Available online at www.sciencedirect.com

SciVerse ScienceDirect

Geochimica et Cosmochimica Acta 109 (2013) 51–61

Geochimica et Cosmochimica Acta

www.elsevier.com/locate/gca

Metaproteomic characterization of high molecular weight dissolved organic matter in surface seawaters in the South China Sea

Hong-Po Dong [a,b], Da-Zhi Wang [a,*], Zhang-Xian Xie [a], Min-Han Dai [a], Hua-Sheng Hong [a]

[a] *State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen, China*
[b] *Research Center for Harmful Algae and Aquatic Environment, Jinan University, Guangzhou, China*

Received 8 August 2012; accepted in revised form 29 January 2013; Available online 8 February 2013

Abstract

Dissolved organic matter (DOM) is an important reservoir of carbon and energy in the marine environment and plays a key role in regulating the global carbon cycle. This study characterized proteins of high-molecular-weight DOM (size between 5 kDa and <0.2 μm, HMW-DOM) collected from the surface seawaters in the South China Sea using a shotgun proteomic approach in combination with the global ocean sampling combined assembly protein database. A total of 367 protein groups matched by 993 unique peptides from 1991 spectra were identified from four surface HMW-DOM samples. Proteins with unknown taxonomic classification and function dominated the dissolved protein pool (43–53%) while the remaining proteins presented close similarity in biological origin among the four sampling sites. Rhodospirillaceae, *Prochlorococcus*, SAR11 clade and viruses were the major contributors to dissolved proteins in the HMW-DOM from surface seawaters while very few proteins were from the eukaryotic phytoplankton and no archaeal proteins were detected. Transporters with substrate specificities for nitrogen- and carbon-containing compounds (1.5% of the total spectra for each) were highly detected while no phosphate transporters were found, suggesting that carbon and nitrogen might be more limiting than phosphorus in the surface seawater. Viral proteins were assigned into three families: Myoviridae, Podoviridae and Siphoviridae, and the Myoviridae proteins were the most abundant. Among them, structure proteins were the most abundant viral proteins. This study indicated that the dissolved proteins of HMW-DOM presented compositional and biologically original homogeneity in the surface seawaters of the South China Sea, and bacteria and viruses dominated the dissolved protein pool.
© 2013 Elsevier Ltd. All rights reserved.

1. INTRODUCTION

Oceanic dissolved organic matter (DOM) is one of the largest dynamic reservoirs of organic carbon on earth (Ducklow, 2002), and plays important roles in chemical, biological and even physical oceanography. It is estimated that high-molecular-weight DOM (size between 1 kDa and <0.2 μm, HMW-DOM) accounts for 20–35% of the total DOM in the ocean (Guo et al., 1995), and plays a more active role than low-molecular-weight DOM (size less than 1 kDa, LMW-DOM) in the biogeochemical cycling of organic carbon and many biogenic elements in marine systems (Guo et al., 1997). A considerable fraction of the HMW-DOM released by phytoplankton is taken up almost exclusively by bacteria and archaea (Azam and Malfatti, 2007). The HMW-DOM is relatively reactive while the LMW-DOM is the major form of recalcitrant DOM (RDOM) in the ocean (Ogawa and Tanoue, 2003). Recently, a new conceptual framework, the microbial carbon

* Corresponding author. Tel.: +86 592 2186016; fax: +86 592 2180655.
E-mail addresses: wangdz66@yahoo.com, dzwang@xmu.edu.cn (D.-Z. Wang).

0016-7037/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.gca.2013.01.041

pump (MCP) is proposed, which emphasizes the microbial transformation of DOM from labile to recalcitrant states (Jiao et al., 2010; Hansell, 2012), suggesting that the MCP may play important roles in transforming HMW-DOM into LMW-DOM. Therefore, knowledge of HMW-DOM will help to elucidate the production, transformation and fate of DOM in the ocean.

In the past few decades, our knowledge of the chemical nature of HMW-DOM has increased rapidly. Aluwihare et al. (1997) reported that carbohydrates are the major constituents of HMW-DOM. Another study indicated that approximately 70–90% of organic nitrogen in HMW-DOM is in the form of amide (McCarthy et al., 1997). In addition, a series of compounds such as peptidoglycan, muramic acid, N-acetyl amino sugars, fatty acid components, carboxyl-rich alicyclic molecules (CRAM) and proteins have been successively identified from HMW-DOM, and some of them, such as CRAM and proteins, are widely distributed throughout the water column (Benner and Kaiser, 2003; Zou et al., 2004; Hertkorn et al., 2006; Wang et al., 2011), which provides valuable information concerning the composition and chemical nature of HMW-DOM.

Proteins are the most important class of biochemical molecules for life, and they are the actual machinery that brings about cell growth, proliferation and homeostasis. It is logical, therefore, that the study of proteins should provide new insights into the source and biogeochemical processes of DOM in the ocean. Tanoue (1995) first reported the presence of dissolved protein in seawater using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). Subsequently, using two-dimensional electrophoresis (2-DE), Yamada and Tanoue (2006) distinguished more than 412 protein spots on 2-DE electrophoretograms of oceanic waters. In an investigation of dissolved proteins in seawater samples collected from a coastal area using SDS–PAGE and 2-DE, Yamada and Tanoue (2009) noted that the electrophoretic patterns are similar between the coastal and pelagic samples, suggesting compositional homogeneity of the dissolved proteins in seawater throughout a broad range of marine environments. More recently, using the shotgun proteomic approach, a total of 182 proteins were identified in three DOM samples (size between 5 kDa and <0.7 μm) collected from the surface and bathypelagic water layers of the South China Sea and, among them, 27 proteins were identified from HMW-DOM (Wang et al., 2011). These studies provide new insights into the source, composition and protection of DOM. However, protein information from HMW-DOM, which is the most active fraction of DOM, is still seriously lacking. So far, very few proteins have been identified from HMW-DOM, which hinds our comprehensive, in-depth understanding of marine DOM including its source, production, remineralization and fate.

Recently, Rusch et al. (2007) reported a metagenomic study of the marine planktonic microbiota in surface (mostly marine) water samples as part of the *Sorcerer II* Global Ocean Sampling (GOS) expedition. This study produced enormous metagenomic data sets, which has not only improved greatly our understanding of the diversity and genetic functions of many microorganisms, but made possible the identification of proteins in the marine system. In a metaproteomic study of marine microorganisms, *in situ* expression of a light-dependent proton pump (proteorhodopsin) from the SAR11 clade was detected off the Oregon coast (Giovannoni et al., 2005). Comparative analyses of microbial membrane proteome revealed a variation in proteins involved in nutrient utilization along an environmental gradient in South Atlantic surface waters (Morris et al., 2010). More recently, using metaproteomic analysis, Sowell et al. (2011) found that transporters for amino acids, taurine, polyamines and glutamine synthetase are among the most highly detected proteins in Oregon coast summer surface waters. These studies mainly focus on the metaproteomics of surface microbial communities in the ocean. However, no specific effort has been devoted to the metaproteomics of surface HMW-DOM in the ocean.

