Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites

Kim, Hyun Uk; Charusanti, Pep; Lee, Sang Yup; Weber, Tilmann

Published in:
Natural Product Reports

Link to article, DOI:
10.1039/C6NP00019C

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites†

Hyun Uk Kim,*ab Pep Charusanti,b Sang Yup Leeab and Tilmann Weber*b

1 Introduction

Microorganisms serve as an important source of secondary metabolites that have various medicinal and industrial uses.1 According to Newman and Cragg,2 who investigated the source of new drugs within a 30 year period from 1981–2010, 69% and 75% of newly introduced anti-infective and anti-cancer drugs, respectively, were natural products or their derivatives; this indicates the importance of natural products and the huge potential for the discovery of novel drug leads.3,4 Bacteria of the family Actinomycetaceae, in particular, are prolific producers of secondary metabolites. These bacteria are the biological source of many drugs that, for example, are used in the treatment of infections (e.g., streptomycin and erythromycin), as immunosuppressants (e.g., rapamycin), and as anthelmintics (e.g., ivermectins). In addition, many other genera, such as mycobacteria, bacilli, and many marine bacteria, have the capability to produce complex bioactive secondary metabolites. Bacteria, mainly of the Streptomyces genus, appear to be the fourth largest source of FDA-approved drugs at 16% overall.5 This amount is greater than the number of drugs derived from fungi (12% overall).

In most described cases, biosynthetic pathways for secondary metabolites are organized as biosynthetic gene clusters (BGCs), which means that genes encoding core biosynthetic enzymes (e.g., polyketide synthase and non-ribosomal peptide synthetase), genes encoding tailoring enzymes, genes encoding for specific precursor biosynthesis pathways, cluster-situated regulators and often genes encoding transporters or resistance factors are physically clustered on the chromosome. Thus, all required genes for the production of a secondary metabolite are encoded within such BGCs (Fig. 1).6

Recently, the discipline of metabolic engineering has increasingly been applied to the secondary metabolite studies to help boost commercial production of target molecules.7–9 The general objective of metabolic engineering is to overproduce chemicals that are valuable to mankind from microbial or mammalian cells, and was first coined in the field of biochemical engineering.9 By its definition, this discipline attempts to systematically understand and engineer a cell’s metabolic network at a systems level.9–11 Although conventional metabolic engineering takes a rational approach, a random approach such as adaptive laboratory evolution is also considered to be a part of metabolic engineering in recent years.12 Moreover, due to advances in systems and synthetic biology, a suite of high-throughput molecular and computational tools are increasingly deployed in the practice of metabolic engineering such that the field is now often called ‘systems metabolic engineering’.11 Systems metabolic engineering has been rigorously applied to platform production strains such as Escherichia coli and Saccharomyces cerevisiae for the production of various chemicals, biofuels and biopolymers.11 In this Highlight, both ‘metabolic engineering’ and ‘systems metabolic engineering’ are used interchangeably.

Traditionally, the metabolic engineering and secondary metabolite communities have focused on different goals. In
conventional metabolic engineering, quantitative values such as titer, yield and productivity for a target biochemical from a production host are more heavily emphasized. These quantitative values constitute a production host’s performance metrics in strain development. In contrast, the secondary metabolite studies seem to have focused more on the discovery of novel and diverse chemical structures and any possible bioactivity they might have. This focus stems partly from the unique biochemical logic of secondary metabolite BGCs, whose manipulation can give rise to diverse chemical structures as output. Manipulation of BGCs, including inactivation or exchange of domains in polyketide synthase and non-ribosomal peptide synthetase, engineering of a domain active site and tailoring enzymes, and shuffling of modules, can all lead to the production of secondary metabolites with novel structures, and has been a classical topic in the secondary metabolite community. With this background in mind, it has become important to view the production of secondary metabolites from the metabolic engineer’s perspective. A motivation is that once a structurally novel molecule is sufficiently determined to have commercial value, its production titer and yield need to be enhanced in order to implement larger-scale experiments, including (pre)clinical trials, and ultimately industrial production (Fig. 1). Such production optimization has recently become even more important because some secondary metabolites and their precursors have been identified by the metabolic engineering community to be important sources of industrial chemicals and biofuels that have conventionally been produced from petrochemical processes.

