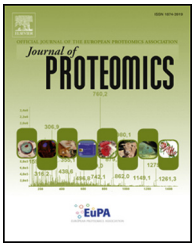
JOURNAL OF PROTEOMICS XX (2013) XXX–XXX

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

Article history:
Received 17 April 2013
Accepted 30 August 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
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,

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

Scientific) and Progeneis LC/MS (NonLinear Dynamics) are also available for the label-free quantitative proteomic analysis. Finally, the last stage is the visualization of the complex functional information [27].

3. Metaproteomic analyses of marine samples

Since the metaproteomic approach was first applied to study the mixed microbial communities in 2004 [3], it has subsequently been extended to the marine ecosystem. Although metaproteomics is still in its infancy, it has shown great potential to unveil novel functional genes and their interactions with the ambient environment as well as their microgeochemical functions. Here, we have grouped the metaproteomic studies in the marine science literature (Table 1) into three main thematic blocks, and some from each block are discussed below.

3.1. *Protein expression in microbial communities and their in situ physiological states*

The first metaproteomic study of aquatic microbial assemblages compares the protein expression in various samples from Chesapeake Bay using the 2-DE method and identifies the excised protein spots using LC-MS/MS coupled with MS–BLAST searching [1]. Spot numbers on 2-DE gel range from 155 to 207, with the most similar protein files sharing 70% of the common spots to the less similar ones sharing 30% of the common spots in Chesapeake Bay metaproteome samples. However, the protein information is too limited in that only three proteins were tentatively identified owing to the lack of a substantial DNA sequence database. The study, for the first time, demonstrates the raw power of metaproteomics using the 2-DE approach for exploring natural microbially expressed proteins to open up the natural marine microbial world to closer scrutiny, and also presents its difficulties in application. Later, another study investigated the SAR11 metaproteome at low-nutrient extremes in the up-welling system of the Sargasso Sea using the high- throughput non-gel based strategy of 2-D LC coupled to LTQ MS/M [28]. A total of 6533 peptides matched to 1042 proteins are identified, among them mass spectra from SAR11 transporters (such as periplasmic substrate-binding proteins for phosphate, amino acids, phosphonate, sugars and spermidine) are highly abundant, which support the view of extreme competition for multiple nutrients in oligotrophic systems and the competitive capability of cells via maximizing their nutrient uptake activity, particularly those involving phosphorus uptake, in nutrient-depleted environments.

In contrast to oceanic and periodic upwelling in the Sargasso Sea, coastal upwelling is the most common system in the ocean, and is usually characterized with the high primary production caused by nutrient enrichment and high-intensity mixing. Different from the study of the Sargasso Sea which focuses on the transporters of SAR11 clade, another study targets this highly productive region during an upwelling period on the Oregon shelf to reveal the dominant metabolic processes occurring within marine bacterial communities in their natural environment [29]. In this latter study, 7151 distinct peptide sequences mapped to 13469 eCDSs are detected and 481 unique protein families are identified using 2-D LC MS/MS. Among them, transport proteins again are highlighted and account for 11 out of the 25 most abundantly represented protein families. Interestingly, SAR11 transport proteins also contribute a significant proportion to the total spectral count, but even excluding SAR11 eCDSs, there are still nine transport protein families ranked in the top 25. These results support the hypothesis concerning the importance of effective nutrient scavenging for the adaptation of microbial cells either in nutrient replete or deplete conditions, and a determining factor in microbial competition and survival. However, differences between oceanic and coastal upwelling systems should be expected. Results from the two studies show that SAR11 phosphate and phosphonate transporters are abundant in the ocean gyre ecosystem but nearly absent from the Oregon shelf upwelling system. Meanwhile, other studies show that the proteins involved in phosphorus transport are more abundant in terms of phosphate-starvation [30,31], this further demonstrates that oceanic oligotrophic gyre surface water has more chance to be subjected to phosphorus-limitation than productive coastal surface water. In addition, the substrates of SAR11 transport proteins are mainly carbon- and nitrogen-containing compounds, but not phosphate, suggesting that carbon and nitrogen, rather than phosphorus, are the major factors for the niche differentiation and productivity limitation in this productive region. The metaproteome of the Oregon shelf microbial community is mainly matched to the SAR11 clade, *Roseobacter* clade, oligotrophic marine gammaproteobacteria group and OM43 clade, and therefore one of the most important characteristics of proteomics might be that metaproteomics provides taxon-specific confirmation of the gene expression in the natural community, for example the expression of the OM43 clade methanol dehydrogenase in the above studies, suggesting that the metabolism of one-carbon compounds by these methylotrophs might play a key role in biogeochemical processes in the coastal ocean.

