Natural product discovery from the human microbiome

Matthew R. Wilson¹, Li Zha, and Emily P. Balskus²
From the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Edited by Ruma Banerjee

Human-associated microorganisms have the potential to biosynthesize numerous secondary metabolites that may mediate important host-microbe and microbe-microbe interactions. However, there is currently a limited understanding of microbiome-derived natural products. A variety of complementary discovery approaches have begun to illuminate this microbial “dark matter,” which will in turn allow detailed mechanistic studies of the effects of these molecules on microbiome and host. Herein, we review recent efforts to uncover microbiome-derived natural products, describe the key approaches that were used to identify and characterize these metabolites, discuss potential functional roles of these molecules, and highlight challenges related to this emerging research area.

Context for secondary metabolite discovery

Secondary or specialized metabolites (natural products) represent a vast array of small molecules produced by many organisms, including microorganisms (1). Since the mid-20th century, activity-based screening of cultivatable microbes, mostly Actinomycetes soil bacteria, and subsequent isolation efforts enabled the discovery of numerous bioactive secondary metabolites, including various polyketides (PKs),³ nonribosomal peptides (NRPs), and hybrids thereof (2). Advances in DNA sequencing, molecular biology, and cultivation methods have also enabled the more recent discovery of secondary metabolites from unculturable microorganisms (3–6). Both microbial natural products and semisynthetic derivatives have significantly impacted human health, serving as a major source of clinically used antibiotics, antifungals, anticancer agents, immunosuppressants, and other drugs (7). Although natural product discovery has predominantly focused on medicinal applications, recent studies have begun to shed light on the fascinating ecological roles of secondary metabolites in producing organisms and their surrounding environments, including the human body (8, 9).

Secondary metabolites from the human microbiota

Trillions of microorganisms colonize the human body and significantly influence human health and disease. These organisms are collectively known as the human microbiome and possess 100-fold more genes than the human genome (10, 11). A recent analysis of genomic and metagenomic sequencing data from the Human Microbiome Project (HMP) showed that human-associated bacteria encode the biosynthetic machinery to synthesize a vast array of secondary metabolites (12). However, the identities of nearly all of these natural products are unknown, and their biological roles are not well understood. Uncovering the mechanisms by which these molecules mediate host-microbe and microbe-microbe interactions could reveal factors influencing the microbiome’s effect on the host, a source of narrow-spectrum therapeutics, and inspire new strategies for treating human disease.

In this Minireview, we describe several recent efforts to discover natural products from human-associated bacteria to highlight strategies for further exploring the biosynthetic capacity of the human microbiome. We also briefly discuss potential biological roles of these metabolites. Our goal is not to provide a comprehensive overview of all small molecules isolated from the human microbiome as this topic has been reviewed recently (13–18). Instead, we showcase how distinct approaches for natural product discovery can be used alone or in combination to unravel the complex metabolic interplay within the human microbiome.

Sequence-based metagenome mining: Lactocillin

Advances in DNA sequencing technologies and the accompanying increase in the available (meta)genomic sequencing data have enabled computational strategies for natural product discovery (19, 20). Genome mining approaches take advantage of the fact that the genes responsible for constructing microbial secondary metabolites are typically co-localized in biosynthetic gene clusters (BGCs). Automated programs such as Cluster-Finder, a hidden Markov model-based probabilistic algorithm, have been developed to rapidly and accurately identify BGCs from DNA sequencing data (21). Recently, Fischbach and co-workers (12) used ClusterFinder to identify more than 14,000 BGCs from HMP reference genomes. These human-associated...
BGCs are predicted to synthesize a broad range of natural product classes, including saccharides, NRPs, PKs, and ribosomally encoded and post-translationally modified peptides (RiPPs). Based on the widespread distribution of RiPPs across multiple body sites and their potential clinical relevance, Fischbach and co-workers (12) searched HMP shotgun metagenomic assemblies for pathways encoding thiopeptides, a well-studied group of RiPPs. Of the 13 identified thiopeptide BGCs, bcg66 was selected for further study as it resembled the well-characterized BGC that produces the antibiotic thiocillin (22). Comparative metabolomics between a wild-type Lactobacillus gasseri strain harboring the bcg66 cluster and an isogenic mutant successfully guided the isolation of lactocillin (Fig. 1). As L. gasseri is a vaginal isolate, this work represents the first structural characterization of a thiopeptide natural product from a member of the human microbiome.