To this end, this Highlight discusses recent studies on the metabolic optimization of native producers and other heterologous hosts for enhanced production of secondary metabolites. In particular, we focus on prokaryotic secondary metabolites because of their high contribution to currently marketed drugs, and their potential as a source of industrial chemicals and fuels. First, we analyze recent trends in metabolic engineering and their potential for enhanced production of secondary metabolites. We then discuss future challenges, such as the development of mathematical models to predict the production of novel and diverse chemical structures and their potential bioactivities. Finally, we conclude with an outlook on the future of metabolic engineering of secondary metabolites.
engineering conducted to enhance the production of various secondary metabolites with emphasis on host selection and different types of engineering approaches used (i.e., rational, random or combined). This analysis is based on a survey of relevant studies reported since 2012. Second, we review various systems biology tools that have been applied to microbial hosts for the enhanced production of secondary metabolites. These tools are discussed in the context of current challenges encountered during the production of secondary metabolites.

2 Recent trends in metabolic engineering for production of secondary metabolites

2.1 Different considerations for host selection when producing primary and secondary metabolites

Selection of a production host determines the suite of tools to be deployed for strain optimization, and therefore is a very important starting point of metabolic engineering. While variables considered for selecting microbial host to produce primary metabolites are also applicable to secondary metabolite production studies, different priorities appear to exist when optimizing the production of primary and secondary metabolites as target bioproducts (Fig. 2). Representative primary metabolites considered for bioproducts include those biosynthesized in central carbon metabolism (e.g., ethanol, lactic acid and succinic acid), amino acids and recently fatty acids. Different priorities are largely caused by stark differences in the biochemistry of primary and secondary metabolites. Primary metabolites, in particular fermentation products and their derivatives, are often produced in high titers at the scale of grams per liter, whereas secondary metabolites are secreted at much lower levels, typically at the scale of micrograms or milligrams per liter. However, production of secondary metabolites can also reach the scale of grams per liter upon metabolic engineering. For this reason, units of titer, yield and productivity for production of primary metabolites are usually defined to be ‘g L⁻¹’, ‘g g⁻¹’ and ‘g L⁻¹ h⁻¹’, respectively, but can be varied for secondary metabolites.

A key consideration for host selection in conventional metabolic engineering (e.g., overproducing primary metabolites using model organisms) has been whether it is possible to maintain both high growth and production rates using defined minimal or industrial media, which are directly linked to the operation cost of a microbial bioprocess; cheaper nutrient utilization is always more favored from a bioprocess perspective. The availability of biosynthetic reactions for the production of a target biochemical is also an important criterion for host selection, but it has been somewhat overcome with state-of-the-art synthetic biology tools for model organisms. Industrially or medicinally valuable compounds, such as 1,4-butanediol and opioid compounds, have been successfully produced from engineered heterologous production hosts E. coli and S. cerevisiae, respectively, by constructing heterologous biosynthetic pathways.

Fig. 1 Three major stages that lead to the optimized production of secondary metabolites from secondary metabolite-producing microorganism. In the context of rational engineering, each indicated component (e.g., metabolite precursors, regulations, medium design, etc.) can be engineering targets. Five colored circles indicate different systems biology tools discussed, and are positioned near each component name where most applicable. Be noted, however, that applications of the shown systems biology tools are not necessarily confined to the indicated components. The word "Signals" in the figure can be any environmental conditions (e.g., aeration, co-culturing with another microorganism and temperature) or chemical elicitors (e.g., antibiotics at sub-lethal concentration and quorum sensing-dependent signaling molecules) that can influence the expression of secondary metabolite biosynthetic gene clusters.
Despite possession of secondary metabolite BGCs of interest, the challenge of capturing or cloning the BGC, which can be 100 kb in length or more, and optimally expressing genes within the BGC, advanced synthetic biological tools greatly aid this process.\(^\text{26-27}\) Finally, because general gene manipulation tools, such as knockout and overexpression of genes, have actively been applied to secondary metabolite-producing microorganisms as manifested in next section (Table S1†), they no longer appear to be a limiting factor for host selection when producing secondary metabolites. In the same line, state-of-the-art synthetic biology tools mostly developed with model organisms (e.g.,\(\text{E. coli}\)) are increasingly available to secondary metabolite producers.\(^\text{28,29}\)

### 2.2 Recent trends in host selection and engineering approaches: native versus heterologous hosts and rational versus random approaches