Different from the above studies which lacked quantitative information, a semi-quantitative method using peptide spectral counting was introduced to perform a comparative metaproteomic study targeting at the membrane proteins expressed by the in situ microbial consortia of surface water from an oligotrophic gyre to a productive coastal upwelling region [32]. Along a natural gradient in nutrient concentrations from the open ocean to coastal region in the South Atlantic, a suite of 10 surface seawater samples targeting the size fraction less than 0.8 μm, was analyzed using LC-MS/MS to compare the metaproteomic profiles. Totally, 5389 peptides hitting 2273 proteins are identified, with an average of 428 ± 158 distinct membrane proteins per sample, which is fewer than the 1042 proteins detected in a previous whole-cell metaproteomic study in the Sargasso Sea [28]. More interestingly, the comparative membrane metaproteomics revealed ocean-scale shifts in microbial nutrient utilization and energy transduction along an environmental gradient. Compound-specific transport proteins showed a regular distribution along the gradient, for example, urea and ammonia transporters dominate in the open ocean. Two types of energy transduction are characterized: one is light energy transduction represented by proteorhodopsin, a light-driven proton pump providing cells with a way to generate energy from sunlight. Four conserved peptide groups from it are

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

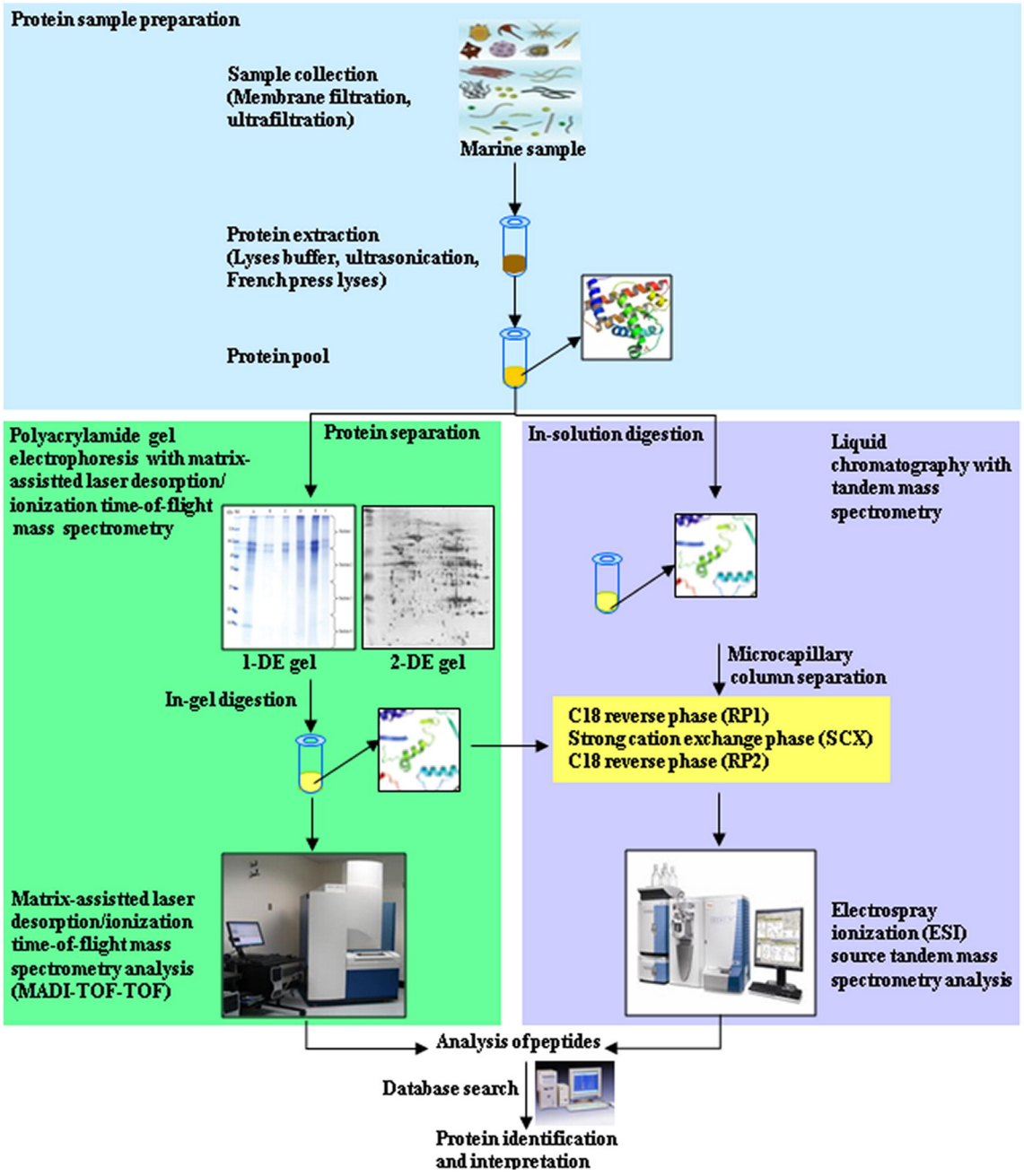


Fig. 1 – Typical workflow for marine metaproteomic analysis. Following common sample preparation, two strategies have been developed: 1) polyacrylamide gel electrophoresis coupled either with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis or with electrospray ionization source tandem mass spectrometry analysis, and 2) liquid chromatography coupled with electrospray ionization source tandem mass spectrometry.

described and different peptide groups have a different distribution feature. In addition, the first verification of in situ protein expression of light-driven proton pumps outside the alpha-proteobacterial SAR11 clade is an intriguing finding in this study. Another one is the mechanical energy transduction exhibited by TonB-dependent transporters (TBDTs), which is known to utilize energy from the cytoplasmic proton motive force resulting in conformational shift of TonB's structure to transport nutrients across the outer membrane of Gram-negative bacteria. TBDTs account for 19% of all detected spectra and are found in each sample, suggesting their activities are important for microbial nutrient acquisition. Furthermore, archaeal ammonia monooxygenase proteins found in the upwelling region suggest archaeal nitrification in nutrient-rich surface waters. Meanwhile, the viral proteins identified in each sample further confirm ubiquitous viral