The structural similarity of this human-associated thiopeptide to thiocillin suggested that lactocillin might exhibit antibacterial activity against Gram-positive organisms. As expected, purified lactocillin was active against Gram-positive pathogens, including residents of the urogenital tract such as Staphylococcus aureus, Enterococcus faecalis, and Corynebacterium aurimucosum. Interestingly, lactocillin did not show antibacterial activity against other vaginal Lactobacillus strains, raising the possibility that lactocillin production may have evolved to prevent the colonization of harmful organisms without affecting nearby beneficial bacteria. Additional colonization experiments with the wild-type L. gasseri strain and the lactocillin isogenic mutant are needed to test this proposal.

Overall, the lactocillin case study illustrates how sequence-based discovery approaches can rapidly identify BGCs from both genomes and complex metagenomes. The ongoing application of these tools, along with continued advances in DNA sequencing and synthesis as well as improved approaches for heterologous expression of BGCs, will continue to provide access to the chemical output of the human microbiome (23).

Functional metagenome mining: Commendamide

As detailed above, sequence-based approaches can provide structural information about natural products prior to isolation. However, a drawback to this approach is its inability to yield functional information about secondary metabolites that are not linked to cultivatable strains. Functional metagenomics is an alternative strategy that can address this issue. This approach involves extracting DNA from an environmental sample (eDNA) and heterologously expressing this genetic information in an easily cultivable host such as Escherichia coli (Fig. 2) (24). The resulting clone libraries are then screened for a particular phenotype (for example antibacterial activity) or the presence of specific genes (for example biosynthetic genes), and the resulting clones of interest are then analyzed for the production of clone-specific metabolites (25). Although functional metagenomic studies have successfully identified bioactive metabolites from soil eDNA (26, 27), this approach was not applied to the human microbiome until relatively recently.

In 2015, Brady and co-workers (28) reported the identification of commendamide, a gut microbial metabolite involved in a host-microbe interaction (Fig. 2). They discovered this metabolite by screening E. coli clone libraries constructed from human stool metagenomic DNA. A cell-based reporter assay was used to identify clones that modulated nuclear factor-H9260B (NF-H9260B), an important transcription factor that responds to a variety of cellular processes. A total of 26 unique bacteria effector genes (Cbegs) were identified from examining a 75,000-member cosmid library. One effector gene family, Cbeg12, was annotated as an N-acyltransferase. Heterologous expression of Cbeg12 and metabolite profiling revealed the production of N-acyl-3-hydroxy palmitoyl glycine, a new natural product that was named commendamide. Cbeg12 was identified in several commensal Bacteroides genomes, and cultivation of Bacteroides vulgatus confirmed production of this metabolite.
Commendamide resembles various long-chain N-acyl amino acids identified in soil microbes and humans (29). These mammalian and bacterial signaling molecules exhibit a broad range of activities, including antibacterial (29), N-type calcium channel activation (30), and agonist activity against diverse G-protein-coupled receptors (GPCR) (31). A broad screen for agonist activity against 242 GPCRs revealed that commendamide specifically activates GPCR132/G2A, a critical receptor that regulates cellular proliferation and immunity (32). Disruption or dysregulation of G2A activity is associated with autoimmune syndrome (33) and atherosclerosis (34). Thus, it is possible that the commensal B. vulgatus produces commendamide to interact with the host immune system via G2A modulation.

The discovery of commendamide showcases the ability of functional metagenomics to rapidly link a single open reading frame present in stool metagenomic DNA to a biologically active metabolite. In the near future, the development of additional heterologous hosts better suited for expression of DNA from prominent members of the human microbiome will undoubtedly increase the efficiency of this natural product discovery strategy (35).

**Ecological approach: Lugdunin**

Studying the chemical ecology of members of the human microbiome is another powerful approach for identifying natural products that mediate interactions in this environment (8, 9). Instead of isolating a natural product and then searching for its biological function, an ecology-based approach starts with a relevant biological activity and then identifies the molecule(s) responsible. Several methods exist for connecting phenotypes to metabolites, including traditional bioassay-guided fractionation, targeted gene knock-out studies, and metabolomics. Ecology-guided strategies have led to the discovery of many biologically active secondary metabolites from microbes living together with marine sponges, plants, fungi, invertebrate animals, and humans (8, 36).

