. Process.* 1, 2. doi: 10.1186/2043-7129-1-2
- De Groot, N. S., Castillo, V., Grana-Montes, R., and Ventura, S. (2012). AGGRESCAN: method, application, and perspectives for drug design. *Methods Mol. Biol.* 819, 199–220. doi: 10.1007/978-1-61779-465-0_14
- De Groot, N. S., Espargaro, A., Morell, M., and Ventura, S. (2008). Studies on bacterial inclusion bodies. *Future Microbiol.* 3, 423–435. doi: 10.2217/17460913.3.4.423

- De Groot, N. S., Sabate, R., and Ventura, S. (2009). Amyloids in bacterial inclusion bodies. *Trends Biochem. Sci.* 34, 408–416. doi: 10.1016/j.tibs.2009.03.009
- De Groot, N. S., and Ventura, S. (2006). Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* 580, 6471–6476. doi: 10.1016/j.febslet.2006.10.071
- De Groot, N. S., and Ventura, S. (2010). Protein aggregation profile of the bacterial cytosol. *PLoS ONE* 5:e9383. doi: 10.1371/journal.pone.0009383
- Dworeck, T., Petri, A. K., Muhammad, N., Fioroni, M., and Schwaneberg, U. (2011). FhuA deletion variant Delta1-159 overexpression in inclusion bodies and refolding with Polyethylene-Poly(ethylene glycol) diblock copolymer. *Protein Expr. Purif.* 77, 75–79. doi: 10.1016/j.pep.2010.12.006
- García-Fruitos, E., Rodríguez-Carmona, E., Díez-Gil, C., Ferraz, R. M., Vázquez, E., Corchero, J. L., et al. (2009). Surface cell growth engineering assisted by a novel bacterial nanomaterial. *Adv. Mater.* 21, 4249–4253. doi: 10.1002/adma.200900283
- García-Fruitos, E., Sabate, R., De Groot, N. S., Villaverde, A., and Ventura, S. (2011). Biological role of bacterial inclusion bodies: a model for amyloid aggregation. *FEBS J.* 278, 2419–2427. doi: 10.1111/j.1742-4658.2011.08165.x
- García-Fruitos, E., Vázquez, E., Díez-Gil, C., Corchero, J. L., Seras-Franzoso, J., Ratera, I., et al. (2012). Bacterial inclusion bodies: making gold from waste. *Trends Biotechnol.* 30, 65–70. doi: 10.1016/j.tibtech.2011.09.003
- Gonzalez-Montalban, N., Natalello, A., García-Fruitos, E., Villaverde, A., and Doglia, S. M. (2008). *In situ* protein folding and activation in bacterial inclusion bodies. *Biotechnol. Bioeng.* 100, 797–802. doi: 10.1002/bit.21797
- Heiker, J. T., Kloting, N., Blüher, M., and Beck-Sickinger, A. G. (2010). Access to gram scale amounts of functional globular adiponectin from *E. coli* inclusion bodies by alkaline-shock solubilization. *Biochem. Biophys. Res. Commun.* 398, 32–37. doi: 10.1016/j.bbrc.2010.06.020
- Ignatova, Z., Thakur, A. K., Wetzel, R., and Gierasch, L. M. (2007). In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. *J. Biol. Chem.* 282, 36736–36743. doi: 10.1074/jbc.M703682000
- Invernizzi, G., Aprile, F. A., Natalello, A., Ghisleni, A., Penco, A., Relini, A., et al. (2012). The relationship between aggregation and toxicity of polyglutamine-containing ataxin-3 in the intracellular environment of *Escherichia coli*. *PLoS ONE* 7:e51890. doi: 10.1371/journal.pone.0051890
- Jürgen, B., Breitenstein, A., Urlacher, V., Buttner, K., Lin, H., Hecker, M., et al. (2010). Quality control of inclusion bodies in *Escherichia coli*. *Microb. Cell Fact.* 9, 41. doi: 10.1186/1475-2859-9-41
- Kudou, M., Ejima, D., Sato, H., Yumioka, R., Arakawa, T., and Tsumoto, K. (2011a). Refolding single-chain antibody (scFv) using lauroyl-L-glutamate as a solubilization detergent and arginine as a refolding additive. *Protein Expr. Purif.* 77, 68–74. doi: 10.1016/j.pep.2010.12.007
- Kudou, M., Yumioka, R., Ejima, D., Arakawa, T., and Tsumoto, K. (2011b). A novel protein refolding system using lauroyl-L-glutamate as a solubilizing detergent and arginine as a folding assisting agent. *Protein Expr. Purif.* 75, 46–54. doi: 10.1016/j.pep.2010.08.011
- Lethanh, H., Neubauer, P., and Hoffmann, F. (2005). The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant *Escherichia coli* and delay degradation of inclusion bodies. *Microb. Cell Fact.* 4, 6. doi: 10.1186/1475-2859-4-6
- Liovic, M., Ozir, M., Zavec, A. B., Peternel, S., Komel, R., and Zupancic, T. (2012). Inclusion bodies as potential vehicles for recombinant protein delivery into epithelial cells. *Microb. Cell Fact.* 11, 67. doi: 10.1186/1475-2859-11-67
- Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., and Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3076–3081. doi: 10.1073/pnas.0708931105
- Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., et al. (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* 325, 328–332. doi: 10.1126/science.1173155
- Miller, J., Arrasate, M., Shaby, B. A., Mitra, S., Masliah, E., and Finkbeiner, S. (2010). Quantitative relationships between huntingtin levels, polyglutamine length, inclusion body formation, and neuronal death provide novel insight into huntington's disease molecular pathogenesis. *J. Neurosci.* 30, 10541–10550. doi: 10.1523/JNEUROSCI.0146-10.2010
- Orr, H. T., and Zoghbi, H. Y. (2007). Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575–621. doi: 10.1146/annurev.neuro.29.051605.113042
- Peternel, S., Jevsevar, S., Bele, M., Gaberc-Porekar, V., and Menart, V. (2008). New properties of inclusion bodies with implications for biotechnology. *Biotechnol. Appl. Biochem.* 49, 239–246. doi: 10.1042/BA20070140
- Radchenko, E., Rogoza, T., Khokhrina, M., Drozdova, P., and Mironova, L. (2011). SUP35 expression is enhanced in yeast containing [ISP+], a prion form of the transcriptional regulator Sfp1. *Prion* 5, 317–322. doi: 10.4161/pri.18426
- Rane, A. M., Jonnalagadda, S., and Li, Z. (2013). On-column refolding of bone morphogenetic protein-2 using cation exchange resin. *Protein Expr. Purif.* 90, 135–140. doi: 10.1016/j.pep.2013.05.008
- Rinas, U., Hoffmann, F., Betiku, E., Estape, D., and Marten, S. (2007). Inclusion body anatomy and functioning of chaperone-mediated *in vivo* inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. *J. Biotechnol.* 127, 244–257. doi: 10.1016/j.jbiotec.2006.07.004
- Rosano, G. L., and Ceccarelli, E. A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb. Cell Fact.* 8, 41. doi: 10.1186/1475-2859-8-41
- Sabate, R., De Groot, N. S., and Ventura, S. (2010). Protein folding and aggregation in bacteria. *Cell. Mol. Life Sci.* 67, 2695–2715. doi: 10.1007/s00018-010-0344-4
- Sabate, R., Espargaro, A., Saupe, S. J., and Ventura, S. (2009). Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion. *Microb. Cell Fact.* 8, 56. doi: 10.1186/1475-2859-8-56
- Sabate, R., Villar-Pique, A., Espargaro, A., and Ventura, S. (2012). Temperature dependence of the aggregation kinetics of Sup35 and Ure2p yeast prions. *Biomacromolecules* 13, 474–483. doi: 10.1021/bm201527m
- Schrodel, A., Volz, J., and De Marco, A. (2005). Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. *J. Biotechnol.* 120, 2–10. doi: 10.1016/j.jbiotec.2005.04.028
- Singh, S. M., Sharma, A., Upadhyay, A. K., Singh, A., Garg, L. C., and Panda, A. K. (2012). Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form. *Protein Expr. Purif.* 81, 75–82. doi: 10.1016/j.pep.2011.09.004
- Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A., and Hamodrakas, S. J. (2013). A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS ONE* 8:e54175. doi: 10.1371/journal.pone.0054175
- Tyedmers, J., Mogk, A., and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* 11, 777–788. doi: 10.1038/nrm2993
- Upadhyay, A. K., Murmu, A., Singh, A., and Panda, A. K. (2012). Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. *PLoS ONE* 7:e33951. doi: 10.1371/journal.pone.0033951
- Vázquez, E., Corchero, J. L., Burgueno, J. F., Seras-Franzoso, J., Kosoy, A., Bosser, R., et al. (2012). Functional inclusion bodies produced in bacteria as naturally occurring nanopills for advanced cell therapies. *Adv. Mater.* 24, 1742–1747. doi: 10.1002/adma.201104330
- Vera, A., Aris, A., Carrio, M., Gonzalez-Montalban, N., and Villaverde, A. (2005). Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. *J. Biotechnol.* 119, 163–171. doi: 10.1016/j.jbiotec.2005.04.006
- Villar-Pique, A., Espargaro, A., Sabate, R., De Groot, N. S., and Ventura, S. (2012). Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. *Microb. Cell Fact.* 11, 55. doi: 10.1186/1475-2859-11-55
- Villar-Pique, A., and Ventura, S. (2012). Modeling amyloids in bacteria. *Microb. Cell Fact.* 11, 166. doi: 10.1186/1475-2859-11-166
- Villaverde, A., García-Fruitos, E., Rinas, U., Seras-Franzoso, J., Kosoy, A., Corchero, J. L., et al. (2012). Packaging protein drugs as bacterial inclusion bodies for therapeutic applications. *Microb. Cell Fact.* 11, 76. doi: 10.1186/1475-2859-11-76
- Walsh, D. J., Noble, G. P., Piro, J. R., and Supattapone, S. (2012). Non-reducing alkaline solubilization and rapid on-column refolding of recombinant prion protein. *Prep. Biochem. Biotechnol.* 42, 77–86. doi: 10.1080/10826068.2011.564256
- Walther, C., Mayer, S., Treflov, A., Sekot, G., Hahn, R., Jungbauer, A., et al. (2014). Prediction of inclusion body solubilization from shaken to stirred reactors. *Biotechnol. Bioeng.* 111, 84–94. doi: 10.1002/bit.24998
- Wang, L. (2009). Towards revealing the structure of bacterial inclusion bodies. *Prion* 3, 139–145. doi: 10.4161/pri.3.3.9922
- Wasmer, C., Benkemoun, L., Sabate, R., Steinmetz, M. O., Coulary-Salin, B., Wang, L., et al. (2009). Solid-state NMR spectroscopy reveals that *E. coli* inclusion bodies of HET-s(218-289) are amyloids. *Angew. Chem. Int. Ed Engl.* 48, 4858–4860. doi: 10.1002/anie.200806100

- Winkler, J., Seybert, A., König, L., Pruggnaller, S., Haselmann, U., Sourjik, V., et al. (2010). Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.* 29, 910–923. doi: 10.1038/emboj.2009.412
- Zheng, H., Miyakawa, T., Sawano, Y., Yamagoe, S., and Tanokura, M. (2013). Expression, high-pressure refolding and purification of human leukocyte cell-derived chemotaxin 2 (LECT2). *Protein Expr. Purif.* 88, 221–229. doi: 10.1016/j.pep.2013.01.008

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

Received: 01 December 2013; paper pending published: 14 January 2014; accepted: 28 January 2014; published online: 14 February 2014.

Citation: Ramón A, Señoralé-Pose M and Marín M (2014) Inclusion bodies: not that bad. . . *Front. Microbiol.* 5:56. doi: 10.3389/fmicb.2014.00056

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ramón, Señoralé-Pose and Marín. 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.



Expression of codon optimized genes in microbial systems: current industrial applications and perspectives

Claudia Elena, Pablo Ravasi, María E. Castelli, Salvador Peirú and Hugo G. Menzella*

Genetic Engineering and Fermentation Technology, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario-Conicet, Rosario, Argentina

Edited by:

Eduardo A. Ceccarelli, Universidad Nacional de Rosario, Argentina

Reviewed by:

Blaine Pfeifer, The State University of New York at Buffalo, USA

Christopher Desmond Reeves, Amyris, USA

*Correspondence:

Hugo G. Menzella, Genetic Engineering and Fermentation Technology, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario-Conicet, Suipacha 590, Rosario 2000, Argentina
e-mail: hmenzella@fbioyf.unr.edu.ar

The efficient production of functional proteins in heterologous hosts is one of the major bases of modern biotechnology. Unfortunately, many genes are difficult to express outside their original context. Due to their apparent “silent” nature, synonymous codon substitutions have long been thought to be trivial. In recent years, this dogma has been refuted by evidence that codon replacement can have a significant impact on gene expression levels and protein folding. In the past decade, considerable advances in the speed and cost of gene synthesis have facilitated the complete redesign of entire gene sequences, dramatically improving the likelihood of high protein expression. This technology significantly impacts the economic feasibility of microbial-based biotechnological processes by, for example, increasing the volumetric productivities of recombinant proteins or facilitating the redesign of novel biosynthetic routes for the production of metabolites. This review discusses the current applications of this technology, particularly those regarding the production of small molecules and industrially relevant recombinant enzymes. Suggestions for future research and potential uses are provided as well.

Keywords: synthetic biology, gene design, codon optimization, strain engineering, microbial systems

INTRODUCTION

Microorganisms are at the core of the production of pharmaceuticals, industrial enzymes, and fine chemicals. In many cases, heterologous expression of genes is required to meet commercial-level demands of target proteins and/or metabolites. In this context, variation in codon usage is considered as one of the major factors affecting protein expression levels, since the presence of rare codons can reduce the translation rate and induce translation errors with a significant impact on the economics of recombinant microbe-based production processes (Ikemura, 1981; Gustafsson et al., 2004). The generation of massive genome sequencing data and cost-effective custom DNA synthesis are foundational technologies for synthetic biology, an emerging discipline that aims to create novel organisms containing designer genetic circuits for the production of drugs, industrial enzymes, biofuels, and chemicals (Endy, 2005; McDaniel and Weiss, 2005; Heinemann and Panke, 2006; Leonard et al., 2008). These circuits are built from standard biological parts, including vectors, promoters, ribosomal binding sites (RBSs), transcriptional terminators, and other gene expression regulatory elements. These parts were initially borrowed from nature and nowadays engineered, to adapt their performance to a particular application, or combined to create sophisticated devices (Shiue and Prather, 2012).

Over the past decade, synthetic biology has contributed to significantly reduce the cost of many products manufactured in microbial systems where only one gene needs to be over-expressed. In many cases, the production of a target protein can be boosted by several orders of magnitude by replacing a native sequence with its optimized counterpart (Gustafsson et al., 2004, 2012). This seemingly simple adjustment is of remarkable importance, since many of these products are now traded as commodities and thus there

is a continuous need to reduce manufacturing costs in order to remain competitive in the global markets (Menzella, 2011). The ambitious next step of synthetic biology is to further reduce the cost and time involved in developing recombinant organisms by using pre-assembled parts that provide stable, predictable protein expression (Dellomonaco et al., 2010; Nielsen and Keasling, 2011).

So far, most of the progress made in synthetic biology was achieved in *Escherichia coli*, a preferred host for the production of recombinant proteins because it combines fast growth rate, inexpensive fermentation media and well understood genetics (Burgess-Brown et al., 2008; Welch et al., 2009; Menzella, 2011). However, efforts have been recently expanded to other hosts including *Streptomyces* species (Medema et al., 2011), *Corynebacterium glutamicum* (Becker and Wittmann, 2012), yeast (Krivoruchko et al., 2011; Siddiqui et al., 2012; Furukawa and Hohmann, 2013), and algae (Wang et al., 2012; Gimpel et al., 2013). This expanded landscape seeks to take advantage of the natural capabilities to synthesize precursors and cofactors required to produce a particular target, exploit secretion abilities, or utilize natural tolerance to over-accumulated metabolites (Zhu et al., 2012). In this review we summarize the current state of the technology for the expression of codon optimized genes in microbial systems. Examples of its application for the production of small molecules and recombinant enzymes of industrial interest are presented, and suggestions for future research and uses are provided.

GENE DESIGN

Choosing a gene for optimal expression requires selection from a large number of sequences. For example, a protein with an average size of 30 kDa may, in theory, be encoded by 10^{100} possible DNA sequences (Welch et al., 2009). Historically, two approaches

have been used for codon optimization. The first, designated “one amino acid-one codon,” uses the most abundant codon of the host to encode all occurrences of a given amino acid in the optimized sequence (Fuglsang, 2003; Gao et al., 2004; Supek and Vlahovicek, 2004; Villalobos et al., 2006; Feng et al., 2010; Marlatt et al., 2010; Wang et al., 2010). This simple strategy, the most popular in the early days of gene synthesis technology, has a major drawback: a strongly transcribed mRNA from a gene with this design will contain a high concentration of a subset of codons, resulting in an imbalance in the tRNA pool, which in turn may reduce growth due to tRNA depletion (Gong et al., 2006; Villalobos et al., 2006).

The second approach, named “codon randomization,” uses translation tables based on the frequency distribution of the codons in an entire genome or a subset of highly expressed genes. These tables attach weights to each codon, thus codons are assigned randomly with a probability given by the weights (Kodumal et al., 2004; Jayaraj et al., 2005; Menzella et al., 2005; Welch et al., 2009; Wang et al., 2010). This strategy was shown to be superior and was quickly adopted by the synthetic biology community. In addition to improving the yield of the desired product, the “codon randomization” strategy offers many further advantages. For example, flexibility in codon selection facilitates gene design by avoiding: (i) repetitive elements that may lead to gene deletions; (ii) internal RBSs, polyadenylation signals, or transcriptional terminators; (iii) secondary mRNA structures (Luisi et al., 2013); and (iv) by facilitating elimination of unwanted restriction sites to assist the assembly of larger constructs (Villalobos et al., 2006). Several large-scale systematic studies describing variations on this strategy have been conducted in recent years to provide data on the effect of sequence variables (Kudla et al., 2009; Welch et al., 2009; Allert et al., 2010).

Besides codon optimization, other parameters need to be considered to design a gene for efficient translation, including the global GC content (Gustafsson, 2009), local context of a given codon (Villalobos et al., 2006), the presence of mRNA sequence motifs (Pertzev and Nicholson, 2006), and the sequence of the region including the first 10 codons (Goodman et al., 2013). Many web-based free softwares, with features ranging from basic to advanced, were created for gene design during the last decade. Examples include: DNA Works (Hoover and Lubkowski, 2002), GeMS (Jayaraj et al., 2005); Optimizer (Puigbo et al., 2007); Synthetic Gene Designer (Wu et al., 2006); and Gene Designer (Villalobos et al., 2006). Currently, the majority of synthetic DNA suppliers (including GenScript, DNA2.0, GeneArt and Genewiz) offer sequence optimization services using proprietary algorithms at no additional cost.

PARTS AND VECTORS

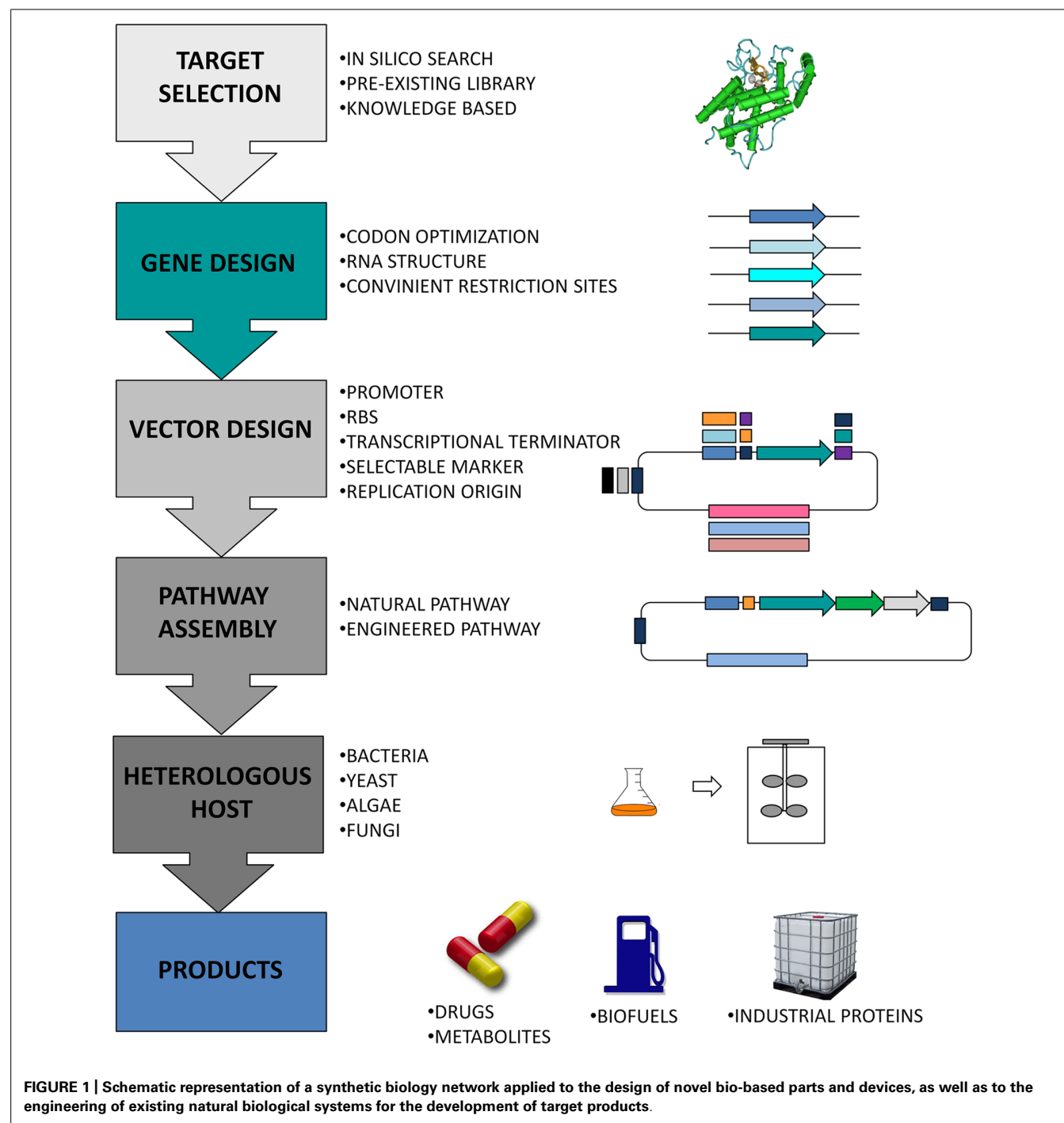
The application of synthetic DNA technology in engineered microorganisms is not restricted to redesigned genes. Classic expression vectors widely used in strain engineering derive from natural sources and were never optimized for robust production. Recently, great interest has arisen in the systematic engineering and standardization of gene expression parts such as promoters, translation initiation signals, transcriptional terminators, selectable markers, and replication origins to allow fast and predictable combination of these elements.

Some applications, such as metabolic engineering, require optimal levels of each enzyme to maximize production. This is typically achieved by modulating gene expression by, for example, varying transcription or translation levels. Synthetic biology can offer collections of promoters and RBSs capable of providing different levels of gene expression for this purpose (Boyle and Silver, 2012; Meng et al., 2013; Vogl et al., 2013). So far, most of the available promoters have been taken from the natural sequences driving the expression of highly expressed genes. Typical examples are the widely used AOX promoter from *Pichia pastoris* (Tschopp et al., 1987) for yeast and the bacteriophage T7 promoter for *E. coli* (Studier and Moffatt, 1986), which provide high transcription levels. Nowadays, synthetic promoter libraries for tunable gene expression are available for many industrially relevant microorganisms including *E. coli* (Wu et al., 2013), *P. pastoris* (Hartner et al., 2008; Ruth et al., 2010; Vogl et al., 2013), *C. glutamicum* (Yim et al., 2013), and *Bacillus subtilis* (Hansen et al., 2009). Likewise, synthetic RBSs can be used to regulate gene expression (Basu et al., 2005; Pfleger et al., 2006). Furthermore a novel method for automatic design of artificial RBSs to control gene expression has been recently described, expanding the toolbox of artificial sequences to be used in custom genetic circuits (Salis et al., 2009).

Despite current efforts, accurate predictions of the response of any given promoter or RBS have often remained elusive. It is possible that unknown interactions among isolated components may significantly affect the optimal level of gene expression needed to achieve a particular flux through a biosynthetic pathway (Keasling, 2012). In a recent work, Kosuri et al. (2013) provided an alternative strategy to screen the behavior of gene expression regulatory elements. They synthesized 12,563 combinations of common promoters and RBSs and simultaneously measured DNA, RNA, and protein levels from the entire library. They found that RNA and protein expression were within twofold of expected levels 80 and 64% of the time, respectively, and that the worst 5% of constructs deviated from prediction by 13-fold on average, which could hinder large-scale genetic engineering projects. This comprehensive study provides a means to test standard part combinations to optimize production of a particular target molecule.

Genes are usually introduced into production microorganisms using plasmid vectors (Figure 1). Synthetic biology provides the means to speed up this process by using designer plasmid vectors, where all the components are synthesized with standard formats to facilitate exchange and testing of parts, as well as the assembly of multi-gene constructs (Leonard et al., 2008; Shetty et al., 2008). Several designs for the construction of synthetic plasmids and for the assembly of parts have been proposed (Menzella et al., 2005, 2007; Reisinger et al., 2006; Shetty et al., 2008; Sarrion-Perdigones et al., 2011). The most popular format among the synthetic biology community was created by Knight and co-workers (Shetty et al., 2011). They proposed the BioBrick standard, where all parts are flanked by a common set of restriction sites that allow the joining, combination, and rapid assembly of genetic parts to create functional gene expression units.

So far, most of the work to create synthetic vectors reported in the literature has been done in *E. coli*. Recently, we created a



plasmid-based platform for the rapid engineering of *C. glutamicum*, a microorganism of great industrial interest. The approach uses reporter genes to examine and classify promoters and RBSs and permits the easy assembly of operons and genes clusters for co-expression of heterologous genes to facilitate metabolic engineering. Similarly, Constante and co-workers described a platform to engineer eukaryotic hosts by using the BioBrick principle. Interestingly, the system contains a variety of novel parts and implements a recombinase-mediated DNA insertion, allowing

chromosomal site-directed exchange of genes in eukaryotic cell lines (Constante et al., 2011).

PRACTICAL APPLICATIONS

The list of products obtained by the expression of codon optimized genes in microorganisms is constantly growing and includes biofuels, pharmaceuticals, novel bio-based materials and chemicals, industrial enzymes, amino acids, and other metabolites (Table 1).

Table 1 | Expression of redesigned genes of industrial interest in microbial systems.

Application	Product	Host	Titer	Reference
Biofuels	Short-chain alcohols			
	Isopropanol	<i>E. coli</i>	4.9–13.6 g/L	Hanai et al. (2007)
	Butanol	<i>E. coli</i>	1.2 g/L	Inui et al. (2008)
		<i>E. coli</i>	4 g/L	Bond-Watts et al. (2011)
	Fatty acid derivatives			
	C12–C18 fatty acid ethyl esters	<i>E. coli</i>	700 mg/L	Steen et al. (2010)
	Isoprenoids			
	Bisabolene	<i>E. coli/S. cerevisiae</i>	>900 mg/L	Peralta-Yahya et al. (2011)
Biopolymers	Polyhydroxybutyrate	<i>S. cerevisiae</i>	180 mg/L	Kocharin et al. (2013)
Chemical precursors	Methyl halides	<i>S. cerevisiae</i>	860 mg/L	Bayer et al. (2009)
Industrial enzymes	Cellulases	<i>S. cerevisiae</i>	NR ^(*)	Heinzelman et al. (2009)
	Phytases	<i>A. oryzae</i>	NR ^(*)	Lichtenberg et al. (2011)
		<i>P. pastoris</i>	12.2 g/L	Xiong et al. (2006)
	Cutinases	<i>P. pastoris</i>	NR ^(*)	Liu et al. (2009)
	Lignocellulases	<i>P. pastoris</i>	1–6 g/L	Mellitzer et al. (2012)
	Lipases	<i>P. pastoris</i>	4 g/L	Chang et al. (2006)
	Prochimosin	<i>E. coli</i>	20 g/L ^(**)	Menzella (2011)
	Steryl glycosidase	<i>E. coli</i>	NR ^(*)	Aguirre et al. (2013)
Metabolites	L-amino acids	<i>C. glutamicum</i>	30–100 g/L	Becker and Wittmann (2012)
Therapeutics	Artemisinic acid	<i>S. cerevisiae</i>	25 g/L	Paddon et al. (2013)
	Polyketide precursors	<i>E. coli</i>	Up to 90 mg/L	Menzella et al. (2007, 2010)
	Multiepitope antigenic proteins	<i>E. coli</i>	NR ^(*)	Talha et al. (2010), de Souza et al. (2013)

(*) Not Reported; (**) Inclusion bodies.

Production of novel biofuels is one of the most attractive applications for synthetic biology. Fuels like ethanol, biodiesel, butanol, and terpenoid compounds are currently produced using engineered microbes (Table 1). In fact, the main obstacle for the production of these molecules at commercial level is the development of robust microbes and processes (Fischer et al., 2008). Synthetic biology provides tools to achieve optimal expression of pathway genes to ensure the efficient conversion of feedstock materials to target molecules, which is critical to the success of any metabolic engineering strategy. There has been considerable progress recently in the production of different biofuels, and some of the processes have reached promising yields. Hanai and co-workers combined enzymes from *Clostridium acetobutylicum* (Thl, CtfAB, and ADC), *Clostridium beijerinckii* (ADH), and *E. coli* (AtoAD) to assemble a fermentative pathway in *E. coli* that resulted in production of isopropanol at titers ranging from 4.9 to 13.6 g/L (Hanai et al., 2007). Butanol production was achieved in *E. coli* using the biosynthetic pathway from *C. acetobutylicum* and other related clostridial species, reaching titers up to 1.2 g/L (Inui et al., 2008). This was further improved to more than 4 g/L of butanol production by replacing enzymes that are naturally reversible with those that drive the reaction toward butanol formation, expressed from codon optimized genes from different bacterial species (Bond-Watts et al., 2011).

Fatty acid derivatives are other promising biofuel candidates, due to their high energy density and low water solubility. Stenn et al. engineered *E. coli* to produce C₁₂–C₁₈ fatty acid ethyl esters (FAEEs) directly from glucose at a titer of ~700 mg/L (Steen et al., 2010). Five engineering strategies were combined to achieve this titer, including the elimination of the β -oxidation pathway and the expression of several synthetic genes from different microorganisms. Monoterpene and sesquiterpene hydrocarbons such as limonene, pinene, and farnesene, are isoprenoid compounds with promising fuel applications that have been produced in *E. coli* and *S. cerevisiae*. Expression in *E. coli* of a codon-optimized bisabolene synthase from the fir tree *Abies grandis*, in conjunction with the introduction of an optimized heterologous mevalonate pathway, resulted in sesquiterpene bisabolene production of 900 mg/L.

A *S. cerevisiae* strain that overproduces farnesyl pyrophosphate also gave bisabolene titers higher than 900 mg/L using the same bisabolene synthase (Peralta-Yahya et al., 2011). The mevalonate pathway expression was further improved in *E. coli* by introducing codon-optimized versions of the mevalonate kinase and phosphomevalonate kinase genes after they were identified as potential pathway bottlenecks (Redding-Johanson et al., 2011).

Codon optimized genes have been extensively used to produce pharmaceuticals in microbial platforms. Polyketides are a class of natural products with a high number of well-established clinical

applications. The development of a variety of methods for polyketide synthases (PKS) engineering (Menzella and Reeves, 2007; Peiru et al., 2009, 2010) led to a pioneer synthetic biology project conducted at Kosan Biosciences. The goal was to obtain polyketide precursors for the synthesis of novel drugs. First, a generic design for type I PKS genes was created to enable easy assembly and expression of chimeric enzymes (Kodumal et al., 2004; Menzella et al., 2005). The sequences of the synthetic genes were then redesigned with custom made software to optimize codon usage in order to maximize expression in *E. coli* and provide a standard set of restriction sites to allow combinatorial assembly into unnatural enzymes. Next, more than three million bases of PKS genes were tested to validate the platform. These efforts produced a variety of novel valuable compounds (Menzella et al., 2007; Menzella et al., 2010).

Another remarkable contribution of synthetic biology is the microbial production of artemisinin, a sesquiterpene endoperoxide used to treat malaria (Paddon et al., 2013). This compound is naturally produced by the plant *Artemisia annua*, but the production of plant-derived artemisinin is expensive; which limits its access to many patients. Recently, Paddon and coworkers engineered strains of *S. cerevisiae* for production of artemisinic acid, a precursor of artemisinin by fermentation. The simultaneous co-expression of synthetic genes provided an efficient biosynthetic route to artemisinic acid, with fermentation titers of 25 g/L.

Production of proteins for therapeutic use also takes advantage of the use of synthetic genes; a comprehensive review describing progress in this field has been recently published by Mitchell (2011). An elegant synthetic biology approach was used to create designer antigenic proteins for immunoassay-based diagnosis. By designing synthetic genes encoding tandem combinations of epitopes joined by flexible peptide linkers, chimeric proteins were obtained for the detection of antibodies in sera with higher sensitivity and specificity (Talha et al., 2010; de Souza et al., 2013).

The global market for industrial enzymes exceeded \$4 billion in 2012 and is therefore an attractive target for cost reduction using synthetic biology tools (Zhou et al., 2004; Menzella, 2011). The use of codon optimized genes allowed notable increases in the production of many enzymes in a variety of hosts, including cellulases in *S. cerevisiae* (Heinzelman et al., 2009), phytases in *Aspergillus oryzae* (Lichtenberg et al., 2011), cutinases (Liu et al., 2009), lignocellulases (Mellitzer et al., 2012), and lipases (Chang et al., 2006) in *P. pastoris* and calf prochymosin in *E. coli* (Menzella, 2011). In the last example, a strain developed in our laboratory harboring a codon optimized gene produced 70% more prochymosin than that obtained with the wild type sequence, with the concomitant reduction in production costs.

In addition to the contribution to achieve more competitive production processes, synthetic genes provide an attractive alternative for the discovery of enzymes for new applications. For example, in order to search for thermostable enzymes to hydrolyze sterol glucosides (major contaminants of oil-derived biodiesel), we screened a library of archeal genes by retrieving the sequences *in silico*, synthesizing codon optimized genes for expression in *E. coli* and assessing their activity against the target. The approach was very successful and resulted in excellent candidates for industrial

use (Aguirre et al., 2013). Other products of commercial interest recently obtained from strains carrying codon optimized genes include L-amino acids in *C. glutamicum* and *E. coli* (Becker and Wittmann, 2012), and polyhydroxybutyrate and methyl halides in *S. cerevisiae* (Bayer et al., 2009; Kocharin et al., 2013).

