

VIEWPOINT OPEN ACCESS

A Non-Metaproteomics Researchers' View on Metaproteomics in Microbiome Research

Velma T. E. Aho¹ | Laure-Alix Clerbaux² | Anne Kupczok³ | Bree Tillett⁴ | Neha Garg⁵ | Jannie G. E. Henderickx⁶ 

¹Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway | ²Laboratory of Hepato-Gastroenterology, Institute of Experimental and Clinical Research, UCLouvain, Brussels, Belgium | ³Bioinformatics Group, Wageningen University & Research, Wageningen, the Netherlands | ⁴Frazer Institute, Faculty of Medicine, The University of Queensland Australia, Brisbane, Australia | ⁵School of Chemistry and Biochemistry, Center for Microbial Dynamics and Infection, Georgia Institute of Technology, Atlanta, Georgia, USA | ⁶Center for Microbiome Analyses and Therapeutics, Leiden University Center of Infectious Diseases (LUCID), Leiden University Medical Center, Leiden, the Netherlands

Correspondence: Jannie G. E. Henderickx (J.G.E.Henderickx@lumc.nl)

Received: 16 January 2025 | **Revised:** 11 April 2025 | **Accepted:** 14 April 2025

Funding: This viewpoint article originated in discussions at a Lorentz Center workshop, which was funded by the Dutch Research Council (NWO) with support of the Federation of European Microbiological Societies (FEMS), Bruker Daltonics, the Human Proteome Organization (HUPO), University of Ottawa School of Pharmaceutical Sciences, New England BioLabs, Qiagen, and Thermo Fisher Scientific.

Keywords: functional analysis | metagenomics | metaproteomics | microbiome | multiomics

ABSTRACT

Metaproteomics, an emerging field among the omic techniques, holds great promise for unraveling the function of microbiomes in host health and our environment. Metaproteomics can also be a valuable addition to multiomics studies of the microbiome, complementing genome-resolved metagenomics, metatranscriptomics, and metabolomics. The potential advancements from metaproteomics and multiomics research touch a breadth of disciplines, including ecology, biochemistry, immunology, medical microbiology, cell physiology, and medicine, and could lead to both fundamental and applied discoveries. However, there are significant roadblocks to widespread adoption of metaproteomics among microbiome researchers. In this Viewpoint article, we highlight the pivotal role of metaproteomics in microbiome research by showcasing its advantages, exploring opportunities to overcome challenges, and paving the way for its broader adoption as a mainstream technique. We hope that the recommendations provided in this Viewpoint article will inspire new, beneficial collaborations between proteomics experts, algorithm and infrastructure developers, biochemists, cell biologists, and microbiologists, enabling the construction of a knowledge base of microbiome function that can have an immediate and direct impact on host health and the environment.

1 | Introduction

In recent years, the exploration of microbiomes and their functional implications has gained significant attraction in many fields such as ecology, human health, agriculture, and bioeconomy [1]. To this end, microbiome studies have predominantly focused on metagenomic next-generation sequencing (NGS) techniques. Metagenomic shotgun sequencing is currently the most widely used method and provides information on the func-

tional potential of a microbiome in addition to the information on phylogenetic distance and taxonomic composition. However, the taxonomic profiles are often yet to be associated with specific functional outcomes, which requires a comprehensive understanding of the proteins effectively translated. Moreover, species composition alone might not be always informative since the same functions can be encoded by different bacterial species [2]. Besides metagenomics, there is an array of omic techniques with which the function of the microbiome can be queried at various

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *PROTEOMICS* published by Wiley-VCH GmbH.

levels. For example, metatranscriptomics offers an additional layer of insight by enabling the study of the complete set of RNA transcripts expressed within a microbial community at a given time. Metaproteomics identifies and quantifies protein levels by mass spectrometry. Hence, metaproteomics provides direct insights into translated proteins within a microbial community, reflecting real-time biological processes and enzymatic activities beyond what genomics or transcriptomics data can reveal. Metabolomics focuses on microbial metabolites and differs from metaproteomics by providing insights into the end products of metabolic pathways rather than the proteins that catalyze those processes. With the development of an array of omic techniques, standardization of microbiome studies is required to increase the comparability and integration of different types of microbiome studies.

During an interdisciplinary Lorentz Center workshop in Leiden (the Netherlands) in March 2024, which focused on “*Deciphering microbiome functions*,” metaproteomics was highlighted as a promising tool. At the same time, challenges in its application were acknowledged in a survey conducted amongst participating scientists who had less experience using metaproteomics. Leveraging omics technologies might bridge microbial identity and functionality, enabling the annotation of proteins of unknown functions and establishing more profound insights into microbial ecology. We believe that this is essential for advancing our knowledge in various domains, including personalized medicine, environmental sustainability, and biotechnology.

