Scientific and Social Research, 2026, Volume 8, Issue 5
http://ojs.bbwpublisher.com/index.php/ssr
Online ISSN: 2981-9946
Print ISSN: 2661-4332

Research Progress of Novel Characterization Techniques in Structural Analysis of Microbial Proteomics

Zekun Duan*

University of Queensland, Brisbane, QLD 4072, Australia

**Author to whom correspondence should be addressed.*

Abstract: Microbial proteomics is a key approach to understanding and analysing microbial physiology, metabolism, pathogenic mechanisms, and environmental adaptation. Due to limitations in sample purity, throughput, and cost, traditional structural analysis techniques struggle to fully detect the vast number of unknown proteins within the microbial proteome. In recent years, structural elucidation of the microbial proteome has been driven by innovations in cryo-electron microscopy, high-resolution mass spectrometry, AI-based structure prediction, and multi-model integration techniques. This field has evolved from single-protein studies to the entire proteome, from in vitro purification to in situ dynamics, and from static structures to functional networks. The following article provides a systematic review of the latest advances in cryo-EM visualisation of proteomics, in situ mass spectrometry, AI-based structure prediction, and multi-technique integration, offering a reference for the in-depth decoding of microbial protein functional networks.

Keywords: Microbial proteomics; Structural analysis; Cryo-electron microscopy; In situ mass spectrometry; Artificial intelligence; Multimodal integration

Online publication: June 5, 2026

1. Introduction

Microorganisms are the most widespread and functionally diverse group of life on Earth. The microbial proteome is central to functions such as metabolism, signal transduction, and environmental adaptation. The three-dimensional (3D) structure and dynamic assembly of the proteome directly determine its function and intermolecular interactions, and serve as the link between the genome and physiological phenotypes. Traditional analysis methods rely on X-ray crystallography, which requires multi-step sample purification and is associated with issues such as long turnaround times, low throughput, and limited applicability. Furthermore, this method can only determine the sequence structure of less than 0.1% of the microbial proteome [2]. With the advancement of metagenomics, the “dark proteome”—comprising a vast number

of proteins with known sequences but unknown structural and functional characteristics—has become a bottleneck in research. In recent years, breakthroughs in new technologies have reshaped traditional analytical methods. For example, Cryo-EM has overcome limitations on sample purity, enabling high-resolution structural analysis [3]; high-resolution mass spectrometry can accurately capture the dynamic characteristics and modifications of proteins [4]; and AlphaFold is capable of rapidly predicting protein structures. The integration of these multiple technologies has established a closed-loop process encompassing prediction, analysis, and validation.

2. Cryo-electron microscopy (Cryo- EM)

2.1. Single-particle cryo-electron microscopy (SPA)

SPA achieves near-atomic resolution structural analysis by rapidly freeze-fixing protein samples and combining electron microscopy with 3D reconstruction. This has revolutionised the research paradigm for microbial membrane proteins and multimeric complexes [5]. By focusing on microbial membrane proteins—key targets in antibiotic resistance and substance transport—Croy-EM has overcome the conventional challenges of crystallisation and achieved high-resolution structural analysis. In 2024, Yan Zhaofeng's team analysed the structure of Cdr1, a drug-resistant efflux pump in Candida albicans, identifying the mechanism by which azole substrates are recognised and inhibited, thereby providing a specific target for the development of antifungal drugs. For microbial macromolecular complexes, Cyro-EM combines automated modelling workflows to identify oligoprotein complexes directly from crude bacterial extracts [6]. Using the CryoID, DeepTracer, and ModelAngelo automated workflows developed by a team at the California Institute of Technology, six oligomeric complexes were characterised from extracts of the nitrogen-fixing bacterium Azotobacter vinelandii, including novel oligomeric states such as the phospho-glucose isomerase (Pgi1) decamer and soluble pyridine transhydrogenase (SthA), with resolutions ranging from 1.9 to 3.7 Å, thereby realising "visual proteomics" [7].

The combination of microfluidics and cryo-electron microscopy (MISO) further reduces sample requirements, making it suitable for the analysis of trace proteins in microorganisms. This technology integrates microfluidic purification, capillary deposition, and automated sample preparation, enabling structural characterisation to be completed in just four hours using only 0.1–1 μg of protein, while reducing sample requirements by a factor of 100–500 compared with conventional methods [8]. A β-galactosidase isolated from a single E. coli colony was purified using MISO, resulting in a 2.2 Å high-resolution structure that was fully consistent with conventional results, thereby providing a viable approach for proteomic characterisation of uncultivable microorganisms and rare microbial communities.