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

Table 1 – Summary of metaproteomic studies in marine science from a literature search.

Sample description (sampling depth)	Peptides/proteins identified	Protein/peptide separation method	MS platform	Database used	Quantification	Reference/ year
Microbial community						
Estuary surface samples (1 m)	11/3	2-D PAGE + LC	MALDI-TOF MS, Q-TOF MS/MS	Protein sequence database from the Matrix Science Mascot web	Yes	[1]/2005
Oceanic surface samples (5 m)	6533/1042	2-D LC	LTQ MS/MS	SAR11/Prochlorococcus/Synechococcuse eCDS from Sargasso Sea metagenome as well as genomes from sequenced isolates	No	[28]/2008
Coastal upwelling samples (10 m)	7151/481[a]	2-D LC	LTQ-Orbitrap MS/MS	Translated GOS2-11 eCDS database as well as genomics from two Oregon coastal isolates [31]	No	[29]/2011
Oceanic and coastal upwelling samples (5–8 m)	5389/2273	1-D LC (reverse-phase chromatography)	LTQ-Orbitrap MS/MS	Extensive database from the GOS metagenomics project containing over 600,000 predicted proteins	Yes	[32]/2010
Marine organic matter						
POM samples (41, 200, 500, 1000 m)	–/505	1-D PAGE + LC	LTQ MS/MS	Constructed protein database form National Center for Biotechnology Information (NCBI) containing 780573 protein entries from the major planktonic communities	No	[34]/2010
DOM samples (10, 75, 3000 m)	286/182	1-D PAGE + LC	LTQ-Orbitrap MS/MS	Combined protein database downloaded from NCBI and the Moore Foundation marine microbial peptide dataset	Yes	[35]/2011
DOM samples (10 m)	993/367	1-D PAGE + LC	LTQ-Orbitrap MS/MS	GOS combined assembly protein database	Yes	[33]/2013
POM and sediment samples (4, 50, 100, 40, 60, 100 m)	–/207, 11, 22, 136, 53, 82 -/52,24,23	1-D PAGE + LC	LTQMS/MS	Constructed database containing proteomes of two diatom isolates and an SAR11 clade isolate	No	[36]/2012
Microbial community from special habitats						
Cold seep sample	–/356	1-D PAGE + LC	LTQ-Orbitrap MS/MS	Custom Mascot database based on ORF predictions according to the metagenomic data	No	[37]/2012
Symbiotic sample	–/2819	1-D PAGE + LC, 2-D LC	LTQ-Orbitrap MS/MS	Databases composed of the symbiotic metagenomes and the genomes of related organisms	Yes	[38]/2012
Algal bloom water sample	–/–	1-D PAGE + LC	LTQ MS/MS	Database created using corresponding metagenomics database	Yes	[39]/2012

'Separation'-the method to separate proteins: one dimensional electrophoresis (1-DE), two-dimensional electrophoresis (2-DE); 1/2-D LC, two-dimensional liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization-time of flight; LTQ MS/MS, linear trap quadrupole tandem mass spectrometer.