In this study, we combined ‘shotgun’ mass spectrometry (MS)-based proteomics with a GOS combined assembly protein database search to characterize dissolved proteins in marine surface HMW-DOM collected from the coastal and pelagic areas of the South China Sea. To our knowledge, this study is the first attempt to characterize dissolved proteins of HMW-DOM at a metaproteomic level, and to compare the features of dissolved proteins from coastal and pelagic surface seawater. Our goal was to identify dissolved proteins in surface HMW-DOM, trace their sources, and reveal the dominant metabolic processes occurring in summer surface waters in the South China Sea. We found that bacteria and viruses were the two major contributors to the dissolved proteins of the surface HMW-DOM, and the protein composition and biological origin presented close similarity among the four sampling sites.

2. MATERIALS AND METHODS

2.1. Sampling and preconcentration of surface dissolved proteins

Samples from the 10 m water depth were collected at four sites during a summer cruise in the northern South China Sea from 17 July 2009 to 17 August 2009 (Fig. 1). Detailed information concerning the stations and samples

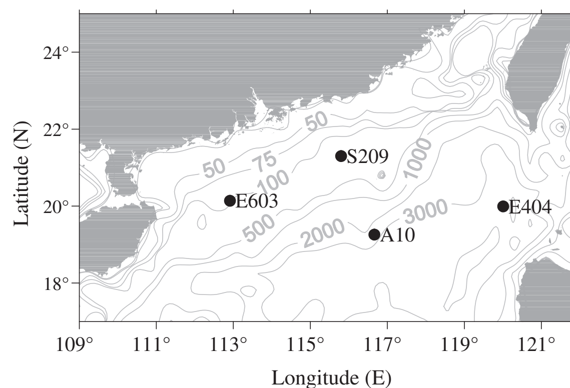

Fig. 1. Map of the South China Sea showing E603, A10, E404 and S209, the sampling locations for HMW-DOM.

Table 1
Basic data of sampling stations and seawater samples in this study (nd is below detectable limit).

Stations	Latitude (N)	Longitude (E)	Distance to the nearest coast (km)	Water depth(m)	Sampling date	Temperature (°C) at 10 m	Salinity at 10 m
E603	20°06.9′	112°54.6′	199	96	July 25, 2009	29.1	33.6
A10	19°15.5′	116°39.7′	417.7	2823	Aug. 25, 2009	29.8	33.4
E404	19°51.0′	120°01.2′	166.8	3618	Aug. 14, 2009	29.3	33.3
S209	21°18.0′	115°47.9′	148	135	July 30, 2009	29.4	33.6

Stations	$NO_3^- + NO_2^-$ (μM) at 5 m	$NO_3^- + NO_2^-$ (μM) at 25 m	PO_4^{3-} (μM) at 5 m	PO_4^{3-} (μM) at 25 m	Chl *a* (mg L^{-3}) at 5 m	Chl *a* (mg L^{-3}) at 25 m	Chl *a* (mg L^{-3}) at 50 m	Chl *a* (mg L^{-3}) at 75 m
E603	nd	nd	nd	nd	0.09	0.12	0.07	0.55
A10	0.14	0.14	0.10	0.10	0.08	0.07	0.25	0.25
E404	nd	nd	0.09	0.09	0.51	0.49	0.80	1.02
S209	0.11	0.10	nd	nd	0.08	0.07	0.60	0.49

is presented in Table 1. The procedures for preconcentration of dissolved proteins were modified following the method of Tanoue (1995). Seawater samples were collected with a rosette sampler fitted with 12-L Niskin Go-flo bottles (General Oceanics) and the volume of each sample was 40 L. Each sample was filtered through a GF/F glass fiber filter (Whatman, with a nominal pore size of 0.7 μm) immediately after sampling. A final concentration of 0.01% (w:v) SDS was added to the filtrates to help solubilize the proteins. The filtrates were further filtered through a 0.2 μm filter (Millipore, Durapore membrane filter) and sodium azide (NaN_3, 5 mM final concentration) was added to the filtrates to prevent bacterial growth.

All the filtrates (<0.2 μm) were preconcentrated using the Pelicon 2 cross-flow ultrafiltration system (Millipore; 1 m^2 of filter area) with a filter having a nominal molecular weight cut-off of 1 kDa (1000 NMMCO filter, low protein-binding regenerated cellulose) onboard ship at room temperature. Retention characteristics of the 1 kDa cross-flow ultrafiltration membrane in the system were evaluated in a previous study (Wang et al., 2011). The system was pre-cleaned with 0.1 M NaOH and washed with deionized water following the manufacturer's instructions. Before use of each new cross-flow ultrafiltration membrane and tubing, a 3.5% solution of NaCl (20 L) was concentrated in the same manner as the seawater sample to serve as a control. The preconcentrated samples (approximately 600–700 mL) were kept frozen (−80 °C) until further concentration using a 5 kDa ultrafiltration cell (Millipore).

2.2. Preparation of the HMW-DOM dissolved proteins

Procedures for further concentration and desalting of the preconcentrated samples from the cross-flow ultrafiltration were previously reported (Wang et al., 2011). Briefly, the preconcentrated samples were further concentrated and desalted using a stirred ultrafiltration cell with a filter having a nominal molecular mass cut-off of 5 kDa (filter area, 13.4 cm^2; low protein-binding regenerated cellulose, Millipore). When the sample was concentrated to approximately 25 mL, the same volume of the desalting buffer was added, and the mixture was concentrated continuously. This step was repeated three times to desalt the retentate. The retentate was then precipitated with ice-cold 20% trichloroacetic acid in acetone. The mixture was centrifuged at 20,000*g* for 30 min at 4 °C, and then the pellets were rinsed twice with ice-cold acetone, and air-dried. Finally the protein pellets were solubilized in rehydration buffer, and stored in a freezer (−80 °C) until further SDS–PAGE analysis.

