Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

A reinvigorated era of bacterial secondary metabolite discovery

Ashley M Sidebottom[1] and Erin E Carlson[1,2,3]

Secondary metabolite discovery from bacteria has become increasingly successful in the last decade due to the advancement of integrated genetic-based, spectrometric-based and informatics-based techniques. Microbes and their unique metabolic outputs have been widely studied since the beginning of modern medicine; however, it is well known that the current repertoire of secondary metabolites, or more commonly natural products, is incomplete and the understanding of natural product-mediated intracellular dialog is in its infancy. Here, we highlight the present state of bacterial metabolomics including compound discovery approaches and new strategies for probing the role of these molecules within communication networks.

Addresses
[1] Department of Chemistry, Indiana University, 800 East Kirkwood Avenue, Bloomington, IN 47405, USA
[2] Department of Molecular and Cellular Biochemistry, Indiana University, 212 South Hawthorne Drive, Bloomington, IN 47405, USA
[3] Present permanent address: Department of Chemistry, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455, USA

Corresponding author: Carlson, Erin E (carlsone@umn.edu)

Current Opinion in Chemical Biology 2015, **24**:104–111

This review comes from a themed issue on **Omics**

Edited by **Benjamin F Cravatt** and **Thomas Kodadek**

For a complete overview see the Issue and the Editorial

Available online 2nd December 2014

http://dx.doi.org/10.1016/j.cbpa.2014.10.014

Introduction

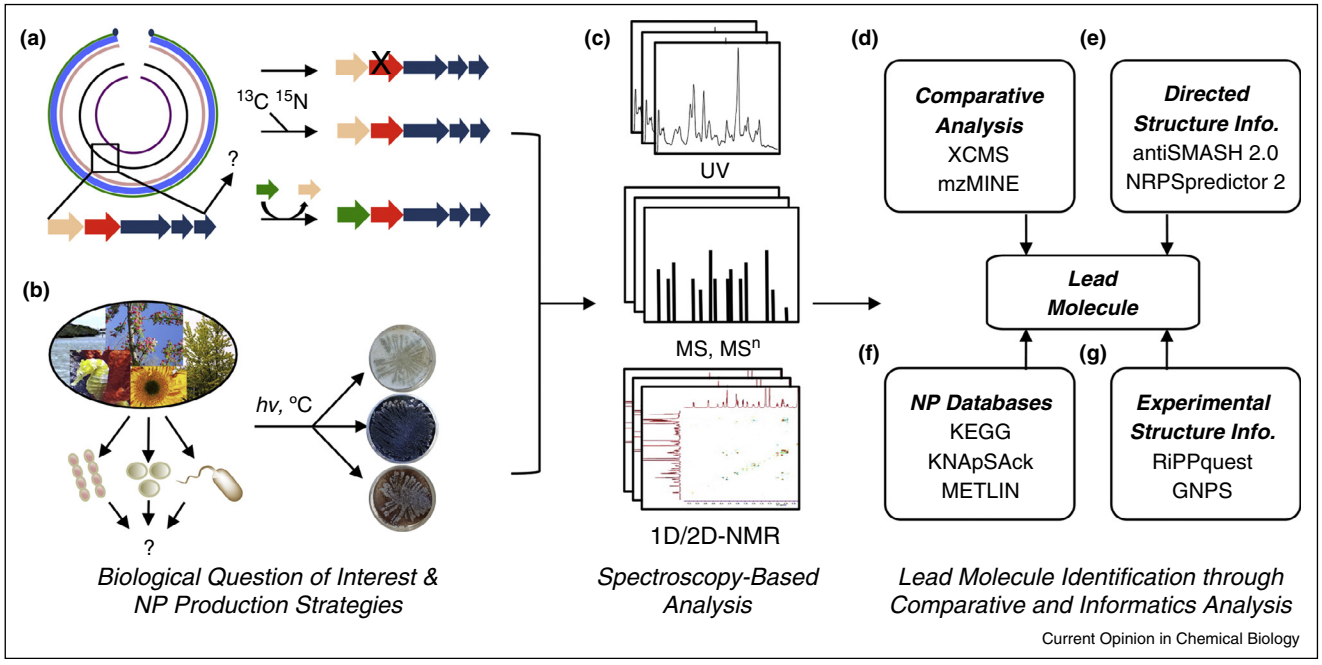
Microbial species are exceptional sources of therapeutically relevant secondary metabolites, or natural products (NPs). These compounds are produced by bacteria for functions ranging from defense to nutrient acquisition and have been harvested by modern medicine for use as antibacterials, antifungals, anticancer agents and other treatments. While NPs have been deemed non-essential for organism survival, they are involved in relationships that promote environmental advantages for both the producing and surrounding organisms and their continued elucidation is key to uncovering novel compound scaffolds and pathways.