[a] Refers to protein families;

infection and biogeochemical roles of viruses in the marine ecosystem.

3.2. Metaproteomic characterization of marine organic matter

Marine OM plays an important role in oceanic and global carbon cycling, which in turn impacts weather. Understanding the processes involved in the production, transformation, remineralization and protection of OM in the oceans has been a primary goal of marine biogeochemists and oceanographers over the past few decades. Proteins, as the actual machinery that brings about cell growth, proliferation and homeostasis, are a major component of OM, and therefore, the study of proteins should provide new insights into the biogeochemical cycling of OM in the ocean. Recently, a metaproteomic study focusing on the dissolved organic matter (DOM) in the surface seawater of the South China Sea reveals protein composition and origin and its protection mechanism [33]. Using a shotgun proteomic approach, a total of 993 unique peptides matching to 367 protein groups are identified from four surface DOM samples within the size fraction between 5 kDa and 0.2 μm. Based on bioinformatics analysis, about half of the proteins did not have an exact taxonomic classification or function, suggesting that more effort should be devoted to explore gene pools in the marine environment. In terms of the proteins identified, the Rhodospirillaceae, *Prochlorococcus*, SAR11 clade and viruses are the major contributors to the dissolved protein pool. Similarly, viral proteins are also reported to exist in surface water samples [29,32]. All of these again demonstrate the notable roles of viruses in ocean carbon and nutrient cycling. Interestingly, the proteins involved in transport and other cellular processes are highly detected, which might result from microbial secretion or detrital proteins.

Particulate organic matter (POM) represents the link between marine surface primary production and the burial of OM in marine sediments, and plays important roles in regulating marine carbon cycling and global climate. A recent report characterizes the metaproteome of POM within a water column to see what happens from the surface down to deep water [34]. The authors collected POM of over 0.7 μm size from near the surface (41 and 200 m) and mesopelagic layers (500 and 1000 m) in an up-welling system in the South China Sea, and confidently identify 505 proteins. Proteins derived from the Cyanobacteria dominate throughout the whole water column, although particulate protein composition varies at different layers. Their study further confirms that living marine organisms are an important source of POM. More importantly, specific proteins are detected, such as archaeal ammonia monooxygenases, in the 200, 500 and 1000 m POM samples, together with the same proteins that were found at a coastal upwelling region in South Atlantic surface waters, probably demonstrating that the Archaea are important nitrifiers and widely distributed in the oceans, even from the surface down to the deep ocean. Certainly, proteins involved in transport or nutrient utilization, stress responses or adaptation, and energy generation are detected and the distribution at different layers are described, indicating that different biological processes occur along the water column. Another finding of this study is that it provided some insights into the fate of particulate proteins, such as remineralization of POM and mechanisms of protecting OMs from degradation. Recently, the protein profiles of DOM are targeted in the water column at the oligotrophic region [35]. In this study, three samples with size fraction less than 0.7 μm were collected at the surface (10 and 75 m) and bathypelagic (3000 m) layers. Each sample was separated into two fractions with a cut-off of 0.2 μm. In total, 182 proteins and 286 peptides are identified. The highest protein number is detected in the samples at the 75 m layer while the other two samples show an insignificant difference. In addition, the protein number in each sample with size less than 0.2 μm is comparable. When examining the functional classification (excluding the non-functional annotation proteins) for the fraction over 0.2 μm, proteins involved in lipid, amino acid and inorganic ion transport and metabolism, photosynthesis, and cell wall or membrane or envelope biogenesis are more highly detected in the 75 m layer sample than in the other two samples, while proteins involved in translation, ribosomal structure and biogenesis, and defense mechanisms are more abundant in the 3000-m layer. For the fraction less than 0.2 μm, urea ABC transporter is the frequently detected cyanobacterial protein only in the 10 m sample, and the MR involved in energy production and conversion is the abundant archaeal protein in samples from the 75 and 3000 m layers. These results indicate the diverse and dynamic feature of dissolved proteins along the water column and the biological activities at different water depths. The above two studies also discuss the inputs and preservation mechanism of proteins from the upper water layer down to the deep sea. More recently, a study was conducted to identify and track proteins through the water column in the Bering Sea, during and after a diatom bloom [36]. In the surface water, 207 proteins are identified, decreasing through the water column to 52 in the post-bloom shelf surface sediment, to finally 24 in the deeper basin sediment. Most of the proteins identified in POM samples are of diatom origin, which might be closely related to the spring bloom. The preferentially retained proteins during recycling are the organelle-bound, transmembrane, photosynthetic, and light harvesting related proteins, which are also found in POM in the mesopelagic layers in the western South China Sea [34]. All these findings suggested that organelle and membrane protection play important roles in protein preservation.