Here, we aim to report the advantages of metaproteomics in uncovering microbiome functions, while highlighting the challenges we encounter as non-experts in implementing this approach within our genome-resolved microbiome research. We conclude with recommendations for using metaproteomics to improve our understanding of the functional capabilities of microbial communities.

1.1 | Metaproteomics: A Key Player in the Multiomic Toolkit

Metaproteomics can be a valuable addition to multiomics studies of the microbiome, complementing genome-resolved metagenomics, metatranscriptomics, and metabolomics. First, metaproteomics can corroborate the functional potential of genomes based on sequence information by showing that the corresponding proteins are indeed synthesized [3]. This is useful as the genome-based functional repertoire of the microbiome might not be representative of the synthesized proteins, which is dependent on, for example, resources available or the presence of other community members. Metaproteomics can reveal the variation of protein abundances across time and under different conditions, providing detailed insights into the functional dynamics and ecological niche of an organism [4]. Thus, even when bacterial communities are stable in their taxonomic composition or in their genome-based functional repertoire due to functional redundancy, metaproteomics can yield useful insights into the functional activity. While metabolomics also uncovers microbial community dynamics and metabolic processes, the advantage of metaproteomics over metabolomics is that metaproteomics data can be resolved taxonomically, which is more challenging for

metabolomics. Moreover, metaproteomics can also improve gene and functional annotation by detecting open reading frames that were missed by de novo protein prediction [5].

Metaproteomics also has several advantages compared to transcriptomic measurements, since it measures the protein abundances in situ where post-transcriptional regulation is considered [6]. Additionally, proteins are more stable in the cell than RNA. This is particularly important for validating and improving metabolic reconstructions derived from genomic data and confirming if the synthesized proteins align with these predictions [7, 8]. Furthermore, metaproteomics can provide insights into co-regulation and help to understand the regulation of changes in metabolic profiles observed via metabolomics, delineating co-dysregulated pathways.

Finally, metaproteomics of host-associated communities can measure host and microbial protein abundances directly without the need for adjusting protocols for transcriptomics that might rely on polyA enrichment (for the host) and rRNA depletion (for metatranscriptomics). For host-microbe interactions, it can be crucial to measure the synthesized proteins directly (compared to transcript expression) to understand the bidirectional interactions between a host and its microbiome, the impact of microbial molecules on host receptors and metabolism, and the ways in which the host can modify and control its microbiome. In the context of the human gut microbiome, metaproteomics might help to assess the impact of the host on the changes in the microbiome functions and vice versa, for example, for enhanced hepatic lipid metabolism, impaired intestinal permeability, altered energy metabolism, or disrupted bile acid metabolism. Research that incorporates both host and microbiome layers has been described as “holo-omics” [9, 10] and metaorganism research [11]. The use of metaproteomics can be highly beneficial in this expanding field of research [12].

1.2 | Roadblocks for Widespread Application of Metaproteomics

The insight into biological processes and function provided by metaproteomic analysis is appealing for many researchers, particularly for translational research. However, there are significant roadblocks to widespread uptake among scientists. Although the roadblocks below are discussed in light of metaproteomics, not all apply to metaproteomics per se but rather reflect general problems related to omic techniques.

One roadblock is that metaproteomics is perceived to be more expensive than other functional omics. Recent advancements, such as multiplexing and short-gradient methods, have reduced per-sample costs [13], although instrument costs remain high. Unlike metagenomics and metatranscriptomics, there is a lack of commercial metaproteomic sequencing facilities providing metaproteomic analyses, so that research laboratories which lack expertise are often deterred from engaging in these types of experiments.

Moreover, metaproteomic sample preparation and analysis pipelines are often complex and overwhelming. Sample preparation is technically complex since the availability of commercial

kits is limited. This often leads to batch effects in larger studies, as well as biases toward either Gram-positive or Gram-negative bacteria, depending on bacterial lysis protocols. Sample contamination from the environment (e.g., fatty acids, metabolites) is another challenge, as microbiome research has shifted from analyzing bacterial proteins in isolation to complex communities. Although this problem is not unique to metaproteomics, peptide acquisition is highly susceptible to sample degradation and to blockages from a build-up of contamination over time, which can be more common in these complex environments. Automation and adaptations of standard approaches to high throughput, such as filter-aided sample preparation (FASP), can help overcome these challenges [14].