Investigation of a potential ecological interaction within the human nasal microbiome led to the recent discovery of lugdunin, a secondary metabolite produced by the commensal bacterium *Staphylococcus lugdunensis* (37). The potentially beneficial properties of this organism were first illuminated when Zipperer *et al.* (37) made the key observation that *S. lugdunensis* strongly antagonized the human pathogen *S. aureus* (Fig. 3). Transposon mutagenesis traced this observed activity back to a 30-kb gene cluster encoding NRPS and additional biosynthetic enzymes, suggesting that an NRP was likely responsible for the observed antibacterial activity. Overcoming initial cultivation hurdles, Zipperer *et al.* (37) isolated lugdunin, an NRP antibiotic containing several additional features that could not be predicted using current bioinformatic analysis methods (Fig. 3).

These isolation efforts, along with chemical synthesis, provided sufficient quantities of lugdunin to further explore its functional role in the nasal cavity. Minimum inhibitory concentration assays demonstrated that lugdunin exhibits potent antimicrobial activity against Gram-positive bacteria, including vancomycin-resistant *Enterococcus* sp. and *S. aureus*, with no detectable signs of resistance. This commensal bacterial metabolite also reduced and in some cases completely eradicated *S. aureus* colonization rates in both skin and animal infection models. Moreover, epidemiological studies showed that *S. aureus* colonization rates are significantly lower in humans colonized with *S. lugdunensis* as compared with *S. lugdunensis*-negative patients. These studies suggest that lugdunin production by *S. lugdunensis* may interfere with *S. aureus* colonization in vivo.

Interspecies competition is a relatively common occurrence between human-associated microbes, particularly in nutrient-poor body sites such as the skin microbiota (38). Analyzing the composition of microbial communities from these sites using new statistical tools could reveal additional pairs of microbes that co-occur or are differentially present, revealing symbiotic, mutualistic, or antagonistic interactions to guide further natural product discovery (39).

**In vitro biochemistry: Colibactin**

The full or partial in vitro reconstitution of a cryptic natural product biosynthetic pathway is an unconventional strategy for
gaining structural information about a secondary metabolite that can be uniquely enabling in certain situations. This approach involves characterizing the products that arise from combining purified biosynthetic enzymes with potential cofactors and substrates (40, 41). Unlike heterologous expression of a biosynthetic gene cluster, an \textit{in vitro} approach can facilitate an understanding of the detailed chemical mechanisms involved in synthesizing a natural product (42). This strategy also bypasses potential problems with microbial cultivation and regulating gene cluster expression \textit{in vivo}. Finally, partial \textit{in vitro} reconstitution yields structural information that can assist with isolation efforts as exemplified by the ongoing characterization of colibactin, a secondary metabolite made by commensal and pathogenic \textit{E. coli}.

In 2006, Oswald and co-workers (43) reported that certain \textit{E. coli} strains cause cell cycle arrest and DNA double-strand breaks in mammalian cell lines. Transposon mutagenesis linked genotoxicity to a 54-kb biosynthetic gene cluster (the \textit{pks} island) that encodes both polyketide synthase (PKS) and NRPS enzymes. This biosynthetic machinery was hypothesized to produce a small molecule genotoxin, colibactin. Although colibactin is produced by strains of the most well-studied bacterial species, the active genotoxin has eluded all isolation attempts, and its chemical structure is currently unknown.

The first insights into colibactin's structure originated from comparative bioinformatic analyses of the \textit{pks} gene cluster and \textit{in vitro} biochemical characterization of the "prodrug" resistance mechanism and aminomalonate-forming and -utilizing enzymes. Bioinformatics suggested that colibactin was initially made as an inactive \textit{N}-acylated precursor (precolibactin) that could be processed by a peptidase in the final stages of biosynthesis, releasing the active natural product. Based on this hypothesis, we biochemically characterized ClbN and ClbB, the NRPS modules predicted to initiate colibactin biosynthesis, and we showed that they synthesize and elongate an \textit{N}-acyl-D-Asn "prodrug motif." Moreover, we demonstrated that a periplasmic peptidase, ClbP, removes this scaffold from precolibactin...
MINIREVIEW: Natural products from the human microbiome

mimics. The relevance of this reactivity for colibactin biosynthesis was later supported by the successful isolation and structural characterization of the hydrolyzed “prodrug” motif from colibactin-producing E. coli (45) and in vivo studies showing that ΔclbP mutants lack genotoxicity (46).

The characterization of this self-resistance mechanism led to the hypothesis that ΔclbP mutant strains would accumulate inactive precolibactins that might be more readily isolated. Using this approach, several groups have identified candidate precolibactins from E. coli ΔclbP mutant strains, including metabolites that contain a spirocyclopropane ring and thiazole heterocycle(s) (47–53). These structural features are found in other DNA-alkylating natural products, suggesting that colibactin may directly interact with DNA (54, 55).

