Delta University Scientific Journal
Journal home page: https://dusj.journals.ekb.eg

Role of Metaproteomic in the Investigating Marine Life

Esraa Elsaeed[1], Shymaa Enany[2,3], Samar Solyman[2,4], Mohamed Shohayeb[1] and Amro Hanora[2*]

[1] Department of Microbiology and Immunology, Faculty of Pharmacy, Delta University for Science and Technology, Gamasa 11152, Egypt.
[2] Department of Microbiology and Immunology, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt.
[3] Biomedical Research Department, Armed Force College of Medicine, Cairo, Egypt.
[4] Department of Microbiology and Immunology, Faculty of Pharmacy, Sinai University-Elkantara branch, Egypt.

Correspondence: Amro Mohamed Said Hanora; Ismailia; 41522; Tel.: +20-10-0032-3406; ahanora@yahoo.com

ABSTRACT

Marine ecosystems continue to be a mystery. Metaproteomic is the large-scale characterization of the complete protein complex in environmental microbiota at a particular point in time. Although the use of Metaproteome in marine environments is still in its infancy, numerous researchers have used metaproteomic analysis on a wide range of animals, water, and sediments. The use of metaproteomic analysis on seawater and organisms allows for the molecular interpretation of cell phenotypic information as well as the identification of changes in the expression of individual genes. Some metaproteomic studies that were done on marine organisms were briefly reviewed here, with an emphasis on those using alternative omics methodologies rather than metaproteomic approaches. We also reviewed challenges that metaproteomic research faces, as well as potential solutions.

Keywords: *Metaproteomic; marine organisms; metagenomics; LC-MS/MS*

1. Introduction

Metaproteomic is defined as the large-scale characterization of the complete protein complex in environmental microbiota at a particular point in time (Wilmes et al., 2015). Metagenomic is nontargeted that applies massive sequencing to genetic material obtained directly from environmental samples (Branda et al., 2005).
Between the blueprint proteins detected by metagenomics and the true phenotype, there was still a protein detection gap (Seifert et al., 2013). Genomic sequence annotation is very prone to mistakes because of failures in sequencing, wrong open reading frames (ORFs) recognitions, wrong start codon annotations, and poor sequence similarity to the database's genes (Kircher and Kelso, 2010). In addition, some gene entries are marked as heterologous and lack biochemical evidence, and if only a small number of ORFs are sequenced, the annotation is much worse (Seifert et al., 2013). Because of all the previous reasons, metaproteomic was developed.
Some of a species' ability to work in a consortium can be checked by looking at the proteins that are already there. The fact that a protein is there is not enough to prove anything; more information is needed about how it works (Seifert et al., 2013). Using metaproteomic data obtained for functional analysis, the structure of communities can be analyzed. In the majority of organisms, proteins are the most abundant component of cells, making it possible to calculate the relative biomass contributions of different species (Kleiner, 2019). Using conventional metaproteomic data, a technique that employs biomass rather than gene/genome copy counts (e.g., 16S rRNA amplicon or metagenome sequencing) have devised and validated to study community structure (Kleiner, 2019).

2. Metaproteomic studies of marine organisms

Metaproteomic started with research on activated sludge. It is the most popular way to treat wastewater in industry and improve our understanding of the complex microbial ecosystems that make the process work better (Armengaud et al., 2013). The latter study used two dimensions polyacrylamide gel electrophoresis (2D-PAGE) to separate the proteins and liquid chromatography-mass spectrometry (LC-MS/MS) and *de novo* protein sequencing to figure out what the removed proteins were.

Biofilms are bacterial communities embedded in a matrix of extracellular polymeric compounds attached to a surface (Branda et al., 2005). Living in biofilms protects bacteria from deleterious conditions (Davey and O'toole, 2000) and it plays an important role in the disease cycle of bacterial pathogens in both animals and plants. Metaproteomic has been used in studying biofilms and acid mine drainage by exploring many kinds of microbial communities. Acidophilic biofilms from the Richmond Mine in Northern California are by far the most deeply sampled. In a first shotgun proteomic analysis, high protein coverage (48%) was found for the dominant *Leptospirillum rubarum* (Wilmes and Bond, 2004). This meant that over 2000 proteins were found. One of these proteins was found to be an iron-oxidizing cytochrome, which is a key part of how biofilms make energy (Wilmes and Bond, 2004).