2.3. One dimensional SDS–PAGE and enzymatic digestion of dissolved proteins

A rehydrated dissolved protein sample (40–60 μL), equivalent to ca. 40 L of the original seawater, was applied to each lane of a 5–12% Bis-Tris gel (13 × 13 cm). Electrophoresis was performed at a constant voltage of 180 V in the electrode buffer solution (25 mM Tris, 192 mM glycine and 0.1% SDS) on a Hoefer SE 600 apparatus (Amersham).

Low molecular weight standards were used for reference. After electrophoresis, the proteins on the gel were visualized with colloidal Coomassie Brilliant Blue G-250 (CBB, Bio-Rad). The gel was cut into four equal pieces to isolate molecular weight fractions, followed by reduction, alkylation, and in-gel digestion with trypsin as described by Wilm et al. (1996).

2.4. Mass spectrometry analysis

Mass spectrometry analysis followed the method described previously (Wang et al., 2011). Briefly, the nano-LC MS^2 experiment was performed on an HPLC system connected to an LTQ-Orbitrap mass spectrometer (Thermo Fisher). The peptides extracted from the gel bands were resuspended with 5% acetonitrile in 0.1% formic acid and loaded onto a Captrap column (0.5 × 2 mm, Michrom Bioresources) for 5 min at a flow rate of 20 μL min^{-1}. The sample was subsequently separated using a C18 reverse-phase column (0.075 × 150 mm, packed with 3 μm Magic C18-AQ particles, Michrom Bioresources) at a flow rate of 500 nL min^{-1}. The mobile phases were 5% acetonitrile with 0.1% formic acid (phase A and the loading phase) and 95% acetonitrile with 0.1% formic acid (phase B). To achieve proper separation, a 60-min linear gradient from 5% to 45% phase B was employed. The separated sample was introduced into the mass spectrometer via an Advance 30 μm silica tip (Michrom Bioresources). The spray voltage was set at 1.8 kV and the heated capillary was at 180 °C. The mass spectrometer was operated in data-dependent mode and each cycle of duty consisted of one full-MS survey scan at the mass range 400–2000 Da with a resolution power of 60,000 using the Orbitrap section, followed by MS/MS experiments for the 10 strongest peaks using the LTQ section. The AGC expectation during full-MS and MS/MS were 500,000 and 10,000, respectively. Peptides were fragmented in the LTQ section using collision-induced dissociation with helium and the normalized collision energy value was set at 35%. Previously fragmented peptides were excluded for 60 s.

2.5. Protein identification

Tandem mass spectra were extracted using BioWorks version 3.3.1 SP1 (Thermo Fisher). All MS^2 spectra were searched against the GOS Combined Assembly protein (P) database (Rusch et al., 2007; Yooseph et al., 2007) using the SEQUEST algorithm (Thermo Fisher, version 28) available through the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) (https://portal.camera.calit2.net/) downloaded on 16 May, 2010. The database contained 6,120,000 proteins predicted from microbial environmental genomic data through clustering and hidden Markov model approaches. Data searches were completed, with the following parameters: enzyme type trypsin; Parent Ion Tolerance, 50 ppm; Fragment Ion Mass Tolerance, 1.00 Da. Modifications of cysteine residues by 57.021 Da (resulting from the iodoacetamide modification) and methionine by 15.999 Da (oxidation) were allowed. The Trans-Proteomic Pipeline (version 4.1.1, Institute for Systems Biology) was used to validate MS^2 based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller et al., 2002). Protein identifications were accepted if they could be established at greater than 99.0% probability. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides but could not be differentiated based on MS^2 analysis alone were grouped to satisfy the principles of parsimony. To evaluate false positive rates, all MS^2 spectra were also searched against a decoy database containing corresponding reversed protein sequences of the GOS database. The false positive rate (FPR) for peptide identification was computed using the formula (Verberkmoes et al., 2009): FPR = (2 × reverse)/(forward + reverse). The FPR for the total identified peptides from the four samples was 1.9%. Protein groups matched by only one unique peptide were excluded. The FPR of protein groups identified with two or more peptides was 0.5%.

2.6. Bioinformatics analysis

Relative protein abundance was assessed at the peptide level using a semi-quantitative method: peptide spectral counting (Ryu et al., 2008). Peptide spectral counts were determined based on the number of times a peptide that correlated with a given protein was selected for collision induced dissociation, including all repeated selections of the same peptide.

Protein annotations and taxonomic assignments were performed as described previously (Morris et al., 2010). Briefly, all GOS proteins identified using interrogating peptide tandem mass spectra were searched against CAMERA's Non-Identical Peptide Sequences database using the Basic Local Alignment Search Tool (BLAST) available through CAMERA. BLAST expect scores $>10^{-4}$ were annotated as 'unknown'. The best BLAST hit for all GOS protein-coding sequences (GCDS) identified by the same peptide or peptides was first evaluated to determine the level of consensus among all taxonomic identifications. Taxonomic identities and protein function assignments were made at the most specific level that applied to all of the GOS protein annotations. For example, if all the best BLAST hits of GOS proteins identified by the same peptide or peptides were from different bacterial lineages, then the taxonomic assignment was Bacteria. However, if all the best BLAST hits of GOS proteins identified by the same peptide or peptides were from alphaproteobacteria, then the taxonomic assignment was alphaproteobacteria. The process was repeated for all proteins. An analogous approach was used to determine the level of consensus among functional annotations. Owing to the complexity of viral taxonomy, viral proteins were assigned to family uniformly. The gene ontology (GO) annotation of other proteins, except for viral proteins, was performed using the GO annotation database of Non-Identical Peptide Sequences database from CAMERA.

3. RESULTS AND DISCUSSION

3.1. Separation and identification of dissolved proteins in surface HMW-DOM

An electrophotogram of the dissolved proteins from the four HMW-DOM samples separated using one-dimensional SDS–PAGE is shown in Fig. 2. The lanes of all samples visualized using CBB stain showed a smeared staining pattern throughout the molecular mass range, although a few faint bands were recognized on the background staining. The staining intensity of the dissolved proteins in HMW-DOM samples from sites E404 and S209 was higher than that from sites E603 and A10 (Fig. 2), showing that there were different amounts of dissolved proteins in the 40 L seawater from the different sampling sites. Notably, the staining intensity of dissolved proteins on gel was inconsistent with chlorophyll *a* (Chl *a*) and nutrient concentration in the upper 25 m (Table 1), suggesting that the dynamics of dissolved proteins might be different from those of nutrients as well as Chl *a*.

