CHEMICAL STRATEGIES TO DECIPHER THE HUMAN MICROBIOME

EMILY P. BALSKUS[*,†]

[*]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA*

[†]*Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA*

My view of the present state of research exploring the chemistry–biology interface

An important objective of research at the interface of chemistry and biology is to elucidate the molecular details of chemical interactions and transformations occurring in living organisms. Gaining this knowledge advances both fundamental biological discovery and the development of therapeutics. A particularly exciting frontier in contemporary chemical biology lies at the intersection of chemistry and microbiology. This interest has been fueled by recent advances in our understanding of microbes (e.g. archaea, bacteria, archaea, and microscopic eukaryotes) and the complex communities they inhabit (microbiomes). We have an increased appreciation for both the phylogenetic and functional diversity of microbes and the importance of the metabolic activities they perform in diverse microbiomes. In particular, the trillions of microbes that live in and on the human body, termed the human microbiome, have captured the interests and imaginations of chemists and biologists alike [1]. The composition of the human microbiome can be strongly correlated to health and disease states [2], and manipulating this community represents an exciting emerging therapeutic strategy [3].

This increase in our understanding of the human microbiome can be traced to the tremendous advances in DNA sequencing technologies that have fueled the current genomic era of biology [4]. Analyzing DNA isolated directly from complex microbiomes (metagenomics) circumvents the challenges of cultivating microbes in a lab setting and provides insights into their identities and functional potential. In combination with other multi-omics approaches, such as the analysis of gene expression (metatranscriptomics) and metabolite production (metabolomics), these techniques hold the potential to provide insights into the molecular mechanisms underlying the functions of the human microbiome and its influences on host biology.

However, application of multi-omics approaches has also revealed tremendous gaps in our understanding of microbial functions that limit our understanding of the human microbiome. An inability to efficiently link genomic information to biochemical function represents a grand challenge in this current era of biology that is especially acute for microbes and microbiomes. Studies of the human gut microbiome have revealed an estimated ~40% of genes in this community cannot be given any annotation [5]. Even our in most well-studied model bacterium, *Escherichia coli* K-12, only ~50% of its genome

is well characterized [6]. Metabolomics analyses of human stool have also highlighted large numbers of uncharacterized metabolites, many of likely microbial origin [7]. To move from descriptive to a mechanistic understanding of the human microbiome and elucidate the microbial chemical processes that play causal roles in host health, it is critical what we address this gap in knowledge.

The challenge of gaining a mechanistic understanding of the human microbiome provides many opportunities for chemists who are increasingly applying their knowledge and skills to this problem [8]. Below, I discuss my group's contributions to this topic, with an emphasis on foundational studies of the chemistry of the human gut microbiome.

I view this area as a particularly exciting research frontier at the interface of chemistry and biology, and one that is ripe for further innovation.

My recent research contributions to exploring the chemistry and biology of the human microbiome

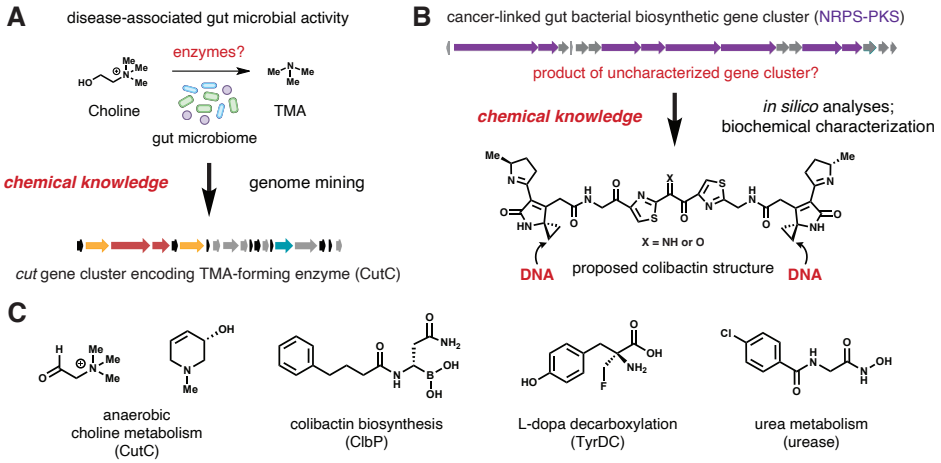

Fig. 1. Deciphering the human microbiome with chemistry. A) Connecting gut microbial metabolic activities to genes and enzymes as exemplified by our studies of anaerobic choline metabolism. B) Connecting uncharacterized genes in the gut microbiome to metabolic activities as exemplified by our studies of the genotoxic natural product colibactin. C) Representative small molecule inhibitors of gut bacterial enzymes.