CONCLUSION AND FUTURE PERSPECTIVES

The benefits of using codon optimized genes in industrial biotechnology have been extensively demonstrated during the past decade and this technology is being rapidly adopted by strain developers in order to remain competitive in the current market. In the examples presented here, just one or a few synthetic genes need to be introduced into a host to generate novel products or to dramatically reduce the cost of producing existing ones. The cost of synthetic genes has been constantly decreasing during the last decade; and technologies to assemble large fragments of DNA and to make multiple simultaneous changes to wild type genomes are becoming available (Montague et al., 2012). Thus, we can envision a future where custom-made microorganisms can be designed for a particular application (Gibson et al., 2010).

One of the fields where these new technologies can make a dramatic contribution is the production of commodity chemicals in microbes. Initial steps toward this ambitious goal have already been taken by industry. For example, an *E. coli* strain has been engineered to produce 1,3-propanediol, where in addition to the introduction of the pathway for the production of this target from glycerol, several changes were made in the genome to increase the final yield (Nakamura and Whited, 2003).

Although tremendous progress has been made, in order to fully harness the potential of synthetic biology we need a deeper understanding of the underlying molecular principles of living systems and further development of bioinformatic tools to assist in the modeling of synthetic genomes behavior. These advances are expected to arrive from the interactions among many scientific disciplines.

REFERENCES

- Aguirre, A., Peiru, S., Eberhardt, F., Vetcher, L., Cabrera, R., and Menzella, H. G. (2013). Enzymatic hydrolysis of sterol glucosides, major contaminants of vegetable oil-derived biodiesel. *Appl. Microbiol. Biotechnol.* doi: 10.1007/s00253-013-5345-4 [Epub ahead of print].
- Allert, M., Cox, J. C., and Hellinga, H. W. (2010). Multifactorial determinants of protein expression in prokaryotic open reading frames. *J. Mol. Biol.* 402, 905–918. doi: 10.1016/j.jmb.2010.08.010
- Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature* 434, 1130–1134. doi: 10.1038/nature03461
- Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., and Voigt, C. A. (2009). Synthesis of methyl halides from biomass using engineered microbes. *J. Am. Chem. Soc.* 131, 6508–6515. doi: 10.1021/ja809461u
- Becker, J., and Wittmann, C. (2012). Bio-based production of chemicals, materials and fuels-*Corynebacterium glutamicum* as versatile cell factory. *Curr. Opin. Biotechnol.* 23, 631–640. doi: 10.1016/j.copbio.2011.11.012
- Bond-Watts, B. B., Bellerose, R. J., and Chang, M. C. (2011). Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* 7, 222–227. doi: 10.1038/nchembio.537
- 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

- Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. (2008). Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. *Protein Expr. Purif.* 59, 94–102. doi: 10.1016/j.pep.2008.01.008
- Constante, M., Grunberg, R., and Isalan, M. (2011). A biobrick library for cloning custom eukaryotic plasmids. *PLoS ONE* 6:e23685. doi: 10.1371/journal.pone.0023685
- Chang, S. W., Lee, G. C., and Shaw, J. F. (2006). Codon optimization of *Candida rugosa* lip1 gene for improving expression in *Pichia pastoris* and biochemical characterization of the purified recombinant LIP1 lipase. *J. Agric. Food Chem.* 54, 815–822. doi: 10.1021/jf052183k
- Dellomonaco, C., Fava, F., and Gonzalez, R. (2010). The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microb. Cell Fact.* 9, 3. doi: 10.1186/1475-2859-9-3
- de Souza, M. Q., Galdino, A. S., Dos Santos, J. C., Soares, M. V., De Nobrega, Y. C., Alvares Ada, C., et al. (2013). A recombinant multiepitope protein for hepatitis B diagnosis. *Biomed. Res. Int.* 2013, 148317. doi: 10.1155/2013/148317
- Endy, D. (2005). Foundations for engineering biology. *Nature* 438, 449–453. doi: 10.1038/nature04342
- Feng, Z., Lanwei, Z., Han, X., and Zhang, Y. (2010). Codon optimization of the calf prothymosin gene and its expression in *Kluyveromyces fragilis*. *World J. Microbiol. Biotechnol.* 26, 895–901. doi: 10.1007/s11274-009-0249-2
- Fischer, C. R., Klein-Marcuschamer, D., and Stephanopoulos, G. (2008). Selection and optimization of microbial hosts for biofuels production. *Metab. Eng.* 10, 295–304. doi: 10.1016/j.ymben.2008.06.009
- Fuglsang, A. (2003). Codon optimizer: a freeware tool for codon optimization. *Protein Expr. Purif.* 31, 247–249. doi: 10.1016/S1046-5928(03)00213-4
- Furukawa, K., and Hohmann, S. (2013). Synthetic biology: lessons from engineering yeast MAPK signalling pathways. *Mol. Microbiol.* 88, 5–19. doi: 10.1111/mmi.12174
- Gao, W., Rzewski, A., Sun, H., Robbins, P. D., and Gambotto, A. (2004). UpGene: application of a web-based DNA codon optimization algorithm. *Biotechnol. Prog.* 20, 443–448. doi: 10.1021/bp0300467
- 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
- Gimpel, J. A., Specht, E. A., Georgianna, D. R., and Mayfield, S. P. (2013). Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr. Opin. Chem. Biol.* 17, 489–495. doi: 10.1016/j.cbpa.2013.03.038
- Gong, M., Gong, F., and Yanofsky, C. (2006). Overexpression of tnaC of *Escherichia coli* inhibits growth by depleting tRNA^{Pro} availability. *J. Bacteriol.* 188, 1892–1898. doi: 10.1128/JB.188.5.1892-1898.2006
- Goodman, D. B., Church, G. M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479. doi: 10.1126/science.1241934
- Gustafsson, C. (2009). “Tools designed to regulate translational efficiency,” in *The Metabolic Pathway Engineering Handbook: Tools and Applications*, ed. C. Smolke (Boca Raton, FL: CRC Press), 678.
- Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22, 346–353. doi: 10.1016/j.tibtech.2004.04.006
- Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., and Welch, M. (2012). Engineering genes for predictable protein expression. *Protein Expr. Purif.* 83, 37–46. doi: 10.1016/j.pep.2012.02.013
- Hanai, T., Atsumi, S., and Liao, J. C. (2007). Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* 73, 7814–7818. doi: 10.1128/AEM.01140-07
- Hansen, M. E., Wangari, R., Hansen, E. B., Mijakovic, I., and Jensen, P. R. (2009). Engineering of *Bacillus subtilis* 168 for increased nisin resistance. *Appl. Environ. Microbiol.* 75, 6688–6695. doi: 10.1128/AEM.00943-09
- Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., et al. (2008). Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res.* 36, e76. doi: 10.1093/nar/gkn369
- Heinemann, M., and Panke, S. (2006). Synthetic biology – putting engineering into biology. *Bioinformatics* 22, 2790–2799. doi: 10.1093/bioinformatics/btl469
- Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., et al. (2009). A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5610–5615. doi: 10.1073/pnas.0901417106
- Hoover, D. M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* 30, e43. doi: 10.1093/nar/30.10.e43
- Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 146, 1–21. doi: 10.1016/0022-2836(81)90363-6
- Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., et al. (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 77, 1305–1316. doi: 10.1007/s00253-007-1257-5
- Jayaraj, S., Reid, R., and Santi, D. V. (2005). GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res.* 33, 3011–3016. doi: 10.1093/nar/gki614
- Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195. doi: 10.1016/j.ymben.2012.01.004
- Kochan, K., Siewers, V., and Nielsen, J. (2013). Improved polyhydroxybutyrate production by *Saccharomyces cerevisiae* through the use of the phosphoketolase pathway. *Biotechnol. Bioeng.* 110, 2216–2224. doi: 10.1002/bit.24888
- Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004). Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15573–15578. doi: 10.1073/pnas.0406911101
- Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., et al. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14024–14029. doi: 10.1073/pnas.1301301110
- Krivoruchko, A., Siewers, V., and Nielsen, J. (2011). Opportunities for yeast metabolic engineering: lessons from synthetic biology. *Biotechnol. J.* 6, 262–276. doi: 10.1002/biot.201000308
- Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258. doi: 10.1126/science.1170160
- Leonard, E., Nielsen, D., Solomon, K., and Prather, K. J. (2008). Engineering microbes with synthetic biology frameworks. *Trends Biotechnol.* 26, 674–681. doi: 10.1016/j.tibtech.2008.08.003
- Lichtenberg, J., Pedersen, P. B., Elvig-Joergensen, S. G., Skov, L. K., Olsen, C. L., and Glitsøe, L. V. (2011). Toxicological studies on a novel phytase expressed from synthetic genes in *Aspergillus oryzae*. *Regul. Toxicol. Pharmacol.* 60, 401–410. doi: 10.1016/j.yrtph.2011.05.010
- Liu, Z., Gosser, Y., Baker, P. J., Ravee, Y., Lu, Z., Alemu, G., et al. (2009). Structural and functional studies of *Aspergillus oryzae* cutinase: enhanced thermostability and hydrolytic activity of synthetic ester and polyester degradation. *J. Am. Chem. Soc.* 131, 15711–15716. doi: 10.1021/ja9046697
- Luisi, I., Pavan, S., Fontanive, G., Tossi, A., Benedetti, F., Savoini, A., et al. (2013). An albumin-derived peptide scaffold capable of binding and catalysis. *PLoS ONE* 8:e56469. doi: 10.1371/journal.pone.0056469
- Marlatt, N. M., Spratt, D. E., and Shaw, G. S. (2010). Codon optimization for enhanced *Escherichia coli* expression of human S100A11 and S100A1 proteins. *Protein Expr. Purif.* 73, 58–64. doi: 10.1016/j.pep.2010.03.015
- McDaniel, R., and Weiss, R. (2005). Advances in synthetic biology: on the path from prototypes to applications. *Curr. Opin. Biotechnol.* 16, 476–483. doi: 10.1016/j.copbio.2005.07.002
- Medema, M. H., Breitling, R., and Takano, E. (2011). Synthetic biology in *Streptomyces* bacteria. *Methods Enzymol.* 497, 485–502. doi: 10.1016/B978-0-12-385075-1.00021-4
- Mellitzer, A., Weis, R., Glieder, A., and Flicker, K. (2012). Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Microb. Cell Fact.* 11, 61. doi: 10.1186/1475-2859-11-61
- Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G., and Wang, Y. (2013). Quantitative design of regulatory elements based on high-precision strength

- prediction using artificial neural network. *PLoS ONE* 8:e60288. doi: 10.1371/journal.pone.0060288
- Menzella, H. G. (2011). Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli*. *Microb. Cell Fact.* 10, 15. doi: 10.1186/1475-2859-10-15
- Menzella, H. G., Carney, J. R., Li, Y., and Santi, D. V. (2010). Using chemobiosynthesis and synthetic mini-polyketide synthases to produce pharmaceutical intermediates in *Escherichia coli*. *Appl. Environ. Microbiol.* 76, 5221–5227. doi: 10.1128/AEM.02961-09
- Menzella, H. G., Carney, J. R., and Santi, D. V. (2007). Rational design and assembly of synthetic trimodular polyketide synthases. *Chem. Biol.* 14, 143–151. doi: 10.1016/j.chembiol.2006.12.002
- Menzella, H. G., and Reeves, C. D. (2007). Combinatorial biosynthesis for drug development. *Curr. Opin. Microbiol.* 10, 238–245. doi: 10.1016/j.mib.2007.05.005
- Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., et al. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* 23, 1171–1176. doi: 10.1038/nbt1128
- Mitchell, W. (2011). Natural products from synthetic biology. *Curr. Opin. Chem. Biol.* 15, 505–515. doi: 10.1016/j.cbpa.2011.05.017
- Montague, M. G., Lartigue, C., and Vashee, S. (2012). Synthetic genomics: potential and limitations. *Curr. Opin. Biotechnol.* 23, 659–665. doi: 10.1016/j.copbio.2012.01.014
- 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
- Nielsen, J., and Keasling, J. D. (2011). Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* 29, 693–695. doi: 10.1038/nbt.1937
- Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532. doi: 10.1038/nature12051
- Peiru, S., Gramajo, H., and Menzella, H. G. (2009). Design and synthesis of pathway genes for polyketide biosynthesis. *Methods Enzymol.* 459, 319–337. doi: 10.1016/S0076-6879(09)04614-X
- Peiru, S., Gramajo, H. C., and Menzella, H. G. (2010). Recombinant approaches to large polyketide molecules as potential drugs. *Drug Discov. Today Technol.* 7, e95–e146. doi: 10.1016/j.ddtec.2010.10.002
- Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* 2, 483. doi: 10.1038/ncomms1494
- Pertzev, A. V., and Nicholson, A. W. (2006). Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res.* 34, 3708–3721. doi: 10.1093/nar/gkl459
- Pfleger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* 24, 1027–1032. doi: 10.1038/nbt1226
- Puigbo, P., Guzman, E., Romeu, A., and Garcia-Vallve, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* 35, W126–W131. doi: 10.1093/nar/gkm219
- Redding-Johanson, A. M., Batth, T. S., Chan, R., Krupa, R., Szmidt, H. L., Adams, P. D., et al. (2011). Targeted proteomics for metabolic pathway optimization: application to terpene production. *Metab. Eng.* 13, 194–203. doi: 10.1016/j.ymben.2010.12.005
- Reisinger, S. J., Patel, K. G., and Santi, D. V. (2006). Total synthesis of multi-kilobase DNA sequences from oligonucleotides. *Nat. Protoc.* 1, 2596–2603. doi: 10.1038/nprot.2006.426
- Ruth, C., Zuellig, T., Mellitzer, A., Weis, R., Looser, V., Kovar, K., et al. (2010). Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*. *Syst. Synth. Biol.* 4, 181–191. doi: 10.1007/s11693-010-9057-0
- Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950. doi: 10.1038/nbt.1568
- Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juarez, P., Fernandez-Del-Carmen, A., Granell, A., et al. (2011). GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE* 6:e21622. doi: 10.1371/journal.pone.0021622
- Shetty, R., Lizarazo, M., Rettberg, R., and Knight, T. F. (2011). Assembly of BioBrick standard biological parts using three antibiotic assembly. *Methods Enzymol.* 498, 311–326. doi: 10.1016/B978-0-12-385120-8.00013-9
- Shetty, R. P., Endy, D., and Knight, T. F. Jr. (2008). Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2, 5. doi: 10.1186/1754-1611-2-5
- Shiue, E., and Prather, K. L. J. (2012). Synthetic biology devices as tools for metabolic engineering. *Biochem. Eng. J.* 65, 82–89. doi: 10.1016/j.bej.2012.04.006
- Siddiqui, M. S., Thodey, K., Trenchard, I., and Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* 12, 144–170. doi: 10.1111/j.1567-1364.2011.00774.x
- Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., et al. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559–562. doi: 10.1038/nature08721
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130. doi: 10.1016/0022-2836(86)90385-2
- Supek, F., and Vlahovicek, K. (2004). INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics* 20, 2329–2330. doi: 10.1093/bioinformatics/bth238
- Talha, S. M., Salminen, T., Chugh, D. A., Swaminathan, S., Soukka, T., Pettersson, K., et al. (2010). Inexpensive designer antigen for anti-HIV antibody detection with high sensitivity and specificity. *Clin. Vaccine Immunol.* 17, 335–341. doi: 10.1128/CVI.00283-09
- Tschopp, J. F., Brust, P. F., Cregg, J. M., Stillman, C. A., and Gingeras, T. R. (1987). Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Res.* 15, 3859–3876. doi: 10.1093/nar/15.9.3859
- Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7:285. doi: 10.1186/1471-2105-7-285
- Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., and Glieder, A. (2013). Synthetic core promoters for *Pichia pastoris*. *ACS Synth. Biol.* doi: 10.1021/sb400091p [Epub ahead of print].
- Wang, B., Wang, J., Zhang, W., and Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol.* 3:344. doi: 10.3389/fmicb.2012.00344
- Wang, X., Li, X., Zhang, Z., Shen, X., and Zhong, F. (2010). Codon optimization enhances secretory expression of *Pseudomonas aeruginosa* exotoxin A in *E. coli*. *Protein Expr. Purif.* 72, 101–106. doi: 10.1016/j.pep.2010.02.011
- Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., et al. (2009). Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* 4:e7002. doi: 10.1371/journal.pone.0007002
- Wu, G., Bashir-Bello, N., and Freeland, S. J. (2006). The synthetic gene designer: a flexible web platform to explore sequence manipulation for heterologous expression. *Protein Expr. Purif.* 47, 441–445. doi: 10.1016/j.pep.2005.10.020
- Wu, Y., Zhang, Y., Tu, R., Liu, H., and Wang, Q. (2013). Construction of synthetic promoters for *Escherichia coli* and application in the biosynthesis of cis, cis-muconic acid. *Sheng Wu Gong Cheng Xue Bao* 29, 760–771.
- Xiong, A. S., Yao, Q. H., Peng, R. H., Zhang, Z., Xu, F., Liu, J. G., et al. (2006). High level expression of a synthetic gene encoding *Peniophora lycii* phytase in methylotrophic yeast *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 72, 1039–1047. doi: 10.1007/s00253-006-0384-8
- Yim, S. S., An, S. J., Kang, M., Lee, J., and Jeong, K. J. (2013). Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* 110, 2959–2969. doi: 10.1002/bit.24954
- Zhou, Z., Schnake, P., Xiao, L., and Lal, A. A. (2004). Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein Expr. Purif.* 34, 87–94. doi: 10.1016/j.pep.2003.11.006
- Zhu, L., Zhu, Y., Zhang, Y., and Li, Y. (2012). Engineering the robustness of industrial microbes through synthetic biology. *Trends Microbiol.* 20, 94–101. doi: 10.1016/j.tim.2011.12.003

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. The Associate Editor declares that despite being affiliated to the same institution as the authors, the review process was handled objectively and no conflict of interest exists.

Received: 10 December 2013; paper pending published: 17 December 2013; accepted: 14 January 2014; published online: 04 February 2014.

Citation: Elena C, Ravasi P, Castelli ME, Peirú S and Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial

applications and perspectives. *Front. Microbiol.* 5:21. doi: 10.3389/fmicb.2014.00021

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Elena, Ravasi, Castelli, Peirú and Menzella. 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.



Flavoprotein monooxygenases for oxidative biocatalysis: recombinant expression in microbial hosts and applications

Romina D. Ceccoli¹, Dario A. Bianchi² and Daniela V. Rial^{1*}

¹ Área Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario; CONICET, Rosario, Argentina

² Instituto de Química Rosario (IQUR, CONICET-UNR), Área Análisis de Medicamentos, Departamento de Química Orgánica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

Edited by:

Eduardo A. Ceccarelli, Universidad Nacional de Rosario, Argentina

Reviewed by:

Pablo D. De María, Sustainable Momentum, Spain
Vicente Gotor-Fernández, Universidad de Oviedo, Spain

*Correspondence:

Daniela V. Rial, Área Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario; CONICET, Suipacha 531, Rosario, S2002LRK, Argentina
e-mail: rial@inv.rosario-conicet.gov.ar;
drial@fbioyf.unr.edu.ar

External flavoprotein monooxygenases comprise a group of flavin-dependent oxidoreductases that catalyze the insertion of one atom of molecular oxygen into an organic substrate and the second atom is reduced to water. These enzymes are involved in a great number of metabolic pathways both in prokaryotes and eukaryotes. Flavoprotein monooxygenases have attracted the attention of researchers for several decades and the advent of recombinant DNA technology caused a great progress in the field. These enzymes are subjected to detailed biochemical and structural characterization and some of them are also regarded as appealing oxidative biocatalysts for the production of fine chemicals and valuable intermediates toward active pharmaceutical ingredients due to their high chemo-, stereo-, and regioselectivity. Here, we review the most representative reactions catalyzed both *in vivo* and *in vitro* by prototype flavoprotein monooxygenases, highlighting the strategies employed to produce them recombinantly, to enhance the yield of soluble proteins, and to improve cofactor regeneration in order to obtain versatile biocatalysts. Although we describe the most outstanding features of flavoprotein monooxygenases, we mainly focus on enzymes that were cloned, expressed and used for biocatalysis during the last years.

Keywords: flavoprotein monooxygenase, Baeyer-Villiger oxidation, biooxidations, biocatalysis, sulfoxidation, epoxidation, hydroxylation, recombinant biocatalyst

FLAVOPROTEIN MONOOXYGENASES

Flavoprotein monooxygenases comprise a family of enzymes that participate in a wide variety of metabolic processes both in prokaryotic and eukaryotic cells. They are involved in pathways of degradation of aromatic compounds, polyketides biosynthesis, antibiotic resistance and, biosynthesis of compounds with relevant biological activities as cholesterol, antibiotics, and siderophores. Some of these enzymes participate in routes that allow microbial utilization of organic compounds as carbon and energy sources. Most flavoprotein monooxygenases are able to use molecular oxygen (O₂) as oxygen donor to oxygenate an organic compound, a reaction that depends on a reduced flavin cofactor to activate O₂ by electron donation. These enzymes are classified as external (EC 1.14.13) and internal monooxygenases (EC 1.13.12). External monooxygenases rely on reduced coenzymes in the form of NADPH or NADH as sources of reducing power for the flavin, whereas in internal monooxygenases the flavin is reduced by the substrate itself. Besides, there are flavin-dependent enzymes that are able to catalyze hydroxylations of organic compounds. In this case, the flavin is required to oxidize the substrate *via* a reaction in which the oxygen atom comes from water while O₂ serves to recycle the flavin (van Berkel et al., 2006; Torres Pazmiño et al., 2010b). External flavoprotein monooxygenases contain non-covalently bound FAD or FMN

and catalyze the NAD(P)H-dependent insertion of a single oxygen atom into an organic substrate while the second atom of oxygen is reduced to water. They are classified in six classes (A–F) according to structural- and sequence-related characteristics (van Berkel et al., 2006). Besides natural flavoprotein monooxygenases, modified flavins have been used as organocatalysts and novel artificial flavoenzymes have been generated by flavin re-design (de Gonzalo and Fraaije, 2013).

In the following sections we review the most representative reactions catalyzed by prototype flavoprotein monooxygenases, highlighting the strategies employed to produce them recombinantly. We describe the most outstanding features of bacterial flavoprotein monooxygenases, albeit we mainly focus on enzymes that were cloned, expressed and used for biocatalysis during the last years.

BAEYER-VILLIGER OXIDATIONS

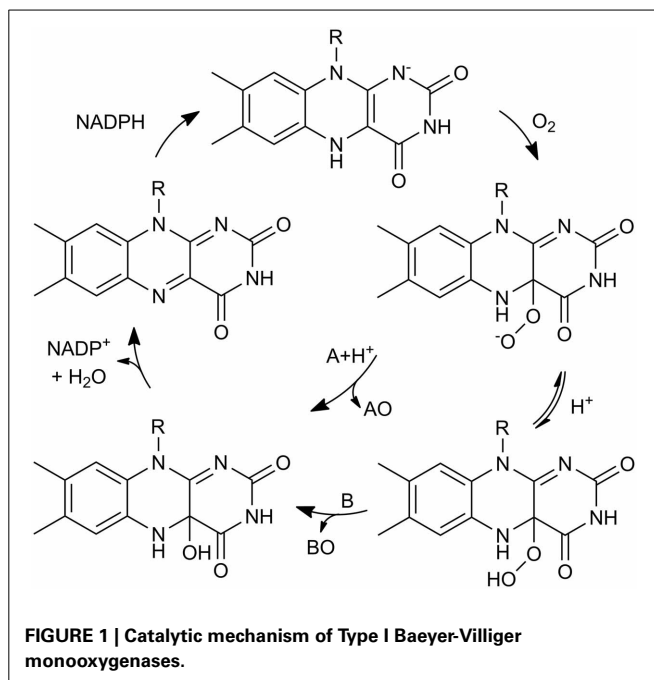
The oxidation of ketones is known in organic chemistry as Baeyer-Villiger oxidation (Baeyer and Villiger, 1899). This reaction involves peracids or hydrogen peroxide to achieve the oxidation of ketones to esters or lactones. Chiral lactones are valuable intermediates toward the synthesis of natural products and analogs (reviewed in de Gonzalo et al., 2010). For Baeyer-Villiger oxidations, the enzyme-mediated transformation

has become the preferred method due to its high enantio-, regio-, and chemoselectivity. Besides, the process takes place in environmentally friendly conditions, avoids the use of toxic reagents, and allows scale-up. The ability of some microorganisms to grow in a certain alcohol or ketone grabbed the attention to the enzymes involved in those metabolic routes and prompted the discovery of Baeyer-Villiger monooxygenases (BVMOs). An increasing number of BVMOs have been identified, cloned, recombinantly expressed, engineered and used for biocatalysis. This topic has been the matter of very comprehensive revisions during the last 5 years (de Gonzalo et al., 2010; Torres Pazmiño et al., 2010a; Leisch et al., 2011; Balke et al., 2012), hence the present section focuses mainly on the strategies used to overcome gene expression problems and cofactor regeneration limitations, scale-up and a summary of the most recent applications.

Different types of BVMOs exist. Type I BVMOs contain FAD, depend on NADPH for catalysis and belong to the class B of flavoprotein monooxygenases while Type II BVMOs are FMN- and NADH-dependent enzymes and belong to class C of flavoprotein monooxygenases. In addition, there are atypical BVMOs that do not share their characteristics (Willettts, 1997; van Berkel et al., 2006; Torres Pazmiño et al., 2010a). Type I BVMOs are flavoenzymes that catalyze the oxidation of a linear or cyclic ketone to an ester or lactone, respectively, at the expense of molecular oxygen and NADPH. For these enzymes NADPH is the required electron donor. As a result, an oxygen atom is inserted into a carbon-carbon bond adjacent to a carbonyl group in the substrate and the other one is reduced to water. The mechanism of this reaction is proposed to proceed *via* formation and stabilization of a covalent bond between oxygen and the C4a of the isoalloxazine ring of reduced FAD. This C4a-peroxyflavin performs a nucleophilic attack on the carbonyl group of the substrate giving rise to a Criegee intermediate that rearranges spontaneously to the product (Figure 1). BVMOs can also oxygenate heteroatoms probably *via* an electrophilic mechanism (reviewed in Mihovilovic, 2006; van Berkel et al., 2006; Torres Pazmiño et al., 2008a).

In 2002, a consensus motif for Type I BVMOs was described (Fraaije et al., 2002) that was very useful for the identification of novel Type I BVMOs by genome mining (Torres Pazmiño et al., 2010a). Recently, a more specific sequence motif was identified that allowed a better distinction between typical Type I BVMOs and flavin-containing monooxygenases (FMOs) (Riebel et al., 2012). More than fifty Type I BVMO genes are currently available for recombinant expression. Most of them are of bacterial origin but the cloning of the coding sequences of few eukaryotic Type I BVMOs has been reported during the last 2 years (Leipold et al., 2012; Beneventi et al., 2013; Mascotti et al., 2013).

The cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB 9871 (CHMO_{Acineto}) is a model Type I BVMO that has been studied in-depth. It showed to be a robust biocatalyst, as expressed in *E. coli* from a pET-22b derived vector, able to catalyze selective oxygenation of a broad variety of ketones in desymmetrizations reactions, regiodivergent oxidations and kinetic resolutions (reviewed in Mihovilovic, 2006; Leisch et al., 2011). Recently, the CHMO from *Rhodococcus* sp. HI-31 was crystallized with its substrate, cyclohexanone, and with NADP⁺ and FAD at 2.4 Å resolution (Yachnin et al., 2012). A benchmark



reaction catalyzed by *E. coli* cells overexpressing CHMO_{Acineto} was the asymmetric oxidation of the racemic bicyclo[3.2.0]hept-2-en-6-one in large scale. For this purpose, the strategies for optimization of the bioconversion included a fed-batch biotransformation, the use of resin-based *in situ* substrate feeding and product removal (SFPR) technology, a fine control of the bioprocess and a proper aeration. This bioconversion was scaled-up to pilot-plant scale (200 L) and 4.5 g/L of lactone were produced (Baldwin et al., 2008). The SFPR methodology allows the use of substrate concentrations beyond toxicity levels and avoids inhibition of the reaction by the product or substrate as their concentrations in the culture remains below inhibitory levels. Scale-up methodologies for Baeyer-Villiger biooxidation of ketones were reviewed in de Gonzalo et al. (2010). An innovative monitoring system was developed based on the use of flow-calorimetry to measure temperature changes due to Baeyer-Villiger oxygenations catalyzed by encapsulated *E. coli* expressing CHMO_{Acineto} (Bučko et al., 2011). Variants of CHMO_{Acineto} with enhanced oxidative and thermal stabilities were obtained by rational and combinatorial mutagenesis at M and C residues without affecting the activity or selectivity of the enzyme (Opperman and Reetz, 2010). In addition, wild-type and mutant CHMO_{Acineto} catalyzed the conversion of 4-ethylidenecyclohexanone into *E*- and *Z*-configured lactones, respectively, and successive reactions catalyzed by transition metals were used to produce different trisubstituted *E*- or *Z*-olefins (Zhang et al., 2013). *E. coli* cells expressing the CHMO from *Xanthobacter* sp. ZL5 (CHMO_{Xantho}) have a very broad substrate acceptance profile and the ability to convert some bulky ketones not accepted by other BVMOs (Rial et al., 2008a,b). More recently, Alexander et al. (2012) reported the cloning and evaluation of a CHMO from the xenobiotic-degrading *Polaromonas* sp. JS666. Initial oxidation assays showed no results due to formation of inclusion bodies but

upon optimization, a detailed screening of the biocatalyst could be performed (**Table 1**).

The cyclopentanone monooxygenase from *Comamonas* sp. NCIMB 9872 (CPMO) is another cycloketone-converting BVMO, which can display enantiodivergent transformations with respect to the CHMO group. *E. coli* cells overexpressing the gene coding for CPMO were used as biocatalysts to oxidize an oxo-bridged ketone in order to obtain a heterobicyclic lactone, a key intermediate in formal total syntheses of various natural products containing a tetrahydrofuran structural motif such as *trans*-kumausyne, goniofufurone analogs and showdomycin (Mihovilovic et al., 2006). In this work, the biotransformation was carried out in a bioreactor using the *in situ* SFPR technology and the desired lactone was obtained in 70 % isolated yield (Mihovilovic et al., 2006). In a recent report, CHMO_{Xantho}- and CPMO-mediated biooxidations of a bridged-bicyclic ketone were performed in shake-flasks scale and allowed access to both antipodal lactones in very good yields and high enantiomeric excess (e.e.). These chiral lactones were key intermediates toward (+) and (–) non-natural carba-C-nucleosides in high optical purity (Bianchi et al., 2013).

Other remarkable Type I BVMOs is the phenylacetone monooxygenase from *Thermobifida fusca* (PAMO) (**Table 1**). Its coding sequence was cloned and expressed in *E. coli* from a pBAD/myc-HisA-derived vector (Fraaije et al., 2005). PAMO can tolerate high temperatures and organic solvents (Fraaije et al., 2005; de Gonzalo et al., 2006a). The enzyme was purified, characterized and it was the first BVMO for which the three-dimensional structure was elucidated by X-ray diffraction (Malito et al., 2004). Some years afterward, Orru et al. (2011) solved the crystal structure of reduced and oxidized PAMO in complex with NADP⁺. Since the substrate profile of wild-type PAMO is mainly limited to some aromatic ketones and sulfides (de Gonzalo et al., 2005b; Rodríguez et al., 2007; Zambianchi et al., 2007), protein engineering strategies were undertaken aiming at expanding the substrate profile of PAMO without affecting its stability (Bocola et al., 2005). By a site-directed mutagenesis approach it was possible to expand the substrate range of the enzyme to some prochiral cyclic ketones, sulfides, and amines (Torres Pazmiño et al., 2007). Thermostable PAMO mutants with high activity and enantioselectivity for the conversion of 2-substituted cyclohexanones derivatives were produced by saturation mutagenesis focused on specific sites of PAMO (Reetz and Wu, 2009; Wu et al., 2010). Directed evolution and rational re-design of PAMO and other BVMOs were thoroughly reviewed recently (Zhang et al., 2012).

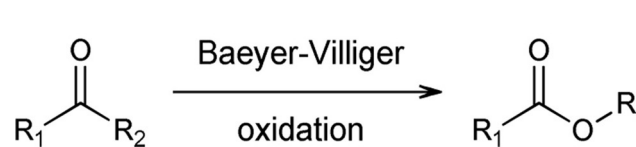
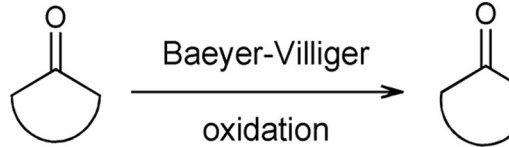
It has been reported that the addition of organic co-solvents to biotransformations can influence conversion and selectivity of reactions catalyzed by wild-type PAMO and variants (Rioz-Martínez et al., 2008, 2009, 2010b). Recently, de Gonzalo et al. (2012) improved the biocatalytic performance of a PAMO mutant (the variant with M446 replaced by G) in hydrophilic organic solvents and, including a weak anion exchange resin, they were able to attain the dynamic kinetic resolution of a range of benzylketones. Further optimization of PAMO biooxidations considered the buffer and ionic strength of the reaction media as well as the coupled reaction for cofactor regeneration (Rodríguez

et al., 2012). To improve the biocatalyst performance, fundamental aspects of protein expression such as host strain, inducer concentration, temperature and length of induction as well as riboflavin addition were considered (van Bloois et al., 2012). This approach also evaluated biotransformation conditions including external sugars as sources of reducing power for NADPH regeneration, substrate concentration and, biotransformation temperature and length. Recently, Dudek et al. (2013b) developed a screening method based on periplasmically expressed PAMO aiming at enabling complete access of substrates to the enzyme and facilitating NADPH recycling by externally added phosphite dehydrogenase (PTDH) from *Pseudomonas stutzeri* WM88. The *pamO* gene was cloned into a pBAD-derived plasmid between an N-terminal Tat-dependent signal sequence of the endogenous *E. coli* protein TorA and a C-terminal Myc epitope/His-tag. The Tat-PAMO protein was functionally expressed in the periplasm of *E. coli* cells and this system was used together with the PTDH-based regeneration system for biotransformations. Just recently, this analysis was extended to the screening of a library of PAMO mutants, which resulted in the isolation of a quadruple mutant with the same thermostability as the wild-type enzyme but with an extended substrate scope (Dudek et al., 2013a) (**Table 1**).

Another available BVMO is the cyclopentadecanone monooxygenase from *Pseudomonas* sp. HI-70 (CPDMO). Its gene was cloned in 2006 and initial assays detected activity toward large ring ketones (C11–C13), substituted cyclohexanones (Iwaki et al., 2006) and ketosteroids (Beneventi et al., 2009). The biocatalytic performance of CPDMO was evaluated extensively in 2011 and showed a behavior similar to CHMO within desymmetrizations and kinetic resolutions, but performed particularly interesting in regiodivergent oxidations (Fink et al., 2011). Another robust biocatalyst is cyclododecanone monooxygenase from *Rhodococcus ruber* SC1 (CDMO), which was used as a case study to show the potentials of a new tool for chiral catalysts assessment (Fink et al., 2012). The 4-hydroxyacetophenone monooxygenase (HAPMO) from *Pseudomonas fluorescens* ACB has been available for many years (Kamerbeek et al., 2001). In 2009, the gene encoding for a HAPMO from *Pseudomonas putida* JD1 was cloned, functionally expressed and characterized (Rehder et al., 2009). Soluble protein production was problematic thus several strategies were undertaken to circumvent this limitation. Expression of the HAPMO-encoding gene was assayed from two different plasmids and in several bacterial hosts, in media with different composition, at various temperatures, in the presence or absence of FMN and by co-expression with molecular chaperones. By biotransformations in crude cell extracts it was found that this enzyme preferentially oxidizes aryl-aliphatic ketones (Rehder et al., 2009). In a following work, the biooxidation of the aromatic ketone 3-phenyl-2-butanone was scaled-up in a bioreactor and yields improved by the use of adsorbent resins for an *in situ* SFPR (Geitner et al., 2010).

