Tetrahedron 74 (2018) 3215–3230

Contents lists available at ScienceDirect

Tetrahedron

journal homepage: www.elsevier.com/locate/tet

Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway

Benjamin A. Schneider, Emily P. Balskus*

Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 18 January 2018
Received in revised form
15 March 2018
Accepted 20 March 2018
Available online 27 March 2018

Keywords:
Gut microbiota
Natural product biosynthesis
Nonribosomal peptide synthetase
Peptide aldehyde
Protease inhibitor
Protease

ABSTRACT

Natural products from the human microbiota may mediate host health and disease. However, discovery of the biosynthetic gene clusters that generate these metabolites has far outpaced identification of the molecules themselves. Here, we used an isolation-independent approach to access the probable products of a nonribosomal peptide synthetase-encoding gene cluster from *Ruminococcus bromii*, an abundant gut commensal bacterium. By combining bioinformatics with in vitro biochemical characterization of biosynthetic enzymes, we predicted that this pathway likely generates an *N*-acylated dipeptide aldehyde (ruminopeptin). We then used chemical synthesis to access putative ruminopeptin scaffolds. Several of these compounds inhibited *Staphylococcus aureus* endoproteinase GluC (SspA/V8 protease). Homologs of this protease are found in gut commensals and opportunistic pathogens as well as human gut metagenomes. Overall, this work reveals the utility of isolation-independent approaches for rapidly accessing bioactive compounds and highlights a potential role for gut microbial natural products in targeting gut microbial proteases.

1. Introduction

Small molecules produced by the human gut microbiota are potential mediators of this microbial community's effects on host health and disease.[1] However, the major inhabitants of the gut have not been extensively investigated as natural product producers. Though genome and metagenome sequencing continues to reveal that human gut microbes have a rich biosynthetic potential, discovering natural products from these organisms has proven challenging, in part because many cannot be cultivated in the laboratory. Moreover, investigations to date have found that gut microbial natural products are often difficult or impossible to isolate or are not produced under standard laboratory conditions.[1–3] Though there are limited examples of isolating small molecules produced by gut microbes in pure culture (e.g., ruminococcin A),[4] alternative strategies such as functional metagenomics[5] and expression of biosynthetic gene clusters in heterologous hosts[6] have also revealed gut microbial natural products (Fig. 1A). Overall, there is clearly a continued need for new approaches that will provide more rapid access to products of gut microbial gene clusters.

We envisioned a strategy for accessing gut microbial secondary metabolites that would combine in vitro characterization of biosynthetic enzymes with chemical synthesis (Fig. 1B). By mining human gut metagenomic sequence data, we could identify small nonribosomal peptide synthetase (NRPS) biosynthetic gene clusters of interest based on metagenomic sequencing data and microbial ecology. These enzymes share a conserved chemical logic and would therefore be amenable to bioinformatic analyses and prediction of their natural product structures. We could then test our predictions and identify key biosynthetic building blocks using in vitro biochemical assays with purified biosynthetic enzymes. Finally, we would access the candidate natural product structures using chemical synthesis and evaluate these focused small molecule libraries for bioactivity. A key advantage of this approach is that it could provide a more rapid way to access bioactive small molecules compared to traditional isolation- or heterologous expression-based natural product discovery. Indeed, Brady and coworkers recently demonstrated the utility of a related strategy (the "synthetic-bioinformatic natural products", or syn-BNPs, approach) in their discovery of humimycin A (Fig. 1A).[7] By mining sequenced genomes from the human microbiota for NRPS gene clusters, predicting the structures of the likely gene cluster products using bioinformatics, and synthesizing the predicted

* Corresponding author.
E-mail address: balskus@chemistry.harvard.edu (E.P. Balskus).