Biofilm samples were studied using proteomics-based genotyping surveys to identify genes involved in niche partitioning at different phases of biofilm development (Denef et al., 2010). It was also possible to distinguish between proteins obtained from active community members and lysed cells based on identifiable amino acid changes caused by hydrolysis (Justice et al., 2012). These findings include species-specific hydrogen isotope fractionation in proteins (Fischer et al., 2013), nitrogen transport patterns through these communities (Justice et al., 2014), and various post-translational modifications specific to different biofilm growth phases (Chen et al., 2013).

One of the early metaproteomic studies of marine environments was the study of microbial communities in the Chesapeake Bay. There are now many such studies about many different marine environments, such as surface waters (Morris et al., 2010), coastal systems (Morris et al., 2010), low oxygen waters, animal gut symbionts (Herbst et al., 2013), ship hull biofilms (Morris et al., 2010), and sediments (Urich et al., 2014).

Metaproteomic analysis has been done on a variety of samples, including nutrient-depleted waters from an open ocean like the South Atlantic (Morris et al., 2010), small bacterioplankton (1.2 m) in nutrient-rich waters (Oregon coast) (Sowell et al., 2011), membrane-enriched fractions of small cells (0.8 m) from coastal waters, seasonally sampled cold coastal waters of the West Antarctic Proteins from the SAR11, and bacterial populations during a phytoplankton bloom (North Sea) (19). Roseobacter and oligotrophic marine gammaproteobacterial clades (Sowell et al., 2011; Williams et al., 2012)often made up most of the proteins in surface water, whereas cyanobacterial proteins made up most of the proteins in ocean water samples that are low in nutrients (Morris et al., 2010).

On a functional level, metaproteomic has given us important information about carbon and nutrient cycling in surface waters. Other two examples of important functions are proteorhodopsin-mediated light-driven proton pumps and methylotrophy (Morris et al., 2010).

Metaproteomic has permitted the identification of main metabolic pathways to dominating taxa from diffuse hydrothermal venting sediments in the Norwegian-Greenland Sea (Urich et al., 2014). An autotrophic sulfur oxidation process, associated with CO_2 fixation and sulphate reduction, was reflected in this study. To improve protein identification, this work used a hybrid sequence database derived from both metagenomic and meta-transcriptomic data.

Microbiological communities can exist in nutrient-poor environments because of their symbiotic relationship with deep-water marine organisms. Researchers examined the metaproteome and meta-metabolome in an intestinally-less oceanic, which is home to an established community of five different bacteria (Kleiner et al., 2012). It has been shown that the symbiont may employ CO oxidation paired with sulphate reduction to generate energy; a pathway for the host's waste fermentation products to be utilized, and a mechanism for rapid symbiotic evolution have all been discovered by the researchers.

In a metaproteomic study of two freshwater lakes (NY, USA), proteins from Betaproteobacteria and Cyanobacteria predominate. Details on photosynthesis and electron transport systems were revealed in the article (Hanson et al., 2014). Researchers found lower amounts of transport proteins in lakes than in marine systems, suggesting that nutrients are not as scarce in these habitats.

Metaproteomic and metagenomics investigations have been performed on Ace Lake, an Antarctic meromictic lake (Ram et al., 2005). Green sulfur-producing bacteria predominate in densely packed colonies near the chemocline, which results in a salty and anoxic lake below the chemocline (Tyson et al., 2004). These cold-weather-adapted bacteriochlorophylls, syntrophic sulfur conversions, and membrane fluidity were discovered by metaproteomic, revealing the significance of these physical and metabolic characteristics for bacterial survival.

One of the most interesting aspects of a new comprehensive study on permafrost soil is that this method of phylogenetic distribution is not biased by the microbes' selection, and it provides greater bioinformatic independence than databases with several millions of samples (Seifert et al., 2013).

3. Challenges in metaproteomic studies

Workflows in the field of metaproteomic investigations of the bacterial community face major challenges (Seifert et al., 2013):

Firstly, proteins are commonly extracted from samples containing large concentrations of eukaryotic cells or samples contaminated with organic and inorganic contaminants.

Secondly, the unavailability of high-resolution mass spectrometers at a reasonable price.

Thirdly, it is hard to identify proteins based on their genetic sequences, whether public or private. Hypothetical proteins make up a large portion of proteomic datasets (40–50%), making it difficult to integrate them into metabolic circuit reconstructions (VerBerkmoes et al., 2009).

Fourthly, challenges in the interpretation of peptide and protein identification results. Metaproteomic data makes it difficult to determine a community's metabolic processes due to the very small number of detected protein samples (Seifert et al., 2013).