Linking microbial metabolism to genes and enzymes

Studies in humans and animal models have revealed numerous metabolic processes that are associated with the gut microbiome, including the metabolism of dietary compounds (e.g. complex carbohydrates, phytochemicals), host-derived metabolites, and xenobiotics. However, the gut microbes, genes, and enzymes responsible for specific transformations are often unknown. Our work deciphering multiple gut microbial metabolic pathways reveals the power of applying a chemical understanding of enzymes and metabolism to guide hypothesis generation and biochemical studies.

Our efforts in this area began with studies of anaerobic choline metabolism (Fig. 1A) Gut microbes metabolize the dietary and host-derived nutrient choline to trimethylamine (TMA), which is further metabolized by the human liver to trimethylamine-*N*-oxide

(TMAO). TMAO had been recently linked to cardiovascular disease in humans and animal models [9], and TMA plays a key role in the inherited metabolic disorder trimethylaminuria [10]. Despite this pathway's strong links to health, the gut organisms responsible for anaerobic choline metabolism and the genes and enzymes involved were completely unknown.

We used an understanding of the chemical logic of bacterial metabolism to rationally identify genes involved in choline utilization (the *cut* gene cluster) in gut bacterial genomes [11]. Recognizing chemical similarities between a proposed radical-based C–N bond cleavage reaction that converted choline to TMA and acetaldehyde and a well-characterized transformation from ethanolamine metabolism in *Salmonella*, we hypothesized that these two pathways may share genes in common. Searches for homologs of ethanolamine utilization (*eut*) genes in a choline metabolizing organism revealed a choline utilization (*cut*) gene cluster. Unexpectedly, the key C–N bond cleaving enzyme encoded by the cut gene cluster (CutC) did not resemble the enzyme from ethanolamine metabolism but was instead a new member of the glycyl radical enzyme (GRE) family.

Discovering the cut gene cluster and CutC enabled additional studies aimed at gaining a mechanistic understanding of this activity. We could accurately identify the gut bacteria possessing this metabolism [12]. In collaboration with the Rey group, we deleted this pathway from model gut microbiomes in gnotobiotic animals, demonstrating causal links between choline metabolism and various host phenotypes, including metabolic disease [13]. Finally, as discussed in more detail below, characterizing the activity [14] and structure [15] of CutC with the Drennan gotup helped us to identify small molecule inhibitors of this gut bacterial GRE.

We have continued to explore innovative strategies to link gut microbial metabolism with genes and enzymes. For multiple activities, we have found monitoring changes in bacterial gene expression to be a particularly powerful approach to enzyme discovery. Specifically, comparing gene expression under conditions in which metabolism is present or absent has frequently revealed upregulation of the enzymes responsible for transformations of interest, including dehydroxylation of neurotransmitters [16] and polyphenol metabolites [17-19].

An exciting recent frontier has been leveraging multi-omics data from clinical cohorts for gut microbial enzyme discovery. With the Huttenhower and Chan groups, we identified gut bacterial enzymes involved in acetylating the important inflammatory bowel disease medication 5-aminosalicylic acid (5-ASA), abolishing its activity [20]. Comparisons of the stool metatranscriptomes and metabolomes of drug metabolizing and non-metabolizing patients revealed two groups of enzymes (thiolases and acyl-CoA *N*-acyltransferases) that correlated with drug metabolism and lack of efficacy. Biochemical characterization revealed these enzymes possesses 5-ASA *N*-acetyltransferase activity.