In addition, lack of reproducibility in data-dependent acquisition (DDA) can discourage non-proteomic scientists from investigating these approaches. In traditional sample acquisition (DDA), fragment ion selection is used to identify peptides [15]. The stochastic nature of fragment ion selection, in which the most abundant ion is selected, biases protein identification. This also creates large amounts of missing data between samples. Determining whether proteins are truly absent from a sample or missing at random can be difficult. Corrective data adjustments will differ depending on the source of these missing values. Exactly which proteins are identified/missing will differ between runs. While samples can be fractionated with increasing depth and sequencing resolution, or run with technical replicates to increase reproducibility, this has the caveat of increasing analysis costs. New data-independent sample acquisition (DIA) approaches reduce fragment ion selection bias by sequencing all ions in a defined range, known as a “dynamic range.” However, this approach produces chimeric spectra, which can challenge peptide assembly into proteins [16–19]. Some strategies, such as the use of neural network algorithms by the tool DIA-NN [20], successfully address this challenge. Recently, the metaproteomics initiative worked on a “Microbiologist’s Guide to Metaproteomics” [21, 22] to provide non-metaproteomics researchers an introduction to the field and strategies on how to handle missing data.

Arguably, the greatest challenge limiting the widespread application of metaproteomics is protein inference. Unlike metagenomic analyses, where nucleotide assembly is defined by highly constrained bonds, peptides can originate from many different proteins with alignment predicted by complex, often proprietary, algorithms, masking exactly how protein lists are assembled. As a result, exactly which proteins are called can differ between assembly algorithms and different commercial alignment software have license-specific formats. An in-depth explanation of this challenge can be found at Schallert et al. [23].

Protein identification is further challenged by the common requirement to search large publicly available databases (>10 million protein sequences), which can lead to false discovery limitations with many proteins remaining unidentified. As expected, there are many millions of bacterial species included in these generic databases. Constructing a study-specific protein database from study genomic data can reduce the false discovery rate (FDR). However, deep metagenomic shotgun sequencing and the additional sample preparation increase cost; the latter being required as there are not yet established par-

allel extraction methods, although these are currently being developed.

Many research groups, which lack the means or the expertise to undertake these studies themselves, are increasingly using publicly available datasets. However, poorly recorded metadata can limit subsequent mining of protein sequences. There are currently initiatives underway to standardize metaproteomic metadata records [24, 25].

1.3 | Future Directions and Opportunities

Despite the challenges, we believe metaproteomics offers many advantages and can be a valuable addition to multiomics studies of the microbiome, complementing genome-resolved metagenomics. The steps that could be taken to pave the way for more widespread application of metaproteomics also provide many opportunities for collaboration across disciplines.

First of all, there is a need for better standardization of methods related to metaproteomics. This holds true for every step of the workflow, starting from optimizing sample storage conditions [26] and extraction protocols [27], all the way to harmonization of data formats [28] and establishing best practices for the repositories where data is archived [29]. Such standardization and harmonization would also be crucial for enhancing the comparability between studies. Furthermore, sharing laboratory and computational workflows through publications, online platforms (such as Zenodo (<https://zenodo.org>), Figshare (<https://figshare.com>), the Open Science Framework (<https://osf.io>), Galaxy [30] (<https://galaxyproject.org/>), WorkflowHub (<https://workflowhub.eu/>), or Protocols.io (<https://www.protocols.io>) and code repositories (e.g., Git-based repositories for scripts and software) improves transparency by fulfilling the FAIR principles [31]. Such resources also offer a good starting point for researchers new to the field. Finally, community efforts to benchmark workflows, such as CAMPI [27] (<https://metaproteomics.org/campi/>), are an asset for the entire field, taking steps toward establishing methodological gold standards.

Another area where improvement is needed is in the incorporation of metaproteomics in multiomic workflows. Steps toward adopting metaproteomics in multiomic workflows by non-experts could include improving methods for parallel extraction of different biomolecules (DNA, RNA, and protein) from the same sample, and robust, open source and easy to use software tools for integrating multiomic data. Collaboration between metaproteomic researchers and other experts in microbial omics is also crucial for establishing the best possible reference databases for metaproteomics. This is important for environments that are less studied, particularly if the option of using a study-specific metagenome-based database is not available due to cost restrictions [32]. Finally, another topic for collaborative research could be performing extensive and systematic comparisons between metatranscriptomic and metaproteomic datasets to better understand the relationships between mRNA and protein abundances in complex environmental samples.