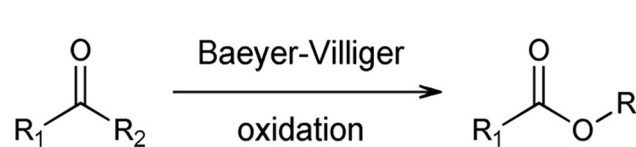
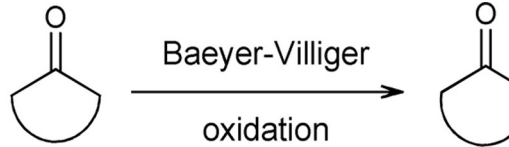
Other BVMOs have been newly reported (**Table 1**). A set of predicted 22 *bvmo*-encoding genes from *Rhodococcus jostii* RHA1 were cloned but only 12 of them could be expressed as soluble active enzymes (Szolkowy et al., 2009). However, by applying a

Table 1 | Baeyer-Villiger oxidation.

<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Baeyer-Villiger oxidation</p> </div> <div style="text-align: center;">  <p>Baeyer-Villiger oxidation</p> </div> </div>			
Enzyme	Gene	Expression system	References ^a
CHMO	<i>chmo</i> from <i>Polaromonas</i> sp. JS666	<i>E. coli</i> BL21 Star (DE3)/pET101/D-TOPO vector	Alexander et al., 2012
	mutants of <i>chnB</i> from <i>A. calcoaceticus</i> NCIMB 9871	<i>E. coli</i> BL21-Gold (DE3)/pET22b	Opperman and Reetz, 2010
	<i>chnB</i> from <i>A. calcoaceticus</i> NCIMB 9871 and <i>pos5</i> (NADH kinase) from <i>S. cerevisiae</i> CEN.PK2-1D	<i>E. coli</i> BL21(DE3)/pET22b (<i>chnB</i>) and ACYCDuet-1 (<i>pos5</i>)	Lee et al., 2013
	<i>chnB</i> from <i>A. calcoaceticus</i> NCIMB 9871 and <i>gapB</i> (NADP ⁺ -dependent GAPDH) from <i>B. subtilis</i>	<i>E. coli</i> (<i>gapA</i> null, NAD ⁺ -dependent glyceraldehyde-3-phosphate dehydrogenase mutant)/pET22b (<i>chnB</i>) and pDHC29 (<i>gapB</i>)	Wang et al., 2013
PAMO	different <i>pamO</i> mutants from <i>T. fusca</i>	<i>E. coli</i> TOP10/pBAD/myc-HisA	Wu et al., 2010
	different mutant <i>pamO</i> from <i>T. fusca</i>	<i>E. coli</i> TOP10/pBAD/myc-HisA	Dudek et al., 2011
	wild-type and mutant <i>pamO</i> from <i>T. fusca</i>	<i>E. coli</i> TOP10/pBAD-Tat-AldO (periplasmic expression)	Dudek et al., 2013a,b
BVMO	22 <i>bvmo</i> genes from <i>R. jostii</i> RHA1 (16 BVMOs with demonstrated activity)	<i>E. coli</i> TOP10/pBADN and pCRE2 (with phosphite dehydrogenase gene)	Riebel et al., 2012
	<i>bvmo</i> from <i>A. radioresistens</i> S13	<i>E. coli</i> BL21(DE3)/pT7 expression vector	Minerdi et al., 2012
STMO	wild-type and mutant <i>stmo</i> from <i>R. rhodochrous</i>	<i>E. coli</i> BL21(DE3) CodonPlus/Champion pET-vector	Franceschini et al., 2012
	wild-type <i>stmo</i> from <i>R. rhodochrous</i>	<i>E. coli</i> BL21(DE3)/pET28b	Leipold et al., 2013
SAPMO	SAPMO from <i>C. testosteroni</i> KF1	<i>E. coli</i> BL21 Star (DE3)/pET101/D-TOPO vector	Weiss et al., 2013
OTEMO	OTEMO gene from <i>P. putida</i> NCIMB 10007	<i>E. coli</i> BL21(DE3)/pET28b	Kadow et al., 2012
	OTEMO gene from <i>P. putida</i> ATCC 17453	<i>E. coli</i> BL21(DE3)/pSD80	Leisch et al., 2012
2,5-DKCMO	2,5-DKCMO gene from <i>P. putida</i> NCIMB 10007	<i>E. coli</i> BL21(DE3)/pET28b	Kadow et al., 2011, 2012
3,6-DKCMO	3,6-DKCMO gene from <i>P. putida</i> NCIMB 10007	<i>E. coli</i> BL21(DE3)/pET28b (coexpression of GroES-GroEL chaperones)	Kadow et al., 2012
2,5- and 3,6-DKCMO	2,5- and 3,6-DKCMO genes from <i>P. putida</i> NCIMB 10007 and Fre gene from <i>E. coli</i>	<i>E. coli</i> BL21(DE3)/pET28b (DKCMOs, coexpression of GroES-GroEL); pGas (<i>Fre</i> or fusion genes, L-rhamnose inducible)	Kadow et al., 2013

(Continued)

Table 1 | Continued

 			
Enzyme	Gene	Expression system	References ^a
BVMO type II	<i>camE</i> ₂₅₋₁ , <i>camE</i> ₂₅₋₂ , <i>camE</i> ₃₆ and Fred gene from <i>P. putida</i> ATCC 17453	<i>E. coli</i> BL21(DE3)/pSD80 (each <i>camE</i> gene and tandem clones); <i>E. coli</i> BL21(DE3) pLysS/pET17b (Fred gene)	Iwaki et al., 2013
FMO (type II)	<i>fmo</i> from <i>S. maltophilia</i> several <i>fmo</i> genes from <i>R. jostii</i> RHA1	<i>E. coli</i> BL21(DE3)/pET-YSBLIC-3C <i>E. coli</i> /pBADN and pCRE2	Jensen et al., 2012 Riebel et al., 2013a,b

^a The information shown corresponds to reports as of 2010.

Abbreviations: 3,6-DKCMO, 3,6-diketocamphane 1,6-monooxygenase; 2,5-DKCMO, 2,5-diketocamphane 1,2-monooxygenase; ACMO, acetone monooxygenase; BVMO, Baeyer-Villiger monooxygenase; CDMO, cyclododecanone monooxygenase; CHMO, cyclohexanone monooxygenase; CPMO, cyclopentanone monooxygenase; FMO, flavin-containing monooxygenase; OT, 2-oxo- Δ^3 -4,5,5-trimethylcyclopentenylacetic acid; OTEMO, 2-oxo- Δ^3 -4,5,5-trimethylcyclopentenylacetyl-CoA monooxygenase; PAMO, phenylacetone monooxygenase; SAPMO, 4-sulfophenyl acetate monooxygenase; STMO, steroid monooxygenase.

high-throughput cloning strategy and optimized expression conditions Riebel et al. (2012) were able to express the 22 probable *bvmo* genes identified in the genome of *R. jostii* RHA1 in soluble form. They cloned the selected genes under the control of *araBAD* promoter directly or as a fusion with the PTDH gene (Riebel et al., 2012). Other recently expressed *bvmo* genes include the *almA* gene from *Acinetobacter radioresistens* S13 that encodes a BVMO involved in the subterminal oxidation of alkanes (Minerdi et al., 2012), a steroid monooxygenase (STMO) from *Rhodococcus rhodochrous* which crystal structure (Franceschini et al., 2012) and substrate profile (Leipold et al., 2013) were determined, a 4-sulfoacetophenone monooxygenase (SAPMO) from *Comamonas testosteroni* KF-1 that is involved in the biodegradation of 4-sulfophenylcarboxylates (Weiss et al., 2013), and the 2-oxo- Δ^3 -4,5,5-trimethylcyclopentenylacetyl-CoA monooxygenase (OTEMO) from *P. putida* NCIMB 10007 (Kadow et al., 2012) and from *P. putida* ATCC 17453 (Leisch et al., 2012). This enzyme participates in the degradation of camphor in the native microbial host but, recombinantly expressed in *E. coli*, it is able to accept α,β -unsaturated monocyclic and bicyclic ketones (Kadow et al., 2012).

The strict dependence of BVMOs on NADPH for catalysis certainly impairs the practical applications of these enzymes due to the high costs of NADPH or to the requirement of a cofactor regeneration system. The possibility to carry out the desired biotransformation in whole-cell systems is a beneficial alternative since the cell itself provides the NADPH. The co-expression of glucose-6-phosphate dehydrogenase or the addition of carbohydrates to the culture media can improve NADPH regeneration by the host cells (Walton and Stewart, 2002; Lee et al., 2007). Besides, several other options are available to regenerate NADPH for BVMO activity (recently reviewed in de Gonzalo et al., 2010). Some coenzyme regeneration systems are

based on a coupled enzymatic reaction that produces NADPH at the expense of an auxiliary substrate. Typical pure enzymes used for the regeneration of NADPH include glucose-6-phosphate dehydrogenase, PTDH, alcohol dehydrogenase and glucose dehydrogenase. However, these systems need to be added to the activity assays. In the last years, an alternative strategy was developed in which fusion proteins between a PTDH and certain BVMOs were produced and evaluated as self-sufficient biocatalysts (Torres Pazmiño et al., 2008b, 2009). Other approaches based on the chemical (de Gonzalo et al., 2005a) or photochemical (Hollmann et al., 2007) regeneration of the flavin bound to the BVMO have also been investigated. Most recently, two strategies to improve NADPH regeneration were presented and tested in BVMO-mediated biotransformations (Table 1). In one approach, a NADH kinase from yeast was used for the direct phosphorylation of NADH to NADPH in *E. coli* cells producing CHMO_{Acineto} (Lee et al., 2013). This approach enhanced the oxidation of cyclohexanone in a fed-batch biotransformation and doubled the productivity of ϵ -caprolactone when compared with the control lacking the NADH kinase (Lee et al., 2013). The other approach proposed a strategy to increase NADPH bioavailability by replacing the native NAD⁺-dependent glyceraldehyde-3-phosphate dehydrogenase *gapA* gene in *E. coli* with a NADP⁺-dependent *gapB* gene from *Bacillus subtilis*, hence producing in *E. coli* a NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenase from a plasmid and the CHMO_{Acineto} from a compatible expression vector (Wang et al., 2013).

Two additional BVMOs (named 2,5-diketocamphane 1,2-monooxygenase (2,5-DKCMO) and 3,6-diketocamphane 1,6-monooxygenase (3,6-DKCMO)) that participate in the camphor-degrading metabolic route in *P. putida* NCIMB 10007 are Type II BVMOs (Kadow et al., 2011, 2012). They are two-component systems consisting of a monooxygenase and a

reductase, and depend on FMN- and NADH for activity. The genes encoding the monooxygenase subunit of 2,5-DKCMO and 3,6-DKCMO were recombinantly expressed in *E. coli* (Kadow et al., 2011, 2012). The expression of the gene encoding the oxygenase subunit of 3,6-DKCMO required the assistance of molecular chaperones for enhanced soluble expression (Kadow et al., 2012) (Table 1). These biocatalysts were able to convert mainly bicyclic ketones. Three *camE* genes from *P. putida* ATCC 17453 coding for different monooxygenase subunits of DKCMO isoenzymes were cloned and expressed (Iwaki et al., 2013). In addition, one FMN reductase (*Fred*) gene from the same bacteria was identified and cloned individually or in tandem with the respective 2,5-, or 3,6-DKCMO-coding genes. Pairs DKCMO-Fred were able to convert bicyclic ketones with enantiomeric specificity in recombinant whole-cell systems (Iwaki et al., 2013) (Table 1). Recently, a flavin-reductase Fre from *E. coli* was reported as an appropriate partner for providing reduced FMN to either 2,5- or 3,6-DKCMO from *P. putida* NCIMB 10007. Couples DKCMO-Fre were able to oxidize camphor and norcamphor in the presence of NADH generated by formate dehydrogenase (FDH) from *Candida boidinii* (Kadow et al., 2013).

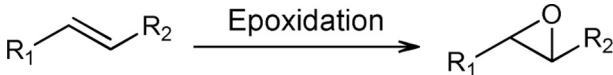
Besides BVMOs, FMOs are capable of catalyzing Baeyer-Villiger oxidations (Table 1). Jensen et al. (2012) reported the ability of an FMO from *Stenotrophomonas maltophilia* (SMFMO) to catalyze some Baeyer-Villiger oxidations as well as sulfoxidations and to use both NADH and NADPH. The codon-optimized synthetic gene was cloned and SMFMO was produced in *E. coli*, purified and its crystal structure elucidated. Just recently, Riebel et al. (2013a) cloned, expressed in *E. coli* and explored the catalytic potential of several novel flavoprotein monooxygenases from *R. jostii* RHA1 with homology to FMOs. The authors studied the ability of the novel enzymes, classified as Type II FMOs, to convert phenylacetone, (\pm)-bicyclo[3.2.0]hept-2-en-6-one and methyl phenyl sulfide (thioanisole). These results were further extended by in-depth screening in whole-cell systems (Riebel et al., 2013b).

EPOXIDATIONS

Epoxides are valuable precursors for synthetic applications toward bioactive compounds, thus in this section we describe attractive epoxidations carried out by recombinant flavoprotein monooxygenases.

Two-component styrene monooxygenases belong to the class E of flavoprotein monooxygenases (Table 2). The first step in the metabolic utilization of styrene in *Pseudomonas* sp. VLB120 is catalyzed by an oxygenase (StyA) and a NADH-flavin oxidoreductase (StyB). The genes coding for this enzyme (StyAB) were identified, cloned and expressed in *E. coli* (Panke et al., 1998). In order to investigate the relationships between styrene epoxidation, StyAB production, cell growth and carbon metabolism, two-liquid-phase continuous cultures of *E. coli* expressing *styAB* genes of *Pseudomonas* sp. VLB120 were performed in a 3 L-stirred reactor (Bühler et al., 2008). The two-phase system made possible the operation of the biocatalyst at subtoxic non-inhibitory substrate and product concentrations. It also allowed control of the epoxidation rate by varying styrene feed concentration (Bühler et al., 2008). In order to improve styrene biotransformation by *E. coli* expressing the two-component styrene monooxygenase from *P. putida* CA-3, the corresponding coding gene was subjected to *in vitro* evolution followed by an indole bioconversion-based screening (Gursky et al., 2010). Ukaegbu et al. (2010) reported the X-ray crystal structure of the N-terminally His-tagged oxygenase subunit of the styrene monooxygenase from *P. putida* S12. Based on this structural data and aiming at improving the preference of the enzyme toward α -substituted styrene, point mutations were introduced in the styrene monooxygenase from *Pseudomonas* sp. LQ26 (Lin et al., 2010, 2011a,b) by site-directed mutagenesis (Qaed et al., 2011). This procedure allowed the development of mutants with increased reactivity toward α -substituted styrene derivatives. By a similar rational design approach, a different set of mutations was found to exhibit increased epoxidation activity toward styrene and *trans*- β -methyl styrene compared with the wild-type enzyme. Interestingly, one of these mutants showed

Table 2 | Epoxidation.

			
Enzyme	Gene	Expression system	References ^a
SMO	wild-type and mutant <i>styAB</i> fragment from <i>P. putida</i> CA-3	<i>E. coli</i> XL-10 Gold/pBluescript II KS; <i>E. coli</i> BL21(DE3)/pRSET-B	Gursky et al., 2010
	wild-type and several mutant <i>styAB2</i> from <i>Pseudomonas</i> sp. LQ26	<i>E. coli</i> BL21(DE3)/pET28a	Lin et al., 2010, 2011a, 2012
	different mutant <i>styAB2</i> from <i>Pseudomonas</i> sp. LQ26	<i>E. coli</i> BL21(DE3)/pET28a	Qaed et al., 2011
	<i>styA1/StyA2B</i> from <i>R. opacus</i> 1CP	<i>E. coli</i> BL21(DE3) pLysS/pET16bP	Tischler et al., 2009, 2010

^aThe information shown mainly corresponds to reports as of 2010.

Abbreviation: SMO, styrene monooxygenase.

reversed enantiomeric preference toward 1-phenylcyclohexene (Lin et al., 2012) (**Table 2**).

van Hellemond et al. (2007) reported the discovery of a novel styrene monooxygenase (SmoA) in a metagenomic library derived from loam soil. *In vitro* activity assays using crude cell extracts of bacteria producing SmoA evidenced epoxidation of styrene and styrene derivatives to the corresponding (S)-epoxides with excellent e.e. In *Rhodococcus opacus* 1CP, a self-sufficient styrene monooxygenase was reported that harbors in the same polypeptide chain a monooxygenase and a NADH-flavin oxidoreductase (StyA2B) (Tischler et al., 2009). Moreover, a multifunctional monooxygenase system (StyA1/StyA2B) was recently described in the same microorganism as composed of a single styrene monooxygenase (StyA1) and the StyA2B polypeptide (Tischler et al., 2010) (**Table 2**). Purified StyA2B was able to oxidize styrene, 2-chlorostyrene, 3-chlorostyrene, 4-chlorostyrene, 4-methylstyrene, and dihydronaphthalene.

Surprisingly, the epoxidation of an electron-rich C=C functionality in the *oxo*-bridged bicyclic ketone (1R,5S)-8-oxabicyclo[3.2.1]oct-6-en-3-one was catalyzed by CHMO_{Xantho}, representing the first report of a BVMO involved in this conversion (Rial et al., 2008a).

SULFOXIDATIONS

Enantioselective sulfoxidations are difficult to accomplish chemically and, therefore enzyme-mediated sulfoxidations have attracted the attention of chemists and biochemists during the last decades.

In 2007, it was shown that crude cell extracts of bacteria producing SmoA can catalyze sulfoxidation reactions with high enantioselectivity toward aromatic sulfides (van Hellemond et al., 2007). The multifunctional monooxygenase system StyA1/StyA2B from *R. opacus* 1CP (Tischler et al., 2009, 2010) is also capable of oxidizing thioanisole (**Table 3**). In addition, oxidation of sulfides can be readily catalyzed by *E. coli* cells expressing the two-component NADH-dependent styrene monooxygenase from *P. putida* CA-3 (Nikodinovic-Runic et al., 2013). Boyd et al. (2012) have demonstrated that this biocatalyst is able to S-oxidize benzo[b]thiophene and, other nine sulfur-containing compounds, including thioanisole and some substituted analogs, benzo[b]thiophene, and 2-methylbenzo[b]thiophene were accepted as substrates as well (Nikodinovic-Runic et al., 2013). The enzyme that had been previously engineered for improved alkene epoxidation (Gursky et al., 2010) showed an increased S-oxidation capability when compared with the wild-type form, being the sulfur atom in the thiophene ring a better target than the sulfur atom in an alkyl chain (Nikodinovic-Runic et al., 2013).

The BVMOs are also capable of stereoselective sulfoxidations. In 2005, the ability of PAMO to oxidize aromatic sulfides was proved but the sulfoxides displayed poor e.e. (de Gonzalo et al., 2005b; Fraaije et al., 2005). The enzymatic oxidation of sulfides mediated by pure PAMO from *T. fusca*, HAPMO from *P. fluorescens* ACB and ethionamide monooxygenase (EtaA) from *Mycobacterium tuberculosis* was evaluated in several aqueous-organic media (de Gonzalo et al., 2006a). More recently, the HAPMO from *P. putida* JD1 was challenged in the sulfoxidation of methyl-4-tolyl sulfide using crude cell extracts (Rehdorf

et al., 2009) and, the repertoire of chiral sulfoxides accessed was extended by using both PAMO and HAPMO from *P. fluorescens* ACB as crude cell-free extracts (Rioz-Martínez et al., 2010a). The stereoselectivity of PAMO-mediated oxidation of several prochiral thio-ethers has been enhanced by mutagenesis of M446 to G (Torres Pazmiño et al., 2007) and solvent engineering methodologies were explored in order to expand the applications of wild-type and M446G PAMO (de Gonzalo et al., 2012). The oxidation of benzyl methyl sulfide was evaluated in seventeen combinations of buffer/co-solvent and compared with the reaction in aqueous medium. The reaction in Tris-HCl pH 9.0 containing 5 % methanol rendered the corresponding sulfoxide in high conversion and good e.e. with only low levels of sulfone as by-product (de Gonzalo et al., 2012). Besides, Rodríguez et al. (2012) analyzed the effect of several enzymatic cofactor regeneration systems and cofactor concentrations in the oxidation of thioanisole by PAMO. By protein engineering, amino acidic positions were identified in PAMO that alter conversion and selectivity of S-oxidations (Dudek et al., 2011). In this investigation, the bulky prochiral benzyl phenyl sulfide, which is only very poor substrate for wild-type PAMO, was readily oxidized by the M446G mutant (Dudek et al., 2011). Therefore, it was selected as substrate to evaluate periplasmic expression of PAMO variants in order to establish a whole-cell screening method for the assessment of libraries of PAMO toward the identification of mutants with altered biocatalytic performances (Dudek et al., 2013a,b).

Nine BVMOs from *R. jostii* RHA1 that were cloned and expressed by Riebel et al. (2012) showed confirmed S-oxidation activity on thioanisole, benzyl phenyl sulfide, benzyl ethyl sulfide or ethionamide. Besides, the BVMO coded by the gene *almA* from *A. radioresistens* S13 is also able to act on ethionamide to give the corresponding S-oxide (Minerdi et al., 2012) (**Table 3**). Most recently, the oxidation of thioanisole by CHMO_{Acineto} was investigated using FDH as NADPH recycling system (Zhai et al., 2013). The authors constructed a whole-cell biocatalyst able to co-express the *chmo*_{Acineto} and a modified *fdh* gene from *C. boidinii* that can utilize NADP⁺ efficiently. Several conditions including concentration of the biocatalyst and substrate, pH, temperature and time of reaction as well as addition of dimethylsulfoxide and NADP⁺ were evaluated to optimize the biocatalytic reaction (Zhai et al., 2013).

The mFMO from *Methylophaga* sp. SK1 was the first bacterial Type I FMO reported in the literature (Choi et al., 2003). The authors cloned, expressed and characterized the recombinant enzyme. They determined its activity on N- and S-containing compounds such as trimethylamine and thiourea and emphasized its ability to produce indigo blue. Alfieri et al. (2008) noted an unexpected activity of this enzyme on dimethylsulfoxide. Few years later, Rioz-Martínez et al. (2011) reported the fusion of an optimized and thermostable PTDH with the mFMO from *Methylophaga* sp. SK1 and evaluated its ability to act on several prochiral sulfides using NADPH as electron donor. Amongst the accepted sulfides, thioanisole was the best substrate but the chiral sulfoxides were obtained in moderate e.e. Some substituted thioanisole derivatives as well as other (hetero)aromatic sulfides and alkyl butyl sulfides were also oxidized to the corresponding sulfoxides, showing moderate to very

Table 3 | S-oxidation.

$$\text{R}_1-\text{S}-\text{R}_2 \xrightarrow{\text{S-oxidation}} \text{R}_1-\overset{\text{O}}{\underset{\text{||}}{\text{S}}}-\text{R}_2$$

Enzyme	Gene	Expression system	References ^a
SMO	<i>styA1/StyA2B</i> from <i>R. opacus</i> 1CP wild-type and mutant <i>styAB</i> fragment from <i>P. putida</i> CA-3	<i>E. coli</i> BL21(DE3) pLysS/pET16bP <i>E. coli</i> BL21(DE3)/pRSET-B	Tischler et al., 2009, 2010 Gursky et al., 2010; Boyd et al., 2012; Nikodinovic-Runic et al., 2013
PAMO	different mutant <i>pamO</i> from <i>T. fusca</i> wild-type and mutant <i>pamO</i> from <i>T. fusca</i>	<i>E. coli</i> TOP10/pBAD/myc-HisA <i>E. coli</i> TOP10/pBAD-Tat-AldO plasmid (periplasmic expression)	Dudek et al., 2011 Dudek et al., 2013a,b
BVMO	9 <i>bvmo</i> genes from <i>R. jostii</i> RHA1 <i>bvmo</i> gene from <i>A. radioresistens</i> S13	<i>E. coli</i> TOP10/pBADN and pCRE2 (phosphite dehydrogenase gene) <i>E. coli</i> BL21(DE3)/pT7 expression vector	Riebel et al., 2012 Minerdi et al., 2012
FMO (type I)	<i>fmo</i> from <i>Methylophaga</i> sp. SK1 <i>tmm</i> gene from <i>M. silvestris</i>	<i>E. coli</i> TOP10/pCRE2 (soluble NADPH regenerating phosphite dehydrogenase) <i>E. coli</i> BLR(DE3) pLysS/pET28a	Rioz-Martínez et al., 2011 Chen et al., 2011
FMO (type II)	<i>fmo</i> gene from <i>S. maltophilia</i> several <i>fmo</i> genes from <i>R. jostii</i> RHA1	<i>E. coli</i> BL21(DE3)/pET-YSBL-LIC-3C <i>E. coli</i> /pBADN and pCRE2	Jensen et al., 2012 Riebel et al., 2013a
CHMO	<i>chmo</i> gene from <i>A. calcoaceticus</i> NCIMB 9871 and mutant <i>fdh</i> (formate dehydrogenase) gene from <i>C. boidinii</i>	<i>E. coli</i> BL21(DE3)/pET22b (<i>chmo</i>) and pACYCDuet-1 (<i>fdh</i> or both genes in tandem)	Zhai et al., 2013

^aThe information shown mainly corresponds to reports as of 2010.

Abbreviations: BVMO, Baeyer-Villiger monooxygenase; FMO, flavin-containing monooxygenase; CHMO, cyclohexanone monooxygenase; *fdh*, formate dehydrogenase; HAPMO, 4-hydroxyacetophenone monooxygenase; PAMO, phenylacetone monooxygenase; SMO, styrene monooxygenase; *tmm*, trimethylamine monooxygenase.

good enantioselectivity (Rioz-Martínez et al., 2011). Another bacterial FMO, the trimethylamine monooxygenase TMM from *Methylocella silvestris*, was cloned, functionally expressed in *E. coli* and evaluated in the oxidation of dimethylsulfide and dimethylsulfoxide (Chen et al., 2011).

The recently described Type II FMOs are also able to oxidize sulfides, as it was shown for the SMFMO from *S. maltophilia* on thioanisole, *p*-tolyl methyl sulfide, *o*- and *p*-chlorophenyl methyl sulfide, benzyl methyl sulfide, and phenyl ethyl sulfide and, for the set of *R. jostii* Type II FMOs on thioanisole mainly (Jensen et al., 2012; Riebel et al., 2013a).

It is worth noting that de Gonzalo et al. (2011) created artificial flavoenzymes that behaved as self-sufficient flavoprotein monooxygenases capable of stereocomplementary hydrogen

peroxide-driven sulfoxidations by reconstitution of the apo form of a riboflavin-binding protein isolated from eggs with modified flavin derivatives.

N-HYDROXYLATIONS AND N-OXIDATIONS

N-hydroxylating flavoprotein monooxygenases (NMOs) mediate the FAD-dependent oxidation of amines using NADPH as electron donor in the presence of molecular oxygen. Typical representatives are L-ornithine hydroxylases and L-lysine hydroxylases.

In *Pseudomonas aeruginosa*, L-ornithine hydroxylase catalyzes the hydroxylation of the side chain amine of L-ornithine to produce the corresponding hydroxylamine, the initial step in the biosynthesis of the siderophore pyoverdine. After further modifications the hydroxylamine produces a hydroxamate functional

group that is able to chelate ferric ions. The gene coding for L-ornithine hydroxylase (*PvdA*) from *P. aeruginosa* PAO1 was cloned and overexpressed in *E. coli* as a His-tagged fusion (Ge and Seah, 2006; Meneely and Lamb, 2007). The authors characterized it biochemically and its specificity toward several amino acids was investigated. In 2011, two structures of this enzyme, one in its oxidized state and the other in its reduced state, were presented for the first time (Olucha et al., 2011). An L-ornithine hydroxylase was identified in the proteome of *R. jostii* RHA1 (Bosello et al., 2012) and, its coding sequence was cloned and expressed from a pBAD-based expression vector just recently (Riebel et al., 2013a).

The Type I mFMO from *Methylophaga* sp. SK1 is capable of *N*-oxygenations as it was demonstrated for trimethylamine, cysteamine, thiourea, and other *N*-containing compounds (Choi et al., 2003; Alfieri et al., 2008). This activity on trimethylamine was further explored upon fusion of mFMO with the PTDH for self-sufficient cofactor regeneration (Rioz-Martinez et al., 2011). The TMM from *M. silvestris* was also active on methylated amines such as trimethylamine and dimethylamine (Chen et al., 2011). This enzyme is proposed to catalyze the oxidation of trimethylamine to trimethylamine *N*-oxide both in eukaryotes and prokaryotes. In the same work, the authors extended the substrate analysis to four additional bacterial TMMs from *Roseovarius* sp. 217, *Ruegeria pomeroyi* DSS-3, *Pelagibacter ubique* HTCC1002, and *P. ubique* HTCC7211, which were cloned from genomic DNA or synthetic genes, recombinantly expressed and purified from *E. coli* (Chen et al., 2011).

HYDROXYLATIONS

The 4-hydroxybenzoate hydroxylases (PHBH) are classified as class A of flavoprotein monooxygenases, they are encoded by a single gene and contain tightly bound FAD to the sole dinucleotide binding domain present in these enzymes. They depend on NADPH or NADH as electron donors for flavin reduction. Prototype enzymes are PHBH from *P. fluorescens* and *P. aeruginosa*. Physiologically, PHBH participates in routes of degradation of aromatic carbon compounds in soil bacteria by catalyzing the hydroxylation in position 3 of the activated 4-hydroxybenzoate to generate 3,4-dihydroxybenzoate (protocatechuate) that finally enters the β -ketoadipate pathway. The genes encoding the PHBH from *P. aeruginosa* and *P. fluorescens* were cloned and expressed in *E. coli* more than 20 years ago (Entsch et al., 1988; van Berkel et al., 1992). More recently, by means of a combinatorial mutagenesis approach starting from available single mutations of the PHBH from *P. fluorescens* NBRC 14160 multiple properties of the enzyme were simultaneously improved (Suemori and Iwakura, 2007). Subsequently, 53 conserved residues from 92-aligned PHBH primary sequences and 19 non-conserved but presumable functional residues from *P. fluorescens* NBRC 14160 PHBH were substituted with each of the natural amino acids and, activity as well as NADPH reaction specificity were evaluated (Suemori, 2013). Recently, the genes coding for a 4-hydroxybenzoate hydroxylase (*pobA*) and a 3-hydroxybenzoate hydroxylase (*mobA*) from the moderate halophyte *Chromohalobacter* sp. HS-2 were cloned and overexpressed in *E. coli* (Kim et al., 2012). They are part of a cluster containing the genes responsible for the metabolism of benzoate

and hydroxybenzoate in this bacterium (Kim et al., 2008, 2012). Both genes were cloned into the pET-28a(+) vector in order to obtain a fusion to a carboxyl-terminal His-Tag. Initial overexpression experiments gave mostly insoluble His-tagged proteins. Therefore, the culture was subjected to heat shock to induce the expression of *E. coli* DnaK and DnaJ molecular chaperones prior to IPTG-dependent expression of the hydroxylase genes in an attempt to improve protein solubility. Bioconversion of 4- or 3-hydroxybenzoates to protocatechuate was tested in resting cells producing the recombinant 4- or 3-hydroxybenzoate hydroxylase, respectively. The authors reported an increase in product formation, reflected in enhanced bioconversion efficiency, when the reaction was carried out after heat-induction of molecular chaperones (Kim et al., 2012). In 2008, the *mobA* gene from *C. testoteroni* GZ39 coding for 3-hydroxybenzoate hydroxylase (3HB4H) was cloned and subjected to directed evolution by error-prone PCR. With only a single point mutation, the enzyme able to hydroxylate phenolic acids was transformed into an enzyme that can also act on phenol (Chang and Zylstra, 2008).

The recombinant expressions of the 3-hydroxybenzoate 6-hydroxylase (3HB6H) from *Pseudomonas alcaligenes* NCIMB 9867 P25X and from *Polaromonas naphthalenivorans* CJ2 were reported in 2005 and 2007, respectively (Gao et al., 2005; Park et al., 2007). Recently, a 3HB6H from *R. jostii* RHA1 was cloned in the pBAD/Myc-His vector, overexpressed in *E. coli* and characterized biochemically (MonTERSINO and van Berkel, 2012). This FAD-dependent enzyme introduces the hydroxyl group in *p*-position with respect to the previous OH on a series of *o*- or *m*-substituted 3-hydroxybenzoate derivatives. In this study, the authors performed a survey for flavin-dependent hydroxylases in *R. jostii* RHA1 genome and found several hydroxylases that belong to class A of flavoprotein monooxygenases (MonTERSINO and van Berkel, 2012). The crystal structure of the recombinant 3HB6H from *R. jostii* RHA1 was solved recently (MonTERSINO et al., 2013).

Due to their ability to oxidize monophenol to *o*-diphenol compounds, 4-hydroxyphenylacetate 3-hydroxylases (HPAH) are attractive biocatalysts (Lee and Xun, 1998). The enzymes are two-component systems formed by reductase and hydroxylase subunits. Thotsaporn et al. (2004) cloned each subunit in pET-11 derived vectors, expressed them in *E. coli*, and purified and characterized the recombinant enzymes. Later, the *E. coli* W HPAH was cloned in a pETDuet vector, expressed in *E. coli* BL21(DE3) and the whole-cell system was used for the biotransformation of 4-substituted halophenols to the corresponding catechols in shake-flasks and in a 5 L bioreactor (Coulombel et al., 2011). Another group of two-component flavin-dependent monooxygenases that catalyze the oxygenation of 4-hydroxyphenylacetate is represented by the HPAH from *P. aeruginosa*. It was cloned, expressed in *E. coli* and characterized. It was able to oxidize tyrosol to hydroxytyrosol and various phenols (Chakraborty et al., 2010). Recently, both genes coding for HPAH from *P. aeruginosa* PAO1 were cloned into a pETDuet-1 vector and expressed in *E. coli* cells. Biotransformation of several compounds was tested in whole-cell systems and the oxidation of *p*-coumaric acid to caffeic acid was scaled-up both in the absence and in the presence of glucose or glycerol by stepwise increases in substrate concentration in order

to avoid substrate inhibition of the enzymatic activity (Furuya and Kino, 2013).

OTHER REACTIONS CATALYZED BY FLAVOPROTEIN MONOOXYGENASES

The oxidation of indole by microbial oxygenases has been studied during the last 30 years (Ensley et al., 1983). Amongst the recombinant enzymes utilized for this purpose, phenol hydroxylases and styrene monooxygenases are worth mentioning (Doukyu et al., 2003; Gursky et al., 2010). An interesting application of indole biooxidation by styrene monooxygenases was the development of a colorimetric method for the screening of a directed evolution library of *styAB* from *P. putida* CA-3 (Gursky et al., 2010).