2.2. Cryo-electron tomography

Cryo-ET enables 3D imaging of intact microbial cells, allowing for the in situ localisation and structural characterisation of protein complexes in their native cellular environment, and capturing dynamic assembly and interaction networks under physiological conditions [9]. The team led by Cao Qi at Shanghai Jiao Tong University optimised the strategy for particle selection and symmetry prediction using Cryo-EM, characterised the fibre structure of the Bacillus spore coat protein GerQ, and determined its 8nm fibre assembly pattern, establishing a structural foundation for research into spore germination mechanisms and the discovery of antimicrobial targets [10].

2.3. Technical limitations and areas for optimisation

Cryo-EM still faces three major challenges in the analysis of the microbial proteome: the first is the limitation of sample complexity, as protein abundances vary greatly in microbial cell extracts, making it difficult to capture low-abundance proteins [11]; the second is the difficulty of data processing, as single-particle picking and 3D reconstruction are time-consuming and difficult to adapt to the whole-proteome scale; and thirdly, insufficient in situ resolution, as interference from the intracellular environment in cryo-EM often results in resolutions of 3 nm or higher, making it difficult to resolve atomic-level structures [12]. Current optimisation efforts are focused on: developing highly specific enrichment probes to target low-abundance proteins [13]; integrating AI algorithms to automate particle sorting, classification, and reconstruction, thereby increasing throughput; and combining phase plates, direct electron detectors, and cryo-FIB ultrathin sectioning techniques to improve in-situ resolution.

3. High-resolution mass spectrometry (High-resolution MS)

3.1. In situ mass spectrometry (nMS)

nMS dissociates intact protein complexes under non-denaturing conditions, allowing precise determination of molecular weight, stoichiometry, and assembly heterogeneity; and is suitable for analysing microbial multi-subunit metabolic complexes, ribosomes, and secretory systems [14]. A triple-combination technique comprising "nMS-IRMP-top-down proteomics" has been employed to characterise ribosomal heterogeneity in E. coli. This approach identifies dynamic differences in protein variants, post-translational modifications (PTMs), and the strength of subunit interactions between the 30S and 50S subunits, thereby offering a novel perspective for the study of translational regulatory mechanisms. About nitrogen-fixing bacterial nitrogenase complexes, the stoichiometric ratio of molybdoferrin to ferrin was precisely determined using nMS, and metal cofactor binding sites were elucidated using collision-induced dissociation (CID), thereby demonstrating the structural basis for the regulation of nitrogen-fixing activity [15].

3.2. Ion mobility spectrometry (IMS)

IMS separates proteins based on the shape and charge of their ions, distinguishing between protein variants (proteoforms) that share the same amino acid sequence but differ in conformation, modification, or assembly state, and captures the conformational dynamics of microbial proteins under environmental stress [16]. Brukertims TOF series combines dual TIMS and PASEF scanning modes to achieve high-coverage analysis of microbial proteomes, identifying over 3,000 proteins in a single run and simultaneously distinguishing between conformational isomers. Research on acid stress in E. coli has enabled IMS to precisely capture the conformational transition of the DNA-binding protein HU, clarifying the mechanistic link between its conformational changes and its DNA-protective function.

3.3. Post-translational modification (PTM) mass spectrometry

Post-translational modifications (PTMs) of microbial proteins—including phosphorylation, glycosylation, acetylation, and methylation—are key mechanisms for rapid functional regulation; high-resolution MS coupled with enrichment techniques enables the precise localisation and quantification of PTM sites [17]. Using TiO_2 enrichment and Orbitrap mass spectrometry to analyse Mycobacterium tuberculosis, 236 phosphorylation sites were identified, revealing the regulatory network of metabolic pathways mediated by

serine/threonine kinases. Research on the gut microbial proteome, utilising lectin enrichment coupled with mass spectrometry, has identified 128 glycosylation sites in Bifidobacterium, revealing the mechanisms by which glycosylation regulates bacterial adhesion and host interactions.