For each HMW-DOM sample, four equal slices were cut from the SDS–PAGE gel and each slice was analyzed using an LTQ-Orbitrap mass spectrometer. The 16 runs produced 46517 MS/MS spectra from the four HMW-DOM samples after removing contaminant spectra (keratins, trypsin) by searching against human database plus trypsin peptide sequences. Using the GOS combined assembly protein database, 7–11% of total MS/MS spectra of each sample were assigned to peptide sequences in this database, whereas 89–93% of MS/MS spectra could not match peptide sequences in the GOS database with high confidence. This result was consistent with that of a study by Powell et al. (2005). Several reasons might account for the low MS/MS assignments: (1) the sources of dissolved proteins in seawater was very diverse and complicated, and their peptide sequences might be not present in the GOS database; (2) the abundance of many dissolved proteins were very low and so it was difficult to generate high-quality spectra for good matches; and lastly, (3) dissolved proteins were extensively modified such as *in situ* abiotic glycosylation or biotical glycosylation in the cells (Yamada and Tanoue, 2003; Powell et al., 2005)(), which might have hindered matches of spectra with protein sequences in the GOS database. When at least two peptide matches per protein group were regarded as a positive identification, a total of 367 protein groups matched by 993 unique peptides from 1991 spectra (spectra assigned to proteins identified with two or more peptides) were identified from the four HMW-DOM samples.

For the GCDSs which were included in each protein group, the consensus function and taxonomic identities were obtained by comparison with CAMERA's Non-Identical Peptide Sequences database using BLAST. Some protein groups contained multiple GCDSs which might have been identical at the conserved domains of sequences or due to subtle sequence variation. In a previous study, it is reported that of the 1146 peptides that map to SAR11 CDSs, 14% also map to CDSs which has a different organism as its closest relative (Sowell et al., 2011). Given that fewer peptides are shared between more distant relatives (Sowell et al., 2011), taxonomic assignment is thus more accurate at higher taxonomic level. It is also noted that the maximum number of spectra matching each protein group in our study was less than one hundred, while for protein identification from a microbial community (Sowell et al., 2011), this value can reach more than seven hundred. This might be caused by the lack of high-abundance proteins as well as the relatively homogeneous distribution of many proteins in low abundance.

The accession number in the GOS database, protein probability, sequence coverage (%), number of unique peptides, total spectra used in the identification, protein name, and taxonomic assignment are listed in the Appendix.

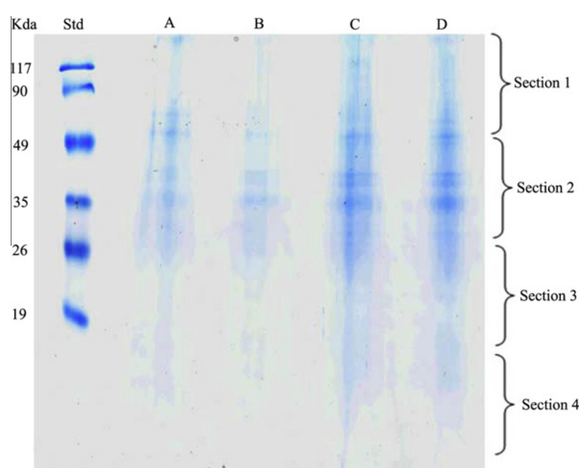

Fig. 2. One-dimensional SDS–PAGE electrophoretogram of the dissolved proteins in surface HMW-DOM of the South China Sea. The left lane is a molecular weight standard. Lanes A, B, C and D represent dissolved proteins from Stations E603, A10, E404 and S209, respectively. The gel was stained with colloidal Coomassie blue and cut into four pieces, i.e., Sections 1, 2, 3 and 4.

3.2. Sources of dissolved proteins in surface HMW-DOM

Each detected GCDS in the protein group was queried against CAMERA's Non-Identical Peptide Sequences database. Protein group was assigned to phylogenetic classification based on consensus of GCDS taxonomic identities. The percentage of protein groups and spectra matching bacteria, eukaryota, viruses and unknown taxonomy is shown in Table 2. For the four HMW-DOM samples, 43 to 53% of protein groups could not be assigned a taxonomic annotation because of poor BLAST expect scores. Except for the protein number, spectra from 'unknown' proteins also presented the highest percentage in the four samples (Table 2). These results suggested that sequences of many proteins in the HMW-DOM from surface seawaters lacked similarity to currently known protein families because of the lack of knowledge concerning the source organisms. In addition, for the GOS database predicted from shotgun sequencing of microbial populations, it is hard to assign some protein sequences to organisms.

Bacterial protein groups accounted for 22.6% to 32.6% of the total groups of each surface HMW-DOM sample,

Table 2
The number of spectra and protein groups that matched phylogenetic groups contributing to dissolved proteins in the four surface HMW-DOM samples in the South China Sea. The number in parenthesis refers to the number or percentage of protein groups.

Species groups	E603	A10	E404	S209
Number of spectra and protein groups at each station				
Alphaproteobacteria	41 (8)	42 (8)	67 (12)	78 (18)
Unknown	163 (38)	225 (49)	257 (47)	206 (41)
Cyanobacteria	41 (4)	37 (3)	35 (7)	16 (3)
Virus	109 (17)	141 (23)	193 (28)	79 (18)
Actinobacteria	7 (2)	3 (1)	14 (2)	3 (1)
Gammaproteobacteria	8 (2)	15 (3)	17 (4)	3 (1)
Firmicutes	0	8 (1)	0	5 (1)
Chloroflexi	23 (2)	16 (2)	33 (3)	19 (1)
Dinophyceae	11 (1)	0	0	0
Betaproteobacteria	9 (1)	3 (1)	3 (1)	17 (3)
Verrucomicrobia	5 (1)	7 (2)	16 (3)	4 (1)
Deltaproteobacteria	0	0	8 (2)	0
Prasinophyceae	0	0	0	4 (1)
The percentage of spectra and protein groups (%)				
Bacteria	32.1 (26.3)	26.4 (22.6)	30.0 (31.2)	35.5 (32.6)
Virus	26.1 (22.4)	28.4 (24.7)	30.0 (25.7)	18.2 (20.2)
Eukaryota	2.6 (1.3)	0 (0)	0	0.9 (0.1)
Unknown	39.1 (50)	45.3 (52.7)	40.0 (43.1)	47.5 (46.1)