We employed a similar approach in our studies of gut bacterial cholesterol metabolism with the Xavier group, comparing protein encoding genes in metagenomes with matched stool metabolomics datasets to reveal predicted steroid dehydrogenases that correlated with the product, the poorly absorbed sterol coprostanol [21]. One of these enzymes was found to catalyze the first and final steps in coprostanol production. Intriguingly, this intestinal steroid metabolism enzyme (IsmA) is found in human gut bacteria that are

currently uncultivated, making biochemical studies central to understanding this pathway and its influence on host cholesterol levels.

Connecting uncharacterized genes to microbial metabolites

Gut microbes also possess important metabolic capabilities that are not yet appreciated. Studying uncharacterized microbial genes and metabolites can guide the discovery of such activities. However, a critical bottleneck in this endeavor is prioritizing the most promising candidates for further study among the numerous genes of unknown function and metabolites of unknown structure. Our efforts to address this problem have revealed multiple potential approaches and further reinforced the importance of leveraging knowledge of enzymatic chemistry.

A particularly understudied aspect of gut microbial metabolism is the production of bioactive natural products. This may be because human-associated bacteria are phylogenetically distinct from the environmental bacteria that are traditionally important sources of medicinally relevant natural products. Nonetheless, studying bioactive natural products from the human microbiome and their impacts on the host and other microbes offers fascinating opportunities for biological discovery, and surveys of metagenomes for natural product biosynthetic gene clusters (BGCs) indicate considerable potential for production in these communities [22,23].

Our work in this area began with our studies of the uncharacterized BGC that produces the genotoxic natural product colibactin (Fig. 1B). The *pks* (or *clb*) gene cluster encodes a hybrid non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) biosynthetic pathway [24]. It is found in various gut Enterobacteriaceae and was shown to promote DNA damage in human cells and animal models. Many subsequent studies revealed correlations between *pks*$^+$ gut bacteria and colorectal cancer (CRC) and demonstrated a causal role for the *pks* genes in tumor development in animal models. Discovering the connection between this uncharacterized BGC and such a striking biological phenotype provided strong motivation for characterization of the associated genotoxic natural product, which had been named colibactin.

Linking the *pks* genes to an associated natural product was unexpectedly challenging due to the chemical instability of colibactin. Rather than pursue traditional natural product isolation approaches, our group and other labs (Crawford, Müller, Herzon, Piel, Zhang, Qian, and Watanabe) used a combination of strategies to gain clues about colibactin, ultimately enabling the prediction of candidate structures.

The approach we pursued focused largely on elucidating colibactin's assembly by biosynthetic enzymes. Merging an understanding of assembly line biosynthetic logic with bioinformatic analyses enabled us to predict the order of enzymes and the transformations they catalyzed, allowing for subsequent testing of hypotheses using enzyme characterization. A critical early discovery was the involvement of a 'prodrug strategy' in colibactin biosynthesis involving installation of an *N*-acyl-D-Asn motif by an initiating NRPS module, elaboration to a complex, inactive biosynthetic precursor termed precolibactin, and removal of this 'prodrug motif' in a final biosynthetic step performed by a periplasmic serine peptidase (ClbP) [25]. This discovery enabled us and others to isolate and structurally characterize products of assembly line derailment from *clbP* mutant strains. Together with our studies of colibactin biosynthetic enzymes and chemical synthesis of putative precolibactin and colibactin structures, these efforts provided enough information for our group [26] and the Herzon and Crawford labs [27]

to propose putative chemical structures for colibactin. These structures, which contains two electrophilic cyclopropane rings, account for all biosynthetic enzymes and explains colibactin's genotoxicity, including its ability to form DNA interstrand crosslinks. Gaining a molecular understanding of colibactin's structure and bioactivity has been critical to elucidating its connections to cancer, including inspiring efforts to identify mutational signatures linked to this genotoxin [28,29].