3.4. Technical challenges and pathways to breakthroughs

The core limitations of MS lie in the low ionisation efficiency of intact proteins, the difficulty in fragmenting large molecular complexes, and the insufficient sensitivity for detecting low-abundance PTMs [18]. Strategies for overcoming these challenges include: developing novel non-denaturing ionisation sources (such as nano-electrospray and laser desorption) to improve the ionisation efficiency of intact complexes; combining these with mild fragmentation methods such as electron transfer dissociation (ETD) and ultraviolet photo-dissociation (UVPD) to preserve information on the complexes and PTMs; and developing highly specific enrichment materials (such as metal-organic frameworks and covalent organic frameworks) to enhance the sensitivity of detection for low-abundance modifications and complexes.

4. Artificial intelligence (AI) structure prediction

4.1. Core technologies and performance breakthroughs in AI prediction

AlphaFold2 utilises the Transformer architecture and attention mechanisms, integrating multiple sequence alignment (MSA), evolutionary and co-evolutionary information, and physical constraints to achieve atomic-level structural prediction, with 98% of protein predictions reaching experimental-level accuracy (CαRMSD < 2 Å) [19]. AlphaFold-Multimer has been extended to predict protein-protein complexes, successfully constructing a complete protein interaction network model for *Bacillus thuringiensis*, with some predictions validated by site-directed mutagenesis. Published in 2024, AlphaFold3 represents a further breakthrough, enabling the prediction of structures for protein-nucleic acid complexes, protein-metabolite complexes, and post-translational modification (PTM) sites, thereby covering the core molecular interactions underlying microbial life processes [20].

The challenge posed by the vast volume of microbial metagenomic data and the scarcity of homologous sequences is resolved through optimised algorithms that enable highly efficient and accurate predictions. The Meta Source model utilises 4.25 billion microbial sequences to construct a niche-to-protein-family association model that predicts the origin of homology sequences for target proteins. It reduces computational resource requirements by more than three times compared to traditional methods, whilst simultaneously improving prediction accuracy. Foldseek enables ultra-fast structural similarity searches, achieving speeds 1,000 times faster than traditional methods and providing an efficient tool for the clustering and functional annotation of microbial protein structures [21].

4.2. Strategies for the application of AI in microbial proteomics

4.2.1. Functional annotation of pathogenic microorganisms and target identification

AI prediction provides a key tool for deciphering the functions of the “dark proteome” of pathogenic bacteria. For Mycobacterium avium complex (MAB), by combining AlphaFold2 and Foldseek, the system annotated 374 new GO terms and optimised 885 existing annotations, identifying structural features of drug-resistance-associated proteins and discovering six new essential proteins, which provide novel targets for the development of anti-tuberculosis drugs. Using AI-based prediction, a large number of lysins with significant

sequence diversity but structurally conserved were identified in intestinal bacteriophages; their cell wall-cleaving activity and specificity were characterised, providing a foundation for the development of precision antimicrobial tools [22].

4.2.2. Microbial enzyme engineering and applications in synthetic biology

AI predicts rapid identification of structural features in microbial extreme enzymes, guiding rational modification to enhance performance. A team led by Wu Bian at the Institute of Microbiology, Chinese Academy of Sciences, has developed the Pythia deep learning framework, which utilises structure-based self-supervised learning to enable ultra-rapid prediction of protein mutation stability, thereby providing an efficient tool for the engineering of thermal stability and acid-base tolerance in industrial microbial enzymes [23]. Research on cellulose-degrading bacteria has seen AI used to predict the structure of glycosidases, clarify the spatial arrangement of catalytic and binding domains, and guide site-directed mutagenesis to increase enzyme activity by a factor of 3.2.

4.2.3. Construction of a microbial community structure map

Predictive AI is driving the transition of the microbiome from the "sequome" to the "structome."

A research team has established a proteome database of the human gut microbiota. After structural alignment, they discovered structural conservation between bacterial isoenzymes and their host homologues. This study has, for the first time, identified the structure of the key enzyme responsible for melatonin synthesis in gut bacteria, clarifying its molecular mechanism in regulating host physiology.

4.3. Technical limitations and development trends

AI prediction has three key limitations: firstly, low accuracy in predicting proteins without homologous sequences (orphan proteins); secondly, insufficient prediction of dynamic conformations and complex assembly; and thirdly, imprecise modelling of the effects of PTMs and ligand binding on structure. Directions for Future Development: Integrating molecular dynamics simulations to enhance dynamic conformation prediction capabilities; incorporating experimental data (cryo-EM maps, MS cross-linking data) to optimise complex prediction; developing microbe-specific training models to improve the prediction accuracy of orphan proteins and proteins from extreme environments.

