JOURNAL OF PROTEOMICS 97 (2014) 27–35

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Review

Marine metaproteomics: Current status and future directions☆

*Da-Zhi Wang**, *Zhang-Xian Xie*, *Shu-Feng Zhang*

State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen 361005, China

ARTICLE INFO

Available online 13 September 2013

Keywords:
Metaproteomics
Marine ecosystem
Microbial community
Ecological function
Biogeochemical cycle

ABSTRACT

Metaproteomics is a new field within the 'omics' science which investigates protein expression from a complex biological system and provides direct evidence of physiological and metabolic activities. Characterization of the metaproteome will enhance our understanding of the microbial world and link microbial communities to ecological functions. Recently, the availability of extensive metagenomic sequences from various marine microbial communities has extended the postgenomic era to the field of oceanography. Although still in its infancy, metaproteomics has shown its powerful potential with regard to functional gene expression within microbial habitats and their interactions with the ambient environment as well as their biogeochemical functions. However, the application of metaproteomic approaches to complex marine samples still faces considerable challenges. This review summarizes the recent progress in marine metaproteomics and discusses the limitations of and perspectives for this approach in the study of the marine ecosystem.

This article is part of a Special Issue entitled: Trends in Microbial Proteomics.

Contents

☆ This article is part of a Special Issue entitled: Trends in Microbial Proteomics.

* *Corresponding author*. Tel.: +86 592 2186016; fax: +86 592 2180655.
E-mail address: dzwang@xmu.edu.cn (D.-Z. Wang).

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.08.024

1. Introduction

Marine ecosystem is the largest aquatic system on the planet. It is reported that coastal habitats alone account for approximately 1/3 of all marine biological productivity, and play important roles in regulating the global climate. It is estimated that the huge body of seawater, with an average of around 2.5×10^6 cells mL^{-1}, harbors significant microbial populations [1], and with a wide range of habitat diversity, including coastal and ocean waters with different influences by human activity; up- and down-welling systems with different nutrient transportation; and surface, intermediate and deep waters with gradient changes of light, temperature and pressure. The native microorganisms play crucial roles in the biogeochemical cycling of elements, such as carbon, phosphorus and nitrogen, as well as organic matter (OM) decomposition and remineralization [2]. Therefore, the study of mixed microbial communities within their natural marine environment is the key to the investigation of the diverse roles played by microorganisms, and to identify the microbial potential for specific environmental stresses.

Metaproteomics is a new field within the 'omics' science which attempts to identify all the proteins expressed at a given time within an ecosystem, and plays a key role in the determination of microbial function [3]. Metaproteomics has been applied in a variety of environments [4–16] as well as human health [17–19]. Recently, with the extensive metagenomic sequences from various marine microbial communities becoming available, metaproteomics has also attracted considerable attention in the field of marine science. Up to now, there are 76 marine metagenomic projects available online, 23 of which have been completed based on the Genomes Online Database (http://www.genomesonline.org/cgi-bin/ OLD). Additionally, more and more marine microorganisms are becoming subjected to whole genome sequencing since the first marine archaeon, *Methanocaldococcus jannaschii*, was sequenced [20]. Recently, a metagenomic study of the marine planktonic microbiota yields an extensive dataset consisting of 7.7 million sequencing reads (6.3 billion bp) which predicts 6.12 million proteins [21]. These predictions add tremendous diversity to known protein families and cover nearly all known prokaryotic protein families, which provide a powerful protein database for identifying proteins in the marine ecosystem. This study has made metaproteomics better available in the field of marine science by providing a more relevant database. In this paper, we review the advancement of metaproteomics in the marine environment, and discuss the challenges of this approach in the study of the marine ecosystem.

2. Metaproteomic analysis strategies

With the rapid development of mass spectrometry (MS) technology over the past few decades, several strategies have been applied in marine metaproteomic study (Fig. 1). Typically, the metaproteomic approach involves up to four main steps, namely sample collection; protein extraction, purification and fractionation; MS analysis; and finally protein interpretation with further bioinformatics analysis. Two major work flows have been developed: 1) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) coupled either with matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF-TOF) MS analysis or with electrospray ionization source-tandem MS (ESI-MS/MS) analysis; and 2) liquid chromatography coupled with electrospray ionization source-tandem MS (LC-ESI-MS/MS). However, because of the wide range of protein expression and perturbing matrix compounds likely to challenge its application in complex marine samples, as well as its being restricted by molecular sizes, p*I* ranges, and hydrophobicity of the proteins, the usage of two-dimensional gel electrophoresis (2-DE) in combination with MALDI-TOF-TOF MS has waned and is much less important as an identification technique in general. As a result, LC-MS/MS approaches have become more popular and successful in recent studies. Regardless of any work flow, it should be considered whether quantification is needed in the experiment design, either using labeling or label-free approaches. After that, metaproteomics can begin with protein extraction.

Another important issue is that during protein extraction, including purification and concentration, care should be taken to avoid bias and loss at each step. Care is also necessary to avoid the introduction of additional interfering compounds in the final extract, which might decrease the effectiveness of digestion or hamper MS analysis. Usually, proteins are separated using either 1-DE or 2-DE, and then subjected to digestion into peptides using trypsin or other enzymes. After that, the peptides are brought to MS analysis or further LC separation using a microcapillary column of C18 reversed-phase (RP) or strong-cation-exchange phase (SCX). Times of peptide separation depend on the complexity of the sample. Often, it is sufficient to use 2D-GE-MS or 1-D SDS PAGE plus LC-MS/MS or 2-D (SCX-RP) LC-MS/MS. Sometimes, multidimensional peptide separation is needed when the sample composition is too complex. The next most important step is MS analysis. In this step the raw data are output and the following analysis is totally based on the data. The sensitivity, accuracy and scanning speed of MS determine the quality of the raw data as well as the availability of the data. Frequently used MS including Q-TOFs or LTQ-orbitrap or FT-ICR are state-of-the-art systems for high performance tandem MS measurement. The most frequently used ionization techniques are MALDI and ESI, the former where ionization of peptides is triggered with a laser beam matrix-embedding the peptides, while the latter disperses a peptide-containing liquid using electrospray to achieve ionization. Such improvements of the mass spectrometer technique facilitate better protein identification, helping greatly in the detection of low abundant proteins, even making the possibility of single-cell proteomics come true in the future.

Raw data are further submitted to interpretation with several software packages such as Mascot [22,23], SEQUEST [23] and de novo software (PEAKS [24] for example) to achieve confident identification. Two strategies have been developed, direct mass spectra based and *de novo* peptide sequence based and, after that, a quantitative step can be carried out based on the raw data. Recently some software packages such as DTAselect [25] or Scaffold [26] have been applied to sort and filter the raw data, and to conduct quantitative comparison between samples by counting the peptide spectra, a process termed semi-quantitative proteomics. In addition, other commercial software such as SIEVE (version 2, ThermoFisher,