3.3. Metaproteomic analysis of microbial communities from special habitats

Marine microorganisms live in complex and/or extreme ecosystems, and can be exposed to high levels of salinity, temperature and pressure gradients, as well as oxidative conditions, etc. and, therefore, they have developed various ways of protecting themselves against these environmental challenges. Several metaproteomic studies on special marine habitats, for example, a marine cold seep [37], a symbiotic system [38], or sudden events, i.e. a phytoplankton bloom [39] are reported. A dynamic succession of populations at genus-level in the North Sea is traced and the bacterioplankton response to a diatom bloom using both metagenomic and metaproteomic approaches with emphasis on the expression of carbohydrate-active enzymes and phosphate acquisition is investigated [39]. The results indicate that the distinct populations of *Bacteroidetes*, *Gammaproteobacteria* and *Alphaproteobacteria* are

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

specialized for successive decomposition of algal-derived OM, suggesting that a series of ecological niches resulting in a bloom of special populations is provided by the algal substrate availability. This revealed the mechanism whereby planktonic species compete with each other in such a seemingly homogeneous habitat, thus again demonstrating metaproteomics as a powerful tool to illustrate the relationship between microbial communities and their habitats. Another study investigates an unusual marine habitat (cold seep sediments) also using a combination of the metagenomic and metaproteomic approach [37]. These powerful tools allow an assessment of the major metabolic pathways involved in sulfate dependent anaerobic oxidation of methane in cold seep sediments, where the associated microbial community is dominated by free-living anaerobic methanotrophic archaea-1 (ANME-1). In this environment, 245 of the 356 proteins identified are expressed by ANME-1. The key enzymes involved in the reverse methanogenesis pathway are identified and a complete dissimilatory sulfate reduction pathway is detected in the sulfate-reducing *Deltaproteobacteria*. One more report involves the study of a symbiotic microbial community in the deep sea, which is characterized by low nutrient and energy availability [38], and metaproteomics combined with metabolomics is applied to investigate the intricate network of metabolic interactions between the host and its symbionts. A total of 2819 proteins and 97 metabolites are identified and quantified, and some previously undescribed pathways are proposed, which include pathways for symbiont assimilation of the host waste products (acetate, propionate, succinate and malate), the potential use of carbon monoxide and hydrogen, CO_2 fixation and sulfate reduction, and thus indicate that the strong stress in such an oligotrophic environment shapes the closed relationship between the host and symbionts through strict natural selection.

4. Challenges and perspectives of marine metaproteomics

Although metaproteomics has been applied to answer a variety of scientific questions concerning marine ecosystems and has shown its power in revealing special features of mixed marine samples and overall microbial ecosystem function, there are still obstacles standing in the way of the application of metaproteomics in marine science.