Fifthly, Biomass production is a common limitation in genome-scale metabolic models to ensure growth-related chemical biosynthesis. Naphthalene-enriched communities derived from soils indicated that 200 extra enzymes were needed to generate biomass. Knowing about active and sequenced species in these soils helps guide these reconstructions. It is better to finish metabolic functions with enzymes identified in these organisms (Guazzaroni et al., 2012).

Sixthly, even relatively little alterations in the protein sequence of closely related bacteria can have a significant bearing on the identification of proteins. One mutation in every tenth amino acid results in completely new tryptic peptides, which makes it difficult to identify any peptide that belongs to the protein that is being studied. As a result, researchers started sequencing metagenomes simultaneously with meta-proteomics experiments (Ram et al., 2005).

4. Metaproteomic workflow

The most typical metaproteomic workflow includes the collection and preservation of samples, cellular lysis, protein extraction, tryptic digestion, peptide separation by liquid chromatography (LC), and mass spectrometric analysis. LC, which is capable of separating highly complicated peptide mixtures, has aided in the development of metaproteomic. Also, high-resolution mass spectrometers that can capture vast amounts of accurate mass spectrometric data and computer programmers that can process and analyze large amounts of data enable the identification of tens of thousands of peptides that can be detected and quantified using liquid chromatography-tandem mass spectrometry (LC-MS/MS).

If a researcher is interested in metaproteomic, four steps must be taken into consideration: removing chemicals and reagents that may interfere with restriction digestion and mass spectrometry identification; removing host cells and enhancing microbial cells; classifying samples before mass spectrometry detection; extracting protein efficiently from the complete microbial community.

5. Conclusion

The primary advantage of metaproteomics over metagenomics and metatranscriptomics is that it is based on "functional" information. With the use of a precise database of protein sequences, we may determine which species or higher taxa, such as genus, family, etc. are there. We can gain a deeper understanding of the functional process and relationships of various community members.

Metaproteomic investigations may be challenging in the absence of genetic databases of the microbes or the host; therefore, it is preferred to conduct a metaproteomic study in conjunction with metagenomic analysis.

Disclosure

The authors report no conflicts of interest in this work.

References

Armengaud, J., Marie Hartmann, E., Bland, C., 2013. Proteogenomics for environmental microbiology. Proteomics 13, 2731–2742. https://doi.org/10.1002/PMIC.201200576

Branda, S.S., Vik, Å., Friedman, L., Kolter, R., 2005. Biofilms: The matrix revisited. Trends in Microbiology