Most uncharacterized gut microbial genes do not have such obvious connections to bioactivity, complicating efforts to prioritize them for study. We have explored several strategies to address this issue that rely on analysis of uncharacterized genes in multi-omics datasets from clinical cohorts. Our first example involved GREs, which are a particularly abundant enzyme family within the human gut microbiome. Classifying GREs using sequence similarity network (SSN) analysis and identifying individual enzyme clusters in metagenomes from the Human Microbiome Project revealed a prominent uncharacterized GRE in the human gut microbiome [30]. Using information from its genomic context to generate functional hypotheses, we discovered this enzyme is a 4-hydroxyproline dehydratase (HypD), a previously unrecognized metabolic activity. The prominence of HypD in commensal Clostridia and the human pathogen *Clostridioides difficile* have inspired efforts to elucidate its role in pathogenesis and colonization resistance [31]. We have more recently applied this chemically guided functional profiling approach to gut metatranscriptomes, prioritizing an uncharacterized polyphenol dehydroxylase from *Gordonibacter* for further study [18]. Interestingly, its presence in metagenomes correlated with consumption of foods that are sources of its substrate, hydrocaffeic acid, suggesting further opportunities to integrate information about host phenotypes when identifying uncharacterized gut bacterial enzymes.

Developing chemical tools to control metabolism in microbiomes

In addition to informing efforts to discover enzymes and metabolic activities, the knowledge and skills of chemists are critical in developing strategies to control and study microbial metabolism in the human microbiome. Current strategies to achieve this goal in native, complex microbiomes are limited in their precision, impacting either multiple organisms or multiple functions. Genetic approaches are also not yet broadly applicable to this setting. We envision small molecule inhibitors of gut bacterial enzymes, along with other types of chemical probes, providing an ideal strategy for studying the gut microbiome [32]. In particular, inhibitors that target non-essential gut bacterial activities may allow for precise gut microbiome manipulation and may represent an exciting, overlooked therapeutic strategy. A pioneering proof of concept for this approach was reported by Redinbo who developed inhibitors of gut bacterial β-glucoronidase enzymes, which prevent toxicity associated reactivation of the cancer chemotherapeutic Irinotecan [33]. Our efforts to develop inhibitors of gut bacterial enzymes have revealed multiple effective routes for inhibitor discovery and the promise of such tools to modulate disease-associated metabolism (Fig. 1C).

We initially explored this concept in the context of anaerobic choline metabolism. Solving a crystal structure of CutC bound to its substrate choline inspired the design and screening of choline analogs to uncover promising inhibitors, as well as subsequent structure-guided medicinal chemistry to access improved analogs [34,35]. These

compounds are active across multiple species that harbor this pathway. Rational design also led to the discovery of inhibitors of colibactin biosynthesis that engage the catalytic serine of the prodrug-activating enzyme ClbP, preventing production of the active genotoxin [36].

More recently, we have explored phenotypic high-throughput screening as a strategy to identity inhibitors of anaerobic choline metabolism, comparing anaerobic growth on choline- vs glycerol-containing media in the presence of small molecule libraries [37]. This led to the discovery of drug-like compounds that differentially impact growth and could be optimized to provide more potent inhibitors that lower TMAO levels *in vivo*.

Finally, repurposing of drugs and drug candidates may be a promising strategy to rapidly identify inhibitors of gut bacterial enzymes. We first explored this for gut bacterial tyrosine decarboxylase, a gut bacterial enzyme that may contribute to metabolism of the Parkinson's disease medication L-dopa in the periphery [16]. To inhibit this pyridoxal phosphate-dependent enzyme, we turned to α-fluoromethyltyrosine (AFMT), a mechanism-based inhibitor developed by Merck in the 1970s. Including AFMT in L-dopa based drug cocktails improved L-dopa pharmacokinetics in an animal model. Recently, we also tested the repurposed hydroxamic acid-based urease inhibitors lithostat and benurestat for their ability to modulate urease activity in gut bacteria [38]. Metabolism of urea to ammonia by this enzyme is thought to contribute to circulating ammonia levels and becomes detrimental when ammonia metabolism by the liver is impaired. While both inhibitors were effective in bacterial cultures, only benurestat lowered ammonia levels *in vivo*, likely due to its increased hydrophobicity and poorer oral bioavailability. Benurestat treatment completely protected mice from lethality in a model of acute liver disease, illustrating the power of gut microbiome-targeted inhibitors to impact host phenotypes and their therapeutic potential.