First of all, protein sample preparation for MS analysis. The success of metaproteomic analysis relies on unbiased protein extraction from a complex environmental sample [40]. For microbial isolates, the dynamic range of expressed proteins can reach the order of 10^4–10^6, even larger with a complex environmental community [41]. Due to this, there are problems such as low abundant proteins and effectiveness of cell lysis among different species as well as the presence of minorities in a microbial community which are hard to include in the extracts. Thus, the challenges in complete proteomic characterization due to organismal complexity and the wide-range of protein expression force us to improve the protocols of protein extraction. Many studies have made great efforts on this aspect, such as soil [42,43], sediment [7], wastewater treatment biofilm [44], and marine biofilm [45]. For the marine environment, up to now, few studies report on a comparison of methods. One study compares three procedures to extract proteins from GF/F filter, and finds that sonication is good to help protein release to the lysis buffer, and finally compares 2-DE and 1-D SDS PAGE combining LC-MS/MS [46]. The results show that 2-DE is not a good method for protein separation due to the highly smeared staining background, similarly seen in other studies [1,47]. The other reasons for the disadvantages involved in using 2-DE for metaproteomic studies include limitations in the molecular size, p*I* range, and hydrophobicity of the proteins that can be analyzed, as well as low throughput and its time consuming nature even with automatic techniques. Standard proteomics based on the LC-MS system greatly fill the gap, especially increasing by 1–2 orders of magnitude the dynamic range so as to further enhance the detection of low abundance proteins in samples, and the strategy is known to be cost-effective. On the other hand, pre-fixation as soon as the samples are collected might often be ignored. Intracellular protein expression is a dynamic process, so either using chemical reagents such as chloramphenicol or protease inhibitor cocktail to stop protein synthesis and inhibit protease activities [1] or collecting samples as fresh as possible [32,39] to keep microbial communities at an in situ status are alternatives. However, care should be taken because the former probably alters the physiologic status thus lowering the power of metaproteomics to investigate the real world of microbial consortia, while the latter will be restricted when a large volume of water is needed. Both of these problems will be addressed by improving the effectiveness of protein extraction and enhancing the sensitivity of MS.

Secondly, the effectiveness of protein identification is directly and greatly restricted by the correlation between the community composition and the databases for searching. For example, using the GOS metagenomic library, a database closely related to the sample suite, 6.2 times as many peptides were identified than when searching against the GenBank nonredundant database [32]. Another study extends the database using the Thaps database plus the GOS Combined Assembly Protein database, which contains over 6 million marine microbial proteins sequenced from genomic data, to identify proteins from diatom post-bloom shelf sediment, but no unique bacterial protein is identified with confidence [36]. The evidence is strong because the sample for testing is thought to be microbial protein dominated. Nowadays, more and more studies suggest using a database dependent on the generation of environment-specific metagenomes [32,37,39]. Another way to resolve the problem is to devote time and effort to the development of *de novo* methods [48–50]. Efforts using the traditional method continue to bring additional microbial lineages into culture, but this is difficult and time-consuming because it is well known that more than 90% of microorganisms are hard to be cultured successfully. In addition, it is impossible to sequence all the organism genomes in a microbial community. However, whole genomic sequences of isolates will serve as a complement for the environment-specific metagenomes [28,29]. For genome annotation, although the great advantages of the new DNA sequencing technologies can undoubtedly generate vast quantities of DNA sequences of high quality, the use of fragmentary metagenomic datasets can still be restricted if sequences cannot be accurately assigned to the correct

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

organism. Many studies show a fraction of unknown proteins without any functional annotation or taxonomic assignment [29,33–35], and the reasons are probably first the fitness of the database chosen for searching; second, the low quality of the mass spectra to be matched to the distinct sequence in order to distinguish its ortholog belonging to other organisms; and the lack of a definite annotation of the protein or DNA sequence, the homologies of which do not exist in the database.

Thirdly, due to the high-throughput protein datasets in a typical metaproteomic experiment, how to pick up the bio-information available and present the proteomic data for visualization by readers is difficult for a biologist without good computer science training. Fortunately, a good review now provides detailed information concerning the visualization of proteomic data [27]. However, more effort should be devoted to resolve this problem and to make software more available and easier-to-use.