In order to gain insights on recent trends in host selection and engineering approaches deployed to optimize the production of secondary metabolites, we conducted a literature survey on relevant studies reported since 2012 (Table S1†). Our survey reveals that production studies using native producers outnumber those using heterologous hosts (Fig. 3A). For native producers, diverse hosts such as\(\text{Saccharopolyspora spinosa}\),\(\text{Streptomyces bingchengensis}\),\(\text{Streptomyces chattanoogensis}\) and\(\text{Streptomyces roseosporus}\), have actively been subjected to production optimization of their native secondary metabolites. These hosts are not necessarily model actinomycetacean species. Meanwhile, heterologous production hosts were not confined to actinomycetes such as\(\text{S. coelicolor}\)\(^\text{30}\) and\(\text{Streptomyces venezuelae}\)\(^\text{11}\) because biologically distant and/or model organisms such as\(\text{E. coli}\),\(\text{Bacillus subtilis}\)\(^\text{31}\) and\(\text{Myxococcus xanthus}\)\(^\text{32}\) were also considered (Table S1†). Production performances of these heterologous hosts often appeared to be worse or at least not better than native producers, and the secondary metabolite production studies currently seem to favor the use of native producers.\(^\text{33}\) However, a potential advantage of using a model organism as a heterologous production host (e.g.,\(\text{E. coli}\)) would be better access to the state-of-the-art synthetic biology approach in comparison with native secondary metabolite producers. In a recent study using\(\text{E. coli}\) for the deoxyviolacein production,\(^\text{34}\) deoxyviolacein biosynthetic pathway was divided into two modules with\(\text{L-tryptophan}\) as a key intermediate, and heterologous expression of each module was independently optimized. In this optimization procedure, a molecular biosensor was newly developed and used to detect the intracellular pool of\(\text{L-tryptophan}\) along with fluorescence-activated cell sorting. This approach led to a titer of 1.92 g L\(^{-1}\).

Meanwhile, studies on the production of secondary metabolites using native hosts (e.g., actinomycetes) or heterologous hosts have additional considerations in addition to those already considered for the production of primary metabolites often using model organisms (Fig. 2). Culturability of a native production host can be a major issue because some secondary metabolite-producing microorganisms are unculturable, and do not grow fast enough or with sufficient reproducibility despite possession of secondary metabolite BGCs of interest.\(^\text{35}\) If a target BGC never gets expressed in a native host, a heterologous host should be considered. Doing so introduces the challenge of capturing or cloning the BGC, which can be 100 kb

---

Fig. 2 Flowchart of systems metabolic engineering considered for secondary metabolite production. Steps in red are specific to the secondary metabolite production, and can be considered altogether as examining the candidate host strain’s native production capacity for a target bioproduct in the case of primary metabolite production. BAC and BGC stand for bacterial artificial chromosome and biosynthetic gene cluster, respectively. The step “Systems metabolic engineering” is also applicable to the primary metabolite production, and further details can be found elsewhere.\(^\text{11}\)
3 Systems biology tools to optimize production of secondary metabolites

As just mentioned, many metabolic engineering efforts to boost the production of secondary metabolite are centered on engineering native producers. In this case, systems biology tools tend to be more important than synthetic biology tools initially because one needs to gain insights into the fundamental biochemistry of native producers before actually engineering them. Therefore, we next discuss systems biology tools that have been used in the study on optimization of secondary metabolite production (Table S1†). We also discuss additional systems tools that can be further considered to overcome current challenges associated with secondary metabolite production.

3.1 Genome-scale metabolic models

Genome-scale metabolic modeling has become a popular rational approach to enhance the production of secondary metabolites.46,47 Genome-scale metabolic models continue to be an important tool in systems biology by predicting global metabolic flux distributions under given genetic and environmental conditions. A genome-scale metabolic model is a large-scale stoichiometric model that describes all the metabolic pathways experimentally and/or theoretically characterized through stoichiometric coefficients and mass balances of participating metabolites, and is simulated using numerical optimization.9 This modeling approach takes an assumption of pseudo-steady state, which can be best applied to simulating primary metabolism in an exponential growth phase.9 Metabolic questions that can be best addressed with genome-scale metabolic models include, but are not limited to, prediction of the most efficient pathway that leads to the maximal production yield of a target bioproduct, and optimization of precursor supply and intracellular redox balances, typically through prediction of the effects of gene knockouts and overexpressions.48,49 Due to its ease of implementation and relatively high predictive power, this modeling approach has contributed to a diverse array of applications,53,58 for example prediction of gene manipulation targets in metabolic engineering for enhanced biochemical production,21 and prediction of drug targets in microbial pathogens51 and cancers (e.g., hepatocellular carcinoma).52 Genome-scale metabolic models can be relatively easily created using the genome sequence of an organism.53–55 Several software programs have been introduced to automate a large part of the metabolic modeling procedure, which enable the rapid reconstruction of draft genome-scale metabolic models of multiple species.56,57