Outlook to future developments of research exploring the chemistry and biology of the human microbiome

The application of chemical knowledge and tools to study the human microbiome is in its infancy but has already played a critical role in advancing our understanding of important microbial functions and influencing the direction of this exciting research field. Moving forward, chemistry will become increasingly integrated into studies of microbes and microbiomes and the efforts of chemical biologists will be enabled by exciting new tools from other disciplines. In particular, efforts to study uncharacterized microbial genes and metabolites will be impacted by emerging artificial intelligence (AI)-based technologies, including methods for protein structure prediction (AlphaFold) [39], prediction of enzyme substrates and activities [40], and prediction of metabolite structures from BGCs and/or mass spectrometry data [41]. Application of AI to analysis of microbiome data may also help to identify microbial functions linked to particular activities and phenotypes. These computational tools may synergize with advances in experimental approaches for biochemical and structural characterization, including microfluidics systems for high-throughput biochemistry [42] and improved analytical methods for metabolite structural characterization such as microcrystal electron diffraction (micro-ED) [43]. Synthetic chemistry can play an increasing role in efforts to elucidate uncharacterized metabolites via comparison of natural samples to synthetic standards (reverse metabolomics) [44] or the synthesis of predicted natural product structures [45].

Finally, further efforts to develop inhibitors that target metabolic activities from the human microbiome and their application *in vivo* will assess the promise of this approach and its therapeutic potential, perhaps opening the door to a new class of targets for small molecule drug discovery.

References and citations

1. Human Microbiome Project Consortium, *Nature* **486**, 207 (2012).
2. J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, R. Knight, *Nat. Med.* **24**, 392 (2018).
3. M. Pitashny, I. Kesten, D. Shlon, D. B. Hur, H. Bar-Yoseph, *Drugs* **85**, 117 (2025).
4. V. van Heyningen, *Heredity* **123**, 58 (2019).
5. A. Almeida, S. Nayfach, M. Boland, F. Strozzi, M. Beracochea, et al., *Nat. Biotechnol.* **39**, 105 (2021).
6. L. R. Moore, R. Caspi, D. Boyd, M. Berkmen, A. Mackie, et al., *Nucleic Acids Res.* **52**, 12201 (2024).
7. A. Vich Vila, J. Zhang, M. Liu, K. N. Faber, R. K. Weersma, *Gut* **73**, 1909 (2024).
8. X. Yang and H. C. Hang. *Science* **2024***, 386*, eado8548.
9. Z. Wang, E. Klipfell, B. J. Bennett, R. Koeth, B. S. Levison, et al., *Nature* **472**, 57 (2011).
10. J. Messenger, S. Clark, S. Massick, M. Bechtel, *J. Clin. Aesthet. Dermatol.* **6**, 45 (2013).
11. S. Craciun, E. P. Balskus, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21307 (2012).
12. A. Martínez-del Campo, S. Bodea, H. A. Hamer, J. A. Marks, H. J. Haiser, et al., *mBio* **6**, e00042 (2015).
13. K. A. Romano, A. Martínez-del Campo, K. Kasahara, C. L. Chittim, E. I. Vivas, et al., *Cell Host Microbe* **22**, 279 (2017).
14. S. Craciun, J. A. Marks, E. P. Balskus, *ACS Chem. Biol.* **9**, 1408 (2014).
15. S. Bodea, M. A. Funk, E. P. Balskus, C. L. Drennan, *Cell Chem. Biol.* **23**, 1206 (2016).
16. V. Maini Rekdal, E. N. Bess, J. E. Bisanz, P. J. Turnbaugh, E. P. Balskus, *Science* **364**, eaau6323 (2019).
17. V. Maini Rekdal, P. N. Bernardino, M. U. Luescher, S. Kiamehr, P. J. Turnbaugh, et al., *eLife* **9**, e50845 (2020).
18. M. Bae, C. Le, R. S. Mehta, X. Dong, L. M. Pieper, et al., *Cell Host Microbe* **32**, 1 (2024).
19. X. Dong, M. Bae, C. Le, M. A. Aguilar Ramos, E. P. Balskus, *J. Am. Chem. Soc.* **147**, 7231 (2025).
20. R. S. Mehta, J. R. Mayers, Y. Zhang, A. Bhosle, N. R. Glasser, et al., *Nat. Med.* **29**, 700 (2023).
21. D. J. Kenny, D. R. Plichta, D. Shungin, N. Koppel, A. B. Hall, et al., *Cell Host Microbe* **28**, 245 (2020).
22. M. S. Donia, P. Cimermancic, C. J. Schulze, L. C. Wieland Brown, J. Martin, et al., *Cell* **158**, 1402 (2014).