with proteins from the Alphaproteobacteria, Cyanobacteria and Chloroflexi being abundant in each sample (Table 2). Viral proteins accounted for 20.2% to 25.7% of the total proteins in each sample while spectra matching viral proteins accounted for 18.2% to 30% of the total spectra in each sample. These results indicated that a considerable number of proteins in the HMW-DOM from surface seawater were derived from bacteria and viruses, and very few proteins were from the eukaryotic phytoplankton. Although the GOS database consists mainly of prokaryotic sequences (90.8%), 2.8% of total sequences in the GOS database are from the eukaryotic kingdom (Yooseph et al., 2007). In addition, we note in our previous study on the HMW-DOM in the South China Sea (Wang et al., 2011), that very few of the dissolved proteins identified are from eukaryotic phytoplankton. These studies imply that eukaryotic phytoplankton-origin proteins are labile and may not accumulate in the HMW-DOM to a detectable level. On the other hand, in the GOS database, 3.7% of total sequences are from viruses (Yooseph et al., 2007), and viral proteins form a significant fraction of the marine surface microbial plankton proteome (Morris et al., 2010; Sowell et al., 2011). In our study, the detection of abundant viral proteins in the HMW-DOM suggested that the HMW-DOM might play a key role in the microbial loop, and supported the conclusion that viruses can affect the efficiency of the biological pump through causing the lysis of bacterial cells, converting them into particulate and dissolved organic carbon (Suttle, 2005). It is interesting that proteins from the archaea were not detected in the four HMW-DOM samples in our present study, nor in the surface DOM in our previous work (Wang et al., 2011), although 2.7% of the sequences in the GOS database were from the archaea (Yooseph et al., 2007). These results further demonstrated that the activity of the archaea mainly occurs in the meso- and bathypelagic zones (Karner et al., 2001). Although archaea are also distributed in the surface waters of the ocean, their proteins might not accumulate in seawater to a detectable level.

Interestingly, for each sample, 37–88% of spectra of the Alphaproteobacteria were hits to the Rhodospirillaceae, suggesting the important contribution of this group to the HMW-DOM. Members of the Rhodospirillaceae mainly consist of non-sulfur purple photosynthetic bacteria with nitrogen-fixing ability being found in almost all species (Ludden and Robert, 1995). During the summer in the South China Sea, the Rhodospirillaceae forms the second most abundant cluster within the Alphaproteobacteria cluster in surface waters (Zhang et al., 2011), and the highest potential for nitrogen fixation is also reported (Loick et al., 2007). The prevalence of spectra for the Rhodospirillaceae in our study suggested that active nitrogen fixation might occur during the summer in the South China Sea. Moreover, very low concentrations of dissolved inorganic nitrogen (DIN, nitrate + nitrite) were observed in the upper 25 m (Table 1), which might provide an ideal environment for the growth of nitrogen-fixing bacteria.

The SAR11 clade is a group of Alphaproteobacteria that is ubiquitous in both open-ocean and coastal waters, which is thought to have a significant function in nutrient cycling in the oceans. Analysis of 16S rDNA clones showed that the SAR11 cluster dominates within the Alphaproteobacteria and is ubiquitous in the South China Sea (Zhang et al., 2011). In our study, 12–39% of spectra of Alphaproteobacteria were attributed to the SAR11 clade, showing that the SAR11 group was one of major contributors to dissolved proteins in the HMW-DOM. Meanwhile, it is also noteworthy that the spectra from *Prochlorococcus* accounted for 22–51% of cyanobacterial spectra, suggesting that proteins from *Prochlorococcus* were the main component of dis-

solved proteins in the HMW-DOM. *Prochlorococcus* is the dominant species among the photosynthetic populations in the South China Sea, and its abundance is generally two orders of magnitude higher than that of picoeukaryotes (Jiao and Yang, 2002). Taken together, these results showed that the source of dissolved proteins retained in the HMW-DOM was indicative of the dominant bacteria in the microbial community, and that specific bacterial groups regulated the composition of dissolved proteins in the surface HMW-DOM of the oceanic region.

All the viral proteins identified were assigned into three viral families: the Myoviridae, Podoviridae and Siphoviridae (Fig. 3), suggesting that these three families might be more abundant than others in the oceanic region. Moreover, all of these viruses were dsDNA tailed bacteriophages that could infect bacteria. This result was in agreement with the observation that the predominant GOS viral sequences identified within microbial fractions originate from tailed bacteriophages (Williamson et al., 2008). It should be pointed out that GOS database is produced from metagenomic studies of the marine microbial communities which focus on analysis of DNA sequences. In our study, thus, RNA viruses were not detected in that no RNA viral sequences are present in the GOS database.

It is noteworthy that the abundance of proteins matching the Myoviridae was significantly higher than that of proteins from the other two families. About 82–89% of total viral GCDSs at each site were matched to Myoviridae proteins while spectra matching Myoviridae proteins accounted for about 71–87% of the total spectra in each sample (Fig. 3). This result was consistent with previous laboratory observations that myoviruses are often the most commonly isolated phages from natural marine viral communities (Suttle, 2005). Myoviruses are typically lytic and often have a broader host range than other tailed phages, even infecting different species of bacteria (Sullivan et al., 2003). By contrast, podoviruses are also typically lytic, but have very narrow host ranges. Siphoviruses are lysogenic, and need to be triggered into the lytic cycle although they also have a relatively broad range (Suttle, 2005). Therefore, based on difference of lifestyle of the three viral

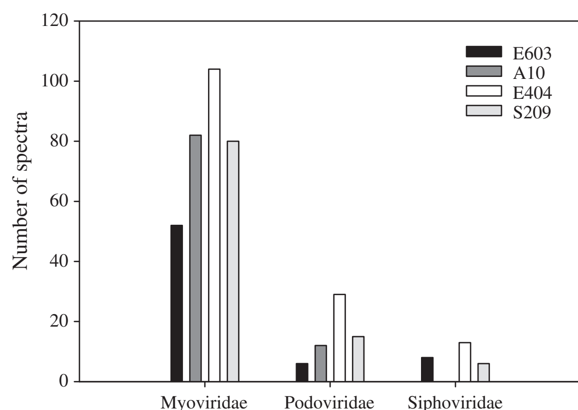

Fig. 3. The distribution of spectra that matched three viral families contributing to viral proteins from the four HMW-DOM samples in the South China Sea.

families, it can be postulated that myoviruses may possess short generation times and high reproductive rates, and so can quickly become the dominant phages in virus communities in the ocean (Suttle, 2005). In our study, the high abundance of proteins matching Myoviridae in the surface HMW-DOM probably verified the above hypothesis concerning the selective pressures facing viral communities which were obtained during early studies of viral morphotypes using transmission electron microscopy.