Fig. 3 Number of reported studies aimed at enhancing the production of prokaryotic secondary metabolites using (A) different types of production hosts and (B) engineering approaches (i.e., rational, random or combined). Summaries of each study are available in Table S1†. Search words in PubMed were: production[tiab] AND engineering[tiab] AND (secondary metabolites OR polypeptides OR nonribosomal peptides OR lactam OR glycopeptide OR macrolide) NOT plant AND 2012:2016(dp); (secondary metabolite[tiab] OR natural product[tiab] OR actinomycetes [tiab] OR streptomyces[tiab]) AND (design[tiab] OR engineering[tiab]) AND 2012:2016(dp).
Plackett–Burman design and response surface methodology, leading to identification of key medium components along with their optimal concentrations. Recently, this statistical optimization approach has been applied to the production of daptomycin from a S. roseosporus mutant strain,\(^6\) neomycin from Streptomyces fradiae,\(^6\) nosiheptide from Streptomyces actuosa,\(^6\) and pikromycin from S. venezuelae.\(^6\) In all these studies, minerals and/or carbon sources that most affected the secondary metabolite production were selected, and their optimal concentrations were determined using the Plackett–Burman design and response surface methodology. The DOE approach will continue to play an important role in optimizing multiple bioprocess variables, including microbial cultivation media.

### 3.3 \(^{13}\)C-Metabolic flux analysis

A similar approach to the aforementioned genome-scale metabolic modeling is \(^{13}\)C-metabolic flux analysis, which also uses information on mass balance of metabolites with their stoichiometric coefficients.\(^7\) However, the two differ in that \(^{13}\)C-metabolic flux analysis uses \(^{13}\)C-labelling data from isotopic labelling experiments in addition to a stoichiometric metabolic network model in order to estimate in vivo metabolic fluxes under a given condition. Despite its precise measurement of in vivo flux values, \(^{13}\)C-metabolic flux analysis has not been frequently deployed to analyze metabolism of secondary metabolite-producing microorganisms in the past when compared to model organisms.\(^6\) Recently, Coze et al. investigated differences in metabolic flux distributions of an actinorhodin-producing S. coelicolor wild-type and its mutant, in which its native four BGCs were removed.\(^6\) \(^{13}\)C-Metabolic flux analysis revealed a few insights, including a more active pentose phosphate pathway in the mutant, and a competition for common precursors such as acetyl-CoA between primary and secondary metabolism. Although this study was not intended for production optimization, the \(^{13}\)C-metabolic flux analysis used in this study could be useful in analyzing the metabolic status of engineered strains towards enhanced production. Finally, use of dynamic \(^{13}\)C-metabolic flux analysis should be useful in analyzing metabolic status during the secondary metabolite production phase that is in non-steady state.\(^6\) Because many complex regulations take place during the secondary metabolite production, conventional \(^{12}\)C-metabolic flux analysis based on the pseudo-steady state assumption, in the strict sense, is not an ideal approach to analyze this specific metabolic status. Also, dynamic \(^{12}\)C-metabolic flux analysis can be more advantageous for the analysis of microbial metabolism in fed-batch fermentations because this fermentation mode has very dynamic cultivation conditions. Fed-batch fermentations are predominantly conducted in industry.

### 3.4 A challenge of identifying metabolic and regulatory gene manipulation targets and further systems approaches to be considered

Identifying metabolic and/or regulatory genes responsible for the enhanced production of secondary metabolites stands as an
important challenge. We expect that additional systems biology tools available in the metabolic engineering community, which have not been applied to the secondary metabolite production yet, will help to meet this challenge.

First, the aforementioned genome-scale metabolic model contains information about the reactions that lead to secondary metabolite biosynthesis, but it does not contain any regulatory information. Probabilistic modeling approach is likely to be a good option to model regulation associated with secondary metabolite biosynthesis. As a relevant recent example, a regulatory network describing relationships between genes encoding transcription factors and their target metabolic genes was modeled by calculating conditional probabilities for *Mycobacterium tuberculosis*. In this modeling approach, conditional probabilities are used to describe relationship between transcription factors and their target metabolic genes; they reveal the probability that a metabolic gene gets expressed or inactivated, depending on the expression status of a transcription factor. This probabilistic information is used to calculate more accurate metabolic flux values using the genome-scale metabolic model of *M. tuberculosis*. A similar integrative modeling approach can be considered to identify regulatory genes that can boost the expression of metabolic genes related to the biosynthesis of a target secondary metabolite.