5. Integration of multimodal technologies

5.1. Combining AI predictions with Cryo-EM

AI prediction, when combined with cryo-EM, forms a cyclical "prediction–screening–validation" workflow: first, AI is used to annotate the microbial proteome at scale to identify key unknown proteins; then, cryo-EM is employed to characterise experimental structures, refine the AI model, and validate its functionality. For membrane proteins, AI-assisted cryo-EM model building has resolved the challenge of sequence alignment in low-resolution maps, reducing the time required for structural determination from several months to a few weeks [24].

5.2. Mass spectrometry (MS) and Cryo-EM

MS and Cryo-EM complement each other: MS precisely determines complex composition, PTMs, and

stoichiometric ratios [25]; Cryo-EM provides spatial structure and in situ localisation. In the characterisation of *E. coli* ribosomes, in situ MS analysis was used to identify the composition and modifications of subunit proteins, with Cryo-EM characterising the overall conformation; this integrated approach has revealed the mechanisms linking ribosomal heterogeneity to the regulation of translation. Research on nitrogen-fixing bacteria has utilised MS to determine the composition and modification of the metal cofactors in nitrogenase, with Cryo-EM analysis identifying the metal cluster binding sites, thereby jointly elucidating the mechanism of nitrogen fixation.

5.3. Computational simulation and laboratory analysis

Molecular dynamics (MD) and Monte Carlo simulations are combined with experimental structures to characterise the dynamic conformations, ligand binding, and catalytic mechanisms of microbial proteins. The mechanism of resistance in antibiotic resistance proteins; characterisation of the substrate-bound structure using Cryo-EM; and MD simulations revealing drug entry and exit channels and conformational transition pathways. Research into microbial chemoreceptors, combining AI predictions with MD simulations, has demonstrated the conformational cascades involved in signal transduction, thereby providing a theoretical foundation for the regulation of signalling pathways.

6. Conclusion

New characterisation techniques have become central to research into microbial proteome structures; Cryo-EM enables high-resolution in-situ structural visualisation, mass spectrometry provides precise identification of dynamic modifications and complex assembly, and AI technology has resolved the challenge of annotating vast numbers of unknown proteins, with multi-technique integration forming a looped system for structure-function characterisation. Current research has transitioned from single proteins to the entire proteome, from in vitro to in situ, and from static to dynamic studies. It has achieved breakthrough progress in pathogen control, synthetic biology, and microbiome research. In the future, it will be necessary to further overcome technical challenges, optimise data integration, and expand the application of these technologies in extreme conditions. It is imperative to drive the advancement of microbial proteomics research from structural characterisation to functional decoding, and from basic research to industrial application, in order to provide core technological support for human health, industrial production, and environmental management.

Disclosure statement

The author declares no conflict of interest.

References

[1] Sun CC, Li XY, Huang WR, et al., 2025, Large Models for Protein Prediction and Generation: From Sequence, Structure to Function. Life Sciences, 37(12): 1534–1548.

[2] Ma H, Li LW, Bi XY, et al., 2025, Surface-enhanced Raman Spectroscopy (SERS) over 50 years: Theory, Applications, and Prospects. The Journal of Light Scattering, 37(3): 357–514.

[3] Chao TY, Zhao Z, 2025, Research Progress and Development Trends in Cryo-electron Microscopy Sample Preparation and Cryo-thinning Techniques. Hi-Tech & Industrialization, 31(11): 34–36.

[4] Cai F, Ling YY, Li L, et al., 2026, Research Progress in Mass Spectrometry-based Clinical Body Fluid Glycoproteomics. Journal of Chinese Mass Spectrometry Society, preprint, 1–32.

[5] Yin GL, Sun WH, Pang XY, et al., 2022, Application of Cryo-electron Microscopy Technology in Molecular Botany Research. Biotechnology Bulletin, 38(1): 15–32.

[6] Peng Y, Lu Y, Sun H, et al., 2024, Cryo-EM Structures of *Candida albicans* Cdr1 Reveal Azole-substrate Recognition and Inhibitor Blocking Mechanisms. Nature Communications, 15(1): 7722.