It should be noted that it is difficult to distinguish living or dead dissolved proteins from the mixture of the HMW-DOM using the present methodology. Our previous study suggests that dissolved proteins in the DOM may be derived from detrital proteins originating from bacteria and phytoplankton rather than from the living organism retained on a 0.2 μm filter (Wang et al., 2011). In this present study, seawater was pre-filtered using 0.2 μm filter before the HMW-DOM sample was concentrated. Thus, other dissolved proteins, except for viral proteins, in the HMW-DOM might result from cellular lysis or the extracellular releases of bacteria and phytoplankton which accumulate in the HMW-DOM as detrital proteins. Detection of abundant viral proteins expanded the DOM pool, suggesting that viral particles were present in the HMW-DOM. However, whether they are viral proteins or other dissolved proteins in the HMW-DOM, their turnover time in seawater depends on their structure and ambient environment, as discussed previously (Wang et al., 2011).

3.3. Features of dissolved proteins in surface HMW-DOM

Proteins with unknown functions were predominant in the surface HMW-DOM. About 53–65% of protein groups of each sample matched hypothetical or conserved hypothetical proteins, accounting for 48–62% of the total spectra in each sample. In a study of 12 Roseobacter exoproteomes (Christie-Oleza et al., 2012), over 40% of CDS predicted as secreted proteins were annotated as “hypothetical proteins”. These data suggested that knowledge concerning proteomes or exoproteomes of marine microbes is still scarce.

Transporters for organic compounds, such as amino acids and sugars, were frequently detected in the surface HMW-DOM. TRAP dicarboxylate transporter- DctP matching the SAR11 clade, which was a tripartite ATP-independent periplasmic dicarboxylate transporter for mannitol and/or chloroaromatic compounds, was found in three sites (Table 3). This transporter is a secondary carrier using the electrochemical H^+ gradient as the driving force (Janausch et al., 2002). The SAR11 transporters involved in the uptake of spermidine/putrescine, taurine and glycine betaine were also detected (Table 3). Similar findings have been reported from the proteomic analysis that focuses on the SAR11 metaproteome in the oligotrophic Sargasso Sea (Sowell et al., 2008) and the microbial planktonic metaproteome in Oregon coastal surface waters (Sowell et al., 2011). In our study, the dissolved proteins in the surface HMW-DOM might have been mainly derived from highly expressed proteins in dominant bacteria or viruses. In combination with coastal shelf and pelagic distri-

Table 3
Important proteins detected in surface HMW-DOM samples in the South China Sea.

Protein family	Station distribution	No. of peptides	No. of spectra	No. of GCDSs	Potential organisms
TRAP dicarboxylate transporter – DctP subunit (mannitol/chloroaromatic compounds)	E603	2	5	665	
	A10	3	6	24	SAR11 cluster
	E404	4	18	24	
Glutamate/glutamine/aspartate/asparagine ABC transporter, periplasmic substrate-binding protein	E603	2	4	1	Rhodobacteraceae
	E404	4	8	3	
	S209	2	2	1	
Spermidine/putrescine-binding periplasmic protein	E404	3	5	1	SAR11 cluster
	S209	2	3	1	
Taurine transport system periplasmic protein	S209	2	3	457	SAR11 cluster
Substrate-binding region of ABC-type glycine betaine transport system	S209	2	3	128	SAR11 cluster
ABC transporter	A10	3	3	3	SAR11 cluster
	S209	5	11	1	
Ammonium transporter family	S209	2	3	68	*Prochlorococcus*
Urea ABC transporter	E603	2	2	31	*Prochlorococcus*
Proteorhodopsin	E404	2	3	332	SAR11 cluster
	E209	2	3	332	
Bacterioferritin	E603	2	3	1	Pseudomonadaceae
	E404	2	3	1	

bution of the four DOM samples, our data suggested that adaptation for effective nutrient scavenging was vital for the microbial community in the surface waters of the oligotrophic South China Sea. In addition, it is noteworthy that the majority of transporters detected were involved in the transport of nitrogen-containing compounds, such as urea, ammonium and amino acids (Table 3) and their spectra accounted for 1.5% of the total spectra, but no phosphate transporters were detected, suggesting that nitrogen might be more limiting than phosphorus in surface seawaters. This finding was consistent with studies from the low-nutrient seawater of the South Atlantic open ocean (Morris et al., 2010) and from highly productive Oregon coastal seawaters (Sowell et al., 2011). In the oligotrophic South China Sea, phosphorus is relatively abundant and does not restrain phytoplankton growth, while nitrogen is limited and thus renders the South China Sea a potential environment for ideal growth of diazotrophic phytoplankton (Chen, 2005). In our study, integrated DIN concentration in the upper 25 m waters during the sampling period was 0.11 μM at site A10 and 0.08 μM at site S209, and below the detectable limit at sites E603 and E404 (Table 1). At sites A10, the ratio of DIN: DIP (dissolved inorganic phosphorus) (1.4:1) in the upper 25 m waters were much lower than the Redfield ratio of 16:1, implying nitrogen deficiency for plankton growth. Our data supported the conclusion drawn from metaproteomic analyses of the microbial community that metaproteomic analysis can help reveal the nutrient status of cells in marine environments.

Proteorhodopsin from the SAR11 cluster was found in the surface HMW-DOM from sites E404 and S209. Proteorhodopsins are light-dependent proton pumps that are predicted to have an important role in the ecology of the oceans by supplying energy for microbial metabolism (Beja et al., 2000). Proteorhodopsin genes have been discovered from seawater through the cloning and sequencing of large genomic DNA fragments (Beja et al., 2000). Giovannoni et al. (2005) reported that SAR11 strain HTCC1062 (*Pelagibacter ubique*) expresses its proteorhodopsin gene in either diurnal light or in darkness when exposed to the natural environment. Since bacteria belonging to the SAR11 clade can contribute up to 35% of the bacterial abundance (Morris et al., 2002), the photo-organoheterotrophic bacteria with proteorhodopsins may dominate the world's oceans. In our study, detection of proteorhodopsins from the SAR11 cluster in the surface HMW-DOM confirmed the presence and expression of a proteorhodopsin gene in the South China Sea.

Bacterioferritins originating from *Pseudomonas* constitute a broad superfamily of iron-storage proteins, and were detected in the surface HMW-DOM from sites E603 and E404. These proteins form a roughly spherical, hollow shell from 24 identical subunits, incorporating 12 haem groups, and iron is stored as a hydrated ferric oxide mineral in its central cavity (Carrondo, 2003). The basic role performed by bacterioferritin is to provide sufficient iron to cells. However, when ambient iron concentration is high, bacterioferritin also has a protective 'anti-oxidat-

ing' function by sequestering the iron inside the cavity (Carrondo, 2003). In addition, siderophore–iron complexes and TonB-dependent receptors have been detected in POM from the water column (Dong et al., 2010). To date, a series of proteins involved in iron uptake, transport and storage have been found in marine organic matter, implying that iron might play an important role in the marine plankton in this area.