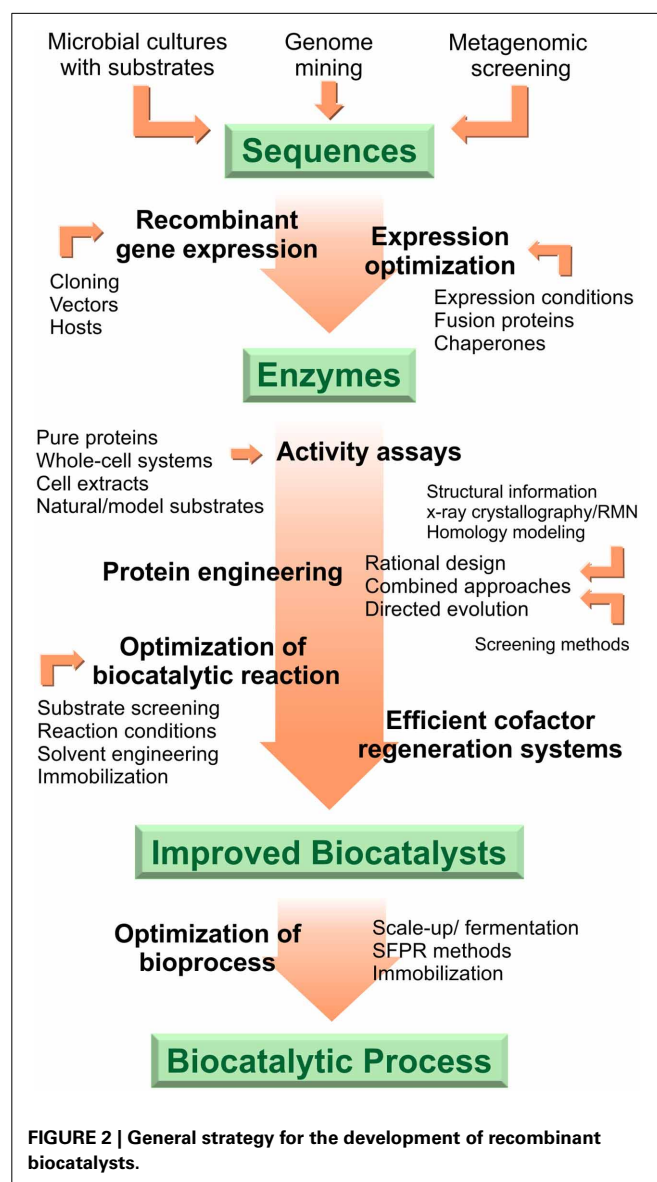
Choi et al. (2003) showed the ability of the mFMO from *Methylophaga* sp. SK1 to produce indigo in *E. coli* and, in the presence of tryptophan, they could increase the production of indigo up to 160 mg/L. To further improve this process, the original plasmid was subjected to deletions in the upstream region of the *fmo* ORF from *Methylophaga aminisulfidivorans* MP^T chromosomal DNA previously cloned. The best producing strain was selected and the composition of the medium as well as pH and temperature for the production of indigo were optimized. As result, the authors could produce 920 mg/L of bio-indigo from the recombinant *E. coli* cells (Han et al., 2008). In 2011, the same research group reported the production of indigo in large scale batch fermentation and in continuous cultivation from the above-mentioned recombinant strain. The latter fermentation mode allowed them to accumulate 23 g of bio-indigo in 110 h (Han et al., 2011). The observation that cultures of cells producing the fusion PTDH-mFMO turned blue motivated an investigation of the ability of isolated PTDH-mFMO to synthesize some indigoid derivatives and the results were readily visualized as different colors of the reaction mixtures (Rioz-Martínez et al., 2011). The first oxidation of indole catalyzed by a BVMO was reported in 2007 for the M446G mutant of PAMO (Torres Pazmiño et al., 2007), when cultures expressing this recombinant enzyme turned blue due to the formation of indigo blue.

Some Type I BVMOs have also the ability to oxidize boron-containing compounds (Branchaud and Walsh, 1985; Walsh and Chen, 1988). In 2005, PAMO from *T. fusca* was shown to oxidize phenylboronic acid to phenol (de Gonzalo et al., 2005b). A year later the same research group reported a similar result for HAPMO from *P. fluorescens* ACB (de Gonzalo et al., 2006b). These studies were extended to a variety of boron-containing acetophenones, vinyl boron compounds and racemic boron-containing compounds and two other BVMOs were evaluated (i.e., the PAMO M446G mutant and CHMO_{Acineto}) (Brondani et al., 2011). The chemoselectivity of the reactions was variable and depended on the biocatalyst, being the boron oxidation exclusively preferred over Baeyer-Villiger oxidation on 3-substituted acetophenones for wild-type and M446G PAMO. The selectivity between epoxidation and boron oxidation was investigated in vinyl boron compounds and only boron oxidation was reported in some cases. The excellent chemoselectivity of PAMO was employed to attain the kinetic resolution of boron-containing compounds giving chiral alcohols and chiral boron compounds in high e.e. (Brondani et al., 2011). In a following report, the

enantioselectivity of the BVMOs for oxidative kinetic resolutions of racemic cyclopropyl boronic esters, phenylethyl boronates, and β -boronated carboxylic esters was investigated (Brondani et al., 2012a). In addition, recombinant PAMO efficiently mediated the chemoselective oxidation of some organoselenium acetophenones to the corresponding selenoxides (Andrade et al., 2011) and, a chiral selenium compound was afforded by the kinetic resolution of a racemic selenium-containing aromatic compound with high e.e. in a reaction mediated by PAMO (Brondani et al., 2012b).

CONCLUDING REMARKS

Biocatalysis is an environmentally friendly strategy for the elaboration of fine chemicals, natural products or other biologically active compounds. During the last decades enormous efforts have been done to satisfy the demands of biocatalysts for organic synthesis. However, multidisciplinary and coordinate work is



still required to enlarge the repertoire of accessible reactions and compounds. The development of recombinant biocatalysts for organic synthesis and industrial applications involves multiple steps beginning from sequence selection up to bioprocess improvement (Figure 2). Aiming at describing and exemplifying this entire process, in the preceding sections we presented recent work and the state of the art on flavoprotein monooxygenases-mediated reactions for the creation of selective and stable biocatalysts as well as robust biotransformation processes. Bioinformatics analysis, recombinant DNA technology and protein engineering methods are part of the basic toolkit toward optimized redox biocatalysts. The sequence of interest can be derived from natural or synthetic origin. Once it is cloned, selection of convenient expression vectors and improved hosts, optimization of growing and induction media or conditions and assisted protein folding can help reach proper recombinant expression levels. Considering the critical requisite of flavoprotein monooxygenases for cofactor recycling, activity of the biocatalysts can be evaluated in formats ranging from whole-cell systems to pure enzymes. Protein engineering techniques such as directed evolution, rational re-design and *de-novo* design of enzymes allow the expansion of the range of biocatalysts available and the development of tailored enzymes. Innovations on solvent or reaction medium engineering have also a huge impact on the biocatalytic outcome. Immobilization methods, strain improvement by metabolic engineering, and scale-up procedures under fine bioprocess control are further valuable tools for the development of a successful biocatalytic process for the industry.

ACKNOWLEDGMENTS

We would like to thank the financial support from Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT; PRH 24 PICT 2009-0088), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET; PIP 2012-2014 N° 1156) and Universidad Nacional de Rosario (UNR; BIO287, BIO339). Daniela V. Rial and Dario A. Bianchi are staff researchers of CONICET, Argentina. Romina D. Ceccoli is a post-doctoral fellow of the same Institution. Daniela V. Rial and Dario A. Bianchi are Assistant Professors (Profesores Adjuntos) and Romina D. Ceccoli is a Teaching Assistant (Auxiliar de primera categoría) at Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario.

REFERENCES

- Alexander, A. K., Biedermann, D., Fink, M. J., Mihovilovic, M. D., and Mattes, T. E. (2012). Enantioselective oxidation by a cyclohexanone monooxygenase from the xenobiotic-degrading *Polaromonas* sp. strain JS666. *J. Mol. Catal. B Enzym.* 78, 105–110. doi: 10.1016/j.molcatb.2012.03.002
- Alferi, A., Malito, E., Orru, R., Fraaije, M. W., and Mattevi, A. (2008). Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6572–6577. doi: 10.1073/pnas.0800859105
- Andrade, L. H., Pedrozo, E. C., Leite, H. G., and Brondani, P. B. (2011). Oxidation of organoselenium compounds. A study of chemoselectivity of phenylacetone monooxygenase. *J. Mol. Catal. B Enzym.* 73, 63–66. doi: 10.1016/j.molcatb.2011.07.018
- Baeyer, A., and Villiger, V. (1899). Einwirkung des Caro'schen Reagens auf Ketone. *Ber. Dtsch. Chem. Ges.* 32, 3625–3633. doi: 10.1002/cber.189903203151
- Baldwin, C. V. F., Wohlgemuth, R., and Woodley, J. M. (2008). The first 200-L scale asymmetric Baeyer-Villiger oxidation using a whole-cell biocatalyst. *Org. Process Res. Dev.* 12, 660–665. doi: 10.1021/op800046t
- Balke, K., Kadow, M., Mallin, H., Sass, S., and Bornscheuer, U. T. (2012). Discovery, application and protein engineering of Baeyer-Villiger monooxygenases for organic synthesis. *Org. Biomol. Chem.* 10, 6249–6265. doi: 10.1039/c2ob25704a
- Beneventi, E., Niero, M., Motterle, R., Fraaije, M. W., and Bergantino, E. (2013). Discovery of Baeyer-Villiger monooxygenases from photosynthetic eukaryotes. *J. Mol. Catal. B Enzym.* 98, 145–154. doi: 10.1016/j.molcatb.2013.10.006
- Beneventi, E., Ottolina, G., Carrea, G., Panzeri, W., Fronza, G., and Lau, P. C. K. (2009). Enzymatic Baeyer-Villiger oxidation of steroids with cyclopentadecanone monooxygenase. *J. Mol. Catal. B Enzym.* 58, 164–168. doi: 10.1016/j.molcatb.2008.12.009
- Bianchi, D. A., Moran-Ramallal, R., Iqbal, N., Rudroff, F., and Mihovilovic, M. D. (2013). Enantiocomplementary access to carba-analogs of C-nucleoside derivatives by recombinant Baeyer-Villiger monooxygenases. *Bioorg. Med. Chem. Lett.* 23, 2718–2720. doi: 10.1016/j.bmcl.2013.02.085
- Bocola, M., Schulz, F., Leca, F., Vogel, A., Fraaije, M. W., and Reetz, M. T. (2005). Converting phenylacetone monooxygenase into phenylcyclohexanone monooxygenase by rational design: towards practical Baeyer-Villiger monooxygenases. *Adv. Synth. Catal.* 347, 979–986. doi: 10.1002/adsc.200505069
- Bosello, M., Mielcarek, A., Giessen, T. W., and Marahiel, M. A. (2012). An enzymatic pathway for the biosynthesis of the formylhydroxyornithine required for rhodochelin iron coordination. *Biochemistry* 51, 3059–3066. doi: 10.1021/bi201837f
- Boyd, D. R., Sharma, N. D., McMurray, B., Haughey, S. A., Allen, C. C., Hamilton, J. T., et al. (2012). Bacterial dioxygenase- and monooxygenase-catalysed sulfoxidation of benzo[b]thiophenes. *Org. Biomol. Chem.* 10, 782–790. doi: 10.1039/c1ob06678a
- Branchaud, B. P., and Walsh, C. T. (1985). Functional group diversity in enzymic oxygenation reactions catalyzed by bacterial flavin-containing cyclohexanone oxygenase. *J. Am. Chem. Soc.* 107, 2153–2161. doi: 10.1021/ja00293a054
- Brondani, P. B., de Gonzalo, G., Fraaije, M. W., and Andrade, L. H. (2011). Selective oxidations of organoboron compounds catalyzed by Baeyer-Villiger monooxygenases. *Adv. Synth. Catal.* 353, 2169–2173. doi: 10.1002/adsc.201100029
- Brondani, P. B., Dudek, H., Reis, J. S., Fraaije, M. W., and Andrade, L. H. (2012a). Exploiting the enantioselectivity of Baeyer-Villiger monooxygenases via boron oxidation. *Tetrahedron* 23, 703–708. doi: 10.1016/j.tetasy.2012.05.004
- Brondani, P. B., Guilmo, N. M. A. F., Dudek, H. M., Fraaije, M. W., and Andrade, L. H. (2012b). Chemoenzymatic approaches to obtain chiral-centered selenium compounds. *Tetrahedron* 68, 10431–10436. doi: 10.1016/j.tet.2012.09.087
- Bučko, M., Schenkmyerová, A., Gemeiner, P., Vikartovská, A., Mihovilovic, M. D., and Lacík, I. (2011). Continuous testing system for Baeyer-Villiger biooxidation using recombinant *Escherichia coli* expressing cyclohexanone monooxygenase encapsulated in polyelectrolyte complex capsules. *Enzyme Microb. Technol.* 49, 284–288. doi: 10.1016/j.enzmictec.2011.05.013
- Bühler, B., Park, J. B., Blank, L. M., and Schmid, A. (2008). NADH availability limits asymmetric biocatalytic epoxidation in a growing recombinant *Escherichia coli* strain. *Appl. Environ. Microbiol.* 74, 1436–1446. doi: 10.1128/AEM.02234-07
- Chakraborty, S., Ortiz-Maldonado, M., Entsch, B., and Ballou, D. P. (2010). Studies on the mechanism of p-hydroxyphenylacetate 3-hydroxylase from *Pseudomonas aeruginosa*: a system composed of a small flavin reductase and a large flavin-dependent oxygenase. *Biochemistry* 49, 372–385. doi: 10.1021/bi901454u
- Chang, H. K., and Zylstra, G. J. (2008). Examination and expansion of the substrate range of m-hydroxybenzoate hydroxylase. *Biochem. Biophys. Res. Commun.* 371, 149–153. doi: 10.1016/j.bbrc.2008.04.032
- Chen, Y., Patel, N. A., Crombie, A., Scrivens, J. H., and Murrell, J. C. (2011). Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17791–17796. doi: 10.1073/pnas.1112928108
- Choi, H. S., Kim, J. K., Cho, E. H., Kim, Y. C., Kim, J. I., and Kim, S. W. (2003). A novel flavin-containing monooxygenase from *Methylophaga* sp strain SK1 and its indigo synthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 306, 930–936. doi: 10.1016/S0006-291X(03)01087-8
- Coulombel, L., Nolan, L. C., Nikodinovic, J., Doyle, E. M., and O'Connor, K. E. (2011). Biotransformation of 4-halophenols to 4-halocatechols using *Escherichia coli* expressing 4-hydroxyphenylacetate 3-hydroxylase. *Appl. Microbiol. Biotechnol.* 89, 1867–1875. doi: 10.1007/s00253-010-2969-5
- de Gonzalo, G., and Fraaije, M. W. (2013). Recent developments in flavin-based catalysis. *ChemCatChem* 5, 403–415. doi: 10.1002/cctc.201200466

- de Gonzalo, G., Mihovilovic, M. D., and Fraaije, M. W. (2010). Recent developments in the application of Baeyer-Villiger monooxygenases as biocatalysts. *ChemBioChem* 11, 2208–2231. doi: 10.1002/cbic.201000395
- de Gonzalo, G., Ottolina, G., Carrea, G., and Fraaije, M. W. (2005a). $[Cp^*Rh(bpy)(H_2O)]^{2+}$ as a coenzyme substitute in enzymatic oxidations catalyzed by Baeyer-Villiger monooxygenases. *Chem. Commun.* 3724–3726. doi: 10.1039/b504921k
- de Gonzalo, G., Torres Pazmiño, D. E., Ottolina, G., Fraaije, M. W., and Carrea, G. (2005b). Oxidations catalyzed by phenylacetone monooxygenase from *Thermobifida fusca*. *Tetrahedron* 16, 3077–3083. doi: 10.1016/j.tetasy.2005.08.004
- de Gonzalo, G., Ottolina, G., Zambianchi, F., Fraaije, M. W., and Carrea, G. (2006a). Biocatalytic properties of Baeyer-Villiger monooxygenases in aqueous-organic media. *J. Mol. Catal. B Enzym.* 39, 91–97. doi: 10.1016/j.molcatb.2006.01.010
- de Gonzalo, G., Torres Pazmiño, D. E., Ottolina, G., Fraaije, M. W., and Carrea, G. (2006b). 4-Hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens* ACB as an oxidative biocatalyst in the synthesis of optically active sulfoxides. *Tetrahedron* 17, 130–135. doi: 10.1016/j.tetasy.2005.11.024
- de Gonzalo, G., Rodríguez, C., Rios-Martínez, A., and Gotor, V. (2012). Improvement of the biocatalytic properties of one phenylacetone monooxygenase mutant in hydrophilic organic solvents. *Enzyme Microb. Technol.* 50, 43–49. doi: 10.1016/j.enzmictec.2011.09.006
- de Gonzalo, G., Smit, C., Jin, J., Minnaard, A. J., and Fraaije, M. W. (2011). Turning a riboflavin-binding protein into a self-sufficient monooxygenase by cofactor redesign. *Chem. Commun.* 47, 11050–11052. doi: 10.1039/c1cc14039f
- Doukyu, N., Toyoda, K., and Aono, R. (2003). Indigo production by *Escherichia coli* carrying the phenol hydroxylase gene from *Acinetobacter* sp strain ST-550 in a water-organic solvent two-phase system. *Appl. Microbiol. Biotechnol.* 60, 720–725. doi: 10.1007/s00253-002-1187-1
- Dudek, H. M., de Gonzalo, G., Torres Pazmiño, D. E., Stepniak, P., Wyrwicz, L. S., Rychlewski, L., et al. (2011). Mapping the substrate binding site of phenylacetone monooxygenase from *Thermobifida fusca* by mutational analysis. *Appl. Environ. Microbiol.* 77, 5730–5738. doi: 10.1128/AEM.00687-11
- Dudek, H. M., Fink, M. J., Shivange, A. V., Dennig, A., Mihovilovic, M. D., Schwaneberg, U., et al. (2013a). Extending the substrate scope of a Baeyer-Villiger monooxygenase by multiple-site mutagenesis. *Appl. Microbiol. Biotechnol.* doi: 10.1007/s00253-013-5364-1
- Dudek, H. M., Popken, P., van Bloois, E., Duetz, W. A., and Fraaije, M. W. (2013b). A generic, whole-cell-based screening method for Baeyer-Villiger monooxygenases. *J. Biomol. Screen.* 18, 678–687. doi: 10.1177/1087057113480390
- Ensley, B. D., Ratzkin, B. J., Osslund, T. D., Simpson, M. J., Wackett, L. P., and Gibson, D. T. (1983). Expression of the naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222, 167–169. doi: 10.1126/science.6353574
- Entsch, B., Nan, Y., Weaich, K., and Scott, K. F. (1988). Sequence and organization of *pobA*, the gene coding for p-hydroxybenzoate hydroxylase, an inducible enzyme from *Pseudomonas aeruginosa*. *Gene* 71, 279–291. doi: 10.1016/0378-1119(88)90044-3
- Fink, M. J., Fischer, T. C., Rudroff, F., Dudek, H., Fraaije, M. W., and Mihovilovic, M. D. (2011). Extensive substrate profiling of cyclopentadecanone monooxygenase as Baeyer-Villiger biocatalyst reveals novel regiodivergent oxidations. *J. Mol. Catal. B Enzym.* 73, 9–16. doi: 10.1016/j.molcatb.2011.07.003
- Fink, M. J., Rial, D. V., Kapita, P., Lengar, A., Rehdorf, J., Cheng, Q., et al. (2012). Quantitative comparison of chiral catalysts selectivity and performance: a generic concept illustrated with cyclododecanone monooxygenase as Baeyer-Villiger biocatalyst. *Adv. Synth. Catal.* 354, 3491–3500. doi: 10.1002/adsc.201200453
- Fraaije, M. W., Kamerbeek, N. M., van Berkel, W. J., and Janssen, D. B. (2002). Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett.* 518, 43–47. doi: 10.1016/S0014-5793(02)02623-6
- Fraaije, M. W., Wu, J., Heuts, D. P., van Hellemond, E. W., Spelberg, J. H., and Janssen, D. B. (2005). Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. *Appl. Microbiol. Biotechnol.* 66, 393–400. doi: 10.1007/s00253-004-1749-5
- Franceschini, S., van Beek, H. L., Pennetta, A., Martinoli, C., Fraaije, M. W., and Mattevi, A. (2012). Exploring the structural basis of substrate preferences in Baeyer-Villiger monooxygenases: insight from steroid monooxygenase. *J. Biol. Chem.* 287, 22626–22634. doi: 10.1074/jbc.M112.372177
- Furuya, T., and Kino, K. (2013). Catalytic activity of the two-component flavin-dependent monooxygenase from *Pseudomonas aeruginosa* toward cinnamic acid derivatives. *Appl. Microbiol. Biotechnol.* doi: 10.1007/s00253-013-4958-y
- Gao, X., Tan, C. L., Yeo, C. C., and Poh, C. L. (2005). Molecular and biochemical characterization of the *xlnD*-encoded 3-hydroxybenzoate 6-hydroxylase involved in the degradation of 2,5-xynolol via the gentisate pathway in *Pseudomonas alcaligenes* NCIMB 9867. *J. Bacteriol.* 187, 7696–7702. doi: 10.1128/JB.187.22.7696-7702.2005
- Ge, L., and Seah, S. Y. (2006). Heterologous expression, purification, and characterization of an l-ornithine N(5)-hydroxylase involved in pyoverdine siderophore biosynthesis in *Pseudomonas aeruginosa*. *J. Bacteriol.* 188, 7205–7210. doi: 10.1128/JB.00949-06
- Geitner, K., Rehdorf, J., Snajdrova, R., and Bornscheuer, U. T. (2010). Scale-up of Baeyer-Villiger monooxygenase-catalyzed synthesis of enantiopure compounds. *Appl. Microbiol. Biotechnol.* 88, 1087–1093. doi: 10.1007/s00253-010-2724-y
- Gursky, L. J., Nikodinovic-Runic, J., Feenstra, K. A., and O'connor, K. E. (2010). *In vitro* evolution of styrene monooxygenase from *Pseudomonas putida* CA-3 for improved epoxide synthesis. *Appl. Microbiol. Biotechnol.* 85, 995–1004. doi: 10.1007/s00253-009-2096-3
- Han, G. H., Bang, S. E., Babu, B. K., Chang, M., Shin, H. J., and Kim, S. W. (2011). Bio-indigo production in two different fermentation systems using recombinant *Escherichia coli* cells harboring a flavin-containing monooxygenase gene (*fmo*). *Proc. Biochem.* 46, 788–791. doi: 10.1016/j.procbio.2010.10.015
- Han, G. H., Shin, H. J., and Kim, S. W. (2008). Optimization of bio-indigo production by recombinant *E. coli* harboring *fmo* gene. *Enzyme Microb. Technol.* 42, 617–623. doi: 10.1016/j.enzmictec.2008.02.004
- Hollmann, F., Taglieber, A., Schulz, F., and Reetz, M. T. (2007). A light-driven stereoselective biocatalytic oxidation. *Angew Chem. Int. Ed.* 46, 2903–2906. doi: 10.1002/anie.200605169
- Iwaki, H., Grosse, S., Bergeron, H., Leisch, H., Morley, K., Hasegawa, Y., et al. (2013). Camphor pathway redux: functional recombinant expression of 2,5- and 3,6-diketocamphane monooxygenases of *Pseudomonas putida* ATCC 17453 with their cognate flavin reductase catalyzing Baeyer-Villiger reactions. *Appl. Environ. Microbiol.* 79, 3282–3293. doi: 10.1128/AEM.03958-12
- Iwaki, H., Wang, S., Grosse, S., Bergeron, H., Nagahashi, A., Lertvorachon, J., et al. (2006). Pseudomonad cyclopentadecanone monooxygenase displaying an uncommon spectrum of Baeyer-Villiger oxidations of cyclic ketones. *Appl. Environ. Microbiol.* 72, 2707–2720. doi: 10.1128/AEM.72.4.2707-2720.2006
- Jensen, C. N., Cartwright, J., Ward, J., Hart, S., Turkenburg, J. P., Ali, S. T., et al. (2012). A flavoprotein monooxygenase that catalyses a Baeyer-Villiger reaction and thioether oxidation using NADH as the nicotinamide cofactor. *ChemBioChem* 13, 872–878. doi: 10.1002/cbic.201200006
- Kadow, M., Balke, K., Willetts, A., Bornscheuer, U. T., and Bäckvall, J. E. (2013). Functional assembly of camphor converting two-component Baeyer-Villiger monooxygenases with a flavin reductase from *E. coli*. *Appl. Microbiol. Biotechnol.* doi: 10.1007/s00253-013-5338-3
- Kadow, M., Loschinski, K., Sass, S., Schmidt, M., and Bornscheuer, U. T. (2012). Completing the series of BVMOs involved in camphor metabolism of *Pseudomonas putida* NCIMB 10007 by identification of the two missing genes, their functional expression in *E. coli*, and biochemical characterization. *Appl. Microbiol. Biotechnol.* 96, 419–429. doi: 10.1007/s00253-011-3859-1
- Kadow, M., Sass, S., Schmidt, M., and Bornscheuer, U. T. (2011). Recombinant expression and purification of the 2,5-diketocamphane 1,2-monooxygenase from the camphor metabolizing *Pseudomonas putida* strain NCIMB 10007. *AMB Exp.* 1, 13. doi: 10.1186/2191-0855-1-13
- Kamerbeek, N. M., Moonen, M. J., van der Ven, J. G., van Berkel, W. J., Fraaije, M. W., and Janssen, D. B. (2001). 4-Hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens* ACB. A novel flavoprotein catalyzing Baeyer-Villiger oxidation of aromatic compounds. *Eur. J. Biochem.* 268, 2547–2557. doi: 10.1046/j.1432-1327.2001.02137.x
- Kim, D., Kim, S. W., Choi, K. Y., Lee, J. S., and Kim, E. (2008). Molecular cloning and functional characterization of the genes encoding benzoate and p-hydroxybenzoate degradation by the halophilic *Chromohalobacter* sp. strain HS-2. *FEMS Microbiol. Lett.* 280, 235–241. doi: 10.1111/j.1574-6968.2008.01067.x
- Kim, W., Park, Y. R., Im, S., Kim, D., and Kim, S. W. (2012). Expression and bioconversion of recombinant m- and p-hydroxybenzoate hydroxylases from a novel

- moderate halophile, *Chromohalobacter* sp. *Biotechnol. Lett.* 34, 1687–1692. doi: 10.1007/s10529-012-0950-3
- Lee, J. Y., and Xun, L. (1998). Novel biological process for L-DOPA production from L-tyrosine by P-hydroxyphenylacetate 3-hydroxylase. *Biotechnol. Lett.* 20, 479–482. doi: 10.1023/A:1005440229420
- Lee, W.-H., Kim, J.-W., Park, E.-H., Han, N. S., Kim, M.-D., and Seo, J.-H. (2013). Effects of NADH kinase on NADPH-dependent biotransformation processes in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 97, 1561–1569. doi: 10.1007/s00253-012-4431-3
- Lee, W. H., Park, J. B., Park, K., Kim, M. D., and Seo, J. H. (2007). Enhanced production of epsilon-caprolactone by overexpression of NADPH-regenerating glucose 6-phosphate dehydrogenase in recombinant *Escherichia coli* harboring cyclohexanone monooxygenase gene. *Appl. Microbiol. Biotechnol.* 76, 329–338. doi: 10.1007/s00253-007-1016-7
- Leipold, F., Rudroff, F., Mihovilovic, M. D., and Bornscheuer, U. T. (2013). The steroid monooxygenase from *Rhodococcus rhodochrous*; a versatile biocatalyst. *Tetrahedron* 24, 1620–1624. doi: 10.1016/j.tetasy.2013.11.003
- Leipold, F., Wardenga, R., and Bornscheuer, U. T. (2012). Cloning, expression and characterization of a eukaryotic cycloalkanone monooxygenase from *Cylindrocarpus radicola* ATCC 11011. *Appl. Microbiol. Biotechnol.* 94, 705–717. doi: 10.1007/s00253-011-3670-z
- Leisch, H., Morley, K., and Lau, P. C. (2011). Baeyer-Villiger monooxygenases: more than just green chemistry. *Chem. Rev.* 111, 4165–4222. doi: 10.1021/cr1003437
- Leisch, H., Shi, R., Grosse, S., Morley, K., Bergeron, H., Cygler, M., et al. (2012). Cloning, Baeyer-Villiger biooxidations, and structures of the camphor pathway 2-oxo-Delta(3)-4,5,5-trimethylcyclopentenylacetyl-coenzyme A monooxygenase of *Pseudomonas putida* ATCC 17453. *Appl. Environ. Microbiol.* 78, 2200–2212. doi: 10.1128/AEM.07694-11
- Lin, H., Liu, Y., and Wu, Z. L. (2011a). Asymmetric epoxidation of styrene derivatives by styrene monooxygenase from *Pseudomonas* sp. LQ26 effects of α - and β -substituents. *Tetrahedron* 22, 134–137. doi: 10.1016/j.tetasy.2010.12.022
- Lin, H., Liu, Y., and Wu, Z. L. (2011b). Highly diastereo- and enantio-selective epoxidation of secondary allylic alcohols catalyzed by styrene monooxygenase. *Chem. Commun.* 47, 2610–2612. doi: 10.1039/c0cc04360e
- Lin, H., Qiao, J., Liu, Y., and Wu, Z. L. (2010). Styrene monooxygenase from *Pseudomonas* sp. LQ26 catalyzes the asymmetric epoxidation of both conjugated and unconjugated alkenes. *J. Mol. Catal. B Enzym.* 67, 236–241. doi: 10.1016/j.molcatb.2010.08.012
- Lin, H., Tang, D. F., Ahmed, A. A., Liu, Y., and Wu, Z. L. (2012). Mutations at the putative active cavity of styrene monooxygenase: enhanced activity and reversed enantioselectivity. *J. Biotechnol.* 161, 235–241. doi: 10.1016/j.jbiotec.2012.06.028
- Malito, E., Alfieri, A., Fraaije, M. W., and Mattevi, A. (2004). Crystal structure of a Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13157–13162. doi: 10.1073/pnas.0404538101
- Mascotti, M. L., Juri Ayub, M., Dudek, H., Sanz, M. K., and Fraaije, M. W. (2013). Cloning, overexpression and biocatalytic exploration of a novel Baeyer-Villiger monooxygenase from *Aspergillus fumigatus* Af293. *AMB Exp.* 3, 33. doi: 10.1186/2191-0855-3-33
- Meneely, K. M., and Lamb, A. L. (2007). Biochemical characterization of a flavin adenine dinucleotide-dependent monooxygenase, ornithine hydroxylase from *Pseudomonas aeruginosa*, suggests a novel reaction mechanism. *Biochemistry* 46, 11930–11937. doi: 10.1021/bi700932q
- Mihovilovic, M. D. (2006). Enzyme mediated Baeyer-Villiger oxidations. *Curr. Org. Chem.* 10, 1265–1287. doi: 10.2174/13852720677698002
- Mihovilovic, M. D., Bianchi, D. A., and Rudroff, F. (2006). Accessing tetrahydrofuran-based natural products by microbial Baeyer-Villiger biooxidation. *Chem. Commun.* 3214–3216. doi: 10.1039/b606633j
- Minerdi, D., Zgrablic, I., Sadeghi, S. J., and Gilardi, G. (2012). Identification of a novel Baeyer-Villiger monooxygenase from *Acinetobacter radioresistens*: close relationship to the *Mycobacterium tuberculosis* prodrug activator EtaA. *Microb. Biotechnol.* 5, 700–716. doi: 10.1111/j.1751-7915.2012.00356.x
- Montersino, S., Orru, R., Barendregt, A., Westphal, A. H., van Duijn, E., Mattevi, A., et al. (2013). Crystal structure of 3-hydroxybenzoate 6-hydroxylase uncovers lipid-assisted flavoprotein strategy for regioselective aromatic hydroxylation. *J. Biol. Chem.* 288, 26235–26245. doi: 10.1074/jbc.M113.479303
- Montersino, S., and van Berkel, W. J. (2012). Functional annotation and characterization of 3-hydroxybenzoate 6-hydroxylase from *Rhodococcus jostii* RHA1. *Biochim. Biophys. Acta* 1824, 433–442. doi: 10.1016/j.bbapap.2011.12.003
- Nikodinovic-Runic, J., Coulombel, L., Francuski, D., Sharma, N. D., Boyd, D. R., Ferrall, R. M., et al. (2013). The oxidation of alkylaryl sulfides and benzo[b]thiophenes by *Escherichia coli* cells expressing wild-type and engineered styrene monooxygenase from *Pseudomonas putida* CA-3. *Appl. Microbiol. Biotechnol.* 97, 4849–4858. doi: 10.1007/s00253-012-4332-5
- Olucha, J., Meneely, K. M., Chilton, A. S., and Lamb, A. L. (2011). Two structures of an N-hydroxylating flavoprotein monooxygenase: ornithine hydroxylase from *Pseudomonas aeruginosa*. *J. Biol. Chem.* 286, 31789–31798. doi: 10.1074/jbc.M111.265876
- Opperman, D. J., and Reetz, M. T. (2010). Towards practical Baeyer-Villiger monooxygenases: design of cyclohexanone monooxygenase mutants with enhanced oxidative stability. *ChemBioChem* 11, 2589–2596. doi: 10.1002/cbic.201000464
- Orru, R., Dudek, H. M., Martinoli, C., Torres Pazmiño, D. E., Royant, A., Weik, M., et al. (2011). Snapshots of enzymatic Baeyer-Villiger catalysis: oxygen activation and intermediate stabilization. *J. Biol. Chem.* 286, 29284–29291. doi: 10.1074/jbc.M111.255075
- Panke, S., Witholt, B., Schmid, A., and Wubbols, M. G. (1998). Towards a biocatalyst for (S)-styrene oxide production: characterization of the styrene degradation pathway of *Pseudomonas* sp. strain VLB120. *Appl. Environ. Microbiol.* 64, 2032–2043.
- Park, M., Jeon, Y., Jang, H. H., Ro, H. S., Park, W., Madsen, E. L., et al. (2007). Molecular and biochemical characterization of 3-hydroxybenzoate 6-hydroxylase from *Polaromonas naphthalenivorans* CJ2. *Appl. Environ. Microbiol.* 73, 5146–5152. doi: 10.1128/AEM.00782-07
- Qaed, A. A., Lin, H., Tang, D. F., and Wu, Z. L. (2011). Rational design of styrene monooxygenase mutants with altered substrate preference. *Biotechnol. Lett.* 33, 611–616. doi: 10.1007/s10529-010-0472-9
- Reetz, M. T., and Wu, S. (2009). Laboratory evolution of robust and enantioselective Baeyer-Villiger monooxygenases for asymmetric catalysis. *J. Am. Chem. Soc.* 131, 15424–15432. doi: 10.1021/ja906212k
- Rehder, J., Zimmer, C. L., and Bornscheuer, U. T. (2009). Cloning, expression, characterization, and biocatalytic investigation of the 4-hydroxyacetophenone monooxygenase from *Pseudomonas putida* JD1. *Appl. Environ. Microbiol.* 75, 3106–3114. doi: 10.1128/AEM.02707-08
- Rial, D. V., Bianchi, D. A., Kapitanova, P., Lengár, A., van Beilen, J. B., and Mihovilovic, M. D. (2008a). Stereoselective desymmetrizations by recombinant whole cells expressing the Baeyer-Villiger monooxygenase from *Xanthobacter* sp. ZL5: a new biocatalyst accepting structurally demanding substrates. *Eur. J. Org. Chem.* 2008, 1203–1213. doi: 10.1002/ejoc.200700872
- Rial, D. V., Cernuchova, P., van Beilen, J. B., and Mihovilovic, M. D. (2008b). Biocatalyst assessment of recombinant whole-cells expressing the Baeyer-Villiger monooxygenase from *Xanthobacter* sp. ZL5. *J. Mol. Catal. B Enzym.* 50, 61–68. doi: 10.1016/j.molcatb.2007.09.001
- Riebel, A., de Gonzalo, G., and Fraaije, M. W. (2013a). Expanding the biocatalytic toolbox of flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. *J. Mol. Catal. B Enzym.* 88, 20–25. doi: 10.1016/j.molcatb.2012.11.009
- Riebel, A., Fink, M. J., Mihovilovic, M. D., and Fraaije, M. W. (2013b). Type II flavin-containing monooxygenases: a new class of biocatalysts that harbors Baeyer-Villiger monooxygenases with a relaxed coenzyme specificity. *ChemCatChem*. doi: 10.1002/cctc.201300550
- Riebel, A., Dudek, H. M., de Gonzalo, G., Stepniak, P., Rychlewski, L., and Fraaije, M. W. (2012). Expanding the set of rhodococcal Baeyer-Villiger monooxygenases by high-throughput cloning, expression and substrate screening. *Appl. Microbiol. Biotechnol.* 95, 1479–1489. doi: 10.1007/s00253-011-3823-0
- Rioz-Martínez, A., de Gonzalo, G., Torres Pazmiño, D. E., Fraaije, M. W., and Gotor, V. (2009). Enzymatic Baeyer-Villiger oxidation of benzo-fused ketones: formation of regiocomplementary lactones. *Eur. J. Org. Chem.* 2009, 2526–2532. doi: 10.1002/ejoc.200900084
- Rioz-Martínez, A., de Gonzalo, G., Torres Pazmiño, D. E., Fraaije, M. W., and Gotor, V. (2010a). Enzymatic synthesis of novel chiral sulfoxides employing Baeyer-Villiger monooxygenases. *Eur. J. Org. Chem.* 2010, 6409–6416. doi: 10.1002/ejoc.201000890
- Rioz-Martínez, A., de Gonzalo, G., Torres Pazmiño, D. E., Fraaije, M. W., and Gotor, V. (2010b). Synthesis of chiral 3-alkyl-3,4-dihydroisocoumarins by dynamic kinetic resolutions catalyzed by a Baeyer-Villiger monooxygenase. *J. Org. Chem.* 75, 2073–2076. doi: 10.1021/jo902519j