13, 20–26. https://doi.org/10.1016/j.tim.2004.11.006
Chen, L.H., Lin, C.H., Chung, K.R., 2013. A nonribosomal peptide synthetase mediates siderophore production and virulence in the citrus fungal pathogen Alternaria alternata. Molecular plant pathology 14, 497–505. https://doi.org/10.1111/MPP.12021
Davey, M.E., O'toole, G.A., 2000. Microbial biofilms: from ecology to molecular genetics. Microbiology and molecular biology reviews : MMBR 64, 847–867. https://doi.org/10.1128/MMBR.64.4.847-867.2000
Denef, V.J., Kalnejais, L.H., Mueller, R.S., Wilmes, P., Baker, B.J., Thomas, B.C., VerBerkmoes, N.C., Hettich, R.L., Banfield, J.F., 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proceedings of the National Academy of Sciences of the United States of America 107, 2383–2390. https://doi.org/10.1073/PNAS.0907041107
Fischer, C.R., Bowen, B.P., Pan, C., Northen, T.R., Banfield, J.F., 2013. Stable-isotope probing reveals that hydrogen isotope fractionation in proteins and lipids in a microbial community are different and species-specific. ACS Chemical Biology 8, 1755–1763. https://doi.org/10.1021/CB400210Q
Guazzaroni, M.E., Herbst, F.A., Lores, I., Tamames, J., Peláez, A.I., López-Cortés, N., Alcaide, M., Del Pozo, M. V., Vieites, J.M., Von Bergen, M., Gallego, J.L.R., Bargiela, R., López-López, A., Pieper, D.H., Rosselló-Móra, R., Sánchez, J., Seifert, J., Ferrer, M., 2012. Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. The ISME Journal 2013 7:1 7, 122–136. https://doi.org/10.1038/ismej.2012.82
Hanson, B.T., Hewson, I., Madsen, E.L., 2014. Metaproteomic survey of six aquatic habitats: discovering the identities of microbial populations active in biogeochemical cycling. Microbial ecology 67, 520–539. https://doi.org/10.1007/S00248-013-0346-5
Herbst, F.A., Bahr, A., Duarte, M., Pieper, D.H., Richnow, H.H., von Bergen, M., Seifert, J., Bombach, P., 2013. Elucidation of in situ polycyclic aromatic hydrocarbon degradation by functional metaproteomics (protein-SIP). Proteomics 13, 2910–2920. https://doi.org/10.1002/PMIC.201200569
Justice, N.B., Li, Z., Wang, Y., Spaudling, S.E., Mosier, A.C., Hettich, R.L., Pan, C., Banfield, J.F., 2014. (15)N- and (2)H proteomic stable isotope probing links nitrogen flow to archaeal heterotrophic activity. Environmental microbiology 16, 3224–3237. https://doi.org/10.1111/1462-2920.12488
Justice, N.B., Pan, C., Mueller, R., Spaulding, S.E., Shah, V., Sun, C.L., Yelton, A.P., Miller, C.S., Thomas, B.C., Shah, M., VerBerkmoes, N., Hettich, R., Banfield, J.F., 2012. Heterotrophic archaea contribute to carbon cycling in low-pH, suboxic biofilm communities. Applied and environmental microbiology 78, 8321–8330. https://doi.org/10.1128/AEM.01938-12
Kircher, M., Kelso, J., 2010. High-throughput DNA sequencing – concepts and limitations. BioEssays 32, 524–536. https://doi.org/10.1002/BIES.200900181
Kleiner, M., 2019. Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. Msystems 4, e00115-19. https://doi.org/10.1128/MSYSTEMS.00115-19
Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.J., Shah, M., VerBerkmoes, N.C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T., Dubilier, N., 2012. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proceedings of the National Academy of Sciences of the United States of America 109, E1173–E1182. https://doi.org/10.1073/PNAS.1121198109/SUPPL_FILE/SD04.XLS
Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., Rocap, G., 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. The ISME Journal 2010 4:5 4, 673–685. https://doi.org/10.1038/ismej.2010.4
Ram, R.J., VerBerkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., Shah, M., Hettich, R.L., Banfield, J.F., 2005. Community Proteomics of a Natural Microbial Biofilm. Science 308, 1915–1920. https://doi.org/10.1126/SCIENCE.1109070/SUPPL_FILE/RAM.SOM.REV.PDF
Seifert, J., Herbst, F.A., Halkjær Nielsen, P., Planes, F.J., Jehmlich, N., Ferrer, M., Von Bergen, M., 2013. Bioinformatic progress and applications in metaproteogenomics for bridging the gap between genomic sequences and metabolic functions in microbial communities. Proteomics 13, 2786–2804. https://doi.org/10.1002/PMIC.201200566
Sowell, S.M., Abraham, P.E., Shah, M., Verberkmoes, N.C., Smith, D.P., Barofsky, D.F., Giovannoni, S.J., 2011. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. The ISME journal 5, 856–865. https://doi.org/10.1038/ISMEJ.2010.168
Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V. V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43. https://doi.org/10.1038/NATURE02340
Urich, T., Lanzén, A., Stokke, R., Pedersen, R.B., Bayer, C., Thorseth, I.H., Schleper, C., Steen, I.H., Øvreas, L., 2014. Microbial community structure and functioning in marine sediments associated with diffuse

hydrothermal venting assessed by integrated meta-omics. Environmental microbiology 16, 2699–2710. https://doi.org/10.1111/1462-2920.12283

VerBerkmoes, N.C., Denef, V.J., Hettich, R.L., Banfield, J.F., 2009. Functional analysis of natural microbial consortia using community proteomics. Nature Reviews Microbiology 2009 7:3 7, 196–205. https://doi.org/10.1038/nrmicro2080

Williams, T.J., Long, E., Evans, F., Demaere, M.Z., Lauro, F.M., Raftery, M.J., Ducklow, H., Grzymski, J.J., Murray, A.E., Cavicchioli, R., 2012. A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. The ISME journal 6, 1883–1900. https://doi.org/10.1038/ISMEJ.2012.28

Wilmes, P., Bond, P.L., 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environmental Microbiology 6, 911–920. https://doi.org/10.1111/J.1462-2920.2004.00687.X

Wilmes, P., Heintz-Buschart, A., Bond, P.L., 2015. A decade of metaproteomics: Where we stand and what the future holds. PROTEOMICS 15, 3409–3417. https://doi.org/10.1002/PMIC.201500183