Interestingly, most of the viral proteins were derived from viral outer structural proteins, including capsid proteins, head proteins, tail sheath proteins, baseplate wedge, and tail tube proteins. Spectra from structural proteins presented the highest percentage in each sample (Fig. 4). Cluster analyses of the viral sequences revealed hundreds to thousands of viral genes encoding various metabolic and cellular functions (Williamson et al., 2008). However, in our study, only a few terminases were detected. Three reasons might result in the detection of a high percentage of viral structural proteins in our study: (1) the gene sequences of structural proteins account for 40.2% of total viral sequences in the GOS database (Williamson et al., 2008), which probably provides more opportunities for the detection of viral structural proteins; (2) the sampling method we used might lead to enrichment of viral structural proteins, while some metabolic enzymes might be lost during ultrafiltration concentration; and (3) different components from viral particles might possess a different capability of resistance to degradation. No studies on degradation of different components from viral particles have been reported to date, although decay of viral infectivity has been studied with both native and nonnative bacteriophages (Noble and Fuhrman, 1997). Nevertheless, viral decay cannot provide information concerning the rate of disappearance of viral particles. Studies of viral degradation will allow us to elucidate the controlling factors of the dynamics of viruses within the marine microbial food web.

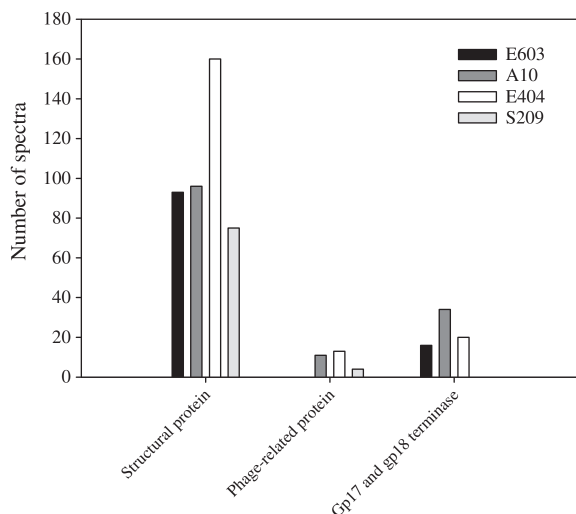

Fig. 4. Types of viral proteins from the four HMW-DOM samples in the South China Sea. *Note:* structural protein includes capsid, head and coat protein.

4. CONCLUSIONS

In this study, hundreds of proteins were first identified from the four HMW-DOM samples using the shotgun proteomic approach combined with the GOS combined assembly protein database. Proteins with unknown taxonomic classification and function dominated the protein pool while other proteins presented close similarity in biological origin among the four sampling sites. The Rhodospirillaceae, SAR11 clade, *Prochlorococcus* and viruses were the major contributors to dissolved proteins in the HMW-DOM from the surface seawaters of the South China Sea, indicating that the protein composition of the HMW-DOM was regulated by specific dominant microbial groups at each sampling site. Detection of transporters with substrate specificities for nitrogen- and carbon-containing compounds suggested that carbon and nitrogen might be more limiting than phosphate in surface seawater. The majority of viral groups in the surface HMW-DOM were matched to the Myoviridae while spectra for viral structural proteins were prevalent.

It must be pointed out that a large number of identified proteins with unknown functions need to be unraveled, and this will rely on further fundamental studies of the protein functions of single bacterial isolates. In addition, partial spectra of many proteins could not be matched to any protein sequences owing to the limitations of the present marine genomic and proteomic databases. As these problems are solved, step by step we will gain more information concerning the protein composition of the HMW-DOM, which will improve our understanding of the sources, dynamics and remineralization mechanisms of DOM, and provide insight into the nutrient status of the microbial community in the ocean. Furthermore, the study of the abundance and distribution of dissolved proteins using quantitative methods will provide new insights into their roles in biogeochemical cycling in the ocean. Given that viruses play important roles in the marine ecosystem, a more specialized viral database is needed for studying the distribution and abundance of some particular viruses in different oceanic regions. Given these advances, it is likely that we will gain more information concerning the roles of viruses in the biogeochemistry of marine DOM.

ACKNOWLEDGEMENTS

We thank the captain and crew of the R/V *DongFangHong 2* for their assistance during the sampling cruise. This study was supported by the Ministry of Science and Technology through grant 2009CB421203, and the National Natural Science Foundation of China through grants 40821063, 40776068, and 40876059. D.-Z. Wang was also supported by the Program for New Century Excellent Talents in University of China. Professor John Hodgkiss is thanked for polishing the English in this paper.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gca.2013.01.041.

REFERENCES

Aluwihare L. I., Repata D. J. and Chen R. C. (1997) A major biopolymeric component to dissolved organic carbon in surface seawater. *Nature* **387**, 166–169.

Azam F. and Malfatti F. (2007) Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* **5**, 782–791.

Beja O., Aravind L., Koonin E. V., Suzuki M. T., Hadd A. and Nguyen L. P., et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906.

Benner R. and Kaiser K. (2003) Abundance of amino sugars and peptidoglycan in marine particles and dissolved organic matter. *Limnol. Oceanogr.* **48**, 118–128.

Carrondo M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J.* **22**, 1959–1968.

Chen Y. L. (2005) Spatial and seasonal variations of nitrate-based new production and primary production in the South China Sea. *Deep-Sea Res. I* **52**, 319–340.

Christie-Oleza J. A., Pina-Villalonga J. M., Bosch R., Nogales B. and Armengaud J. (2012) Comparative proteogenomics of twelve Roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol. Cell. Proteomics*. http://dx.doi.org/10.1074/mcp.M111.013110.

Dong H. P., Wang D. Z., Dai M. H. and Hong H. S. (2010) Characterization of particulate organic matter in the water column of the South China Sea using shotgun proteomic approach. *Limnol. Oceanogr.* **55**, 1565–1578.

Ducklow H. W. (2002) Foreword, p. xv–xviii. In *Biogeochemistry of Marine Dissolved Organic Matter* (eds. D. A. Hansell and C. A. Carlson). Academic Press.

Giovannoni S. J., Bibbs L., Cho J. C., Stapels M. D., Desiderio R., Vergin K. L., Rappe M. S., Laney S., Wilhelm L. J., Tripp J., Mathur E. J. and Barofsky D. F. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **483**, 82–85.