- Rioz-Martínez, A., Kopacz, M., de Gonzalo, G., Torres Pazmiño, D. E., Gotor, V., and Fraaije, M. W. (2011). Exploring the biocatalytic scope of a bacterial flavin-containing monooxygenase. *Org. Biomol. Chem.* 9, 1337–1341. doi: 10.1039/c0ob00988a
- Rodríguez, C., de Gonzalo, G., Fraaije, M. W., and Gotor, V. (2007). Enzymatic kinetic resolution of racemic ketones catalyzed by Baeyer-Villiger monooxygenases. *Tetrahedron* 18, 1338–1344. doi: 10.1016/j.tetasy.2007.05.033
- Rodríguez, C., de Gonzalo, G., and Gotor, V. (2012). Optimization of oxidative bioconversions catalyzed by phenylacetone monooxygenase from *Thermobifida fusca*. *J. Mol. Catal. B Enzym.* 74, 138–143. doi: 10.1016/j.molcatb.2011.09.010
- Rodríguez, C., de Gonzalo, G., Torres Pazmiño, D. E., Fraaije, M. W., and Gotor, V. (2008). Selective Baeyer-Villiger oxidation of racemic ketones in aqueous-organic media catalyzed by phenylacetone monooxygenase. *Tetrahedron* 19, 197–203. doi: 10.1016/j.tetasy.2007.12.008
- Suemori, A. (2013). Conserved and non-conserved residues and their role in the structure and function of p-hydroxybenzoate hydroxylase. *Protein Eng. Des. Sel.* 26, 479–488. doi: 10.1093/protein/gzt026
- Suemori, A., and Iwakura, M. (2007). A systematic and comprehensive combinatorial approach to simultaneously improve the activity, reaction specificity, and thermal stability of p-hydroxybenzoate hydroxylase. *J. Biol. Chem.* 282, 19969–19978. doi: 10.1074/jbc.M610320200
- Szolkowy, C., Eltis, L. D., Bruce, N. C., and Grogan, G. (2009). Insights into sequence-activity relationships amongst Baeyer-Villiger monooxygenases as revealed by the intragenomic complement of enzymes from *Rhodococcus jostii* RHA1. *ChemBioChem* 10, 1208–1217. doi: 10.1002/cbic.200900011
- Thotsaporn, K., Sucharitakul, J., Wongratana, J., Suadee, C., and Chaiyen, P. (2004). Cloning and expression of p-hydroxyphenylacetate 3-hydroxylase from *Acinetobacter baumannii*: evidence of the divergence of enzymes in the class of two-protein component aromatic hydroxylases. *Biochim. Biophys. Acta* 1680, 60–66. doi: 10.1016/j.bbaexp.2004.08.003
- Tischler, D., Eulberg, D., Lakner, S., Kaschabek, S. R., van Berkel, W. J., and Schlomann, M. (2009). Identification of a novel self-sufficient styrene monooxygenase from *Rhodococcus opacus* 1CP. *J. Bacteriol.* 191, 4996–5009. doi: 10.1128/JB.00307-09
- Tischler, D., Kermer, R., Groning, J. A., Kaschabek, S. R., van Berkel, W. J., and Schlomann, M. (2010). StyA1 and StyA2B from *Rhodococcus opacus* 1CP: a multifunctional styrene monooxygenase system. *J. Bacteriol.* 192, 5220–5227. doi: 10.1128/JB.00723-10
- Torres Pazmiño, D. E., Baas, B. J., Janssen, D. B., and Fraaije, M. W. (2008a). Kinetic mechanism of phenylacetone monooxygenase from *Thermobifida fusca*. *Biochemistry* 47, 4082–4093. doi: 10.1021/bi702296k
- Torres Pazmiño, D. E., Snajdrova, R., Baas, B. J., Ghobrial, M., Mihovilovic, M. D., and Fraaije, M. W. (2008b). Self-sufficient Baeyer-Villiger monooxygenases: effective coenzyme regeneration for biooxygenation by fusion engineering. *Angew. Chem. Int. Ed Engl.* 47, 2275–2278. doi: 10.1002/anie.200704630
- Torres Pazmiño, D. E., Dudek, H. M., and Fraaije, M. W. (2010a). Baeyer-Villiger monooxygenases: recent advances and future challenges. *Curr. Opin. Chem. Biol.* 14, 138–144. doi: 10.1016/j.cbpa.2009.11.017
- Torres Pazmiño, D. E., Winkler, M., Glieder, A., and Fraaije, M. W. (2010b). Monooxygenases as biocatalysts: classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.* 146, 9–24. doi: 10.1016/j.jbiotec.2010.01.021
- Torres Pazmiño, D. E., Riebel, A., de Lange, J., Rudroff, F., Mihovilovic, M. D., and Fraaije, M. W. (2009). Efficient biooxidations catalyzed by a new generation of self-sufficient Baeyer-Villiger monooxygenases. *ChemBioChem* 10, 2595–2598. doi: 10.1002/cbic.200900480
- Torres Pazmiño, D. E., Snajdrova, R., Rial, D. V., Mihovilovic, M. D., and Fraaije, M. W. (2007). Altering the substrate specificity and enantioselectivity of phenylacetone monooxygenase by structure-inspired enzyme redesign. *Adv. Synth. Catal.* 349, 1361–1368. doi: 10.1002/adsc.200700045
- Ukaegbu, U. E., Kantz, A., Beaton, M., Gassner, G. T., and Rosenzweig, A. C. (2010). Structure and ligand binding properties of the epoxidase component of styrene monooxygenase. *Biochemistry* 49, 1678–1688. doi: 10.1021/bi901693u
- van Berkel, W. J., Kamerbeek, N. M., and Fraaije, M. W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689. doi: 10.1016/j.jbiotec.2006.03.044
- van Berkel, W., Westphal, A., Eschrich, K., Eppink, M., and de Kok, A. (1992). Substitution of Arg214 at the substrate-binding site of p-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *Eur. J. Biochem.* 210, 411–419. doi: 10.1111/j.1432-1033.1992.tb17436.x
- van Bloois, E., Dudek, H. M., Duetz, W. A., and Fraaije, M. W. (2012). A stepwise approach for the reproducible optimization of PAMO expression in *Escherichia coli* for whole-cell biocatalysis. *BMC Biotechnol.* 12:31. doi: 10.1186/1472-6750-12-31
- van Hellemond, E. W., Janssen, D. B., and Fraaije, M. W. (2007). Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl. Environ. Microbiol.* 73, 5832–5839. doi: 10.1128/AEM.02708-06
- Walsh, C. T., and Chen, Y. C. J. (1988). Enzymic Baeyer-Villiger oxidations by flavin-dependent monooxygenases. *Angew. Chem. Int. Ed Engl.* 27, 333–343. doi: 10.1002/anie.198803331
- Walton, A. Z., and Stewart, J. D. (2002). An efficient enzymatic Baeyer-Villiger oxidation by engineered *Escherichia coli* cells under non-growing conditions. *Biotechnol. Prog.* 18, 262–268. doi: 10.1021/bp1010177c
- Wang, Y., San, K.-Y., and Bennett, G. N. (2013). Improvement of NADPH bioavailability in *Escherichia coli* by replacing NAD⁺-dependent glyceraldehyde-3-phosphate dehydrogenase GapA with NADP⁺-dependent GapB from *Bacillus subtilis* and addition of NAD kinase. *J. Ind. Microbiol. Biotechnol.* 40, 1449–1460. doi: 10.1007/s10295-013-1335-x
- Weiss, M., Kesberg, A. I., Labutti, K. M., Pitluck, S., Bruce, D., Hauser, L., et al. (2013). Permanent draft genome sequence of *Comamonas testosteroni* KF-1. *Stand. Genomic Sci.* 8, 239–254. doi: 10.4056/sigs.387890
- Willets, A. (1997). Structural studies and synthetic applications of Baeyer-Villiger monooxygenases. *Trends Biotechnol.* 15, 55–62. doi: 10.1016/S0167-7799(97)84204-7
- Wu, S., Acevedo, J. P., and Reetz, M. T. (2010). Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2775–2780. doi: 10.1073/pnas.0911656107
- Yachnin, B. J., Sprules, T., Mcevoy, M. B., Lau, P. C., and Berghuis, A. M. (2012). The substrate-bound crystal structure of a Baeyer-Villiger monooxygenase exhibits a Criegee-like conformation. *J. Am. Chem. Soc.* 134, 7788–7795. doi: 10.1021/ja211876p
- Zambianchi, F., Fraaije, M. W., Carrea, G., de Gonzalo, G., Rodríguez, C., Gotor, V., et al. (2007). Titration and assignment of residues that regulate the enantioselectivity of phenylacetone monooxygenase. *Adv. Synth. Catal.* 349, 1327–1331. doi: 10.1002/adsc.200600598
- Zhai, X.-H., Ma, Y.-H., Lai, D.-Y., Zhou, S., and Chen, Z.-M. (2013). Development of a whole-cell biocatalyst with NADPH regeneration system for biosulfoxidation. *J. Ind. Microbiol. Biotechnol.* 40, 797–803. doi: 10.1007/s10295-013-1288-0
- Zhang, Z. G., Parra, L. P., and Reetz, M. T. (2012). Protein engineering of stereoselective Baeyer-Villiger monooxygenases. *Chem. Eur. J.* 18, 10160–10172. doi: 10.1002/chem.201202163
- Zhang, Z. G., Roiban, G. D., Acevedo, J. P., Polyak, I., and Reetz, M. T. (2013). A new type of stereoselectivity in Baeyer-Villiger reactions: access to E- and Z-olefins. *Adv. Synth. Catal.* 355, 99–106. doi: 10.1002/adsc.201200759

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

Received: 20 December 2013; paper pending published: 03 January 2014; accepted: 14 January 2014; published online: 06 February 2014.

Citation: Ceccoli RD, Bianchi DA and Rial DV (2014) Flavoprotein monooxygenases for oxidative biocatalysis: recombinant expression in microbial hosts and applications. *Front. Microbiol.* 5:25. doi: 10.3389/fmicb.2014.00025

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ceccoli, Bianchi and Rial. 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.



Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments

Roslyn M. Bill*

School of Life and Health Sciences, Aston University, Birmingham, UK

Edited by:

Germán Leandro Rosano, Instituto de Biología Molecular y Celular de Rosario, Argentina

Reviewed by:

Oliver Spadiut, Vienna University of Technology, Austria

Eda Celik, Hacettepe University, Turkey

Luis Passarinha, Universidade da Beira Interior, Portugal

*Correspondence:

Roslyn M. Bill, School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
e-mail: r.m.bill@aston.ac.uk

Several host systems are available for the production of recombinant proteins, ranging from *Escherichia coli* to mammalian cell-lines. This article highlights the benefits of using yeast, especially for more challenging targets such as membrane proteins. On account of the wide range of molecular, genetic, and microbiological tools available, use of the well-studied model organism, *Saccharomyces cerevisiae*, provides many opportunities to optimize the functional yields of a target protein. Despite this wealth of resources, it is surprisingly under-used. In contrast, *Pichia pastoris*, a relative new-comer as a host organism, is already becoming a popular choice, particularly because of the ease with which high biomass (and hence recombinant protein) yields can be achieved. In the last few years, advances have been made in understanding how a yeast cell responds to the stress of producing a recombinant protein and how this information can be used to identify improved host strains in order to increase functional yields. Given these advantages, and their industrial importance in the production of biopharmaceuticals, I argue that *S. cerevisiae* and *P. pastoris* should be considered at an early stage in any serious strategy to produce proteins.

Keywords: yeast, *Saccharomyces cerevisiae*, *Pichia pastoris*, recombinant protein, yield optimization, choice of expression host

RECOMBINANT PROTEIN PRODUCTION IN MICROBES: *Escherichia coli* AS THE MOST POPULAR HOST

Proteins are essential components of living organisms and have a role in virtually every cellular process: they are enzymes; form cellular scaffolds and are central to signaling, transport, and regulatory functions. To study these diverse roles, it is necessary to be able to work with sufficient quantities (typically multi-milligram) of suitably stable and functional protein samples. While some proteins can be isolated from native sources for this purpose, many cannot because they are either intrinsically unstable or are present in impractically low quantities (Bill et al., 2011). Moreover, the study of mutant or truncated forms of a given protein is often central to understanding its structure and activity; such mutants must be synthesized recombinantly.

The biotechnological breakthrough required for recombinant gene expression was first demonstrated 40 years ago in the prokaryotic microbe, *Escherichia coli* (Cohen et al., 1973) and was soon followed by the recombinant production of human somatostatin (Itakura et al., 1977) and human insulin (Goeddel et al., 1979) in *E. coli* cultures. These innovations heralded the era of the recombinant biopharmaceutical: Humulin® synthesized in *E. coli* was launched by Eli Lilly and Company in 1982 (Altman, 1982); in 1987, Novo Nordisk started the industrial production of recombinant human insulin, Novolin®, using cultures of the eukaryotic microbe, *Saccharomyces cerevisiae* (Thim et al., 1986). Today, the recombinant production of biopharmaceuticals, particularly recombinant antibodies and vaccines, is a multi-billion dollar global business (Goodman, 2009), with more than 150 having been approved by the United States Food and Drug Administration to

date (Ferrer-Miralles et al., 2009; Zhu, 2012). Approximately 20% of these biopharmaceutical proteins are produced in yeasts (the vast majority in *S. cerevisiae*), 30% in *E. coli* and 50% in mammalian cell-lines and hybridomas (Ferrer-Miralles et al., 2009; Mattanovich et al., 2012).

Research into the science of recombinant protein production is also thriving, both as an academic discipline in its own right and as a means to produce a myriad of proteins for further study (Lee et al., 2012). In 2010, it was reported that the proportion of recombinant genes expressed in *E. coli*, compared with those expressed in all hosts had remained constant, at roughly 60% per year during the 15 year period 1995–2009 (Sørensen, 2010). **Table 1** includes the corresponding data for the other commonly used host cells; it shows that the proportion of recombinant genes expressed in *E. coli* has remained high to date and that approximately half of these genes are eukaryotic. For all other hosts, the absolute numbers are much smaller, but it is notable that the proportion of recombinant genes expressed in *Pichia pastoris* has steadily increased from 1995 to date, in contrast to all other host cells (**Table 1**). Coupled with the beginnings of a decline in usage for *E. coli* over the last 8 years, this could suggest that researchers are beginning to recognize the capacity of *P. pastoris* to produce more challenging recombinant targets.

Escherichia coli stands out as the pre-eminent host cell for producing recombinant proteins in both commercial [50% of proteins; (Ferrer-Miralles et al., 2009; Mattanovich et al., 2012)] and research (>70% of proteins; **Table 1**) laboratories; it is quick and inexpensive to culture, making it ideal in many respects. However, it has been established that producing eukaryotic proteins

Table 1 | Recombinant gene expression in the most commonly used host cells.

Year	All host cells	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>P. pastoris</i>	Insect cells	Mammalian cell-lines
1980	0	0	0	0	0	0
1985	0	0	0	0	0	0
1990	12	75% (9; 4E)	8% (1)	0	17% (2)	0
1995	37	70% (26; 17E)	5% (2)	5% (2)	5% (2)	8% (3)
2000	50	70% (35; 17E)	0	4% (2)	12% (6)	12% (6)
2005	121	85% (103; 53E)	0	5% (6)	6% (7)	2% (2)
2010	172	76% (131; 67E)	0	9% (15)	5% (6)	5% (9)
2013	128	73% (94; 54E)	2% (2)	11% (14)	4% (5)	4% (5)

The proportion of recombinant genes expressed in *E. coli*, *S. cerevisiae*, *P. pastoris*, insect cells, and mammalian cell-lines was calculated according to Sørensen's (2010) methodology; briefly, the PubMed Central database was searched for entries containing "expression purification" in the title field, which returned 1,847 articles. These articles were categorized by year of publication and expression host used and were then examined manually to confirm the categorization. The table shows the percentage of articles reporting recombinant gene expression in a given host cell and year with the actual number in parentheses; for proteins produced in *E. coli* the number of recombinant proteins of eukaryotic origin (E; ranging from unicellular protozoan to human proteins) is also noted. For all other hosts, the target proteins are exclusively eukaryotic. When percentages do not total 100% in a given year, less frequently used hosts (e.g., cell-free systems and other microbes) account for the remainder.

in a prokaryotic host cell often results in inclusion body formation and/or low specific yields (Sørensen, 2010), which may be one reason for the slight decline in its more recent use (Table 1). An explanation for lower success rates with eukaryotic targets is that the rates of protein synthesis and folding are almost an order of magnitude faster in prokaryotes than they are in eukaryotes (Widmann and Christen, 2000). Furthermore, eukaryotic codons are often inefficiently expressed and authentic eukaryotic post-translational modifications cannot yet be achieved in *E. coli* (Sørensen, 2010). However, recent progress has been made in engineering defined glycosylation pathways in *E. coli* (Valderrama-Rincon et al., 2012), while the Keio collection of single-gene knockout mutants offers a route to understanding the molecular bottlenecks to high yields in this prokaryotic host (Baba et al., 2006).

In principle, the use of mammalian cell-lines should overcome the challenges of producing recombinant eukaryotic proteins in *E. coli*, especially with recent advances in stable recombinant gene expression (Bandaranayake and Almo, 2013; Kunert and Casanova, 2013). Furthermore, the authenticity of glycosylation performed by mammalian host cells is an important advantage over all other expression hosts. However, progress in the technologies that enable reproducible gene delivery and selection of stable clones continues to be slow (Bandaranayake and Almo, 2013). Moreover, specific yields from mammalian cell-lines are often low (Zhu, 2012) and Table 1 shows a declining trend in their use.

Eukaryotic microbes offer substantial advantages as host cells, despite their propensity to hyperglycosylate recombinant proteins. For example, an annotated genome sequence has been available for *S. cerevisiae* for almost two decades (Goffeau et al., 1996), an impressive range of deletion and over-expression strains are readily available for *S. cerevisiae* and the *P. pastoris* genome has been available since 2009 (De Schutter et al., 2009). Combining this wealth of molecular and genetic resources, with the fact that yeasts grow an order of magnitude more rapidly than mammalian cell-lines

means that protein production and optimization can be done quickly and efficiently in yeast (Porro et al., 2011). Table 1 shows that for *P. pastoris*, at least, there is an increasing trend in its usage suggesting that these advantages have become more widely known. This is especially notable because *P. pastoris* is a relative new-comer, only having been first developed as a host system in 1985 (Cregg et al., 1985). Less elaborate hyperglycosylation, the availability of strains with humanized glycosylation pathways (Hamilton et al., 2003, 2006) and an increasing repertoire of molecular tools (Prielhofer et al., 2013) make this yeast an excellent alternative to *S. cerevisiae*. In particular, *P. pastoris* has been used with great success to produce challenging targets such as recombinant human G protein-coupled receptors and ion channels (Hedfalk, 2013); in total 19 high resolution structures have been resolved of recombinant eukaryotic membrane proteins produced in *P. pastoris* (Hedfalk, 2013). Table 1 shows that the number of recombinant proteins produced in *S. cerevisiae* is much smaller, despite the fact that this yeast species is an important industrial host for the production of biopharmaceuticals such as hormones (e.g., insulin and human growth hormone), vaccines (against e.g., hepatitis B and human papilloma viruses), and therapeutic adjuncts (human serum albumin) (Martinez et al., 2012); this may be a consequence of the search criteria used in generating Table 1 or possibly a perception that *S. cerevisiae* is not as amenable a host cell as *P. pastoris*.

USING YEASTS TO INCREASE SUCCESS RATES IN RECOMBINANT PROTEIN PRODUCTION EXPERIMENTS

There is no universally applicable solution for the production of all recombinant proteins (Bill, 2001; Sørensen, 2010) and it is not yet possible to predict which host system is most likely to produce a given protein in high functional yields. To be effective, any protein production strategy should therefore encompass more than one host system.

Two main approaches are typically taken to design a new protein production experiment, preferably in combination with

each other: (i) optimizing the corresponding gene sequence so it is more likely to be stably expressed and (ii) minimizing the metabolic burden on the chosen host cell(s) during recombinant protein production (Bonander and Bill, 2012). The first strategy may require that a mutant protein is produced; in support of this protein engineering approach there is an extensive literature on engineering stabilized proteins (Traxlmayr and Obinger, 2012; Scott et al., 2013). Codon optimization is also possible (Oberg et al., 2011) with more recent insights suggesting how this might aid functional expression (Halliday and Mallucci, 2014). In contrast, focusing on the host cell provides an opportunity to optimize the production of the native sequence; the principles of this second approach are broadly similar for all host cells, often requiring straightforward experimentation in the initial stages, such as optimizing culture conditions and induction protocols. Successful bioprocess engineering strategies such as these have been demonstrated to increase recombinant protein yields in cultures of both *P. pastoris* (Rebner et al., 2013; Spadiut et al., 2013) and *E. coli* (Jazini and Herwig, 2013). When a “Design of Experiments” (Bora et al., 2012) approach is used in this context, the effect of multiple parameters on the functional yield of recombinant protein can be examined simultaneously (Holmes et al., 2009); this is important since each input parameter is unlikely to exert an independent effect on functional protein yield (Bora et al., 2012). Successful implementation of such an approach in yeast has been shown to increase the productivity per cell by matching the methanol feed profile to the cellular metabolism (Holmes et al., 2009). In another approach, pulsing *P. pastoris* cells with methanol revealed the potential benefit of stress in increasing productivity (Dietzsch et al., 2011).

In the last few years, significant advances have been made in this second approach by understanding how a yeast cell responds to the stress of producing a recombinant protein at a molecular level, and how this information can be used to identify improved host strains (Bonander et al., 2009; Ashe and Bill, 2011; Bawa et al., 2011; Lee et al., 2012). Since *S. cerevisiae* is particularly amenable to studying the mechanistic basis of high-yielding recombinant protein production experiments using the tools of systems and synthetic biology, its more routine use is an obvious way to produce less tractable proteins recombinantly (Drew et al., 2008). Identifying or engineering yeast strains with improved yield characteristics may either be targeted toward one particular pathway or may take a more global approach (Ashe and Bill, 2011). Examples of the targeted approach are provided by the “humanization” of the yeast glycosylation (De Pourcq et al., 2010) and sterol (Kitson et al., 2011) pathways and modifying membrane phospholipid synthesis to proliferate intracellular membranes (Guerfal et al., 2013). Studies taking a more global approach in both *S. cerevisiae* (Bonander et al., 2005; Bonander and Bill, 2009) and *P. pastoris* (Baumann et al., 2011; Rebner et al., 2013) have identified the importance of the unfolded protein response (UPR; the cellular stress response activated in response to an accumulation of unfolded or misfolded protein) and reduced translational activity in high yielding cultures. In contrast to the mammalian UPR, the simpler UPR of yeast does not lead to down-regulation of translation to reduce protein synthetic load (Patil and Walter, 2001). We have previously noted that reducing protein synthetic capacity in yeast

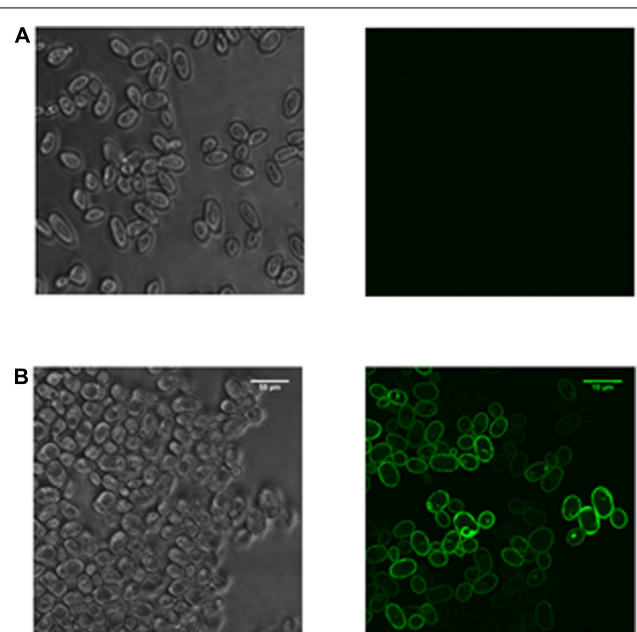


FIGURE 1 | Strain selection enables the production of a human membrane protein in *S. cerevisiae*. Yeast cells were transformed with a plasmid expressing a construct encoding a human membrane protein tagged with green fluorescent protein. Expression was driven from a constitutive promoter and cells were imaged using confocal microscopy with an upright Leica TCS SP5 system. The sample was excited with a visible argon laser at 488 nm and imaged using a 63× oil objective. The panels show confocal images with bright-field and fluorescence for (A) wild-type cells and (B) a mutant *S. cerevisiae* strain selected from a global screen for high yielding strains (Bonander et al., 2005). Only the mutant cells produced correctly localized protein.

might be an effective way to improve recombinant protein yields since this capacity is unregulated in response to unfolded protein in cells (Ashe and Bill, 2011). Such insights, which are not yet possible in higher eukaryotic systems, have been used to select specific yeast strains that can substantially improve recombinant yields compared to wild-type cells (Bonander et al., 2009; Norden et al., 2011; Figure 1). The minimal use of *S. cerevisiae* as a host shown in Table 1 is therefore at odds with this unique potential for optimization; it is possible that the increasing popularity of *P. pastoris* has detracted from the use of *S. cerevisiae*. I suggest that this undervalued host system should therefore be revisited, especially in view of its success in the production of challenging targets (Drew et al., 2008).

YEASTS AS FIRST-CHOICE HOST CELLS IN RECOMBINANT PROTEIN PRODUCTION STRATEGIES

For the majority of researchers, *E. coli* is still the first host cell to be considered in any new protein production experiment; Table 1 shows it has been consistent in its usage for over 30 years, with the beginnings of a decline in the last 8 years. Large protein production initiatives such as NYSGR¹ and OPPF-UK² use *E. coli*, insect, and mammalian cell-lines as routine

¹<http://www.nysgrc.org/psi3-cgi/index.cgi>

²<http://www.oppf.rc-harwell.ac.uk/OPPF/>

hosts; yeast is still employed on an *ad hoc* basis and the reasons for that are unclear. Since individual research teams cannot typically afford the time and investment in the full range of available host systems, I propose that a laboratory with the ability to screen for the expression of recombinant genes in *E. coli*, *S. cerevisiae*, and *P. pastoris* would be well placed to produce most target proteins; **Table 1** shows that since 2005, 85–90% of recombinant genes were expressed in these microbes. Data from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB³) show that, for soluble proteins in particular, the probability of successful expression in *E. coli* is sufficiently high to justify its premier position in **Table 1** (Ferrer-Miralles et al., 2009). Complementing this, yeasts have the capacity to produce the most challenging proteins: **Figure 1** strikingly demonstrates that the selection of a specific *S. cerevisiae* strain enables this type of bespoke optimization for a eukaryotic membrane protein tagged with green fluorescent protein that could not be produced in *E. coli*. The panels show confocal microscopy images with bright-field and fluorescence for wild-type cells and a mutant *S. cerevisiae* strain selected from a global screen for high yielding strains (Bonander et al., 2005). Only the mutant cells produced correctly localized protein. More broadly, it is notable that for eukaryotic membrane proteins, over half of all the structures deposited in the PDB obtained from recombinant material were from proteins synthesized in *P. pastoris* and *S. cerevisiae* (Bill et al., 2011). This lends further support to the use of these eukaryotic microbes alongside their prokaryotic counterpart for producing the majority of target proteins. Such a strategy also makes sense from a practical perspective, since working with bacteria and yeast require similar techniques, equipment, and approaches. Consequently, both hosts can be used within the same laboratory without the need for additional specialist investment. Yeasts should therefore be considered alongside *E. coli* at an early stage in any serious strategy to produce recombinant proteins.

ACKNOWLEDGMENTS

I thank Dr. Hans P. Sørensen, Taconic Europe A/S, Denmark, for his assistance with the analysis of recombinant host cell usage, Dr. Debasmita Sarkar and Charlotte Bland, Aston University, UK, for the images used in **Figure 1** and Dr. Kristina Hedfalk, Gothenburg University, Sweden for critical comments on the manuscript. The confocal microscope used to generate **Figure 1** is supported through the Aston Research Centre for Healthy Ageing (ARCHA).