Although challenges in metaproteomic studies confront marine scientists, the metaproteomic approach has shown its powerful capability and great potential in promoting understanding of in situ microbial communities and their interaction with the environment. Up to now, environmental microbiology lies at the heart of ecology, and integrative approaches (which associate the in situ investigation of the diversity and functions of microbes with physical/chemical characterization of their environment and in vitro molecular and physiological microbiology) are being applied to provide new opportunities for outstanding discoveries. For marine science, where there is little knowledge due to the difficulties in sample collection and the inaccessibility of many regions, applied metaproteomics still has much space to develop. Probably, two types of studies will be prevalent in the future, based on how many species are in an environment sample of concern. On the one hand, there must be a species or group which is dominant in an ecosystem, which can then be considered as the key species or group or special target of interest. Metaproteomic investigations in the Sargasso Sea [28] and marine cold seep [37] provide examples of this aspect. A novel method using microwave cell fixation and flow cytometric sorting is now available for metaproteomic and metagenomic analysis [51], thus making the aim of identifying the metaproteome of a defined oceanic microbial population possible. This might be useful for relatively less complex microbial communities such as communities during an algal bloom. In addition, the physiology of species cultivation in the laboratory might be validated using this approach in the natural environment. Furthermore, a new technique based on single-cell genomic or proteomic analysis has been addressed, which might possibly be a more powerful tool applied in the future. On the other hand, efforts should continue to be devoted to the interactions among populations in environmental communities in order to uncover the functioning of the microbial community in the natural environment, of which some might be focused on unknown or little known ecosystems. Furthermore, quantitative metaproteomics has been used in several studies (see Table 1), mostly using the semi-quantitative method, and this will provide more information concerning the level of protein expression in both the aspects mentioned above. In all, metaproteomics together with other omics approaches will greatly support studies in marine science, thus making the real microbial world and the ecological functions involved more clearly revealed to us in the future.

Acknowledgements

This study was supported by the National Natural Science Foundation of China through grants 41121091 and 41230961, and the Ministry of Science and Technology through grant 2009CB421203. D. Z. Wang was also supported by the Program for New Century Excellent Talents in University of China. We thank Professor John Hodgkiss from The University of Hong Kong for his help with English.

REFERENCES

[1] Kan J, Hanson TE, Ginter JM, Wang K, Chen F. Metaproteomic analysis of Chesapeake Bay microbial communities. Saline Syst 2005;1:7.
[2] Pierre-Alain M, Christophe M, Séverine S, Houria A, Philippe L, Lionel R. Protein extraction and fingerprinting optimization of bacterial communities in natural environment. Microb Ecol 2007;53:426–34.
[3] Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol 2004;6:911–20.
[4] Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, et al. Community proteomics of a natural microbial biofilm. Science 2005;308:1915–20.
[5] Williams MA, Taylor EB, Mula HP. Metaproteomic characterization of a soil microbial community following carbon amendment. Soil Biol Biochem 2010;42:1148–56.
[6] Wang HB, Zhang ZX, Li H, He HB, Fang CX, Zhang AJ, et al. Characterization of metaproteomics in crop rhizospheric soil. J Proteome Res 2010;10:932–40.
[7] Benndorf D, Vogt C, Jehmlich N, Schmidt Y, Thomas H, Woffendin G, et al. Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. Biodegradation 2009;20:737–50.
[8] Bruneel O, Volant A, Gallien S, Chaumande B, Casiot C, Carapito C, et al. Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. Microb Ecol 2011;61:793–810.
[9] Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA, et al. Metaproteogenomic analysis of a dominant green sulphur bacterium from Ace Lake, Antartica. ISME J 2010;4:1002–19.
[10] Habicht KS, Miller M, Cox RP, Frigaard NU, Tonolla M, Peduzzi S, et al. Comparative proteomics and activity of a green sulfur bacterium through the water column of Lake Cadagno, Switzerland. Environ Microbiol 2011;13:203–15.
[11] Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D, et al. An integrative study of a meromictic lake ecosystem in Antarctica. ISME J 2011;5:879–95.
[12] Benndorf D, Balcke GU, Harms H, von Bergen M. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J 2007;1:224–34.
[13] Wilmes P, Wexler M, Bond PL. Metaproteomics provides functional insight into activated sludge wastewater treatment. PLoS One 2008;3:e1778.