23. P. Hirsch, A. Tagirdzhanov, A. Kushnareva, I. Olkhovskii, S. Graf, et al., *Nucleic Acids Res.* **52**, D579 (2024).
24. J. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, *et al.*, *Science* **313**, 848 (2006).
25. C. A. Brotherton, E. P. Balskus, *J. Am. Chem. Soc.* **135**, 3359 (2013).
26. Y. Jiang, A. Stornetta, P. W. Villalta, M. R. Wilson, P. D. Boudreau, et al., *J. Am. Chem. Soc.* **141**, 11489 (2019).
27. M. Xue, C. S. Kim, A. R. Healy, K. M. Wernke, Z. Wang, et al., *Science* **365**, eaax2685 (2019).
28. C. Pleguezuelos-Manzano, J. Puschhof, A. Rosendahl Huber, A. van Hoeck, H. M. Wood, *et al.*, *Nature* **580**, 269 (2020).
29. P. J. Dziubańska-Kusibab, H. Berger, F. Battistini, B. A. M. Bouwman, A. Iftekhar, *et al.*, *Nat. Med.* **26**, 1063 (2020).
30. B. J. Levin, Y. Y. Huang, S. C. Peck, Y. Wei, A. Martínez-del Campo, et al., *Science* **355**, aai8386 (2017).
31. A. D. Reed, J. R. Fletcher, Y. Y. Huang, R. Thanissery, A. J. Rivera, et al., *mSphere* **27**, e0092621 (2022).
32. A. Y. M. Woo, M. A. Aguilar Ramos, R. Narayan, K. C. Richards-Corke, M. L. Wang, et al., *Nat. Rev. Chem.* **7**, 319 (2023).
33. B. D. Wallace, H. Wang, K. T. Lane, J. E. Scott, J. E. Orans, et al., *Science* **330**, 831 (2010).
34. M. Orman, S. Bodea, M. A. Funk, A. Martínez-del Campo, M. Bollenbach, et al., *J. Am. Chem. Soc.* **141**, 33 (2019).
35. M. Bollenbach, M. Ortega, M. Orman, C. L. Drennan, E. P. Balskus, *ACS Med. Chem. Lett.* **11**, 1980 (2020).
36. M. R. Volpe, J. A. Velilla, M. Daniel-Ivad, J. J. Yao, A. Stornetta, et al., *Nat. Chem. Biol.* **19**, 159 (2023).
37. A. Y. M. Woo, W. J. Sandoval-Espinola, M. Bollenbach, A. Wong, M. Sakanaka-Yokoyama, et al., *bioRxiv* 2024.11.08.621386 (2024).
38. K. C. Richards-Corke, Y. Jiang, V. Yeliseyev, Y. Zhang, E. A. Franzosa, et al., *ACS Chem. Biol.* **20**, 48 (2025).
39. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, *et al.*, *Nature* **596**, 583 (2021).
40. T. Yu, H. Cui, J. C. Li, Y. Luo, G. Jiang, H. Zhao, *Science* **379**, 1358 (2023).
41. Y. Hong, Y. Ye, H. Tang, *Annu. Rev. Anal. Chem.* **18**, 193 (2025).
42. M. Gantz, S. Neun, E. J. Medcalf, L. D. van Vliet, F. Hollfelder, *Chem. Rev.* **123**, 5571 (2023).
43. C. G. Jones, M. W. Martynowycz, J. M. Hattne, T. J. Fulton, B. M. Stoltz, et al., *ACS Cent. Sci.* **4**, 1587 (2018).
44. E. C. Gentry, S. L. Collins, M. Panitchpakdi, P. Belda-Ferre, A. K. Stewart, et al., *Nature* 626, 419 (2024).
45. J. Chu, X. Vila-Farres, D. Inoyama, M. Ternei, L. J. Cohen, et al., *Nat. Chem. Biol.* **12**, 1004 (2016).