REFERENCES

- Altman, L. K. (1982). A new insulin given approval for use in U.S. *New York Times*, October 30, 1982.
- Ashe, M. P., and Bill, R. M. (2011). Mapping the yeast host cell response to recombinant membrane protein production: relieving the biological bottlenecks. *Biotechnol. J.* 6, 707–714. doi: 10.1002/biot.201000333
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 1–11. doi: 10.1038/msb4100050
- Bandaranayake, A. D., and Almo, S. C. (2013). Recent advances in mammalian protein production. *FEBS Lett.* doi: 10.1016/j.febslet.2013.11.035
- Baumann, K., Adelantado, N., Lang, C., Mattanovich, D., and Ferrer, P. (2011). Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. *Microb. Cell Fact.* 10, 93. doi: 10.1186/1475-2859-10-93
- Bawa, Z., Bland, C. E., Bonander, N., Bora, N., Cartwright, S. P., Clare, M., et al. (2011). Understanding the yeast host cell response to recombinant membrane protein production. *Biochem. Soc. Trans.* 39, 719–723. doi: 10.1042/BST0390719
- Bill, R. M. (2001). Yeast—a panacea for the structure-function analysis of membrane proteins? *Curr. Genet.* 40, 157–171. doi: 10.1007/s002940100252
- Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., et al. (2011). Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* 29, 335–340. doi: 10.1038/nbt.1833
- Bonander, N., and Bill, R. M. (2009). Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production. *Expert Rev. Proteomics* 6, 501–505. doi: 10.1586/epr.09.65
- Bonander, N., and Bill, R. M. (2012). Optimising yeast as a host for recombinant protein production (review). *Methods Mol. Biol.* 866, 1–9. doi: 10.1007/978-1-61779-770-5_1
- Bonander, N., Darby, R. A., Grgic, L., Bora, N., Wen, J., Brogna, S., et al. (2009). Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb. Cell Fact.* 8, 10. doi: 10.1186/1475-2859-8-10
- Bonander, N., Hedfalk, K., Larsson, C., Mostad, P., Chang, C., Gustafsson, L., et al. (2005). Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci.* 14, 1729–1740. doi: 10.1110/ps.051435705
- Bora, N., Bawa, Z., Bill, R. M., and Wilks, M. D. (2012). The implementation of a design of experiments strategy to increase recombinant protein yields in yeast (review). *Methods Mol. Biol.* 866, 115–127. doi: 10.1007/978-1-61779-770-5_11
- Cohen, S. N., Chang, A. C., Boyer, H. W., and Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 70, 3240–3244. doi: 10.1073/pnas.70.11.3240
- Cregg, J. M., Barringer, K. J., Hessler, A. Y., and Madden, K. R. (1985). *Pichia pastoris* as a host system for transformations. *Mol. Cell Biol.* 5, 3376–3385.
- De Pourcq, K., De Schutter, K., and Callewaert, N. (2010). Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* 87, 1617–1631. doi: 10.1007/s00253-010-2721-1
- De Schutter, K., Lin, Y. C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., et al. (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 27, 561–566. doi: 10.1038/nbt.1544
- Dietzsch, C., Spadiut, O., and Herwig, C. (2011). A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for *Pichia pastoris*. *Microb. Cell Fact.* 10, 14. doi: 10.1186/1475-2859-10-14
- Drew, D., Newstead, S., Sonoda, Y., Kim, H., Von Heijne, G., and Iwata, S. (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.* 3, 784–798. doi: 10.1038/nprot.2008.44
- Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J. L., Vazquez, E., and Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* 8, 17. doi: 10.1186/1475-2859-8-17
- Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., et al. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U.S.A.* 76, 106–110. doi: 10.1073/pnas.76.1.106
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science* 274, 563–547. doi: 10.1126/science.274.5287.546
- Goodman, M. (2009). Market watch: sales of biologics to show robust growth through to 2013. *Nat. Rev. Drug Discov.* 8, 837. doi: 10.1038/nrd3040
- Guerfal, M., Claes, K., Knittelfelder, O., De Rycke, R., Kohlwein, S. D., and Callewaert, N. (2013). Enhanced membrane protein expression by engineering increased intracellular membrane production. *Microb. Cell Fact.* 12, 122. doi: 10.1186/1475-2859-12-122
- Halliday, M., and Mallucci, G. R. (2014). Targeting the unfolded protein response in neurodegeneration: a new approach to therapy. *Neuropharmacology* 76(Pt A), 169–174. doi: 10.1016/j.neuropharm.2013.08.034
- Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., et al. (2003). Production of complex human glycoproteins in yeast. *Science* 301, 1244–1246. doi: 10.1126/science.1088166301/5637/1244
- Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., et al. (2006). Humanization of yeast to produce complex

³<http://www.rcsb.org/pdb/home/home.do>

- terminally sialylated glycoproteins. *Science* 313, 1441–1443. doi: 10.1126/science.1130256
- Hedfalk, K. (2013). Further advances in the production of membrane proteins in *Pichia pastoris*. *Bioengineered* 4, 363–367. doi: 10.4161/bioe.23886
- Holmes, W. J., Darby, R. A., Wilks, M. D., Smith, R., and Bill, R. M. (2009). Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime. *Microb. Cell Fact.* 8, 35. doi: 10.1186/1475-2859-8-35
- Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., et al. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056–1063. doi: 10.1126/science.412251
- Jazini, M., and Herwig, C. (2013). Effects of temperature shifts and oscillations on recombinant protein production expressed in *Escherichia coli*. *Bioprocess Biosyst. Eng.* 36, 1571–1577. doi: 10.1007/s00449-013-0927-1
- Kitson, S. M., Mullen, W., Cogdell, R. J., Bill, R. M., and Fraser, N. J. (2011). GPCR production in a novel yeast strain that makes cholesterol-like sterols. *Methods* 55, 287–292. doi: 10.1016/j.jmeth.2011.09.023
- Kunert, R., and Casanova, E. (2013). Recent advances in recombinant protein production: BAC-based expression vectors, the bigger the better. *Bioengineered* 4, 258–261. doi: 10.4161/bioe.24060
- Lee, S. Y., Mattanovich, D., and Villaverde, A. (2012). Systems metabolic engineering, industrial biotechnology and microbial cell factories. *Microb. Cell Fact.* 11, 156. doi: 10.1186/1475-2859-11-156
- Martinez, J. L., Liu, L., Petranovic, D., and Nielsen, J. (2012). Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* 23, 965–971. doi: 10.1016/j.copbio.2012.03.011
- Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., and Porro, D. (2012). Recombinant protein production in yeasts. *Methods Mol. Biol.* 824, 329–358. doi: 10.1007/978-1-61779-433-9_17
- Norden, K., Agemark, M., Danielson, J. A., Alexandersson, E., Kjellbom, P., and Johanson, U. (2011). Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*. *BMC Biotechnol.* 11:47. doi: 10.1186/1472-6750-11-47
- Oberg, F., Sjöhamn, J., Conner, M. T., Bill, R. M., and Hedfalk, K. (2011). Improving recombinant eukaryotic membrane protein yields in *Pichia pastoris*: the importance of codon optimization and clone selection. *Mol. Membr. Biol.* 28, 398–411. doi: 10.3109/09687688.2011.602219
- Patil, C., and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.* 13, 349–355. doi: 10.1016/S0955-0674(00)00219-2
- Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M., et al. (2011). Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl. Microbiol. Biotechnol.* 89, 939–948. doi: 10.1007/s00253-010-3019-z
- Prielhofer, R., Maurer, M., Klein, J., Wenger, J., Kiziak, C., Gasser, B., et al. (2013). Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microb. Cell Fact.* 12, 5. doi: 10.1186/1475-2859-12-5
- Rebner, C., Graf, A. B., Valli, M., Steiger, M. G., Gasser, B., Maurer, M., et al. (2013). In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnol. J.* doi: 10.1002/biot.201300334 [Epub ahead of print].
- Scott, D. J., Kummer, L., Tremmel, D., and Pluckthun, A. (2013). Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* 17, 427–435. doi: 10.1016/j.cbpa.2013.04.002
- Sørensen, H. P. (2010). Towards universal systems for recombinant gene expression. *Microb. Cell Fact.* 9, 27. doi: 10.1186/1475-2859-9-27
- Spadiut, O., Zalai, D., Dietzsch, C., and Herwig, C. (2013). Quantitative comparison of dynamic physiological feeding profiles for recombinant protein production with *Pichia pastoris*. *Bioprocess Biosyst. Eng.* doi: 10.1007/s00449-013-1087-z [Epub ahead of print].
- Thim, L., Hansen, M. T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., et al. (1986). Secretion and processing of insulin precursors in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 83, 6766–6770. doi: 10.1073/pnas.83.18.6766
- Traxlmayr, M. W., and Obinger, C. (2012). Directed evolution of proteins for increased stability and expression using yeast display. *Arch. Biochem. Biophys.* 526, 174–180. doi: 10.1016/j.abb.2012.04.022
- Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., et al. (2012). An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.* 8, 434–436. doi: 10.1038/nchembio.921
- Widmann, M., and Christen, P. (2000). Comparison of folding rates of homologous prokaryotic and eukaryotic proteins. *J. Biol. Chem.* 275, 18619–18622. doi: 10.1074/jbc.C000156200
- Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* 30, 1158–1170. doi: 10.1016/j.biotechadv.2011.08.022

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

Received: 12 December 2013; accepted: 17 February 2014; published online: 05 March 2014.

Citation: Bill RM (2014) Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.* 5:85. doi: 10.3389/fmicb.2014.00085

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Bill. 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.



Making recombinant proteins in filamentous fungi- are we expecting too much?

Helena Nevalainen* and Robyn Peterson

Biomolecular Frontiers Research Centre, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW, Australia

Edited by:

Eduardo A. Ceccarelli, Universidad Nacional de Rosario, Argentina

Reviewed by:

Liang Shi, Pacific Northwest National Laboratory, USA

Ziyu Dai, Pacific Northwest National Laboratory, USA

*Correspondence:

Helena Nevalainen, Biomolecular Frontiers Research Centre, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia
email: helena.nevalainen@mq.edu.au

Hosts used for the production of recombinant proteins are typically high-protein secreting mutant strains that have been selected for a specific purpose, such as efficient production of cellulose-degrading enzymes. Somewhat surprisingly, sequencing of the genomes of a series of mutant strains of the cellulolytic *Trichoderma reesei*, widely used as an expression host for recombinant gene products, has shed very little light on the nature of changes that boost high-level protein secretion. While it is generally agreed and shown that protein secretion in filamentous fungi occurs mainly through the hyphal tip, there is growing evidence that secretion of proteins also takes place in sub-apical regions. Attempts to increase correct folding and thereby the yields of heterologous proteins in fungal hosts by co-expression of cellular chaperones and foldases have resulted in variable success; underlying reasons have been explored mainly at the transcriptional level. The observed physiological changes in fungal strains experiencing increasing stress through protein overexpression under strong gene promoters also reflect the challenge the host organisms are experiencing. It is evident, that as with other eukaryotes, fungal endoplasmic reticulum is a highly dynamic structure. Considering the above, there is an emerging body of work exploring the use of weaker expression promoters to avoid undue stress. Filamentous fungi have been hailed as candidates for the production of pharmaceutically relevant proteins for therapeutic use. One of the biggest challenges in terms of fungally produced heterologous gene products is their mode of glycosylation; fungi lack the functionally important terminal sialylation of the glycans that occurs in mammalian cells. Finally, exploration of the metabolic pathways and fluxes together with the development of sophisticated fermentation protocols may result in new strategies to produce recombinant proteins in filamentous fungi.

Keywords: filamentous fungi, recombinant proteins, expression, secretion, *Trichoderma reesei*

INTRODUCTION

As scavengers of recalcitrant polymers in nature, filamentous fungi such as the cellulolytic *Trichoderma reesei* are exceptionally good secretors of proteins outside the growing hyphae. Over the years, this property has been improved to the extent that current industrial production strains are well capable of secreting of the order of 100 g/L homologous proteins into the cultivation medium under optimized fermentation conditions (Cherry and Fidantsef, 2003). As these levels are far better than with any other organism, filamentous fungi hold the promise for an ultimate production host for recombinant proteins on an industrial scale. Toward this end, current research is carried out into cellular mechanisms for internal protein quality control, secretion stress, functional genomics relating to protein expression and secretion, post-translational protein modification, application of alternative expression promoters, identification of specific transcription factors and linking the fungal physiology to productivity (reviewed in Punt et al., 2002; Meyer, 2008; Lubertozzi and Keasling, 2009; Sharma et al., 2009; Fleissner and Dersch, 2010; Schuster and Schmoll, 2010; Ward, 2012). Molecular approaches such as optimizing the codon usage and expressing foreign proteins as a fusion to homologous highly secreted proteins, have become a routine

practice in fungal laboratories (Conesa et al., 2001; Nevalainen et al., 2004).

Apart from the yield, it is equally important that a given heterologous protein is produced in an active/functional form. Proteins secreted from filamentous fungi are modified in the secretory pathway by folding, proteolytic processing and addition of glycans as the main modifications. From the point of view of making functional recombinant proteins of mammalian origin in filamentous fungi, the most crucial modification is perhaps glycosylation as it may affect the functionality, serum half-life and immunogenicity of a given protein. The main mode of glycosylation in filamentous fungi is of the high-mannose type and the added sugars do not feature the functionally important terminal sialylation of the glycans that occurs in mammalian cells (Brooks, 2004; Deshpande et al., 2008; De Pourcq et al., 2010). Potential developments in this field will be discussed below.

Experimental evidence suggests that a considerable number foreign proteins expressed in filamentous fungi is lost or stuck in the secretory pathway because of incorrect processing, modification or misfolding that results in their elimination by cellular quality control mechanisms (Archer and Peberdy, 1997; Gouka et al., 1997). Genetic, transcriptomic and proteomic studies into

these mechanisms have revealed several genes and regulatory circuits active in the process (Chapman et al., 1998; Pakula et al., 2003; Al-Sheikh et al., 2004). This knowledge inspired a series of papers involving overexpression of genes encoding cellular foldases and chaperones in *Aspergillus niger*, co-expressed with a heterologous gene of interest (see Nevalainen et al., 2004). The results varied from “no effect” (e.g., *prpA* with calf chymosin; Wang and Ward, 2000) to *pdiA* with plant thaumatin resulting in fivefold yield improvement (Moralejo et al., 2001). The number of published examples is too low to make any wider conclusions but suggests that the yields of heterologous fungal proteins have a better chance for improvement in a filamentous fungal host than proteins of mammalian origin; the same holds true with heterologous proteins that have been produced without a chaperon-boost.

A typical lifecycle for an industrially applied filamentous fungus includes vegetative hyphae and asexual conidia that germinate forming new hyphae, or more rarely, production of sexual spores that undergo meiosis. Proteins are mainly secreted through the growing hyphal tip, making growth and protein secretion intimately linked and thus difficult to study separately (Wessels, 1993). The transport machinery in actively growing hyphae is required to function efficiently to ensure that cell wall material needed for growth is available in hyphal tips, a function that also links to secretion of extracellular proteins. There are many examples of growth rate-associated production of secreted proteins in filamentous fungi, e.g., production of α -amylase by *A. oryzae* (Spohr et al., 1998; Carlsen and Nielsen, 2001) and glucoamylase production by *A. niger* (Schricks et al., 1993; Withers et al., 1998; Pedersen et al., 2000). Low growth rates in chemostat cultures of *T. reesei* generally correlated with an increase in the production of the extracellular proteins. However, protein production decreased at very low growth rates when a rather large part of the carbon source consumed was probably used for maintenance requirements of the cell (Pakula et al., 2005). There is also evidence that starvation will induce enzyme secretion as a quest for the fungus to find food (Gong et al., 1979; Mach et al., 1999). These findings underline the importance of adjusting the cultivation parameters to create physiological conditions that support protein production.

The transforming DNA is typically integrated as part of the fungal genome as alternation between the hyphal growth and formation of uni- or multinuclear conidia makes it hard to maintain a population of autonomously replicating plasmids, for example, to boost the gene copy numbers and thereby product yields. Having said this, some fungi do possess autonomous replication sequence (ARS) elements (e.g., *Fusarium oxysporum*, Powell and Kistler, 1990; *A. nidulans*, Gems et al., 1991; Aleksenko and Clutterbuck, 1995, 1997; *Phanerochaete chrysosporium*, Rao and Reddy, 1984; and *Ashbya gossypii*, Schade et al., 2003) that may be used for increasing gene copy numbers to boost product yields. The recent introduction of *A. gossypii* as a potential production host for recombinant proteins (Ribeiro et al., 2010, 2013; Ribeiro, 2012) has brought these elements back to the limelight.

Filamentous fungi that dominate the scene as recombinant production hosts are the asexually reproducing *A. niger*, *A. oryzae*, and

T. reesei. Consequently, most information on heterologous protein expression has come from studies using these particular fungi as well as the genetically well-characterized *A. nidulans*. A more recent contender for production of homologous and heterologous recombinant gene products is *Chrysosporium lucknowense* developed by Dyadic International Inc (Jupiter, FL, USA). Described advantages of the *C. lucknowense* system are high transformation frequencies, production of proteins at neutral pH, low viscosity of the fermentation broth due to specific strains and short fermentation times (Punt et al., 2001; Emalfarb et al., 2003). In this paper, we will concentrate on the work carried out with *T. reesei* that can be translated to other relevant filamentous fungi with relative ease.

Despite of the considerable amounts of work dedicated to the topic and rapid development of new techniques, there have been no big (published) break-throughs over the last few decades in terms of pushing the yields of heterologous gene products to the level of homologously produced proteins. So, what are we missing and where to look next?

THE UNDERLYING EFFECTS OF RANDOM MUTAGENESIS

Filamentous fungi, have been developed as high-level enzyme producers for over thirty years, first using random mutagenesis and screening and more recently, genetic engineering (reviewed in Nevalainen et al., 2005; Ward, 2012). The history of strain improvement has been well documented for *T. reesei* (Durand et al., 1988; Peterson and Nevalainen, 2012). As classical genetic studies are not possible or not set up for the majority of the popular production hosts such as *A. niger* and *T. reesei*, even though the latter has a sexual stage (*Hypocrea jecorina*) and the former has a close relative *A. nidulans* for which the genetics is well known, the exact nature of the genetic constitution of hyper-producing strains remained unknown until genome sequencing became available. For example, sequencing of the genomes of a series of high cellulase-producing mutant strains of *T. reesei* revealed considerable changes in their genetic makeup compared to the wild-type QM6a (Martinez et al., 2008; Seidl et al., 2008; Le Crom et al., 2009; Vitikainen et al., 2010; reviewed in Peterson and Nevalainen, 2012; and Kubicek, 2013) and explained some of the observed phenotypic and metabolic characteristics; even so, the basis for drastically improved protein secretion remains unresolved.

Some of the high-protein producing strains such as *T. reesei* RutC-30 (Montenecourt and Eveleigh, 1979) are routinely used as expression hosts for homologous and heterologous gene products (reviewed in Mäntylä et al., 1998; Peterson and Nevalainen, 2012). One proposed foundation for efficient protein (cellulase) synthesis and secretion in this strain is an increased content of endoplasmic reticulum (ER), which provides more volumetric space for the synthesis of secreted proteins (Ghosh et al., 1982). This finding ultimately lead to the concept of “freeing up” space in the secretory pathway by deleting the genes encoding the major secreted proteins, Cellobiohydrolase I (CBHI/Cel7A), Cellobiohydrolase II (CBHII/Cel6A), and Endoglucanase I and II (EGI/Cel7B and EGII/Cel5A) in the case of *T. reesei* resulting in strains missing these genes in different combinations (Seiboth et al., 1992; Karhunen et al., 1993; Suominen et al., 1993; Wang et al., 2004;

Rahman et al., 2009). In theory, eliminating the CBHI protein from the secretory pathway should free up about 60% of the capacity (Nummi et al., 1983; Harkki et al., 1991). While the operation has been moderately successful in some cases, in some others deletion of the gene encoding a major secreted protein has made no difference to the yield of a secreted recombinant protein (Miettinen-Oinonen et al., 1997). Deletion of the *cbh1* gene has been executed by targeted replacement of the endogenous *cbh1* locus with the gene of interest (e.g., Karhunen et al., 1993; Joutsjoki, 1994; Saarelainen et al., 1997; Miettinen-Oinonen and Suominen, 2002) with the presumed advantage of expressing the desired gene from a locus with high transcription efficiency. As an example, targeted replacement of the *cbh1* locus of the high cellulolytic strain VTT-D-79125 with the endogenous *egl1*, followed by targeted replacement of the *cbh2* locus with *egl2* resulted in a four-fold increase in EGI activity (Miettinen-Oinonen and Suominen, 2002). There is a lot less published information available for *Aspergillus*, perhaps because of high commercial sensitivity.

It seems evident that identifying individual genes and changes in the genomes will not provide an answer to the pending question of secretion supremacy. More likely, the answer will hide in complex interactions between relevant genes and proteins and their regulation. Also, maybe high cellulase-secreting *T. reesei* strains produced by random mutagenesis and screening have been already “conditioned” for cellulase production and secretion so that introducing a protein of a different nature to be made and secreted in high yields may not be as straightforward as it looks. Considering the above, it might be worthwhile to start again by introducing the gene of interest first and applying random mutagenesis and screening afterward to boost the production levels of the desired protein by mutations “matching” the requirements of this particular protein. Automated high-throughput screening programs will make screening of hundreds of thousands of mutants feasible on a case-by-case basis. The principle of “transformation first, screening second,” has been introduced using *Ashbya*. Random mutagenesis by ethyl methane sulfonate was carried out on *A. gossypii* transformants harboring the *T. reesei egl1* gene, resulting in a global increase in protein secretion and a twofold to threefold increase in extracellular EGI activity (Ribeiro et al., 2013). Banking on random mutagenesis involving the entire genome to achieve a concerted effect to enhance synthesis and secretion of a gene product of interest, made in a transformant host seems attractive. For example, we do not have (as yet) information of all the genes involved in the process of efficient secretion of a gene product and even if we did, the current targeted approach through gene transformation and inactivation together with the requirement for marker recycling would propose a huge effort to deal with hundreds or so of genes probably involved with the process. Doing this by applying the methods of synthetic biology that allows re-engineering of entire pathways seems possible in the not so far future.

ALTERNATIVE PROMOTERS AND TRANSCRIPTION FACTORS

The strong wild-type *cbh1* promoter encoding the major cellulase (CBHI/Cel7A) in *T. reesei* is the “default” promoter for recombinant gene expression (e.g., Harkki et al., 1991; Nyssönen et al.,

1993; Paloheimo et al., 1993; de Faria et al., 2002; Nykänen et al., 2002; Haakana et al., 2004; Nevalainen et al., 2005). While good yields have been obtained using this promoter, it has also turned out that the expression levels, especially those of heterologous proteins, may cause conformational stress to the production organism (Collén et al., 2005; Godlewski et al., 2009; Nykänen et al., in preparation). With a view of feeding a recombinant protein through the secretion pathway in a more uniform manner, some other, mainly constitutive promoters functional on glucose have been explored.

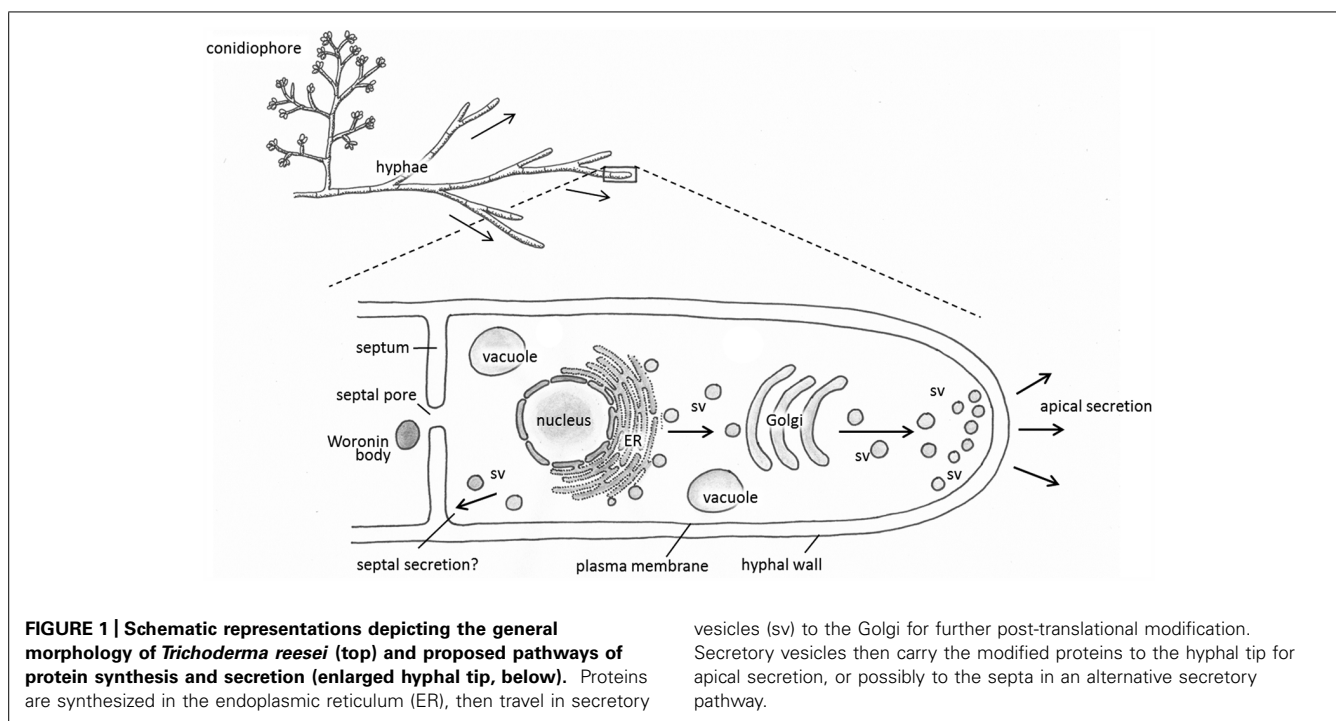
Isolation of *T. reesei* promoters that function on glucose has been described by Nakari et al. (1993; e.g., *tef1* encoding transcription elongation factor 1, and *hfb1* encoding hydrophobin) and Curach et al. (2004; *hex1*); however, there seems to be no published information on the use of these promoters for the expression of recombinant proteins. The *tef1* and *hfb1* promoters were isolated by a cDNA approach while the *hex1* promoter sequence was captured by chromosome walking, based on amino acid sequences from the HEX1 protein identified as one of the major proteins on a secretome of *T. reesei* grown on glucose (Lim et al., 2001). These different approaches introduce proteomic analysis as a tool for discovering and identifying promoters of highly expressed genes though identification of abundant proteins produced under defined conditions. In a recent study, Li et al. (2012) carried out transcriptional RT-qPCR profiling of 13 genes that were part of glucose metabolism in *T. reesei* QM9414 (Mandels et al., 1971). The promoters of *pdh* (pyruvate decarboxylase), *eno* (enolase), *gpd* (glyceraldehyde-3-phosphate dehydrogenase), *tpi* (triose phosphate isomerase), *pda* (pyruvate dehydrogenase), and *kdh* (ketoglutarate dehydrogenase) genes were singled out and proposed as candidates for constitutive expression of recombinant proteins. The *pdh* and *eno* promoters were further used for recombinant expression of the homologous *T. reesei xyn2* gene resulting in the production of 1.61 and 1.52 g/L of xylanase 2 respectively on glucose-containing medium (Li et al., 2012). The result can be considered promising as about similar amounts of xylanase 2 is expressed under its own promoter but on a cellulose-containing medium used for induction of the *xyn2* promoter.

There is continuing interest in studies into transcription factors participating in gene expression (Kiiskinen et al., 2004; Coradetti et al., 2012). While this line of research may help modulating regulation of gene expression and finding suitable and flexible cultivation conditions tailored for the gene promoter in play, it may not solve the problem of loss of the gene product during secretion.

On the note of transcription factors, it has been shown that the presence of multiple copies of one promoter can lead to the depletion of specific transcription factors for that promoter (Verdoes et al., 1994; Margolles-Clark et al., 1996). This situation may be avoided by expression of the gene of interest simultaneously under multiple different promoters inducible under the same conditions but only partly sharing the regulatory factors (Te'o and Nevalainen, 2008; Miyauchi et al., 2013).

TRACKING PROTEIN SECRETION

It has been well established that the majority of secreted proteins including heterologous gene products are secreted through the



growing hyphal tip (Wessels, 1993; Kiep et al., 2008; **Figure 1**). This default pathway is effective, for example, in the secretion of the glucoamylase enzyme in *A. niger* (Wösten et al., 1991) and cellobiohydrolase I in *T. reesei*. The ability of *T. reesei* hyphae to synthesize and secrete tens of grams and more of CBHI per liter of the cultivation medium could perhaps be explained further by assuming that there are supplementary mechanisms operating in the hyphae in addition to secretion via hyphal apices. Indeed, it has been shown that the CBHI enzyme is secreted also from the more mature parts of hyphae (Nykänen, 2002). In support of this view, *T. reesei* EGI and some heterologous enzymes such as *Hormoconis resinae* glucoamylase P and calf chymosin also occurred from mature parts of the *T. reesei* hyphae (Sprey, 1988; Nykänen, 2002). Contrary to these observations, secretion of the heterologous barley cysteine proteinase EPB seemed to occur solely at the hyphal tip thereby following the default pathway in *T. reesei* (Nykänen et al., 1997). These studies imply that there are spatial restrictions in secretion of foreign proteins that may be protein-dependent. Some factors contributing to this include subcellular localisation of the specific mRNA, information printed in the amino acid sequence of a protein and protein glycosylation.

A recent study with EGFP-fused alpha amylase in *A. oryzae* showed constitutive exocytosis also takes place at septa in addition to hyphal tips (Hayakawa et al., 2011). The fusion protein accumulated was shown to rapidly accumulate in the septal periplasm in a process that involved fusion of the secretory vesicles with the septal plasma membrane. Unlike exocytosis through hyphal tips, the process required microtubules but not F-actin whereas secretion through the tips requires both. Exocytosis toward septa may thus provide an interesting alternative to improve the production of secretory enzymes using filamentous fungi considering that in

the industrially exploited species of *Aspergillus* and *Trichoderma*, there are far more septa than hyphal tips.

Recent confocal microscopy and ultrastructural studies into protein secreting *T. reesei* hyphae demonstrate a progressively changing spatial organization of the ER in response to secretion stress (Nykänen et al., in preparation). It has also provided ultrastructural evidence to support information obtained from genome sequencing. For example, the highly cellulolytic *T. reesei* RutC-30 mutant has identified deletions or mutation in genes encoding proteins associated with vesicle trafficking, vacuolar sorting and Golgi associated vacuolar ATPases (Le Crom et al., 2009) which play a role in protein secretion. These types of observations should be taken into account when choosing an expression host with an assessment of the level and type of secretion stress the strain is under before expression of a recombinant protein (Kautto et al., 2013).

To add to the importance of exploring the physiology of the expression host and protein yields comes from the work with *A. oryzae* where disruption of the vacuolar protein sorting receptor gene (*Aovps10*) enhanced production and secretion of both bovine chymosin and human lysozyme by 3- and 2.2-fold respectively (Yoon et al., 2010).

A recently described phenomenon contributing to protein externalization from the fungal hyphae is the “pulsing” mode of secretion noted for the highly expressed CBHI in the high cellulase-producing mutant strain *T. reesei* Rut-C30 (Godlewski et al., 2009). The pulsing may reflect physiological adjustment of the hyphae to the protein overload through membrane recycling and reorganization of the ER subdomains (Godlewski et al., 2009). This view now has support from studies by Nykänen et al. (in preparation) who have described the ER as a highly dynamic organelle of which the subdomains undergo structural changes according to

the protein load in the ER. The pulsing was less evident when the heterologous bacterial enzyme Xylanase B (XynB) was expressed in *T. reesei*; instead, the heterologous protein seemed to continuously accumulate in the hyphae (Godlewski et al., 2009).

The sporadic nature of the studies into visualization and localization of protein accumulation in the fungal hyphae makes drawing broader conclusions hard. However, there are clear indications that, once again, homologous and heterologous proteins are treated differently. According to current practice, the majority of heterologous recombinant proteins are produced as a fusion to an endogenous highly secreted protein such as the main cellobiohydrolase CBHI in *T. reesei*, assumed to function as an aid in the synthesis and secretion of a recombinant protein. One of the less noted, proposed functions of the endogenous fusion protein may be stopping the non-native protein from sticking to the cell wall (Nykänen, 2002).

CONTAINED PROTEIN PRODUCTION

Proteins traveling through the fungal secretory pathway are relatively exposed to their environment. There have been some attempts toward “contained” production of secreted proteins in fungi. Naturally occurring plant protein bodies, derived from vacuoles or ER, represent a stable form of protein accumulation for nutrient storage in seeds. The ability of the maize storage protein Zera to induce the formation of protein bodies has been utilized for recombinant protein production in *T. reesei* (Torrent et al., 2009). A green fluorescent protein (GFP)-Zera peptide fusion accumulated in induced protein body-like organelles in the hyphae, protecting the recombinant fusion protein from cellular degradation whilst also protecting host cell viability. In addition, downstream isolation of the fusion protein was enhanced by the high density of the Zera-induced protein bodies.

Expression of a GFP fusion with the homologous hydrophobin I (HFBI) of *T. reesei* also induced the formation of protein bodies when targeted to the ER using the HDEL ER-retention signal (Mustalahti et al., 2013). A dual benefit was achieved; large ER-derived protein bodies containing soluble fusion protein accumulated in the hyphae, and the hydrophobicity of HFBI enabled effective downstream purification via a simple aqueous two-phase liquid partitioning system (ATPS). Purification by ATPS can be carried out relatively cheaply and simply even in large scale systems, as demonstrated by the successful purification of EGICore-HFBI fusion protein from a 1200 l fermenter culture of a recombinant *T. reesei* strain (Collén et al., 2002; Selber et al., 2004).

DECORATING PROTEINS WITH SUGARS

Glycosylation is one of the most common post-translational modifications and nearly 50% of all known proteins in eukaryotes are glycosylated (Apweiler et al., 1999). Glycans are synthesized by the coordinated action of glycosyltransferases, glycosidases, and other glycan processing enzymes. *N*-linked glycans have a role in many physiological and pathological events including protein and cell trafficking, immunogenicity, cell growth and adhesion, differentiation, tumor invasion, transmembrane signaling and host-pathogen interactions (Zhao et al., 2008).

Studies into the effect of glycosylation on the secretion, stability and activity and binding of secreted proteins have been carried

out mainly with non-recombinant *T. reesei* cellulases. The various approaches and outcomes have been summarized in a recent review by Beckham et al. (2012). In the main, it has turned out that glycosylation has shown to have an effect on enzyme stability (aggregation and thermal stability) and activity. Contribution of glycosylation on protein secretion in filamentous fungi is less well established as some secreted native and recombinant proteins seem not to carry any glycan structures (Ülker and Sprey, 1990; Kurzatkowski et al., 1996; Paloheimo et al., 2003).

It has also been established that different fungi and fungal strains *N*-glycosylate proteins differently (Nevalainen et al., 1998) and that composition of the cultivation medium affects the glycosylation pattern (Stals et al., 2004). Despite of these leads, there are only a handful of papers comparing production of heterologous proteins in different host strains of the same fungal species (e.g., Bergquist et al., 2004) and different growth conditions, not to mention detailed analysis of the glycan structures attached on recombinant proteins. One such paper is that of Miyauchi et al. (2013) where the authors showed that the recombinant Xylanase B protein (from a thermophilic bacterium), produced in *T. reesei* featured multiple forms of the enzyme, decorated with various *N*- and *O*-glycans as assessed by mass spectrometry. One of the *O*-glycans was identified as hexuronic acid, which has not been described previously in the glycosylation patterns of *T. reesei*.

On the note that glycans have an effect on the activity of a protein, one of the problems standing in the way for filamentous fungi becoming effective producers of pharmaceutical proteins targeted for human consumption is the fungal oligo-mannose type glycosylation. While still of high-mannose type *N*-glycosylation patterns, filamentous fungi are far more conservative than yeast that has the tendency to hyperglycosylate proteins (Deshpande et al., 2008). However, they still lack the terminal sialic acid residues, characteristic of human glycosylation and important for defining the function of the glycan. These shortcomings have been addressed in a handful of *in vivo* studies toward modification of the protein glycosylation pathway in filamentous fungi, mainly *A. nidulans*, *A. niger*, *A. oryzae* (Kasajima et al., 2006; Kainz et al., 2008) and *T. reesei* (Maras et al., 1999; Zhong et al., 2011). Compared to yeast, activity in this field is very low.

Alternatively to humanizing the fungal glycosylation pathway, filamentous fungi can be considered a potential option for producing selected glycan-modifying enzymes for *in vitro* modification of glycans attached to recombinant proteins. The “strip and tease” approach is currently being developed in our laboratory (unpublished work). While these scenarios are not fully developed yet, the exceptional protein secretion capacity of filamentous fungi warrants investigation into modification of protein glycosylation in order to make functional therapeutic proteins in these organisms in an economically sustainable manner. Work in yeast provides a guide for these efforts.

CURRENT PRODUCT LEVELS OF RECOMBINANT PROTEINS AND CLOSE COMPETITORS

Recent developments in mammalian cell culture have raised the production levels of heterologous pharmaceutically important proteins to grams per liter (Aldridge, 2006) reaching the yield of 26 g/L of a monoclonal antibody in an industrial setting (Jarvis,

2008). In comparison, published yields for antibodies produced in filamentous fungi are of the order of 0.15 g/L of a CBHI-Fab fusion antibody for *T. reesei* (Nyyssönen et al., 1993) and 0.9 g/L of Trastaztmab for *A. niger* (Ward et al., 2004). While filamentous fungi may have lost this battle, they still hold a good position as producers of various industrial enzymes as the mammalian systems are far too expensive and impractical for the bulk production of low-cost recombinant proteins.

Another near competitor is the methylotrophic yeast *Pichia pastoris* with production levels of 1.6 g/L for a monoclonal antibody in a glycoengineered strain (Ye et al., 2011). As *Pichia* is also capable of efficient secretion and the system is commercially available¹, it offers a competitive edge when choosing a host for the production of heterologous proteins. Overall, it seems ever so important that the matching of the intended product and the production host is done with care and using all available information as the yields for different types of recombinant proteins can vary considerably even within the same host organism (see the review of Demain and Vaishnav, 2009).

THE EFFECT OF FERMENTATION CONDITIONS

A typical workflow aiming at improvement of the product yield has a front-end involving optimization of the gene encoding the product of interest, the expression vector and the transformation method, and adding or removing tags for protein targeting and purification. Nature of these adjustments depends on the chosen production host. The next step is screening of the transformant strains for the desired product and making the recombinant product on a laboratory scale. This involves establishment of the cultivation parameters. It should be noted that different recombinant strains may require slightly different cultivation conditions and a protocol that differs from the transformation host (Sun, 2013). This aspect is often not studied in detail especially at the stage when a high number of genetically modified strains is screened. Therefore, some potentially good producers maybe lost in the “standardized” screening process.