Guo L. D., Santschi P. H. and Warnken K. (1995) Dynamics of dissolved organic carbon (DOC) in oceanic environments. *Limnol. Oceanogr.* **40**, 1392–1403.

Guo L. D. and Santschi P. H. (1997) Composition and cycling of colloids in marine environments. *Rev. Geophys.* **35**, 17–40.

Hansell D. A. (2012) Recalcitrant dissolved organic carbon fractions. *Annu. Rev. Mar. Sci.*. http://dx.doi.org/10.1146/annurev-marine-120710-100757.

Hertkorn N., Benner R., Frommberger M., Kopplin P. S., Witt M., Kaiser K., Kettrup A. and Hedges J. I. (2006) Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010.

Janausch I. G., Zientz E., Tran Q. H., Kroger A. and Unden G. (2002) C4-dicarboxylate carriers and sensors in bacteria. *Biochim. Biophys. Acta* **1553**, 39–56.

Jiao N. Z. and Yang Y. (2002) Ecological studies on *Prochlorococcus* in China Seas. *Chin. Sci. Bull.* **47**, 1243–1250.

Jiao N. Z., Herndl G. J., Hansell D. A., Benner R., Kattner G., Wihelm S. W., Kirchman D. L., Weinbauer M. G., Luo T., Chen F. and Azam F. (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* **8**, 593–599.

Karner M. B., DeLong E. F. and Karl D. M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510.

Keller A., Nesvizhskii A. I., Kolker E. and Aebersold R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392.

Loick N., Dippner J., Nhu-Doan H., Liskow I. and Voss M. (2007) Pelagic nitrogen dynamics in the Vietnamese upwelling area according to stable nitrogen and carbon isotope data. *Deep-Sea Res. I* **54**, 596–607.

Ludden P. W. and Roberts G. P. (1995) The biochemistry and genetics of nitrogen fixation by photosynthetic bacteria. In *Anoxygenic Photosynthetic Bacteria* (eds. R. E. Blankenship, M. T. Madigan and C. E. Bauer). Kluwer Academic Publishing Inc., Dordrecht, pp. 929–947.

McCarthy M. D., Pratum T., Hedges J. I. and Benner R. (1997) Chemical composition of dissolved organic nitrogen in the ocean. *Nature* **390**, 150–154.

Morris R. M., Nunn B. L., Frazar C., Goodlett D. R., Ting Y. S. and Rocap G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME J.* **4**, 673–685.

Morris R. M., Rappe M. S., Connon S. A., Vergin K. L., Siebold W. A., Carlson C. A. and Giovannoni S. J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810.

Nesvizhskii A. I., Keller A., Kolker E. and Aebersold R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658.

Noble R. T. and Fuhrman J. A. (1997) Virus decay and its causes in coastal waters. *Appl. Environ. Microbiol.* **63**, 77–83.

Ogawa H. and Tanoue E. (2003) Dissolved organic matter in oceanic waters. *J. Oceanogr.* **59**, 129–147.

Powell M. J., Sutton J. N., Del Castillo C. E. and Timperman A. I. (2005) Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. *Mar. Chem.* **95**, 183–198.

Rusch D. B., Halpern A. L., Sutton G., Heidelberg K. B., Williamson S. and Yooseph S., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77, doi10.1371/journal/pbio.0050077.

Ryu S., Gallis B., Goo Y. A., Shaffer S. A., Radulovic D. and Goodlett D. R. (2008) Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. *Cancer Inform.* **6**, 243–255.

Sowell S. M., Abraham P. E., Shah M., Verberkmoes N. C., Smith D. P., Bouglas D. F. and Giovannoni S. J. (2011) Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *The ISME J.* **5**, 856–865.

Sowell S. M., Wilhelm L. J., Norbeck A. D., Lipton M. S., Nicora C. D. and Barofsky D., et al. (2008) Transport functions dominate the SAR11 metaproteome at low nutrient extremes in the Sargasso Sea. *The ISME J.* **3**, 93–105.

Sullivan M. B., Waterbury J. B. and Chisholm S. W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051.

Suttle C. A. (2005) Viruses in the sea. *Nature* **437**, 356–361.

Tanoue E. (1995) Detection of dissolved protein molecules in oceanic waters. *Mar. Chem.* **51**, 239–252.

Verberkmoes N. C., Russell A. L., Shah M., Godzik A., Rosenquist M., Halfvarson J., Lefsrud M. G., Apajalahti J., Tysk C., Hettich R. L. and Jansson J. K. (2009) Shotgun metaproteomics of the human distal gut microbiota. *The ISME J.* **3**, 179–189.

Wang D. Z., Dong H. P., Xie Z. X., Dai M. H. and Hong H. S. (2011) Metaproteomic characterization of dissolved organic matter in the water column of the South China Sea. *Limnol. Oceanogr.* **56**, 1641–1652.

Williamson S. J., Rusch D. B., Yooseph S., Halpern A. L., Heidelberg K. B., Glass J. I., Pfannkoch C. A., Fadrosh D., Miller C. S., Sutton G., Frazier M. and Venter J. C. (2008) The sorcerer ‖ global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples.

PLoS One **3**, e1456. http://dx.doi.org/10.1371/journal.pone.0001456.

Wilm M., Shevchenko A., Houthaeve T., Breit S., Schweigerer L., Fotsis T. and Mann M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469.

Yamada N. and Tanoue E. (2003) Detection and partial characterization of dissolved glycoproteins in oceanic waters. *Limnol. Oceanogr.* **48**, 1037–1048.

Yamada N. and Tanoue E. (2006) The inventory and chemical characterization of dissolved proteins in oceanic waters. *Prog. Oceanogr.* **69**, 1–18.

Yamada N. and Tanoue E. (2009) Similarity of electrophoretic dissolved protein spectra from coastal to pelagic seawaters. *J. Oceanogr.* **65**, 223–233.

Yooseph S., Sutton G., Rusch D. B., Halpern A. L., Williamson S. J. and Remington K., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16.

Zhang Y., Jiao N., Sun Z., Hu A. and Zheng Q. (2011) Phylogenetic diversity of bacterial communities in South China Sea mesoscale cyclonic eddy perturbations. *Res. Microbiol.* **162**, 320–329.

Zou L., Wang X. C., Callahan J., Culp R. A., Chen R. F., Altabet M. A. and Sun M. Y. (2004) Bacterial roles in the formation of high-molecular weight dissolved organic matter in estuarine and coastal waters: evidence from lipids and the compound-specific isotopic ratios. *Limnol. Oceanogr.* **49**, 297–302.

Associate editor: Ann Pearson