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024

[14] Jehmlich N, Kleinsteuber S, Vogt C, Benndorf D, Harms H, Schmidt F, et al. Phylogenetic and proteomic analysis of an anaerobic toluene-degrading community. J Appl Microbiol 2010;109:1937–45.
[15] Toyoda A, Iio W, Mitsumori M, Minato H. Isolation and identification of cellulose-binding proteins from sheep rumen contents. Appl Environ Microbiol 2009;75:1667–73.
[16] Burnum KE, Callister SJ, Nicora CD, Purvine SO, Hugenholtz P, Warnecke F, et al. Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. ISME J 2011;5:161–4.
[17] Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J 2009;3:179–89.
[18] Rooijers K, Kolmeder C, Juste C, Dore J, de Been M, Boeren S, et al. An iterative workflow for mining the human intestinal metaproteome. BMC Genomics 2011;12:6.
[19] Rudney JD, Xie H, Rhodus NL, Ondrey FG, Griffin TJ. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. Mol Oral Microbiol 2010;25:38–49.
[20] DeLong EF. Microbial community genomics in the ocean. Nat Rev Microbiol 2005;3:459–69.
[21] Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 2007;5(3):e16.
[22] Cottrell JS, London U. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999;20:3551–67.
[23] Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 1994;5:976–89.
[24] Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 2003;17:2337–42.
[25] Tabb DL, McDonald WH, Yates III JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 2002;1:21–6.
[26] Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun Mass Spectrom 2003;17:2310–6.
[27] Bernhardt J, Michalik S, Wollscheid B, Völker U, Schmidt F. Proteomics approaches for the analysis of enriched microbial subpopulations and visualization of complex functional information. Curr Opin Biotechnol 2013;24:112–9.
[28] Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J 2009;3:93–105.
[29] Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF, et al. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. ISME J 2011;5:856–65.
[30] Dyhrman ST, Haley ST. Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera watsonii*. Appl Environ Microbiol 2006;72:1452–8.
[31] Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci 2006;103:12552–7.
[32] Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J 2010;4:673–85.
[33] Dong HP, Wang DZ, Xie ZX, Dai MH, Hong HS. Metaproteomic characterization of high molecular weight dissolved organic matter in surface seawaters in the South China Sea. Geochim Cosmochim Acta 2013;109:51–61.
[34] Dong HP, Wang DZ, Dai MH, Hong HS. Characterization of particulate organic matter in the water column of the South China Sea using a shotgun proteomic approach. Limnol Oceanogr 2010;55:1565–78.
[35] Wang DZ, Dong HP, Xie ZX, Dai MH, Hong HS. Metaproteomic characterization of dissolved organic matter in the water column of the South China Sea. Limnol Oceanogr 2011;56:1641–52.
[36] Moore EK, Nunn BL, Goodlett DR, Harvey HR. Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. Geochim Cosmochim Acta 2012;83:324–59.
[37] Stokke R, Roalkvam I, Lanzen A, Haflidason H, Steen IH. Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. Environ Microbiol 2012;14:1333–46.
[38] Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Natl Acad Sci 2012. http://dx.doi.org/10.1073/pnas.1121198109.
[39] Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 2012;336:608–11.
[40] Hettich RL, Sharma R, Chourey K, Giannone RJ. Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. Curr Opin Microbiol 2012;15:373–80.
[41] Banfield JF, Verberkmoes NC, Hettich RL, Thelen MP. Proteogenomic approaches for the molecular characterization of natural microbial communities. OMICS 2005;9:301–33.
[42] Chen S, Rillig MC, Wang W. Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection. Proteomics 2009;9:4970–3.
[43] Chourey K, Jansson J, VerBerkmoes N, Shah M, Chavarria KL, Tom LM, et al. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J Proteome Res 2010;9:6615–22.
[44] Abram F, Gunnigle E, O'Flaherty V. Optimisation of protein extraction and 2-DE for metaproteomics of microbial communities from anaerobic wastewater treatment biofilms. Electrophoresis 2009;30:4149–51.
[45] Leary DH, Hervey WJ, Li RW, Deschamps JR, Kusterbeck AW, Vora GJ. Method development for metaproteomic analyses of marine biofilms. Anal Chem 2012;84:4006–13.
[46] Dong HP, Wang DZ, Dai MH, Chan LL, Hong HS. Shotgun proteomics: tools for analysis of marine particulate proteins. Limnol Oceanogr Methods 2009;7:865–74.
[47] Saijo S, Tanoue E. Characterization of particulate proteins in Pacific surface waters. Limnol Oceanogr 2004;49:953–63.
[48] Junqueira M, Spirin V, Balbuena TS, Thomas H, Adzhubei I, Sunyaev S, et al. Protein identification pipeline for the homology-driven proteomics. J Proteomics 2008;71:346–56.
[49] Shevchenko A, Valcu C-M, Junqueira M. Tools for exploring the proteomosphere. J Proteomics 2009;72:137–44.
[50] Wielsch N, Thomas H, Surendranath V, Waridel P, Frank A, Pevzner P, et al. Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches. J Proteome Res 2006;5:2448–56.
[51] Mary I, Oliver A, Skipp P, Holland R, Topping J, Tarran G, et al. Metaproteomic and metagenomic analyses of defined oceanic microbial populations using microwave cell fixation and flow cytometric sorting. FEMS Microbiol Ecol 2010;74:10–8.

Please cite this article as: Wang D-Z, et al, Marine metaproteomics: Current status and future directions, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.08.024