Metaproteomics is still a young and developing discipline. To make sure that its promising methodology is used in the best

possible ways, collaborations should also involve specialists representing the many fields that can benefit from metaproteomic data, ranging from medical and environmental science to biotechnological research and development. Current community efforts in metaproteomics (such as forthcoming CAMPI studies in multiomics, quantitation, and proteins of unknown functions) are already working to address many of the challenges outlined in this article and offer opportunities for the non-metaproteomics scientists to become more involved in the field.

Acknowledgments

This viewpoint article originated in discussions at a Lorentz Center workshop funded by the Dutch Research Council (NWO) with generous support of the Federation of European Microbiological Societies (FEMS), Bruker Daltonics, the Human Proteome Organization (HUPO), University of Ottawa School of Pharmaceutical Sciences, New England BioLabs, Qiagen, and Thermo Fisher Scientific. The authors would like to thank all the Lorentz Center staff and workshop participants who helped make this workshop a success. The authors would also like to thank Dr. Magnus Palmblad, Dr. Pratik D. Jagtap, Dr. Lucia Grenga, Dr. Tim Van Den Bossche, and Dr. Alfredo Blakely-Ruiz for helpful comments and suggestions on the manuscript. The author N.G. acknowledges the financial support from the National Institutes of Health (R35GM142882).

Conflicts of Interest

The authors have declared no conflict of interest.

Data Availability Statement

The authors have nothing to report.