https://doi.org/10.1016/j.tet.2018.03.043

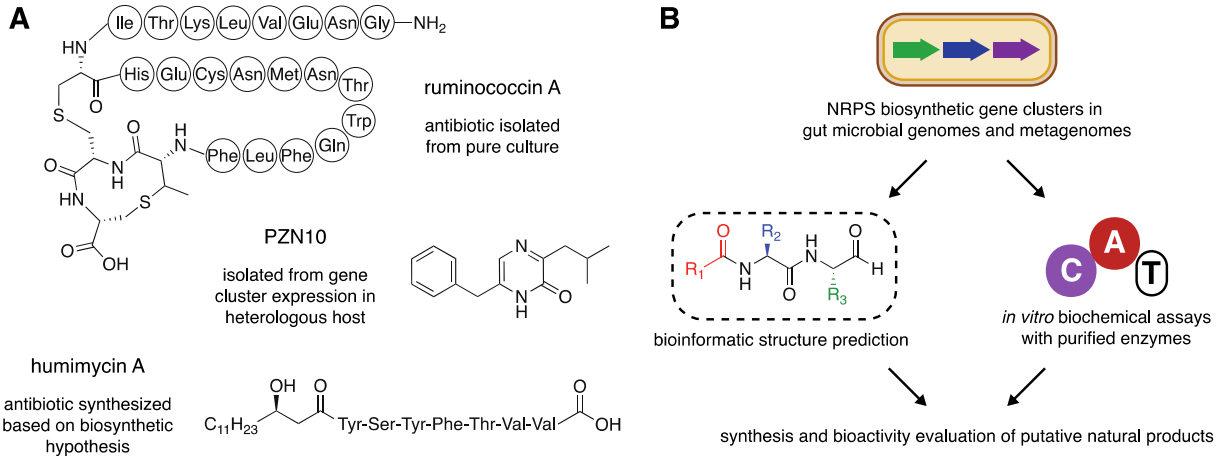

Fig. 1. Isolation-independent approaches may accelerate identification of bioactive natural products from the human gut microbiota. (A) Selected natural products from human gut bacteria, including the proposed structure of the lanthionine-containing bacteriocin (lantibiotic) ruminococcin A, PZN10, and the antibiotic humimycin. (B) Our isolation-independent workflow for characterizing small molecules produced by important gut commensals involves first selecting nonribosomal peptide synthetase (NRPS)-encoding biosynthetic gene clusters of interest based on abundance in metagenomic sequencing data and microbial ecology. Bioinformatic predictions and in vitro biochemical assays then provide structural information that informs the chemical synthesis of candidate natural product structures. These focused small molecule libraries can then be evaluated for bioactivity.

nonribosomal peptides, they accessed a new antibiotic that is active against methicillin-resistant *Staphylococcus aureus* clinical isolates.

Here, we have used our isolation-independent approach to access the putative products of an NRPS gene cluster from *Ruminococcus bromii*, one of the most abundant commensal microbes in the human gut. We first employed bioinformatic analyses to predict the product of this conserved and widely distributed gene cluster (the *rup* gene cluster) as a reactive, *N*-acylated dipeptide aldehyde (ruminopeptin). We then used in vitro biochemical characterization of the NRPS assembly line enzymes to identify the building blocks of ruminopeptin. Using a short, solution phase synthesis, we accessed a library of ruminopeptin analogues and evaluated their bioactivities. We found these molecules inhibit *S. aureus* endoproteinase GluC (SspA/V8 protease), which has been implicated in virulence in a mouse abscess model.[8] The human gut microbe and opportunistic pathogen *Enterococcus faecalis* also produces a virulence-related glutamyl endopeptidase,[9] and further bioinformatics analyses revealed additional homologs of this enzyme in gut microbial genomes and metagenomes. We hypothesize that protease inhibitors of this family may be important for mediating microbe-microbe interactions in the human gut.

2. Results and discussion

2.1. The prominent human gut microbe Ruminococcus bromii possesses an abundant and conserved biosynthetic gene cluster

With the goal of discovering bioactive secondary metabolites from the human gut microbiota, we initially focused on the prominent gut commensal *R. bromii*. This organism is one of the most abundant microbes in the human gut across a diversity of environments and diets,[10–13] and it has an important ecological role in the colon as a keystone species in the degradation of resistant starch.[14,15] *R. bromii* is a member of Clostridium cluster IV, which is significantly less abundant in patients with inflammatory bowel disease (IBD) as compared with healthy subjects.[16] This phylogenetic group of Clostridia contains organisms that are generally considered to be beneficial in the gut environment and includes *Faecalibacterium prausnitzii*, which has a well-studied anti-inflammatory role.[17] Though to our knowledge *R. bromii* has not yet been reported to produce natural products, we hypothesized that this could be a mechanism by which this organism exerts its beneficial effects or maintains its ecological niche in the human gut.

R. bromii encodes a 10.9 kb biosynthetic gene cluster that encodes a single di-modular NRPS, an efflux pump (ABC transporter), two regulatory elements, and two hypothetical proteins (Fig. 2A, Table S1). The *rup* gene cluster (also known as bgc45) has been identified previously by Fischbach and co-workers in a large survey of biosynthetic gene clusters from the human microbiome and is part of a larger family of NRPS gene clusters found in gut microbial genomes and metagenomes.[18] This study also revealed the *rup* gene cluster to be one of the most abundant gene clusters found in human microbiome project (HMP) stool metagenomes. Moreover, a highly similar gene cluster (bgc71, 97.2% nucleotide sequence identity) from a closely related, unisolated *Ruminococcus* species was identified in several RNAseq datasets from stool samples of healthy subjects, indicating that this biosynthetic pathway is likely expressed under physiological conditions.[6] Overall, these findings suggest the product of the *rup* gene cluster is likely produced under physiological conditions. Coupled with the established importance of *R. bromii*, this may indicate a particularly important role for this metabolite in the human gut microbiota.

Based on gene content and NRPS biosynthetic logic, we predicted that the *rup* gene cluster would produce a peptide aldehyde natural product. The NRPS (RupA) features a condensation-starter (C-starter) domain in its first module, indicating that the N-terminus of the product non-ribosomal peptide is likely *N*-acylated,[19] a second complete NRPS module, and a terminal reductase (R) domain (Fig. 2B). This final domain should catalyze release of a nascent thioester intermediate from the NRPS enzyme, generating either an aldehyde or a primary alcohol-containing product.[20] A peptide aldehyde product would likely be able to act as an inhibitor of serine, cysteine, or threonine proteases as has been demonstrated for NRPS-derived peptide aldehydes produced by soil microbes (e.g. fellutamide B[21] and the flavopeptins[22]). Notably, *Ruminococceae* are negatively correlated with protease activity in the colon,[23] and production of small molecule protease inhibitors by these organisms is a potential mechanism by which this association could arise.

If the product of the *rup* gene cluster does play a crucial role in *R. bromii*'s ecology and evolutionary history, we might expect it to be highly conserved in this species. To assess the presence of this