Developing the cultivation conditions including the growth medium is fundamental for the improvement of the yields of recombinant proteins. This has been well established and documented with both mammalian cell cultures (e.g., Aldridge, 2006) and the *Pichia* system (e.g., Gonçalves et al., 2013) where the improvements have been impressive. There is also a wealth of published work on the development of protocols and models for growing fungi in submerged cultures, addressing the nature of the carbon and nitrogen sources, carbon:nitrogen ratio, agitation, aeration, nutrient depletion and feeding, to mention some (reviewed in Workman et al., 2013). Cultivation by solid fermentation and mass screening of fungal strains have also been described in the literature (reviewed e.g., in El-Enshasy, 2006).

Optimization of the production conditions for industrial scale fermentations are typically carried out in-house and patented.

METABOLIC ENGINEERING AND SYNTHETIC BIOLOGY

Metabolic engineering and the application of synthetic biology are heavily reliant on the breadth and depth of large-scale

(“omics”) information available for the organism of interest. This information comes from genome sequences, studies into metabolic pathways and fluxes, transcriptomic and proteomic data and bioinformatic modeling. Amongst these, quantification of metabolic fluxes is of utmost importance to understand biological networks for cellular regulation, and identify bottlenecks in product formation. The potential targets in filamentous fungi may include production of hydrolytic enzymes, organic acids, biofuels and chemicals. For example, the information gathered from the analysis of metabolic fluxes under production conditions will point out locations where the flow of, e.g., carbon in the cell is not going effectively toward the intended product thus proposing a bottleneck. This situation may then be remedied by genetic engineering by redirecting or enhancing the flow resulting in an improved product yield.

Genome sequences are available for a good amount of filamentous fungi including the industrially relevant *A. niger*² and *A. oryzae*³, *P. chrysosporium*⁴ and *T. reesei* (wild-type and mutant strains)⁵. Application of targeted metabolic engineering to improve recombinant protein production has been carried out in bacterial and yeast systems and *Aspergillus* (reviewed by Melzer et al., 2009; Boghigian et al., 2010; Matsuoka and Shimizu, 2010). However, although metabolic engineering has been successfully employed to improve homologous cellulase production in *T. reesei* (Kubicek et al., 2009), little published information is available on its application to recombinant *T. reesei* strains to date.

A high throughput gene deletion in *T. reesei* has recently been developed (Schuster et al., 2012). A series of selection markers were used to provide a primer database for gene deletion, and vector construction was carried out by yeast-mediated recombination. Transformation of the vector into a *T. reesei* strain deficient in non-homologous end joining (NHEJ) was followed by crossing of mutants with sexually competent strains to remove the NHEJ-defect.

Synthetic biology takes matters further with the goal of designing and constructing biological devices and systems. This would involve rearranging and rebuilding large DNA constructs with overlapping DNA fragments and their *in vivo* recombination (Gibson et al., 2009). The strategy has been shown to work well with bacteria and yeast but there are no published reports concerning filamentous fungi as yet. One of the future scenarios may feature mini-cell factories where only the essential functions for making a particular gene product are contained.

CONCLUSION

Filamentous fungi offer enormous potential for efficient and large scale production of recombinant gene products. Importantly, protein secretion provides a platform for the eukaryotic style post-translational modification of proteins. Fungi are cheap to cultivate and down-stream processing is made easy with no need to break cells open for product recovery. In order to capitalize on

¹<http://www.lifetechnologies.com>

²<http://genome.jgi-psf.org/Aspni.home.html>

³<http://www.bio.nite.go.jp/dogan/project/view/AO>

⁴<http://genome.jgi-psf.org/whiterot/whiterot1.home.html>

⁵<http://genome.jgi.doe.gov/genome-projects/pages/projects.jsf?searchText=Trichoderma+reesei>

fungi as recombinant production hosts, research is now directed to revealing the cellular mechanisms for internal protein quality control, secretion stress, functional genomics of protein expression and secretion, protein modification and linking the physiology to productivity. New directions are expected to emerge from overlaying the “omics” data and the rapidly developing technologies of metabolic engineering and synthetic biology. Our long held expectations of filamentous fungi as high-level producers of a wide range of recombinant proteins may well be drawing closer to full realization. Or, perhaps, the future may see a change in the expectations and a focusing toward specific areas of recombinant expression for which the fungal system is particularly adept, such as the production of recombinant enzymes including therapeutic microbial enzymes.

REFERENCES

- Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473, 4–8. doi: 10.1016/S0304-4165(99)00165-8
- Aldridge, S. (2006). Downstream processing needs a boost. *Genet. Eng. News* 26, 1–51.
- Aleksenko, A. Y., and Clutterbuck, A. J. (1995). Recombinational stability of replicating plasmids in *Aspergillus nidulans* during transformation, vegetative growth and sexual reproduction. *Curr. Genet.* 28, 87–93. doi: 10.1007/BF00311886
- Aleksenko, A., and Clutterbuck, A. J. (1997). Autonomous plasmid replication in *Aspergillus nidulans*: AMA1 and MATE elements. *Fungal Genet. Biol.* 21, 373–387. doi: 10.1006/fgbi.1997.0980
- Al-Sheikh, H., Watson, A. J., Lacey, G. A., Punt, P. J., MacKenzie, D. A., Jeenes, D. J., et al. (2004). Endoplasmic reticulum stress leads to the selective transcriptional downregulation of the glucoamylase gene in *Aspergillus niger*. *Mol. Microbiol.* 53, 1731–1742. doi: 10.1111/j.1365-2958.2004.04236.x
- Archer, D. B., and Peberdy, J. F. (1997). The molecular biology of secreted enzyme production by fungi. *Crit. Rev. Biotechnol.* 17, 273–306. doi: 10.3109/07388559709146616
- Beckham, G. T., Dai, Z., Matthews, J. F., Momany, M., Payne, C., Adney, W. S., et al. (2012). Harnessing glycosylation to improve cellulose activity. *Curr. Opin. Biotechnol.* 23, 338–345. doi: 10.1016/j.copbio.2011.11.030
- Bergquist, P. L., Te'o, V. S. J., Gibbs, M. D., Curach, N. C., and Nevalainen, K. M. H. (2004). Recombinant enzymes from thermophilic microorganisms expressed in fungal hosts. *Biochem. Soc. Transact.* 32, 293–297. doi: 10.1042/BST0320293
- Boghigian, B. A., Seth, G., Kiss, R., and Pfeifer, B. A. (2010). Metabolic flux analysis and pharmaceutical production. *Metab. Eng.* 12, 81–95. doi: 10.1016/j.ymben.2009.10.004
- Brooks, S. A. (2004). Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. *Mol. Biotechnol.* 28, 241–255. doi: 10.1385/MB:28:3:241
- Carlsen, M., and Nielsen, J. (2001). Influence of carbon source on alpha-amylase production by *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 57, 346–349. doi: 10.1007/s002530100772
- Chapman, R., Sidrauski, C., and Walter, P. (1998). Intracellular signaling from the endoplasmic reticulum to the nucleus. *Annu. Rev. Cell Dev. Biol.* 14, 459–485. doi: 10.1146/annurev.cellbio.14.1.459
- Cherry, J. R., and Fidantsef, A. L. (2003). Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* 14, 438–443. doi: 10.1016/S0958-1669(03)00099-5
- Collén, A., Persson, J., Linder, M., Nakari-Setälä, T., Penttilä, M., Tjerneld, F., et al. (2002). A novel two-step extraction method with detergent/polymer systems for primary recovery of the fusion protein endoglucanase I-hydrophobin I. *Biochim. Biophys. Acta* 1569, 139–150. doi: 10.1016/S0304-4165(01)00244-6
- Collén, A., Saloheimo, M., Bailey, M., Penttilä, M., and Pakula, T. M. (2005). Protein production and induction of the unfolded protein response in *Trichoderma reesei* strain Rut-C30 and its transformant expressing endoglucanase I with a hydrophobic tag. *Biotechnol. Bioeng.* 89, 335–344. doi: 10.1002/bit.20350
- Conesa, A., Punt, P. J., van Luijk, N., and van den Hondel, C. A. (2001). The secretion pathway in filamentous fungi: a biotechnological view. *Fungal Genet. Biol.* 33, 155–171. doi: 10.1006/fgbi.2001.1276
- Coradetti, S. T., Craig, J. P., Xiong, Y., Shock, T., Tian, C., and Glass, N. L. (2012). Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7397–7402. doi: 10.1073/pnas.1200785109
- Curach, N. C., Te'o, V. S., Gibbs, M. D., Bergquist, P. L., and Nevalainen, K. M. H. (2004). Isolation, characterization and expression of the hex1 gene from *Trichoderma reesei*. *Gene* 331, 133–140. doi: 10.1016/j.gene.2004.02.007
- de Faria, F. P., Te'o, V. S., Bergquist, P. L., Azevedo, M. O., and Nevalainen, K. M. H. (2002). Expression and processing of a major xylanase (XYN2) from the thermophilic fungus *Humicola grisea* var. *thermoidea* in *Trichoderma reesei*. *Lett. Appl. Microbiol.* 34, 119–123. doi: 10.1046/j.1472-765x.2002.01057.x
- Demain, A. L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotech. Adv.* 27, 297–306. doi: 10.1016/j.biotechadv.2009.01.008
- De Pourcq, K., De Schutter, K., and Callewaert, N. (2010). Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* 87, 1617–1631. doi: 10.1007/s00253-010-2721-1
- Deshpande, N., Wilkins, M. R., Packer, N., and Nevalainen, H. (2008). Protein glycosylation pathways in filamentous fungi. *Glycobiology* 18, 626–637. doi: 10.1093/glycob/cwn044
- Durand, H., Baron, M., Calmels, T., and Tiraby, G. (1988). “Classical and molecular genetics applied to *Trichoderma reesei* for the selection of improved cellulolytic industrial strains,” in *FEMS Symposium no. 43: Biochemistry and Genetics of Cellulose Degradation*, eds J.-P. Aubert, P. Beguin, and J. Millet (London: Academic Press), 135–151.
- El-Enshasy, H. A. (2006). “Filamentous fungal cultures-process characteristics, products and applications,” in *Bioprocessing for Value-Added Products from Renewable Resources*, 1st Edn, ed. S.-T. Yang (London, UK: Elsevier), 225–260.
- Emalfarb, M. A., Burlingame, R. P., Olson, P. T., Sinitsyn, A. P., Parriche, M., Bousson, J. C., et al. (2003). Transformation system in the field of filamentous fungal hosts. US Patent, US 6,573,086.
- Fleissner, A., and Dersch, P. (2010). Expression and export: recombinant protein production systems for *Aspergillus*. *Appl. Microbiol. Biotechnol.* 87, 1255–1270. doi: 10.1007/s00253-010-2672-6
- Gems, D., Johnstone, I. L., and Clutterbuck, A. J. (1991). An autonomously replicating plasmid transforms *Aspergillus nidulans* at high frequency. *Gene* 98, 61–67. doi: 10.1016/0378-1119(91)90104-J
- Ghosh, A., Al-Rabiai, S., Ghosh, B. K., Trimiño-Vazquez, H., Eveleigh, D. E., and Montencourt, B. S. (1982). Increased endoplasmic reticulum content of a mutant of *Trichoderma reesei* (RUT-C30) in relation to cellulase synthesis. *Enzyme Microb. Technol.* 4, 110–113. doi: 10.1016/0141-0229(82)90093-X
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Godlewski, M., Kautto, L., and Nevalainen, H. (2009). “The elusive pulse in protein production in *Trichoderma reesei*,” in *Poster Light in Life Sciences (LILS) Conference*, Melbourne, Australia, November 24–27, Poster Session 1, 10.
- Gonçalves, A. M., Pedro, A. Q., Maia, C., Sousa, F., Queiroz, J. A., and Passarinha, L. A. (2013). *Pichia pastoris*: a recombinant microfactory for antibodies and human membrane proteins. *J. Microbiol. Biotechnol.* 23, 587–601. doi: 10.4014/jmb.1210.10063
- Gong, C.-S., Ladisch, M. R., and Tsao, G. T. (1979). Biosynthesis, purification and mode of action of cellulases of *Trichoderma reesei*. *Adv. Chem. Ser.* 181, 261–288. doi: 10.1021/ba-1979-0181.ch013
- Gouka, R. J., Punt, P. J., and van den Hondel, C. A. (1997). Efficient production of secreted proteins by *Aspergillus*: progress, limitations and prospects. *Appl. Microbiol. Biotechnol.* 47, 1–11. doi: 10.1007/s002530050880
- Haakana, H., Miettinen-Oinonen, A., Joutsjoki, V., Mäntylä, A., Suominen, P., and Vehmaanperä, J. (2004). Cloning of cellulase genes from *Melanocarpus albomyces* and their efficient expression in *Trichoderma reesei*. *Enzyme Microb. Technol.* 34, 159–167. doi: 10.1016/j.enzmictec.2003.10.009
- Harkki, A., Mäntylä, A., Penttilä, M., Mutilainen, S., Bühler, R., Suominen, P., et al. (1991). Genetic engineering of *Trichoderma* to produce strains with novel cellulase profiles. *Enzyme Microb. Technol.* 13, 227–233. doi: 10.1016/0141-0229(91)90133-U

- Hayakawa, Y., Ishikawa, E., Shoji, J. Y., Nakano, H., and Kitamoto, K. (2011). Septum-directed secretion in the filamentous fungus *Aspergillus oryzae*. *Mol. Microbiol.* 81, 40–55. doi: 10.1111/j.1365-2958.2011.07700.x
- Jarvis, L. M. (2008). A technology bet. DSM's pharma product unit leverages its biotech strength to survive a tough environment. *Chem. Eng. News* 86, 30–31. doi: 10.1021/cen-v086n029.p030
- Joutsjoki, V. V. (1994). Construction by one-step gene replacement of *Trichoderma reesei* strains that produce the glucoamylase P of *Hormoconis resinae*. *Curr. Genet.* 26, 422–429. doi: 10.1007/BF00309929
- Kainz, E., Gallmetzer, A., Hatzl, C., Nett, J. H., Li, H., Schinko, T., et al. (2008). N-glycan modification in *Aspergillus* species. *Appl. Environ. Microbiol.* 74, 1076–1086. doi: 10.1128/AEM.01058-07
- Karhunen, T., Mäntylä, A., Nevalainen, K. M. H., and Suominen, P. L. (1993). High frequency one-step gene replacement in *Trichoderma reesei*. I. Endoglucanase I overproduction. *Mol. Gen. Genet.* 241, 515–522. doi: 10.1007/BF00279893
- Kasajima, Y., Yamaguchi, M., Hirai, N., Ohmachi, T., and Yoshida, T. (2006). *In vivo* expression of UDP-N-acetylglucosamine: Alpha-3-d-mannoside beta-1,2-N-acetylglucosaminyltransferase I (*GnT-I*) in *Aspergillus oryzae* and effects on the sugar chain of alpha-amylase. *Biosci. Biotechnol. Biochem.* 70, 2662–2668. doi: 10.1271/bbb.60265
- Kautto, L., Grinyer, J., Paulsen, I., Tetu, S., Pillai, A., Pardiwalla, S., et al. (2013). Stress effects caused by the expression of a mutant cellobiohydrolase I and protease inhibition in *Trichoderma reesei* Rut-C30. *N. Biotechnol.* 30, 183–191. doi: 10.1016/j.nbt.2012.07.005
- Kiep, K., Bohle, K., Göcke, Y., Krull, R., and Hempel, D. C. (2008). "Insights in growth and protein formation of *Aspergillus niger* on a molecular level," in *Poster 9th European Conference on Fungal Genetics* Edinburgh, Scotland, April 5–8, *Fungal Biotechnology, Hyphal Growth and Morphogenesis, Evolutionary and Population Biology*, Poster Session 3, PR8.16.
- Kiiskinen, L. L., Kruus, K., Bailey, M., Ylösmäki, E., Siika-Aho, M., and Saloheimo, M. (2004). Expression of *Melanocarpus albomyces* laccase in *Trichoderma reesei* and characterization of the purified enzyme. *Microbiology* 150, 3065–3074. doi: 10.1099/mic.0.27147-0
- Kubicek, C. P. (2013). Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *J. Biotechnol.* 163, 133–142. doi: 10.1016/j.jbiotec.2012.05.020
- Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M., and Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnol. Biofuels* 2:19. doi: 10.1186/1754-6834-2-19
- Kurzatowski, W., Törrönen, A., Filipek, J., Mach, R. L., Herzog, P., Sowka, S., et al. (1996). Glucose-induced secretion of *Trichoderma reesei* xylanases. *Appl. Environ. Microb.* 62, 2859–2865.
- Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J. K., Culley, D. E., Collett, J. R., et al. (2009). Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16151–16156. doi: 10.1073/pnas.0905848106
- Li, J., Wang, J., Wang, S., Xing, M., Yu, S., and Liu, G. (2012). Achieving efficient protein expression in *Trichoderma reesei* by using strong constitutive promoters. *Microb. Cell Fact.* 11:84. doi: 10.1186/1475-2859-11-84
- Lim, D., Hains, P., Walsh, B., Bergquist, P., and Nevalainen, H. (2001). Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach. *Proteomics* 1, 899–909. doi: 10.1002/1615-9861(200107)1:7<899::AID-PROT899>3.0.CO;2-#
- Lubertozzi, D., and Keasling, J. D. (2009). Developing *Aspergillus* as a host for heterologous expression. *Biotechnol. Adv.* 27, 53–75. doi: 10.1016/j.biotechadv.2008.09.001
- Mach, R. L., Peterbauer, C. K., Payer, K., Jaksits, S., Woo, S. L., Zeilinger, S., et al. (1999). Expression of two major chitinase genes of *Trichoderma atroviride* (*T. harzianum* P1) is triggered by different regulatory signals. *Appl. Environ. Microbiol.* 65, 1858–1863.
- Mandels, M., Weber, J., and Parizek, R. (1971). Enhanced cellulase production by a mutant of *Trichoderma viride*. *Appl. Microbiol.* 21, 152–154.
- Mäntylä, A., Paloheimo, M., and Suominen, P. (1998). "Industrial mutants and recombinant strains of *Trichoderma reesei*" in *Trichoderma and Gliocladium*, Vol. 2, eds G. Harman and C. Kubicek (London: Taylor and Francis), 291–309.
- Maras, M., van Die, I., Contreras, R., and van den Hondel, C. A. (1999). Filamentous fungi as production organisms for glycoproteins of bio-medical interest. *Glycoconj. J.* 16, 99–107. doi: 10.1023/A:1026436424881
- Margolles-Clark, E., Hayes, C. K., Harman, G. E., and Penttilä, M. (1996). Improved production of *Trichoderma harzianum* endochitinase by expression in *Trichoderma reesei*. *Appl. Environ. Microbiol.* 62, 2145–2151.
- Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* 26, 553–560. doi: 10.1038/nbt1403
- Matsuoka, Y., and Shimizu, K. (2010). Current status of ¹³C-metabolic flux analysis and future perspectives. *Process Biochem.* 45, 1873–1881. doi: 10.1016/j.procbio.2010.03.025
- Melzer, G., Esfandabadi, M. E., Franco-Lara, E., and Wittmann, C. (2009). Flux design: in silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Syst. Biol.* 3:120. doi: 10.1186/1752-0509-3-120
- Meyer, V. (2008). Genetic engineering of filamentous fungi – progress, obstacles and future trends. *Biotechnol. Adv.* 26, 177–185. doi: 10.1016/j.biotechadv.2007.12.001
- Miettinen-Oinonen, A., Torkkeli, T., Paloheimo, M., and Nevalainen, H. (1997). Overexpression of the *Aspergillus niger* pH 2.5 acid phosphatase gene in a heterologous host *Trichoderma reesei*. *J. Biotechnol.* 58, 13–20. doi: 10.1016/S0168-1656(97)00121-1
- Miettinen-Oinonen, A., and Suominen, P. (2002). Enhanced production of *Trichoderma reesei* endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. *Appl. Environ. Microbiol.* 68, 3956–3964. doi: 10.1128/AEM.68.8.3956-3964.2002
- Miyauchi, S., Te'o, V. S. J., Bergquist, P. L., and Nevalainen, K. M. H. (2013). Expression of a bacterial xylanase in *Trichoderma reesei* under the *egl2* and *cbh2* glycosyl hydrolase gene promoters. *N. Biotechnol.* 30, 523–530. doi: 10.1016/j.nbt.2013.02.005
- Montenecourt, B. S., and Eveleigh, D. E. (1979). "Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei*," in *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis, Advances in Chemistry Series, 181*, eds R. Brown and L. Jurasek (Washington, DC: American Chemical Society), 289–301.
- Moralejo, F. J., Watson, A. J., Jeenes, D. J., Archer, D. B., and Martín, J. F. (2001). A defined level of protein disulfide isomerase expression is required for optimal secretion of thaumatin by *Aspergillus awamori*. *Mol. Genet. Genomics* 266, 246–253. doi: 10.1007/s004380100550
- Mustalahti, E., Saloheimo, M., and Joensuu, J. J. (2013). Intracellular protein production in *Trichoderma reesei* (*Hypocrea jecorina*) with hydrophobin fusion technology. *N. Biotechnol.* 30, 262–268. doi: 10.1016/j.nbt.2011.09.006
- Nakari, T., Alatalo, E., and Penttilä, M. E. (1993). Isolation of *Trichoderma reesei* genes highly expressed on glucose-containing media: characterization of the *tef1* gene encoding translation elongation factor 1 alpha. *Gene* 136, 313–318. doi: 10.1016/0378-1119(93)90486-M
- Nevalainen, H., Harrison, M., Jardine, D., Zachara, N., Paloheimo, M., Suominen, P., et al. (1998). "Glycosylation of cellobiohydrolase I from *Trichoderma reesei*," in *Carbohydrates from Trichoderma reesei and other Microorganisms; Structures, Biochemistry, Genetics and Applications*, eds M. Claeysens, W. Nerinx, and K. Piens (Cambridge, UK: The Royal Society of Chemistry, Thomas Graham House), 335–344.
- Nevalainen, H., Penttilä, M., and Te'o, V. J. S. (2004). "Application of genetic engineering for strain development in filamentous fungi," in *Handbook of Fungal Biotechnology (Mycology Series 20)*, eds D. K. Arora, P. D. Bridge, and D. Bhatnagar (New York, NY: Marcel Dekker), 193–208.
- Nevalainen, K. M. H., Te'o, V. S. J., and Bergquist, P. L. (2005). Heterologous protein expression in filamentous fungi. *Trends Biotechnol.* 23, 468–474. doi: 10.1016/j.tibtech.2005.06.002
- Nummi, M., Niku-Paavola, M. L., Lappalainen, A., Enari, T. M., and Raunio, V. (1983). Cellobiohydrolase from *Trichoderma reesei*. *Biochem. J.* 215, 677–683.
- Nykänen, M. (2002). *Protein Secretion in Trichoderma reesei. Expression, Secretion and Maturation of Cellobiohydrolase I, Barley Cysteine Proteinase and Calf Chymosin in Rut-C30*. Ph.D. thesis, University of Jyväskylä, Jyväskylä. p.107. <https://www.jyu.fi/ajankohtaista/arkisto/2002/03/tiedote-2007-09-18-15-39-04-171749>
- Nykänen, M., Saarelainen, R., Raudaskoski, M., Nevalainen, H., and Mikkonen, A. (1997). Expression and secretion of barley cysteine Endopeptidase B and Cellobiohydrolase I in *Trichoderma reesei*. *Appl. Environ. Microbiol.* 63, 4929–4937.

- Nykänen, M. J., Raudaskoski, M., Nevalainen, H., and Mikkonen, A. (2002). Maturation of barley cysteine endopeptidase expressed in *Trichoderma reesei* is distorted by incomplete processing. *Can. J. Microbiol.* 48, 138–150. doi: 10.1139/w01-144
- Nyyssönen, E., Penttilä, M., Harkki, A., Saloheimo, A., Knowles, J. K., and Keränen, S. (1993). Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Biotechnology (N. Y.)* 11, 591–595. doi: 10.1038/nbt0593-591
- Pakula, T. M., Laxell, M., Huuskonen, A., Uusitalo, J., Saloheimo, M., and Penttilä, M. (2003). The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*. Evidence for down-regulation of genes that encode secreted proteins in the stressed cells. *J. Biol. Chem.* 278, 45011–45020. doi: 10.1074/jbc.M302372200
- Pakula, T. M., Salonen, K., Uusitalo, J., and Penttilä, M. (2005). The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus *Trichoderma reesei*. *Microbiology* 151, 135–143. doi: 10.1099/mic.0.27458-0
- Paloheimo, M., Mäntylä, A., Kallio, J., and Suominen, P. (2003). High-yield production of a bacterial xylanase in the filamentous fungus *Trichoderma reesei* requires a carrier polypeptide with an intact domain structure. *Appl. Environ. Microbiol.* 69, 7073–7082. doi: 10.1128/AEM.69.12.7073-7082.2003
- Paloheimo, M., Miettinen-Oinonen, A., Torkkeli, T., Nevalainen, H., and Suominen, P. (1993). "Enzyme production in *T. reesei* using the *cbhl* promoter," in *Proceedings of the 2nd Tricel Meeting, Majvik, Finland*, Vol. 8, eds T. Reinikainen and P. Suominen. (Fagerpaino Oy, Helsinki: Foundation for Biotechnical and Industrial Fermentation Research), 229–237. doi: 10.1007/s002530050020
- Pedersen, H., Beyer, M., and Nielsen, J. (2000). Glucoamylase production in batch, chemostat and fed-batch cultivations by an industrial strain of *Aspergillus niger*. *Appl. Microbiol. Biotechnol.* 53, 272–277. doi: 10.1099/mic.0.054031-0
- Peterson, R., and Nevalainen, H. (2012). *Trichoderma reesei* RUT-C30 – thirty years of strain improvement. *Microbiology* 158, 58–68. doi: 10.1099/mic.0.054031-0
- Powell, W. A., and Kistler, H. C. (1990). *In vivo* rearrangement of foreign DNA by *Fusarium oxysporum* produces linear self-replicating plasmids. *J. Bacteriol.* 172, 3163–3171.
- Punt, P. J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J., and van den Hondel, C. (2002). Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol.* 20, 200–206. doi: 10.1016/S0167-7799(02)01933-9
- Punt, P. J., Van Zeijl, C., and Van den Hondel, C. (2001). High throughput screening of expressed DNA libraries in filamentous fungi. World patent WO 01/79558.
- Rahman, Z., Shida, Y., Furukawa, T., Suzuki, Y., Okada, H., Ogasawara, W., et al. (2009). Evaluation and characterization of *Trichoderma reesei* cellulase and xylanase promoters. *Appl. Microbiol. Biotechnol.* 82, 899–908. doi: 10.1007/s00253-008-1841-3
- Rao, T. R., and Reddy, C. A. (1984). DNA sequences from a ligninolytic filamentous fungus *Phanerochaete chrysosporium* capable of autonomous replication in yeast. *Biochem. Biophys. Res. Commun.* 118, 821–827. doi: 10.1016/0006-291X(84)91468-2
- Ribeiro, O. (2012). *Physiological Characterization of Ashbya gossypii and Strain Development for Recombinant Protein Production*. Ph.D. thesis, Programa Doutoral em Engenharia Química e Biológica, University of Minho, Portugal. p.171. <http://hdl.handle.net/1822/e0921>
- Ribeiro, O., Magalhães, F., Aguiar, T. Q., Wiebe, M. G., Penttilä, M., and Domingues, L. (2013). Random and direct mutagenesis to enhance protein secretion in *Ashbya gossypii*. *Bioengineered* 4, 322–331. doi: 10.4161/bioe.24653
- Ribeiro, O., Wiebe, M., Ilmén, M., Domingues, L., and Penttilä, M. (2010). Expression of *Trichoderma reesei* cellulases CBHI and EGI in *Ashbya gossypii*. *Appl. Microbiol. Biotechnol.* 87, 1437–1446. doi: 10.1007/s00253-010-2610-7
- Saarelainen, R., Mäntylä, A., Nevalainen, H., and Suominen, P. (1997). Expression of barley endopeptidase B in *Trichoderma reesei*. *Appl. Environ. Microbiol.* 63, 4938–4940.
- Schade, D., Walther, A., and Wendland, J. (2003). The development of a transformation system for the dimorphic plant pathogen *Holleya sinecauda* based on *Ashbya gossypii* DNA elements. *Fungal Genet. Biol.* 40, 65–71. doi: 10.1016/S1087-1845(03)00064-1
- Schrickx, J. M., Krave, A. S., Verdoes, J. C., van den Hondel, C. A., Stouthamer, A. H., and van Verseveld, H. W. (1993). Growth and product formation in chemostat and recycling cultures by *Aspergillus niger* N402 and a glucoamylase overproducing transformant, provided with multiple copies of the *glcA* gene. *J. Gen. Microbiol.* 139, 2801–2810. doi: 10.1099/00221287-139-11-2801
- Schuster, A., Bruno, K. S., Collett, J. R., Baker, S. E., Seiboth, B., Kubicek, C. P., et al. (2012). A versatile toolkit for high throughput functional genomics with *Trichoderma reesei*. *Biotechnol. Biofuels* 5:1. doi: 10.1186/1754-6834-5-1
- Schuster, A., and Schmoll, M. (2010). Biology and biotechnology of *Trichoderma*. *Appl. Microbiol. Biotechnol.* 87, 787–799. doi: 10.1007/s00253-010-2632-1
- Seiboth, B., Messner, R., Gruber, F., and Kubicek, C. P. (1992). Disruption of the *Trichoderma reesei cbh2* gene coding for cellobiohydrolase II leads to a delay in the triggering of cellulase formation by cellulose. *Microbiology* 138, 1259–1264.
- Seidl, V., Gamauf, C., Druzhinina, I. S., Seiboth, B., Hartl, L., and Kubicek, C. P. (2008). The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. *BMC Genomics* 9:327. doi: 10.1186/1471-2164-9-327
- Selber, K., Tjerneld, F., Collen, A., Hyytiä, T., Nakari-Setälä, T., Bailey, M., et al. (2004). Large-scale separation and production of engineered proteins, designed for facilitated recovery in detergent-based aqueous two-phase extraction systems. *Process Biochem.* 39, 889–896.
- Sharma, R., Katoch, M., Srivastava, P., and Qazi, G. (2009). Approaches for refining heterologous protein production in filamentous fungi. *World J. Microbiol. Biotechnol.* 25, 2083–2094. doi: 10.1007/s11274-009-0128-x
- Spohr, A., Carlsen, M., Nielsen, J., and Villadsen, J. (1998). Amylase production in recombinant *Aspergillus oryzae* during fed-batch and continuous cultivation. *J. Ferment. Bioeng.* 86, 49–56. doi: 10.1016/S0922-338X(98)80033-0
- Sprey, B. (1988). Cellular and extracellular localization of endocellulase in *Trichoderma reesei*. *FEMS Microbiol. Lett.* 55, 283–294. doi: 10.1111/j.1574-6968.1988.tb02816.x
- Stals, I., Koen, S., Geusens, S., Contreras, R., Van Beumen, J., and Claeysens, M. (2004). Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Post-secretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology* 14, 713–724. doi: 10.1093/glycob/cwh080
- Sun, A. (2013). *Heterologous Expression of the Mammalian Peptide Hormone Obestatin in Trichoderma reesei*. Ph.D. thesis, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney. p.191. <http://voyager.mq.edu.au/vwebv/holdingsInfo?searchId=593&recCount=50&recPointer=3&bibId=1955601> doi: 10.1007/BF00279894
- Suominen, P. L., Mäntylä, A. L., Karhunen, T., Hakola, S., and Nevalainen, H. (1993). High frequency one-step gene replacement in *Trichoderma reesei*. II. Effects of deletions of individual cellulase genes. *Mol. Gen. Genet.* 241, 523–530. doi: 10.1007/BF00279894
- Te'o, V. S. J., and Nevalainen, K. M. H. (2008). Multiple promoter platform for protein production. World Patent WO 2009076709.
- Torrent, M., Llompart, B., Lasserre-Ramassamy, S., Llop-Tous, I., Bastida, M., Marzabal, P., et al. (2009). Eukaryotic protein production in designed storage organelles. *BMC Biol.* 7:5. doi: 10.1186/1741-7007-7-5
- Ülker, A., and Sprey, B. (1990). Characterization of an unglycosylated low molecular weight 1,4-beta-glucan-glucanohydrolase of *Trichoderma reesei*. *FEMS Microbiol. Lett.* 57, 215–219. doi: 10.1016/0378-1097(90)90068-2
- Verdoes, J. C., Punt, P. J., Stouthamer, A. H., and van den Hondel, C. A. (1994). The effect of multiple copies of the upstream region on expression of the *Aspergillus niger* glucoamylase-encoding gene. *Gene* 145, 179–187. doi: 10.1016/0378-1119(94)90003-5
- Vitikainen, M., Arvas, M., Pakula, T., Oja, M., Penttilä, M., and Saloheimo, M. (2010). Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. *BMC Genomics* 11:441. doi: 10.1186/1471-2164-11-441
- Wang, H., and Ward, M. (2000). Molecular characterization of a PDI-related gene prpA in *Aspergillus niger* var. *awamori*. *Curr. Genet.* 37, 57–64. doi: 10.1007/s002940050009
- Wang, T. H., Liu, T., Wu, Z. H., Liu, S. L., Lu, Y., and Qu, Y. B. (2004). Novel cellulase profile of *Trichoderma reesei* strains constructed by *cbh1* gene replacement with *eg3* gene expression cassette. *Acta Biochim. Biophys. Sin. (Shanghai)* 36, 667–672. doi: 10.1093/abbs/36.10.667
- Ward, M., Lin, C., Victoria, D. C., Fox, B. P., Fox, J. A., Wong, D. L., et al. (2004). Characterization of humanized antibodies secreted by *Aspergillus niger*. *Appl. Environ. Microbiol.* 70, 2567–2576. doi: 10.1128/AEM.70.5.2567-2576.2004
- Ward, O. (2012). Production of recombinant proteins by filamentous fungi. *Biotechnol. Adv.* 30, 1119–1139. doi: 10.1016/j.biotechadv.2011.09.012

- Wessels, J. (1993). Wall growth, protein excretion and morphogenesis in fungi. *New Phytol.* 123, 397–413. doi: 10.1111/j.1469-8137.1993.tb03751.x
- Withers, J. M., Swift, R. J., Wiebe, M. G., Robson, G. D., Punt, P. J., van den Hondel, C. A., et al. (1998). Optimization and stability of glucoamylase production by recombinant strains of *Aspergillus niger* in chemostat culture. *Biotechnol. Bioeng.* 59, 407–418. doi: 10.1002/(SICI)1097-0290(19980820)59:4<407::AID-BIT3>3.0.CO;2-K
- Workman, M., Andersen, M. R., and Thykaer, J. (2013). Integrated approaches for assessment of cellular performance in industrially relevant filamentous fungi. *Ind. Microbiol.* 9, 337–344 doi: 10.1089/ind.2013.0025
- Wösten, H. A., Moukha, S. M., Sietsma, J. H., and Wessels, J. G. (1991). Localization of growth and secretion of proteins in *Aspergillus niger*. *J. Gen. Microbiol.* 137, 2017–2023. doi: 10.1099/00221287-137-8-2017
- Ye, J., Ly, J., Watts, K., Hsu, A., Walker, A., McLaughlin, K., et al. (2011). Optimisation of a glycoengineered *Pichia pastoris* cultivation process for commercial antibody production. *Biotechnol. Prog.* 27, 1744–1750. doi: 10.1002/btpr.695
- Yoon, J., Aishan, T., Maruyama, J., and Kitamoto, K. (2010). Enhanced production and secretion of heterologous proteins by the filamentous fungus *Aspergillus oryzae* via disruption of vacuolar protein sorting receptor gene Aovps10. *Appl. Environ. Microbiol.* 76, 5718–5727. doi: 10.1128/AEM.03087-09
- Zhao, Y.-Y., Takahashi, M., Gu, J.-G., Myyoshi, E., Matsumoto, A., Kitazume, S., et al. (2008). Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci.* 99, 1304–1310. doi: 10.1111/j.1349-7006.2008.00839.x
- Zhong, Y., Liu, X., Xiao, P., Wei, S., and Wang, T. (2011). Expression and secretion of the human erythropoietin using an optimized cbh1 promoter and the native CBHI signal sequence in the industrial fungus *Trichoderma reesei*. *Appl. Biochem. Biotechnol.* 165, 1169–1177. doi: 10.1007/s12010-011-9334-8

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

Received: 25 November 2013; paper pending published: 20 December 2013; accepted: 11 February 2014; published online: 27 February 2014.