The above probabilistic regulatory modeling approach in turn requires a large transcriptome dataset to bestow greater reliability on the calculated conditional probabilities for metabolic genes and transcription factors. Of particular importance is that the large transcriptome (or other omics) dataset needs to be obtained from as many different environmental and/or genetic conditions as possible in order to accurately determine the conditional probabilities that describe relationships between genes. Massive transcriptome analyses have been conducted for *S. coelicolor* under multiple conditions in the past, for example, elucidating nutritional stress response of an antibiotic regulator AfsS. A genome-wide regulatory network, genome-wide gene expression changes during a metabolic switch from exponential to stationary antibiotic production phase, and sigma factor-regulated genes in germination. Despite several massive transcriptome analyses conducted on actinomycetes (or heterologous model hosts), they have barely been deployed to enhance the production of secondary metabolites. Generation of massive omics data in the context of secondary metabolite production should be very important resources for optimizing the whole bioprocess.

3.5 A starting point of systems biology and metabolic engineering studies for secondary metabolites: the Secondary Metabolite Bioinformatics Portal

BGCs of secondary metabolites are a very complex system, and therefore optimization of secondary metabolite production requires relevant insights before actual metabolic engineering for production optimization begins. This step can be particularly challenging for metabolic engineers who are not familiar with the biology of actinomycetes. To this end, we recently released the Secondary Metabolite Bioinformatics Portal (SMBP) available at http://secondarymetabolites.org/, which provides a full list of databases and tools dedicated to secondary metabolite bioinformatics, along with their descriptions and URLs. This portal provides a concise gateway to various bioinformatic resources and tools that can facilitate metabolic engineering of actinomycetes and heterologous model hosts, including databases of secondary metabolites and BGC mining tools (e.g., antiSMASH). There are also other software tools with more generic applications, for example CRISPY-web, an application that supports the design of guideRNAs (sgRNAs) for CRISPR-Cas9 mediated genetic manipulation of microorganisms.

3.6 Perspectives on the direction of secondary metabolite production studies

It should be noted that systems biology tools used in metabolic engineering are complementary to tools and strategies that have been uniquely developed and used for secondary metabolite studies. An obvious reason is that the production of secondary metabolites needs unique considerations that are distinct from those of the primary metabolite production (Fig. 2). For the optimal production and potential commercialization of secondary metabolites, following considerations can be useful. First, processes of systems metabolic engineering need to be taken into account at an early phase of the secondary metabolite production study (e.g., genome mining and host selection). This will enable decision-making not only from a pure biochemistry perspective, but also from an engineering perspective, leading to an integral pipeline from novel secondary metabolite identification to its optimal production. Second, metabolic engineers also need to have a better understanding of the working mechanism of secondary metabolite BGCs. This will facilitate introduction of state-of-the-art tools to secondary metabolite producers, which were initially developed for model organisms. With recent efforts, more relevant achievements are awaiting to be realized.

4 Conclusions

As the discipline of metabolic engineering has expanded into the realm of prokaryotic secondary metabolites, this Highlight aimed to review the current status of metabolic engineering for secondary metabolite production and the relevant systems biology tools used. While our analysis manifested the progress made for the optimal production of secondary metabolites, it also clearly pinpointed room for further development. In particular, state-of-the-art systems biology tools established for the metabolic engineering community, including integrative metabolic and regulatory modeling, (dynamic) C-metabolic flux analysis and omics data generation, have not been fully deployed in the production optimization of secondary metabolites. Upon successful implementation of such tools, more diverse secondary metabolites will be considered for industrial production and commercialization in both medicinal and chemical industries.
5 Acknowledgements

This work was supported by the Novo Nordisk Foundation. H.U.K. and S.Y.L. are also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries from the Ministry of Science, ICT and Future Planning (MSIP) through the National Research Foundation (NRF) of Korea (NRF-2012M1A2A0206556 and NRF-2012M1A2A026557).

6 Notes and references

12 M. Dragosits and D. Mattanovich, Microb. Cell Fact., 2013, 12, 64.