[7] Jamali K, Käll L, Zhang R, et al., 2024, Automated Model Building and Protein Identification in Cryo-EM Maps. Nature, 628(8007): 450–457.

[8] Song Y, Zuo QY, Meng XM, et al., 2025, Frontier Technologies and Development Applications of Cryo-electron Microscopy. Life Sciences, 37(7): 854–867.

[9] Qin CD, Guo Q, Gao N, 2026, Cryo-lift-out Technology for Cryo-electron Tomography of Tissue Samples. Progress in Biochemistry and Biophysics, preprint, 1–20.

[10] Liu W, Han J, Gong W, et al., 2026, Structure of Pancreatic hIAPP Fibrils Derived from Patients with Type 2 Diabetes. Cell, 189(4): 1210.e10.

[11] Du FY, 2025, Research on the Antibacterial Effect and Mechanism of Water-soluble ZnO@APTES Quantum Dots against Methicillin-resistant *Staphylococcus aureus*, thesis, Shenyang Agricultural University.

[12] Pang MW, 2025, Research on Protein Structure Prediction Methods Based on Cryo-electron Microscopy Detection Data, thesis, Jiangsu University of Technology.

[13] Yang ZY, Yang TT, Wang N, et al., 2026, Progress and Challenges in Intact Protein in situ Ionization and Spatial Omics Mass Spectrometry Technologies for Biological Tissues. Journal of Chinese Mass Spectrometry Society, preprint, 1–21.

[14] Huang GM, 2019, Electrophoretic Effects in Electrospray and their Applications in in situ Mass Spectrometry Analysis of Metabolites and Proteins. Chinese Chemical Society. Proceedings of the 22nd National Chromatographic Academic Symposium and Instrument Exhibition of the Chinese Chemical Society (Volume I). School of Chemistry and Materials Science, University of Science and Technology of China; 2019: 65.

[15] Feng YP, Wang JL, Xue DM, 2025, Research Progress in Nitrogenase. Chinese Journal of Grassland, 47(12): 128–138.

[16] Qin ZY, 2024, Functional Analysis of the htpG Gene in *Vibrio mimicus* Infection and Pathogenesis, thesis, Sichuan Agricultural University.

[17] Wang HL, Meng XG, Xiao FP, et al., 2024, Application of High-resolution Liquid Chromatography-mass Spectrometry in Protein Chemical Modifications. Experimental Technology and Management, 41(8): 1–14.

[18] Huang PQ, Zhao Y, Zhu J, et al., 2026, Progress and Application of Imaging Mass Spectrometry Technology in Tumor Spatial Proteomics Research. Journal of Chinese Mass Spectrometry Society, preprint, 1–24.

[19] Ouyang ZZ, Ma YC, Kou YT, et al., 2025, Iterative Breakthroughs and Data Strategy Implications of AlphaFold from the Perspective of Scientific Data. Journal of Agricultural Big Data, 7(4): 485–495.

[20] Gong WB, 2024, Major Breakthroughs in AlphaFold Structure Prediction and Its Impacts and Challenges on Protein Research. Progress in Biochemistry and Biophysics, 51(12): 3073–3083.

[21] Ma H, Li LW, Bi XY, et al., 2025, Surface-enhanced Raman Spectroscopy (SERS) over 50 years: Theory, Applications, and Prospects. The Journal of Light Scattering, 37(3): 357–514.

[22] Li XR, Huang GT, 2026, Research Progress in Phage-antibiotic Synergy (PAS) Therapy. Microbiology China, preprint, 1–24. https://doi.org/10.13344/j.microbiol.china.260098

[23] Sun J, Zhu T, Cui Y, et al., 2025, Structure-based Self-supervised Learning Enables Ultrafast Protein Stability Prediction upon Mutation. The Innovation, 6(1): 100750.

[24] Zeng WY, Zheng SY, Zhao XQ, et al., 2026, Progress in Generative Biology and Biologically Inspired Artificial Intelligence. Life Sciences, 38(2): 248–267.

[25] Yu KR, Zhang NN, Liu YB, 2026, Precise Regulatory Tools for O-GlcNAc Modification Targeting Specific Proteins: Disease Mechanism Research and Therapeutic Opportunities. Acta Chimica Sinica, 84(3): 409–424.