Review

CellPress

Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes

Tilmann Weber[1], Pep Charusanti[1,2], Ewa Maria Musiol-Kroll[1], Xinglin Jiang[1], Yaojun Tong[1], Hyun Uk Kim[1,3], and Sang Yup Lee[1,3]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Alle 6, Hørsholm, Denmark
[2] Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA
[3] Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, BioInformatics Research Center, and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea

Actinomycetes are excellent sources for novel bioactive compounds, which serve as potential drug candidates for antibiotics development. While industrial efforts to find and develop novel antimicrobials have been severely reduced during the past two decades, the increasing threat of multidrug-resistant pathogens and the development of new technologies to find and produce such compounds have again attracted interest in this field. Based on improvements in whole-genome sequencing, novel methods have been developed to identify the secondary metabolite biosynthetic gene clusters by genome mining, to clone them, and to express them in heterologous hosts in much higher throughput than before. These technologies now enable metabolic engineering approaches to optimize production yields and to directly manipulate the pathways to generate modified products.

Actinomycetes as sources for novel drugs

For more than 70 years, actinomycetes (order *Actinomycetales*) have been recognized as important sources for bioactive natural compounds. From the roughly 18 000 known bioactive bacterial compounds, more than 10 000 were described from bacteria of the actinomycete genus *Streptomyces* [1]. Many well-known antibiotics, such as tetracycline, erythromycin, vancomycin, and streptomycin, originate from the secondary metabolism of actinomycetes. Beyond antibiotics, other medically useful natural products that were isolated from this group of bacteria include the immunosuppressant rapamycin, the anticancer agents doxorubicin and bleomycin, the anthelmintic avermectin, and the antifungal compound nystatin.

The traditional approach to small-molecule discovery from microbial sources such as actinomycetes has generally involved cultivation of the microbes under different growth conditions, extraction of the metabolites, and analysis of the extract for bioactivity (e.g., antimicrobial activity) in a chosen assay. Once a bioactive extract is identified, a more-detailed analysis is performed, normally involving chromatography-based separation of the individual constituents, to identify the specific bioactive molecules. Very often, however, this enormous effort leads to the rediscovery of known molecules, a fact that dampened enthusiasm for natural product discovery from actinomycetes over the past two decades.

Although this general strategy is still applied today, several recent developments have renewed enthusiasm for natural product discovery from actinomycetes. Genome sequence analysis from multiple actinomycetes indicates that each bacterium can produce approximately 10-fold more secondary metabolites than has been detected during screening efforts before the availability of the genome sequence data. For this reason, actinomycetes continue to be promising sources of novel bioactive compounds [2]. In addition, the availability of new metabolic engineering strategies now provides alternative approaches to streamline and accelerate the discovery and production of bioactive natural products from microbial or metagenomic sources. Metabolic engineering is a well-established discipline that systematically engineers microbial strains for the overproduction of natural and non-natural chemical compounds that are useful to mankind [3] (Figure 1). Although similar rationales can be applied to actinomycetes, engineering actinomycetes is more difficult than engineering model organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, because actinomycetes possess more diverse genomic content and biochemical machinery [4] (Figure 1). We review tools and methods recently developed for the effective metabolic engineering of actinomycetes, and discuss how these tools enable the generation of microbial cell factories for the production of antibiotics and other secondary metabolites.

Genome mining for the detection and identification of secondary metabolite biosynthetic gene clusters

The protein machinery responsible for the biosynthesis of secondary metabolites in bacteria is encoded by distinct

Corresponding author: Lee, S.Y. (leesy@kaist.ac.kr).
Keywords: actinomycetes; secondary metabolites; metabolic engineering; antibiotics.