Although biosynthetic proposals for the isolated candidate precolibactins involve many of the pks enzymes, several genes that are essential for genotoxicity were not invoked (56). The metabolites isolated to date are therefore not likely precursors to the active genotoxin since additional metabolites isolated to date are therefore not likely precursors that are essential for genotoxicity were not invoked (56). The precolibactins involve many of the pks enzymes play critical roles in biosynthesis. Using in vitro biochemical characterization, Piel and co-workers (57) showed that several of these “missing” enzymes generate the rare PKS extender unit aminomalonate. Recently, we demonstrated that this building block can be accepted by multiple PKS modules from the pks island in vitro (53). As the aminomalonate-forming and -utilizing genes are required for genotoxicity, aminomalonate is likely incorporated into colibactin. In 2016, Qian and co-workers (52) isolated and structurally characterized an aminomalonate-containing candidate precolibactin supporting this hypothesis. Because two essential biosynthetic enzymes remain unaccounted for, future studies will be needed to uncover the structure(s) of the true precolibactin. Because chemical synthesis allows access to many known candidate precolibactins (58), further biochemical experiments with synthetic substrates could provide the information needed to decipher the complete precolibactin structure (59). More broadly, strategically combining bioinformatics and/or in vitro reconstitution of cryptic biosynthetic pathways with chemical synthesis represents an exciting new strategy for accessing microbiome-derived natural products (60).

Outlook and future challenges

Our understanding of the structures and functions of natural products made by the human microbiome is in its infancy. The case studies highlighted here showcase the utility of multiple approaches for identifying and characterizing microbiota-derived metabolites (13). Integrating additional emerging strategies for natural product discovery into such efforts, such as synthetic refactoring of cryptic BGCs (61) and elicitation of cryptic gene cluster expression by small molecules (62), should enable the discovery of additional bioactive secondary metabolites from the human microbiome.

Although the studies discussed here illustrate the tremendous potential of harnessing secondary metabolites from the human microbiome, they also highlight many current challenges. Given the overwhelming amount of sequencing data, how do we prioritize BGCs from the human microbiome for discovery? As in the case of lactocillin, one can focus on BGCs that produce secondary metabolites with structural similarity to known natural products possessing well-defined activities. Resistance gene-guided genome mining may also reveal potential activities of human-associated bacterial BGCs prior to isolation (63, 64). Moreover, one can envision searching metagenomes for BGCs encoding the biosynthetic machinery needed to construct specific “chemical warhead” motifs using either sequence- or function-based approaches (65). Finally, analysis of the distribution of BGCs across human subjects, including various patient populations, may help to identify the pathways most relevant to human biology.

Another major challenge is establishing whether human-associated microbes actually produce particular secondary metabolites in and on the human body. In all of the studies mentioned here, culturable strains were used to isolate and characterize natural products outside of a host context. However, it is unclear whether these molecules are produced in human microbiomes. Sensitive analytical techniques, such as MALDI-imaging mass spectrometry, may help to address this issue (66). Notably, such tools have confirmed the presence of numerous secondary metabolites on the human skin (67), in lung sputum (68), and tissues of various organs (69). Developing better methods for noninvasive sampling of difficult-to-reach body sites, including the respiratory and gastrointestinal tracts, should prove valuable in identifying microbial natural products present in these environments.

Identifying bioactive microbiome-derived natural products and understanding their interactions with the host and other microbes has implications for human health. For instance, the biosynthetic enzymes that produce harmful secondary metabolites could be targeted for inhibition (70). Human-associated microbial metabolites could represent a source of new drugs (12). The bacterial producers of these molecules could also be potential diagnostic biomarkers or probiotic therapies (37). Finally, exploring the activities of human microbiome-derived metabolites may provide insights into important targets within the host that can be modulated for therapeutic benefit (28).

In summary, the human microbiome represents an exciting new frontier for natural product research, but simultaneously it presents substantial challenge to researchers. Efficiently identifying and characterizing microbiome-derived secondary metabolites that mediate important interactions will require collaboration between chemists, engineers, and biologists and additional tools and approaches suitable for exploring the chemistry of this complex microbial habitat.

Acknowledgments—We thank Nitzan Koppel and Vayu Maini Rekda for helpful discussions regarding this manuscript.

References
MINIREVIEW: Natural products from the human microbiome


MINIREVIEW: Natural products from the human microbiome