In 1929, the 'golden age' of NP exploration was founded with the discovery of penicillin G, a β-lactam antibiotic from *Penicillium notatum* [1]. Since that time, a number of drug classes have been deployed including β-lactams (1940s), tetracyclines (1950s), sulfonamides (1940s), aminoglycosides (1950s), glycopeptides (1950s) and cephalosporins (1960s), many of which were isolated from bacterial species. The identification of novel compounds from these sources has declined in recent decades partially due to the generally accepted belief that microbial genomes had been fully exploited. In addition, complicating factors such as molecule rediscovery, a lack of complete compound databases, and limited utilization of sensitive and high-resolution detection methods such as mass spectrometry, reduced the rate of discovery. The field was reinvigorated following sequencing of several microbial genomes, namely *Streptomyces coelicolor* M145 (2002) and *Salinispora tropica* (2007), when it became clear that approximately 70% of the products from putative NP-producing gene clusters (~9% of the genome) remained uncharacterized [2,3]. Current estimates predict that 10^9 NPs remain to be characterized [4]. To access this outstanding biosynthetic potential, targeted (*a priori* knowledge required), semi-targeted (limited knowledge required) and untargeted (no or minimal knowledge required) methods have been applied. These strategies are briefly discussed and their primary roles in secondary metabolomics are highlighted.

A genomic guide to discovery: targeted approaches

Targeted secondary metabolite discovery is driven by prior knowledge of gene organization and their resulting, putative NP structures (Figure 1a). Commonly known as genome-mining, this strategy uses bioinformatics tools to compare conserved sequences across species and strains for the identification of previously uncharacterized compounds [5]. Targeted NP discovery can be pursued using a number of genetic manipulation strategies such as heterologous expression of orphan clusters, overexpression or inactivation of the gene of interest, repressor silencing, and promoter replacement [6–10]. Precursor molecules can also be leveraged either by simple perturbation of the availability of these compounds or use of isotopically labeled derivatives (Figure 1a). Historically, genome-mining of non-ribosomal peptide synthase (NRPS) and polyketide synthase (PKS) clusters has been highly successful due to their homologous, multi-domain and multi-modular arrangement within the genome [11]. These subclasses have remained the primary targets of most studies due to their noteworthy use as therapeutic compounds. Examples

Figure 1



Targeted and untargeted secondary metabolite discovery general scheme. **(a)** Targeted secondary metabolite discovery utilizes prior genetic information to predict gene clusters of interest and their putative products. Methods such as gene inactivation, isotopic precursor labeling, and promoter exchange are commonly used genome-directed methods. **(b)** Untargeted strategies apply a variety of growth conditions to organisms isolated from unique environments or that are members of a prolific genus. **(c)** Spectroscopy-based analysis tools such as UV, MS/MS^n, and 1D/2D-NMR are used for NP detection. Next, a highly varied workflow is employed based on project goal for the elucidation of the lead molecule(s), but most commonly includes a combination of the following: **(d)** comparative analysis (peak-picking, alignment, statistical analysis), **(e)** incorporation of structural information known prior to analysis (directed), **(f)** NP databases, and **(g)** experimentally derived structure information (MS^n fragmentation, UV/MS/NMR scaffold signatures).

of recently characterized compounds and their respective discovery strategies are in Figure 2 [12–16].

The utility of genome-mining tools has increased in the last decade with incorporation of robust spectroscopic analysis methods (Figure 1c). In particular, mass spectrometry has become a powerful tool for analysis of targeted experiments due to the sensitivity, versatility and dynamic range capabilities of this technique. Commonly, putative NP structures are searched for by their predicted precursor components (e.g. amino acids) within tandem mass spectrometry-generated fragmentation spectra (MS^n). For example, peptidogenomics explicitly identifies ribosomal peptides (RPs) and non-ribosomal peptides (NRPs) in an MS^n-guided workflow [13,17,18]. Dorrestein and coworkers have demonstrated the utility of this strategy by detection of 14 unique peptides, including stendomycin II, from several well-studied *Streptomyces* spp. (Figure 2e). In a similar approach, informatipeptin, a lanthipeptide, from *Streptomyces viridochromogenes* DSM 40736 was discovered following the development of a database platform aimed at the identification of ribosomally synthesized and posttranslationally modified (RiPPs) molecules (RiPPquest; Figure 1g) [19••]. Importantly, these approaches are applicable to the analysis of both crude NP samples and genetically engineered secondary metabolomes. This is crucial for continued discovery of NPs from complex and diverse bacterial sources, as it is unlikely to be feasible to design genetically ideal systems for detection of the majority of unexplored NP space.

Significant challenges remain in the characterization of low frequency or unprecedented genetic operons such as hybrid clusters, namely NRPS/PKS, as well as new compound families and those that contain unique modifications and/or transformations such as heterocycles [20]. However, growing information regarding biosynthetic pathways and subsequent compound tailoring has led to increased discovery of compounds such as the methanobactins from *Methylocytis spp*, phenazines from several marine-derived bacteria and ribosomally derived thiopeptides [14,20,21]. Future targeted endeavors will benefit from the inclusion of additional strategies for detection and prioritization as the number of known precursors, subclasses, and post-translational modifications continues to grow. For example, a recent study combined genome-mining with traditional bioactivity-guided methods for the discovery of lomaivitiicin C from *S. tropica* (Figure 2a) [12].