0167-7799/
 http://dx.doi.org/10.1016/j.tibtech.2014.10.009

Glossary

A-domain: adenylation domains of non-ribosomal peptide synthetase (NRPS) modules select, activate amino acids as amino acid adenylates, and transfer them to the peptidyl carrier protein domains of NRPS modules.
AT: acyltransferase domains encoded in modular polyketide synthase specifically select the acyl-CoA building-blocks and transfer them to acyl carrier proteins (ACPs).
ACP: acyl carrier proteins contain a phosphopantetheine prosthetic group tethering the reaction intermediates during polyketide biosynthesis.
BAC: bacterial artificial chromosomes are replicable circular DNA vectors developed from the F-plasmid. They are usually used for cloning long fragments of 150–350 Kb.
CRISPR–Cas9: clustered regularly interspaced short palindromic repeats (CRISPR)-associated nuclease Cas9 is an RNA-guided endonuclease using RNA–DNA base-pairing to cleave target DNA, which is demarcated by protospacer adjacent motif (PAM) sequences. PAM sequences are short nucleotide motifs that are specifically recognized and required by Cas9 for DNA cleavage.
DH: the dehydratase domain of a PKS module is responsible for dehydrating the β-OH group of a polyketide intermediate.
ER: the enoylreductase domain of a PKS module reduces a double bond to a single bond between two adjacent extender units.
HDR: homology-directed repair is a template-dependent pathway for double-strand break (DSB) repair.
KR: the ketoreductase domain of a PKS module is responsible for reducing the β-keto of the polyketide intermediate group to a hydroxyl group.
NRPS: non-ribosomal peptide synthetases specifically activate and condense proteinogenic or non-proteinogenic amino acids in an assembly-line fashion. They are involved in the biosynthesis of many peptide antibiotics such as, for example, vancomycin.
PKS: polyketide synthases condense acyl-CoA units to form a polyketide. There are three types of PKS: type I PKSs are homologous to type I fatty acid synthases that are found for example in mammals. They act either iteratively or in an assembly-line fashion. Type II PKSs are homologous to type II fatty acid synthases as found in many bacteria. Their products usually are aromatic polyketides. Type III PKSs are homologs of plant chalcone/stilbene synthases.
Red/ET: Red/ET recombineering is a method to insert foreign DNA into chromosomes or plasmids based on short homologous sequence regions (<50 bp). The system makes use of the exonucleases/recombinases Redα/Redβ from λ phage or RecE/RecT from Rac prophage.
SMBGC: secondary metabolite biosynthetic gene clusters contain all genes required for biosynthesis, regulation, export, and very often resistance of natural products/secondary metabolites, thus all genes are encoded side-by-side. With only very few exceptions, bacterial secondary metabolite biosynthesis pathways are always organized in SMBGCs. However, SMBGCs can also be found in fungal producers.
TALEN: transcription activator-like effector nucleases are fusions of the *Fok*I cleavage domain and DNA-binding domains derived from TALE proteins. TALEs contain multiple 33–35 amino acid repeat domains, each of which recognizes a single base pair.
TE: the thioesterase domain releases the polyketide or peptide intermediate from the PKS or NRPS assembly line; in many cases the release is combined with a macrolactonization, yielding cyclic molecules.
ZFN: zinc-finger nucleases are fusions of the non-specific DNA cleavage domain from the *Fok*I restriction endonuclease with zinc-finger proteins. ZFN dimers induce targeted DNA DSBs that stimulate DNA damage response pathways.

gene clusters within their genomes, making the identification of these clusters through genome mining an important task [4]. Several major developments have advanced such genome-mining efforts, in particular the development of next-generation sequencing technologies, increased knowledge about the secondary metabolism, and novel mass spectrometry (MS) detection tools. The resulting biological information has been incorporated into several databases and software tools for the genome-wide prediction of gene clusters, the analysis of their corresponding secondary metabolite biosynthetic pathways, and the prediction of substrate specificities [5]. The currently most widely used software to identify such gene clusters is the antibiotics and secondary metabolites analysis shell (antiSMASH) [6]. antiSMASH includes rule-based as well as statistics-based algorithms to identify secondary metabolite biosynthetic gene clusters (SMBGCs; see Glossary) and offers various modules for analyzing the relevant pathways.

In addition to computational tools, genome mining is typically accompanied by proteome and/or metabolome analyses using MS for accurate linking of a target secondary metabolite and its biosynthetic gene cluster [7,8]. The biosynthesis of several classes of secondary metabolites follows a conserved biochemical logic; this is utilized to match secondary metabolite-derived analytical data with genes in the genome of the target organism [9,10]. In addition to the biochemical logic, high-quality genome annotation data for the target organism (e.g., high coverage and accurate open reading frame predictions) and the availability of sufficiently sensitive mass spectrometers determine the success of this MS-based genome mining. In this approach, the analytical data of secondary metabolites generated by MS (i.e., MS^n data) provide specific fragment patterns which contain amino-acyl or glycosyl 'tags'. Amino-acyl tags can be searched against amino acid building blocks predicted from the genome of the target organism using the conserved biochemical logic for ribosomal or non-ribosomal peptides [11–13].

Along the same lines, glycosyl tags from glycosylated natural products can be linked to their corresponding biosynthetic genes among all the glycosylation genes which are initially characterized by mining the genome of the target organism. These methods can be considered either peptidogenomics or glycogenomics, depending on the use of amino-acyl (or peptidyl) or glycosyl 'tags' obtained from the target secondary metabolites, respectively, but the general rationale of their approaches remains the same. Use of peptido- and glycogenomics led to the discovery of novel analogs of the antibiotics stendomycin from *Streptomyces hygroscopicus* ATCC 53653 [11], and arenimycin B from the marine actinobacterium *Salinispora arenicola* CNB-527, an antibiotic effective against multidrug-resistant *Staphylococcus aureus* [12], respectively. Although these methods might need to be tailored for secondary metabolites with hybrid or highly modified structures (e.g., non-ribosomal and polyketide hybrids), they can still be applicable to a variety of compounds to rapidly identify their biosynthetic genes. In another recent example, significant correlations between protein expression levels and the activities of target secondary metabolites under several different growth conditions were validated and used to link target secondary metabolites to their specific gene clusters [14]. Once positive links were identified, the respective proteins were mapped onto the predicted gene clusters of the target organism through quantitative proteomic expression data.

After the gene cluster for a target secondary metabolite has been identified, metabolic engineering can be conducted by using a microbial strain that natively harbors the specific gene cluster provided that it has adequate growth characteristics and is amenable to genetic manipulation. Otherwise, the identified gene cluster can be cloned and expressed in a heterologous host. The recent development of new cloning techniques has now greatly expedited this process.