Citation: Nevalainen H and Peterson R (2014) Making recombinant proteins in filamentous fungi- are we expecting too much? *Front. Microbiol.* 5:75. doi: 10.3389/fmicb.2014.00075

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Nevalainen and Peterson. 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.



Algae-based oral recombinant vaccines

Elizabeth A. Specht and Stephen P. Mayfield*

California Center for Algae Biotechnology, University of California at San Diego, La Jolla, CA, USA

Edited by:

Germán Leandro Rosano, Instituto de Biología Molecular y Celular de Rosario, Argentina

Reviewed by:

Edward Rybicki, University of Cape Town, South Africa
Ruth Elena Soria-Guerra, Universidad Autónoma de San Luis Potosí, Mexico
Pal Maliga, Rutgers University, USA

*Correspondence:

Stephen P. Mayfield, California Center for Algae Biotechnology, University of California at San Diego, Bonner Hall 2150, MC 0368, 9500 Gilman Drive, La Jolla, CA 92093, USA
e-mail: smayfield@ucsd.edu

Recombinant subunit vaccines are some of the safest and most effective vaccines available, but their high cost and the requirement of advanced medical infrastructure for administration make them impractical for many developing world diseases. Plant-based vaccines have shifted that paradigm by paving the way for recombinant vaccine production at agricultural scale using an edible host. However, enthusiasm for “molecular pharming” in food crops has waned in the last decade due to difficulty in developing transgenic crop plants and concerns of contaminating the food supply. Microalgae could be poised to become the next candidate in recombinant subunit vaccine production, as they present several advantages over terrestrial crop plant-based platforms including scalable and contained growth, rapid transformation, easily obtained stable cell lines, and consistent transgene expression levels. Algae have been shown to accumulate and properly fold several vaccine antigens, and efforts are underway to create recombinant algal fusion proteins that can enhance antigenicity for effective orally delivered vaccines. These approaches have the potential to revolutionize the way subunit vaccines are made and delivered – from costly parenteral administration of purified protein, to an inexpensive oral algae tablet with effective mucosal and systemic immune reactivity.

Keywords: oral vaccines, recombinant subunit vaccines, microalgae, plant-produced vaccines, algal engineering

INTRODUCTION

Infectious diseases directly account for nearly 25% of deaths worldwide, and are a predominant cause of morbidity and mortality in the developing world (Fauci et al., 2005). Even for diseases for which vaccines exist, limited access – due to financial as well as infrastructural or medical personnel limitations – is a major contributor to this high infectious disease burden. Many developing world diseases do not yet have vaccines, in part because traditional vaccine production costs present a significant investment hurdle, considering the financial capacity of the intended consumers. Both cost and ease of administration are challenges that must be tackled to address this undue burden on global health and productivity.

Oral vaccination has many distinct advantages over parenteral administration, but has proven difficult to achieve thus far, reflected by the scarcity of licensed oral vaccines. Perhaps the most significant benefit of oral vaccination is the ability to elicit both mucosal and systemic immunity. As most human pathogens enter via mucosal surfaces – either nasally, orally, or by sexual transmission – mucosal immunity can serve as a first line of defense to prevent infection before it reaches the bloodstream (Mason and Herbst-Kralovetz, 2012). Oral vaccines also obviate the need for trained medical personnel to administer them and reduce the risks of infection associated with needles. They also have higher compliance from patients, owing to the lack of fear and resistance associated with injections. Both of these latter aspects are important considerations for successful vaccination campaign coverage in remote or resource-limited settings.

Plant-produced vaccines have two critical advantages: much lower cost than traditional recombinant vaccine platforms, and improved safety because of insusceptibility to mammalian pathogen contamination. The batch costs of plant-produced

vaccines may be as much as a thousand times less than traditional animal cell culture or even bacterial or yeast cell culture, though it has been noted that this will not translate directly to per-dose cost because downstream sales, packaging, and distribution costs are similar regardless of production method (Rybicki, 2009). The current status of plant-produced vaccines in pre-clinical and early phase human clinical trials has been extensively reviewed (Lossel and Waheed, 2011; Mason and Herbst-Kralovetz, 2012; Rosales-Mendoza et al., 2012a,b; Guan et al., 2013; Jacob et al., 2013); despite positive preliminary data, none have made it through to licensing. The only licensed plant-produced vaccine is a veterinary injectable vaccine against Newcastle disease virus in poultry, made from purified antigen expressed in cultured tobacco cells. Dow AgroSciences received Food and Drug Administration (FDA) approval for the vaccine in 2006, but only as a demonstration that plant-produced vaccines can meet the stringent regulatory requirements for approval; it is not currently for sale (Rybicki, 2009).

Plant cells are of particular interest for oral vaccines because their rigid cell walls provide exceptional antigen protection through the stomach into the intestines, where they can access the gut-associated lymphoid tissue (Kwon et al., 2013). Expression within chloroplasts or other storage organelles may also provide additional protection (Khan et al., 2012). While vaccine antigens have been transformed into many edible species including lettuce, tomato, potato, and tobacco, expression in stable transformed crop plants has suffered from low yields, typically less than 1% of total soluble protein (TSP; Lossel and Waheed, 2011). Yields have been increased by transient expression using recombinant viral vectors or *Agrobacterium* infection, but this expression is typically unstable (Rybicki, 2009). Even using these strategies, the most consistently

high-yielding host species is tobacco, which is inedible and therefore would require purification prior to vaccine administration (Lossl and Waheed, 2011).

ALGAE AS A RECOMBINANT PROTEIN PRODUCTION PLATFORM

Green microalgae have proven to be highly useful protein production platforms for a variety of industrial and therapeutic applications, particularly for complex or heavily disulfide-bonded proteins. The chloroplast provides a unique enclosed compartment that facilitates folding (Chebolu and Daniell, 2009), and transgene products have been shown to accumulate to high levels in the algal chloroplast – as high as 10% of TSP (Manuell et al., 2007; Surzycki et al., 2009). Unlike prokaryotes, chloroplasts of algae contain much of the same sophisticated cellular folding machinery as other eukaryotic organisms like yeast. While the algal nuclear genome can also be transformed, to date most transgene expression has been from the chloroplast genome due to reduced gene silencing and higher protein accumulation.

The green alga model organism *Chlamydomonas reinhardtii* has been used to produce a number of human and animal therapeutically relevant proteins, including full-length human antibodies (Tran et al., 2009), signaling molecules such as vascular endothelial growth factor (Rasala et al., 2010), and structural proteins like fibronectin (Rasala et al., 2010). Though expression levels are highly variable by gene, improvements in codon optimization (Franklin et al., 2002; Surzycki et al., 2009) and characterization of ideal gene regulatory elements (Rasala et al., 2011; Specht and Mayfield, 2013) continue to increase levels of transgene expression. *C. reinhardtii*'s success and future potential as a therapeutic protein production platform has been recently reviewed (Rasala and Mayfield, 2011).

ADVANTAGES OF AN ALGAL VACCINE PRODUCTION HOST

Unicellular green algae possess all the positive attributes of plant systems, plus several unique advantages over terrestrial plants as vaccine production hosts. Algal biomass accumulation is extremely rapid, and the entirety of the biomass can be utilized for vaccine production, unlike plants that expend energy producing supporting tissues that do not contain the vaccine antigen or cannot be harvested easily. Algae are also not restricted by growing season or local soil fertility, and concerns of cross-contamination of nearby food crops are non-existent. Enclosed bioreactors can be used for higher biomass yields and to reduce concerns of environmental escape (Franconi et al., 2010), and media can be recycled to minimize water and nutrient loss. The 2002 discovery of transgenic viral capsid protein-expressing maize in food harvests of nearby corn and soybean crops effectively halted efforts to produce vaccines in edible crop plants, making a food crop-based oral vaccine highly unlikely (Rybicki, 2009). Green algae such as *C. reinhardtii* are generally recognized as safe (GRAS) by the FDA, resurrecting hope that unprocessed edible vaccines can be produced in a photosynthetic organism.

Crop plants can contain hundreds of chloroplasts per cell, and each chloroplast harbors dozens of copies of its plastid genome. In contrast, *C. reinhardtii* contains a single chloroplast that occupies about half of the volume of the cell (Franklin and Mayfield,

2005), making stable homoplasmic transformed lines much easier to obtain (a few weeks versus several months) and allowing for increased yields of plastid-expressed vaccine antigens, which account for nearly all antigens expressed to date in algae. This genomic stability, combined with the ability to tightly regulate growth conditions inside contained bioreactors, allows for more consistent expression levels than terrestrial plants, which can vary by several-fold.

Finally, algae can be easily preserved by lyophilization, and two studies of algal-produced vaccine antigens have verified that dried algae stored at room temperature for 6 months (Gregory et al., 2013) or even 20 months (Dreesen et al., 2010) exhibit nearly equivalent antigen effectiveness as freshly harvested algae, though storage at 37° did begin to cause a loss of activity over time (Gregory et al., 2013). The algal cell wall appears sufficient to withstand harsh conditions within the stomach, as very little antigen degradation was observed after whole cells were incubated with pepsin at pH 1.7 (Dreesen et al., 2010). These observations indicate that algae are an ideal host for vaccine transport without cold-chain supply, and that the cells provide adequate protection for antigens en route to the intestinal mucosal lymph tissue, obviating the additional expense associated with encapsulation.

ALGAL VACCINE PROGRESS

The first reported algal-synthesized vaccine antigen was a chimeric molecule comprising the foot-and-mouth disease virus structural protein VP1 and the beta subunit of cholera toxin (CTB), a known mucosal adjuvant (Sun et al., 2003). This antigen had been previously expressed in plants and had demonstrated oral immunity in mice (Wigdorovitz et al., 1999), but advancement of trials was hindered by low expression levels. In *C. reinhardtii*, 3–4% TSP was reported, but higher yields may be possible because the strains examined were not completely homoplasmic (Sun et al., 2003).

The next report of an algal-produced vaccine antigen showed the first *in vivo* data for efficacy conferring immunity. The classical swine fever virus (CSFV) surface protein E2 was expressed from the *C. reinhardtii* chloroplast genome, and total protein extracts were administered subcutaneously with Freud's adjuvant or orally by gavage with no adjuvant. Subcutaneous immunization reportedly induced a significant immune response, but no data for this result was shown. No systemic or mucosal immune response was detected after the oral immunization, and it was suggested that a mucosal adjuvant may be necessary for oral administration to be effective (He et al., 2007).

Wang et al. (2008) expressed the human glutamic acid decarboxylase, a known Type 1 diabetes autoimmune antigen, which reacted with sera from non-obese diabetic mice. Surprisingly, detectable expression was achieved using a non-codon-optimized gene. A more thorough investigation of the factors affecting vaccine antigen expression in algae found that indeed codon optimization is critical for high yield. It has also been noted that yield is highly variable among individual transformants despite the fact that chloroplast transformation proceeds by homologous recombination, eliminating positional effects within the genome (Surzycki et al., 2009).

Oral immunization was finally shown to be effective when the antigen of interest was fused to the B subunit of CTB, which forms a pentameric structure and binds the GM1 ganglioside for internalization into intestinal cells. After feeding freeze-dried algae repeatedly to mice, fecal IgA and systemic IgG antibody titers reached similarly high levels for both the intended *Staphylococcus aureus* antigen and CTB. Significantly, within a week of finishing the 5-week oral vaccination, 80% of immunized mice survived a lethal challenge with *S. aureus* that killed all control mice within 48 h (Dreesen et al., 2010).

Two studies earlier this year reported relatively low yields of two additional algal-produced antigens, but they are still promising compared to previous literature using alternative systems. A human papillomavirus E7 protein, while only accumulated to 0.12% TSP, expressed similar to or better than in other plant systems and did not require fusion to a stabilizing protein to achieve consistent expression. Furthermore, the algal chloroplast-produced E7 was soluble, whereas the plant-produced E7 was found predominantly in the insoluble fraction using multiple solubilization buffers. While the antibody titer elicited by affinity purified protein was much higher, a crude algal extract was shown to be equally effective at preventing tumor development and promoting mouse survival (Demurtas et al., 2013). A chimeric antigen intended to prevent hypertension, consisting of a fusion between angiotensin and a Hepatitis B antigen as a carrier, was the first algal vaccine to be expressed from the nuclear genome without chloroplast targeting. While it only accumulated to 0.05% TSP, it was detectable by Western blot from algal TSP extracts (Soria-Guerra et al., 2014).

Since 2010, several studies have shown that malarial transmission-blocking vaccines can be produced in *C. reinhardtii*. Transmission-blocking vaccines target surface proteins that appear on the sexual and gamete stages of *Plasmodium*, the causative pathogen of malaria. There is some evidence that these vaccines may provide partial protection to individuals, but the main benefit of vaccination with a transmission-blocking vaccine is derived from herd immunity preventing the spread of the disease. Therefore, it is especially critical that transmission-blocking vaccines can be delivered easily and at extremely low cost, to reach threshold coverage of the huge populations living in malaria-endemic regions. One difficulty of producing these *Plasmodium* surface proteins is that they contain multiple EGF-like domains that are heavily disulfide-bonded, rendering them difficult to fold and therefore difficult to accumulate to high levels without forming insoluble aggregates (Gregory et al., 2012). Interestingly, *Plasmodia* appear to not glycosylate their proteins (Gowda and Davidson, 1999), making algal chloroplasts suitable hosts as the chloroplast also does not contain glycosylation machinery.

A total of six algae-produced malarial antigens or fragments thereof – *Pfs25*, *Pfs28*, *Pfs48/45*, *PfMSP1*, *PbMSP1*, and *PbAMA1* – have been shown to fold properly and exhibit antibody recognition akin to that of the native *Plasmodium* surface proteins (Dauvillée et al., 2010; Gregory et al., 2012; Jones et al., 2013). Algal chloroplast-produced *Pfs25* was able to completely prevent malaria transmission, indicated by a total absence of *Plasmodium* oocysts in mosquito midguts after feeding on immunized

mouse sera. Furthermore, feeding lyophilized algae expressing *Pfs25* fused to CTB elicited a mucosal response to both antigens (Gregory et al., 2013). However, systemic IgG response was only observed for the CTB. This is in contrast with the *S. aureus* D2 protein fused to CTB, where systemic immunity was elicited for both domains (Dreesen et al., 2010), suggesting that either the furin protease cleavable linker between the *Pfs25* and CTB domains prevented *Pfs25* from being presented to the systemic immune system, or perhaps that *Pfs25* is inherently less immunogenic. In a different strategy, truncated versions of the malarial proteins AMA1 and MSP1 were fused to the major protein constituent of the chloroplast starch granules, the granule-bound starch synthase (GBSS). Though they were expressed from the nuclear genome, reasonable accumulation was achieved because the proteins were targeted to and sequestered within the chloroplast starch granules. Both oral and injected vaccination using purified starch from these strains reduced parasite load and prolonged mice survival after challenge with *Plasmodium berghei*; in the case of an injected vaccine consisting of both antigens, 30% of mice survived the otherwise-lethal infection (Dauvillée et al., 2010).

All vaccines produced in algae to date are summarized in **Table 1**, along with reported yields and significant pre-clinical findings. Most work thus far has been performed in the green alga model organism *C. reinhardtii*, though one of the earliest reports of an algal-produced hepatitis B antigen was in the marine alga *Dunaliella salina* (Geng et al., 2003) and hepatitis B antigen has also been produced in the diatom *Phaeodactylum tricornutum* (Hempel et al., 2011). In recent years the algal genetic toolkit has been expanded to other algal species, including other green algae, diatoms, and cyanobacteria (Ducat et al., 2011; Georgianna and Mayfield, 2012; Qin et al., 2012), with a goal of broad host range compatibility. Already, over 20 species of algae – including dinoflagellates, red algae, and diatoms – have been transformed, and a suite of promoters and selectable markers have been characterized for many species (see Gong et al., 2011, for a comprehensive review). While the first generation of algal vaccines has been predominantly pioneered in *Chlamydomonas*, these advances can readily be applied to alternative algal species that may be more suitable for large-scale vaccine production.

FUTURE POTENTIAL FOR ALGAL-BASED ORAL RECOMBINANT VACCINES

From the research available to date, it is clear that algae can produce complex vaccine antigens, and that *Chlamydomonas*-produced antigens can elicit immunogenic responses that are appropriate for their intended roles as vaccines. It is also clear that identifying alternative mucosal adjuvants to complement these antigens is critical, whether for co-administration with algal-produced antigens or for incorporation into chimeric fusion proteins. It has been suggested that antigenic fusions with CTB, one of the preferred adjuvants, may interfere with the CTB subunit's ability to form the pentameric structure essential for strong GM1 ganglioside binding (Sun et al., 2003). Many alternatives to CTB are under investigation for oral vaccination in other production platforms, including CpG-containing oligodeoxynucleotides, saponins, and subunits from heat-labile enterotoxin and ricin toxin (Pelosi et al., 2012).

Table 1 | Summary of algal-produced vaccines and significant findings.

Antigen(s)	Host/integration site	Yield	Significant findings	Citation
Foot-and-mouth disease virus (FMDV) structural protein VP1 fused to cholera toxin B subunit (CTB)	<i>C. reinhardtii</i> chloroplast genome	3 to 4% TSP	Protein accumulation was higher than reported in previous plant studies and detectable by Western and ELISA. Binding to GM1 ganglioside was weak but statistically significant. Strains were not completely homoplasmic, so higher yields may be possible.	Sun et al. (2003)
Hepatitis B surface antigen (HBsAg)	<i>D. salina</i> nuclear genome	Up to 3ng/mg soluble protein (0.0003% TSP)	Expression level was quantified by ELISA and reported to be “high” but the results reflect very low expression.	Geng et al. (2003)
Classical swine fever virus (CSFV) structural protein E2	<i>C. reinhardtii</i> chloroplast genome	1.5 to 2% TSP	Subcutaneous immunization of algal extracts induced systemic response, but oral immunization did not result in either systemic or mucosal response.	He et al. (2007)
Human glutamic acid decarboxylase 65 (hGAD65)	<i>C. reinhardtii</i> chloroplast genome	0.25 to 0.3% TSP	Purified hGAD65 reacted with sera from diabetic mice, showing antigenicity. Higher yields may be obtained by codon-optimizing the human gene, which does not mimic the chloroplast AT content.	Wang et al. (2008)
White spot syndrome virus VP28 protein	<i>C. reinhardtii</i> chloroplast genome	Variable, ranging from 0.2 to 20.9% total cellular protein (0.1 to 10.5% TSP)	The authors report high variability in expression levels, despite same-site integration by homologous recombination in the plastid genome. Stronger resistance to antibiotic selection might be useful for identifying high-expressing strains.	Surzycki et al. (2009)
<i>S. aureus</i> D2 fibronectin-binding domain fused to cholera toxin B subunit	<i>C. reinhardtii</i> chloroplast genome	Up to 0.7% TSP	Within whole cells, the antigen can withstand low pH and pepsin treatment, and is stable at room temperature for 20 months. Mice fed with whole algae showed mucosal IgA and systemic IgG responses to both CTB and D2, and 80% survived lethal <i>S. aureus</i> challenge.	Dreesen et al. (2010)
<i>P. falciparum</i> MSP1, <i>P. berghei</i> MSP1, and <i>P. berghei</i> AMA1, C-terminal domains only, fused to granule-bound starch synthase (GBSS)	<i>C. reinhardtii</i> nuclear genome, targeted to the chloroplast starch granules	0.2 to 1.0 µg of protein per mg of purified starch	The <i>P. falciparum</i> gene expressed significantly better than either <i>P. berghei</i> gene. A single dose was as effective as three doses in reducing parasite load after <i>P. berghei</i> challenge in vaccinated mice. Oral immunization with purified starch protected equally well as parenteral administration until day 40.	Dauvillée et al. (2010)
Hepatitis B surface antigen (HBsAg) fused to GFP or ER retention signal	<i>P. tricornutum</i> nuclear genome	0.7% TSP	By inhibitory ELISA, algal-produced HBsAg binds antibodies slightly stronger than commercially available HBsAg produced in yeast.	Hempel et al. (2011)

(Continued)

Table 1 | Continued

Antigen(s)	Host/integration site	Yield	Significant findings	Citation
<i>P. falciparum</i> surface proteins <i>Pfs25</i> and <i>Pfs28</i>	<i>C. reinhardtii</i> chloroplast genome	Not quantified; visible by Western and Coomassie after affinity purification	Conformation-specific antibodies and CD spectroscopy confirmed proper folding of heavily disulfide-bonded antigens. Sera from <i>Pfs25</i> -inoculated mice completely blocked replication within mosquitoes, and therefore malaria transmission.	Gregory et al. (2012)
<i>P. falciparum</i> surface protein <i>Pfs25</i> fused to cholera toxin B subunit	<i>C. reinhardtii</i> chloroplast genome	Up to 0.09% TSP	Mice fed lyophilized algae showed mucosal IgA response to both CTB and <i>Pfs25</i> , but only systemic IgG response to CTB. A furin protease site separated the two domains.	Gregory et al. (2013)
<i>P. falciparum</i> surface protein <i>Pfs48/45</i> C-terminal domain	<i>C. reinhardtii</i> chloroplast genome	Not quantified; visible by Western blot after affinity purification	The antigenic C-terminal domain was recognized by Western blot and ELISA with conformation-specific antibodies. It appeared to accumulate in the insoluble fraction.	Jones et al. (2013)
Human Papillomavirus type 16 E7 protein, attenuated mutant (E7GGG)	<i>C. reinhardtii</i> chloroplast genome	Up to 0.12% TSP. His-tagged version obtained only 0.02% TSP.	In contrast to higher plant expression, <i>C. reinhardtii</i> obtained much higher yields and most of it was soluble. The crude algal extracts elicited lower IgG response than purified protein, but both preparations allowed 70% survival after tumor challenge.	Demurtas et al. (2013)
Angiotensin II fused to Hepatitis B virus capsid antigen (HBcAg)	<i>C. reinhardtii</i> nuclear genome	Up to 0.05% TSP	The chimeric antigen was visible by Western blot and quantitated by ELISA, but no <i>in vivo</i> or immunological activity assays were performed.	Soria-Guerra et al. (2014)

Future work should empirically explore many combinations of antigens, mucosal adjuvants, and even testing multiple linkers and potential translocation domains. As has been noted previously, expression, uptake, and antigenicity are all difficult to predict in the context of plant-produced oral vaccine antigens (Rybicki, 2009), so a high-throughput system like algae is extremely valuable for rapidly testing many versions of potential chimeric vaccine molecules. Furthermore, many antigens will require proper post-translational modifications such as glycosylation to be recognized properly; more work needs to be done to increase expression levels from the nuclear genome, as glycosylation does not occur in the chloroplast.

It has been suggested that the first licensed plant-produced human vaccines likely will not be the first ones tested in humans, many of which targeted pathogens like Hepatitis B for which a relatively inexpensive vaccine already exists (Rybicki, 2009). Stepping stones along the way to human vaccines may include reagents for cheaper diagnostics and development of veterinary vaccines. Several human studies with plant-made vaccines have also indicated a role for oral boosting of an existing immune response conferred by traditional injectable vaccines (Mason and Herbst-Kralovetz, 2012). An algal-produced human vaccine production platform will likely come to fruition as an alternative for very expensive vaccines like HPV, or for novel vaccines against diseases for which no alternative currently exists (Martinez et al., 2012). The cost and logistical considerations

of storage, delivery, and administration in resource-limited settings indicate that plant or algal production may be the only feasible option for large-scale inexpensive vaccination, and thus this avenue deserves increased attention from research funding agencies and investment from the pharmaceutical industry as well.

AUTHOR CONTRIBUTIONS

Elizabeth A. Specht and Stephen P. Mayfield wrote and revised the manuscript. Elizabeth A. Specht developed **Table 1**.

ACKNOWLEDGMENTS

This work was funded by a Department of Energy, Consortium for Algal Biofuels Commercialization grant, DE-EE0003373; and by the California Energy Commission, California Initiative for Sustainable Large Molecule Fuels, 500–10–039. Elizabeth A. Specht was supported by a National Science Foundation graduate research fellowship. We thank Prema Karunanithi for her careful proofreading of the manuscript.

REFERENCES

- Chebolu, S., and Daniell, H. (2009). "Chloroplast-derived vaccine antigens and biopharmaceuticals: expression, folding, assembly and functionality," in *Plant-Produced Microbial Vaccines*, ed. A.V. Karasev. (Berlin: Springer-Verlag Berlin), 33–54. doi: 10.1007/978-3-540-70868-1_3
- Dauvillée, D., Delhaye, 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
- Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., and Giuliano, G. (2013). A *Chlamydomonas*-derived human papillomavirus 16 E7 vaccine induces specific tumor protection. *PLoS ONE* 8:e61473. doi: 10.1371/journal.pone.0061473
- Dreesen, I. A., Charpin-El Hamri, G., and Fussenegger, M. (2010). Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. *J. Biotechnol.* 145, 273–280. doi: 10.1016/j.jbiotec.2009.12.006
- Ducat, D. C., Way, J. C., and Silver, P. A. (2011). Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* 29, 95–103. doi: 10.1016/j.tibtech.2010.12.003
- Fauci, A. S., Touchette, N. A., and Folkers, G. K. (2005). Emerging infectious diseases: a 10-year perspective from the national institute of allergy and infectious diseases. *Emerg. Infect. Dis.* 11, 519–525. doi: 10.3201/eid1104.041167
- Franconi, R., Demurtas, O. C., and Massa, S. (2010). Plant-derived vaccines and other therapeutics produced in contained systems. *Expert Rev. Vaccines* 9, 877–892. doi: 10.1586/erv.10.91
- Franklin, S. E., and Mayfield, S. P. (2005). Recent developments in the production of human therapeutic proteins in eukaryotic algae. *Expert Opin. Biol. Ther.* 5, 225–235. doi: 10.1517/14712598.5.2.225
- Franklin, S., Ngo, B., Efuet, E., and Mayfield, S. P. (2002). Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J. Cell Mol. Biol.* 30, 733–744. doi: 10.1046/j.1365-313X.2002.01319.x
- Geng, D. G., Wang, Y. Q., Wang, P., Li, W. B., and Sun, Y. R. (2003). Stable expression of hepatitis B surface antigen gene in *Dunaliella salina* (Chlorophyta). *J. Appl. Phycol.* 15, 451–456. doi: 10.1023/B:JAPH.0000004298.89183.e5
- Georgianna, D. R., and Mayfield, S. P. (2012). Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* 488, 329–335. doi: 10.1038/nature11479
- Gong, Y., Hu, H., Gao, Y., Xu, X., and Gao, H. (2011). Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. *J. Ind. Microbiol. Biotechnol.* 38, 1879–1890. doi: 10.1007/s10295-011-1032-6
- Gowda, D. C., and Davidson, E. A. (1999). Protein glycosylation in the malaria parasite. *Parasitol. Today* 15, 147–152. doi: 10.1016/S0169-4758(99)01412-X
- Gregory, J. A., Li, F., Tomosada, L. M., Cox, C. J., Topol, A. B., Vinetz, J. M., et al. (2012). Algae-produced Pf525 elicits antibodies that inhibit malaria transmission. *PLoS ONE* 7:e37179. doi: 10.1371/journal.pone.0037179
- Gregory, J. A., Topol, A. B., Doerner, D. Z., and Mayfield, S. (2013). Alga-produced cholera toxin-pf525 fusion proteins as oral vaccines. *Appl. Environ. Microbiol.* 79, 3917–3925. doi: 10.1128/AEM.00714-13
- Guan, Z. J., Guo, B., Huo, Y. L., Guan, Z. P., Dai, J. K., and Wei, Y. H. (2013). Recent advances and safety issues of transgenic plant-derived vaccines. *Appl. Microbiol. Biotechnol.* 97, 2817–2840. doi: 10.1007/s00253-012-4566-2
- He, D. M., Qian, K. X., Shen, G. F., Zhang, Z. F., Li, Y. N., Su, Z. L., et al. (2007). Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chloroplasts. *Colloids Surf. B Biointerfaces* 55, 26–30. doi: 10.1016/j.colsurf.2006.10.042
- Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS ONE* 6:e28424. doi: 10.1371/journal.pone.0028424
- Jacob, S. S., Cherian, S., Sumithra, T. G., Raina, O. K., and Sankar, M. (2013). Edible vaccines against veterinary parasitic diseases-current status and future prospects. *Vaccine* 31, 1879–1885. doi: 10.1016/j.vaccine.2013.02.022
- Jones, C. S., Luong, T., Hannon, M., Tran, M., Gregory, J. A., Shen, Z., et al. (2013). Heterologous expression of the C-terminal antigenic domain of the malaria vaccine candidate Pf548/45 in the green algae *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* 97, 1987–1995. doi: 10.1007/s00253-012-4071-7
- Khan, I., Twyman, R. M., Arcalis, E., and Stoger, E. (2012). Using storage organelles for the accumulation and encapsulation of recombinant proteins. *Biotechnol. J.* 7, 1099–1108. doi: 10.1002/biot.201100089
- Kwon, K. C., Verma, D., Singh, N. D., Herzog, R., and Daniell, H. (2013). Oral delivery of human biopharmaceuticals, autoantigens and vaccine antigens bioencapsulated in plant cells. *Adv. Drug Deliv. Rev.* 65, 782–799. doi: 10.1016/j.addr.2012.10.005
- Loss, A. G., and Waheed, M. T. (2011). Chloroplast-derived vaccines against human diseases: achievements, challenges and scopes. *Plant Biotechnol. J.* 9, 527–539. doi: 10.1111/j.1467-7652.2011.00615.x
- Manuell, A. L., Beligni, M. V., Elder, J. H., Siefker, D. T., Tran, M., Weber, A., et al. (2007). Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. *Plant Biotechnol. J.* 5, 402–412. doi: 10.1111/j.1467-7652.2007.00249.x
- Martinez, C. A., Giuliatti, A. M., and Talou, J. R. (2012). Research advances in plant-made flavivirus antigens. *Biotechnol. Adv.* 30, 1493–1505. doi: 10.1016/j.biotechadv.2012.03.004
- Mason, H. S., and Herbst-Kralovetz, M. M. (2012). “Plant-derived antigens as mucosal vaccines,” in *Mucosal Vaccines: Modern Concepts, Strategies, and Challenges*, ed. P. A. Kozlowski (Berlin: Springer-Verlag Berlin), 101–120. doi: 10.1007/82_2011_158
- Pelosi, A., Shepherd, R., and Walmsley, A. M. (2012). Delivery of plant-made vaccines and therapeutics. *Biotechnol. Adv.* 30, 440–448. doi: 10.1016/j.biotechadv.2011.07.018
- Qin, S., Lin, H., and Jiang, P. (2012). Advances in genetic engineering of marine algae. *Biotechnol. Adv.* 30, 1602–1613. doi: 10.1016/j.biotechadv.2012.05.004
- Rasala, B. A., and Mayfield, S. P. (2011). The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng. Bugs* 2, 50–54. doi: 10.4161/bbug.2.1.13423
- Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., et al. (2010). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* 8, 719–733. doi: 10.1111/j.1467-7652.2010.00503.x
- Rasala, B. A., Muto, M., Sullivan, J., and Mayfield, S. P. (2011). Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5' untranslated region optimization. *Plant Biotechnol. J.* 9, 674–683. doi: 10.1111/j.1467-7652.2011.00620.x
- Rosales-Mendoza, S., Govea-Alonso, D. O., Monreal-Escalante, E., Fragoso, G., and Sciotto, E. (2012a). Developing plant-based vaccines against neglected tropical diseases: where are we? *Vaccine* 31, 40–48. doi: 10.1016/j.vaccine.2012.10.094
- Rosales-Mendoza, S., Rubio-Infante, N., Govea-Alonso, D. O., and Moreno-Fierros, L. (2012b). Current status and perspectives of plant-based candidate vaccines against the human immunodeficiency virus (HIV). *Plant Cell Rep.* 31, 495–511. doi: 10.1007/s00299-011-1194-8
- Rybicki, E. P. (2009). Plant-produced vaccines: promise and reality. *Drug Discov. Today* 14, 16–24. doi: 10.1016/j.drudis.2008.10.002
- Soria-Guerra, R. E., Ramirez-Alonso, J. I., Ibáñez-Salazar, A., Govea-Alonso, D. O., Paz-Maldonado, L. M. T., Bañuelos-Hernández, B., et al. (2014). Expression of an HBCAg-based antigen carrying angiotensin II in *Chlamydomonas reinhardtii* as a candidate hypertension vaccine. *Plant Cell Tissue Organ Cult.* 116, 133–139. doi: 10.1007/s11240-013-0388-x
- Specht, E. A., and Mayfield, S. P. (2013). Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs. *ACS Synth. Biol.* 2, 34–46. doi: 10.1021/sb300069k
- Sun, M., Qian, K. X., Su, N., Chang, H. Y., Liu, J. X., and Chen, G. F. (2003). Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast. *Biotechnol. Lett.* 25, 1087–1092. doi: 10.1023/A:1024140114505
- Surzycki, R., Greenham, K., Kitayama, K., Dibal, F., Wagner, R., Rochaix, J. D., et al. (2009). Factors effecting expression of vaccines in microalgae. *Biologicals* 37, 133–138. doi: 10.1016/j.biologics.2009.02.005
- Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009). Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* 104, 663–673. doi: 10.1002/bit.22446
- Wang, X., Brandsma, M., Tremblay, R., Maxwell, D., Jevnikar, A. M., Huner, N., et al. (2008). A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hGAD65). *BMC Biotechnol.* 8:87. doi: 10.1186/1472-6750-8-87
- Wigdorovitz, A., Carrillo, C., Dus Santos, M. J., Trono, K., Peralta, A., Gomez, M. C., et al. (1999). Induction of a protective antibody response to foot and mouth disease virus in mice following oral or parenteral immunization with

alfalfa transgenic plants expressing the viral structural protein VP1. *Virology* 255, 347–353. doi: 10.1006/viro.1998.9590

Conflict of Interest Statement: The author Stephen P. Mayfield declares a financial interest in Triton Animal Health, a company making orally available nutritional supplements, and potentially orally available vaccines, should these prove to be biologically functional.

Received: 28 November 2013; paper pending published: 23 December 2013; accepted: 30 January 2014; published online: 17 February 2014.

Citation: Specht EA and Mayfield SP (2014) Algae-based oral recombinant vaccines. *Front. Microbiol.* 5:60. doi: 10.3389/fmicb.2014.00060

This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Specht and Mayfield. 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.