References

- G. Berg, D. Rybakova, D. Fischer, et al., "Microbiome Definition Re-Visited: Old Concepts and New Challenges," *Microbiome* 8, no. 1 (2020): 103.
- L. Li, T. Wang, Z. Ning, et al., "Revealing Proteome-Level Functional Redundancy in the Human Gut Microbiome Using Ultra-Deep Metaproteomics," *Nature Communications* 14, no. 1 (2023): 3428.
- J. A. Rodríguez-Ramos, M. A. Borton, B. B. McGivern, et al., "Genome-Resolved Metaproteomics Decodes the Microbial and Viral Contributions to Coupled Carbon and Nitrogen Cycling in River Sediments," *Msystems* 7, no. 4 (2022): e00516–22.
- T. Wang, L. Li, D. Figeys, and Y.-Y. Liu, "Pairing Metagenomics and Metaproteomics to Characterize Ecological Niches and Metabolic Essentiality of Gut Microbiomes," *ISME Communications* 4, no. 1 (2024).
- M. Stambouliau, J. Canderan, and Y. Ye, "Metaproteomics as a Tool for Studying the Protein Landscape of Human-Gut Bacterial Species," *PLOS Computational Biology* 18, no. 3 (2022): 1009397.
- H. Duan, X. Zhang, and D. Figeys, "An Emerging Field: Post-Translational Modification in Microbiome," *Proteomics* 23, no. 3-4 (2023): 2100389.
- C. Tarzi, G. Zampieri, N. Sullivan, and C. Angione, "Emerging Methods for Genome-Scale Metabolic Modeling of Microbial Communities," *Trends in Endocrinology and Metabolism* 35, no. 6 (2024): 533–548.
- F. Zare and R. M. T. Fleming, "Integration of Proteomic Data With Genome-Scale Metabolic Models: A Methodological Overview," *Protein Science* 33, no. 10 (2024): 5150.
- I. Odriozola, J. A. Rasmussen, M. T. P. Gilbert, M. T. Limborg, and A. Alberdi, "A Practical Introduction to Holo-Omics," *Cell Reports Methods* 4, no. 7 (2024): 100820.
- C. M. Kobel, J. Merkesvik, I. M. T. Burgos, et al., "Integrating Host and Microbiome Biology Using Holo-Omics Molecular Omics," *Mol Omics* 20 (2024): 438–452.
- C. Jaspers, S. Fraune, A. E. Arnold, D. J. Miller, T. C. G. Bosch, and C. R. Voolstra, "Resolving Structure and Function of Metaorganisms Through a Holistic Framework Combining Reductionist and Integrative Approaches," *Zoology* 133 (2019): 81–87.
- K. Do, S. Mehta, R. Wagner, et al., "A Novel Clinical Metaproteomics Workflow Enables Bioinformatic Analysis of Host-Microbe Dynamics in Disease," *mSphere* 9, no. 6 (2024): 0079323.
- M. Creskey, L. Li, Z. Ning, et al., "An Economic and Robust TMT Labeling Approach for High Throughput Proteomic and Metaproteomic Analysis," *Proteomics* 23, no. 21-22 (2023): 2200116.
- J. Potriquet, M. Laohaviroj, J. M. Bethony, and J. Mulvenna, "A Modified FASP Protocol for High-Throughput Preparation of Protein Samples for Mass Spectrometry," *PLoS ONE* 12, no. 7 (2017): 0175967.
- J. Zhao, Y. Yang, H. Xu, et al., "Data-Independent Acquisition Boosts Quantitative Metaproteomics for Deep Characterization of Gut Microbiota," *NPJ Biofilms and Microbiomes* 9, no. 1 (2023): 4.
- A. T. Rajczewski, J. A. Blakeley-Ruiz, A. Meyer, et al., "Data-Independent Acquisition Mass Spectrometry as a Tool for Metaproteomics: Interlaboratory Comparison Using a Model Microbiome," *BioRxiv* (2025), <https://doi.org/10.1101/2024.09.18.613707>.
- E. Wu, G. Xu, D. Xie, and L. Qiao, "Data-Independent Acquisition in Metaproteomics," *Expert Review of Proteomics* 21, no. 7-8 (2024): 271–280.
- T. Dumas, R. Martinez Pinna, C. Lozano, et al., "The Astounding Exhaustiveness and Speed of the Astral Mass Analyzer for Highly Complex Samples Is a Quantum Leap in the Functional Analysis of Microbiomes," *Microbiome* 12, no. 1 (2024): 46.
- D. Gómez-Varela, F. Xian, S. Grundtner, J. R. Sondermann, G. Carta, and M. Schmidt, "Increasing Taxonomic and Functional Characterization of Host-Microbiome Interactions by DIA-PASEF Metaproteomics," *Frontiers in Microbiology* 14 (2023): 1258703.
- V. Demichev, C. B. Messner, S. I. Vernardis, K. S. Lilley, and M. Ralser, "DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput," *Nature Methods* 17, no. 1 (2020): 41–44.
- T. Van Den Bossche, J. Armengaud, D. Benndorf, et al., "The Microbiologist's Guide to Metaproteomics," *ChemRxiv* (2025), <https://doi.org/10.26434/chemrxiv-2025-rh2q9>.
- T. Van Den Bossche, M. Ø. Arntzen, D. Becher, et al., "The Metaproteomics Initiative: A Coordinated Approach for Propelling the Functional Characterization of Microbiomes," *Microbiome* 9, no. 1 (2021): 243.
- K. Schallert, P. Verschaffelt, B. Mesuere, D. Benndorf, L. Martens, and T. Van Den Bossche, "Pout2Prot: An Efficient Tool to Create Protein (Sub)Groups From Percolator Output Files," *Journal of Proteome Research* 21, no. 4 (2022): 1175–1180.
- T. Claeys, T. Van Den Bossche, Y. Perez-Riverol, K. Gevaert, J. A. Vizcaino, and L. Martens, "lesSDRF Is More: Maximizing the Value of Proteomics Data Through Streamlined Metadata Annotation," *Nature Communications* 14, no. 1 (2023): 6743.
- C. Dai, A. Füllgrabe, J. Pfeuffer, et al., "A Proteomics Sample Metadata Representation for Multiomics Integration and Big Data Analysis," *Nature Communications* 12, no. 1 (2021): 5854.
- O. Hickl, A. Heintz-Buschart, A. Trautwein-Schult, et al., "Sample Preservation and Storage Significantly Impact Taxonomic and Functional Profiles in Metaproteomics Studies of the Human Gut Microbiome," *Microorganisms* 7, no. 9 (2019): 367.
- T. Van Den Bossche, B. J. Kunath, K. Schallert, et al., "Critical Assessment of MetaProteome Investigation (CAMPI): A Multi-Laboratory Comparison of Established Workflows," *Nature Communications* 12, no. 1 (2021): 7305.

28. E. W. Deutsch, "File Formats Commonly Used in Mass Spectrometry Proteomics," *Molecular & Cellular Proteomics* 11, no. 12 (2012): 1612–1621.
29. A. F. Jarnuczak and J. A. Vizcaino, "Using the PRIDE Database and ProteomeXchange for Submitting and Accessing Public Proteomics Datasets," *Current Protocols in Bioinformatics* 59, no. 1 (2017): 13.31.1–13.31.12.
30. The Galaxy Community. "The Galaxy Platform for Accessible, Reproducible, and Collaborative Data Analyses: 2024 Update," *Nucleic Acids Research* 52, no. W1 (2024): W83–W94.
31. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3, no. 1 (2016): 160018.
32. J. A. Blakeley-Ruiz and M. Kleiner, "Considerations for Constructing a Protein Sequence Database for Metaproteomics," *Computational and Structural Biotechnology Journal* 20 (2022): 937–952.