

Online Nanoflow Two-Dimension Comprehensive Active Modulation Reversed Phase–Reversed Phase Liquid Chromatography High-Resolution Mass Spectrometry for Metaproteomics of Environmental and Microbiome Samples

Matthew R. McIlvin^{*,†} and Mak A. Saito[†]Cite This: *J. Proteome Res.* 2021, 20, 4589–4597

Read Online

ACCESS |



Metrics & More



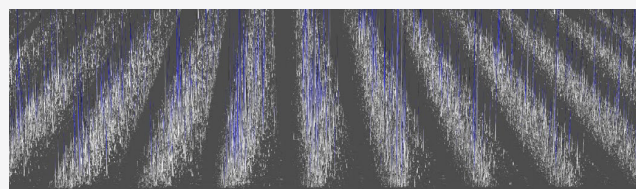
Article Recommendations



Supporting Information

ABSTRACT: Metaproteomics is a powerful analytical approach that can assess the functional capabilities deployed by microbial communities in both environmental and biomedical microbiome settings. Yet, the mass spectra resulting from these mixed biological communities are challenging to obtain due to the high number of low intensity peak features. The use of multiple dimensions of chromatographic separation prior to mass spectrometry analyses has been applied to proteomics previously but can require increased sampling handling and instrument time. Here, we demonstrate an automated online comprehensive active modulation two-dimensional liquid chromatography method for metaproteome sample analysis. A high pH PLRP-S column was used in the first dimension followed by low pH separation in the second dimension using dual modulating C18 traps and a C18 column. This method increased the number of unique peptides found in ocean metaproteome samples by more than 50% when compared to a one-dimension separation while using the same amount of sample and instrument time.

KEYWORDS: metaproteomics, proteomics, microbiome, environmental, comprehensive active modulation, nano 2D LC MS/MS



RPLC active modulation RPLC MS metaproteomics

INTRODUCTION

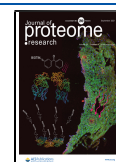
Metaproteomics, the analysis of proteins of mixed microbial populations, is showing itself to be a powerful tool in assessing the functional capabilities deployed by microbial communities in both environmental and biomedical microbiome settings.^{1–15} In natural environments such as the oceans, there are diverse communities of microbial species that change composition over the location and time. Characterizing the biochemical activities of these microbial communities is important in understanding their relationships with biogeochemical cycles, particularly in the context of global carbon cycling and the influence of the marine biological component. In this context, metaproteomics has shown that nutrient and micronutrient transporters, sensors, and storage molecules often increase in abundance within microbes as part of an adaptive response to nutrient scarcity.^{16,17} Moreover, biogeochemically relevant enzyme catalysts can be detected by metaproteomics and can help constrain models and even provide potential chemical reaction rates when coupled with enzyme kinetic parameters.¹⁸ In biomedical microbiome studies, metaproteomics has demonstrated that understanding the functional composition of the microbiome can provide

valuable information about the system that is not revealed by studying the microbial membership composition alone.¹⁹

The biologically diverse mixtures studied in metaproteomic analyses produce extremely rich mass spectra that can be substantially more complex than the human proteome.²⁰ Moreover, the spectra typically need to be mapped to metagenomic or metatranscriptomic data to obtain protein identifications, which may not always be available or be of limited sequencing depth. These biomass samples are often limited in material and difficult to obtain, for example, acquiring dilute microbial biomass in the oceans can require filtration of large volumes of seawater, although some colonial microbiomes and their symbionts are large enough to be physically isolated. Therefore, a method to extract as much information as possible from a single analysis of a small amount of material is useful for metaproteomic workflows.

Received: July 16, 2021

Published: August 12, 2021



A common method for protein identification in metaproteomics currently is bottom-up proteomics using protease digestion to peptides followed by liquid chromatography and high-resolution mass spectrometry. Peak capacity and peptide identification of one-dimension liquid chromatography and mass spectrometry (LC–MS) can be increased by using longer columns and flow gradients but will eventually reach a length and time limit.²¹ Offline two-dimensional liquid chromatography mass spectrometry (2D-LC–MS, collecting fractions of the first LC column and reinjecting them in multiple LC–MS analyses) can achieve higher protein identification than LC–MS due to higher peak capacity, which results in fewer compounds entering the mass spectrometer at any given time.²² However, 2D-LC–MS requires a longer mass spectrometry analysis time and increased sample preparation effort relative to LC–MS and also runs the risk of sample dilution and potential loss of sample due to increased handling or complex analytical systems.^{23–25} A comparison of LC–MS and 2D-LC–MS methods for metaproteomics demonstrated that 2D-LC–MS using strong cation exchange and C18 outperformed LC–MS using C18 alone, although the analysis time was 22 h per sample for 2D-LC–MS in that particular study.²⁶

Other online 2D-LC–MS methods have been developed that use a combination of high and low pH reversed phase columns for proteomics analysis.²⁷ Several approaches have been done to adjust the pH and prepare the eluting sample before analysis on the second dimension, such as active solvent modulation,^{28,29} stop flow chromatography combined with a trap column,³⁰ and online noncontiguous fractionating and concatenating.²³ However, these methods required either higher flow rates, sample size, specialized equipment, or analysis time when compared to the method described here (nanoscale, low sample size, and within a more typical sample analysis time). Environmental samples are often limited in total protein to 10 μ g or less; therefore, nanoscale methods would be beneficial to the analysis.

Online comprehensive active modulation two-dimensional liquid chromatography (LC $\times$ LC) is a method that can combine two separate LC separations that are normally incompatible.³¹ The first LC separation column eluent is modified in-line with a solvent that makes the eluent compatible with the second dimension before loading onto a trap column. The two traps are alternately loaded from the first column and eluted onto the second column throughout the analysis (Figure 1). A single long LC gradient is done in the first LC dimension, while a series of shorter LC gradients are performed in the second LC dimension that match the valve modulation. There are a large variety of LC $\times$ LC combinations to choose from that have the potential to be coupled to mass spectrometry.³¹ Several LC $\times$ LC combinations have been utilized with a variety of detectors, but few have been coupled with high-resolution mass spectrometry (HR-MS).^{32–34}

The nanoscale LC $\times$ LC–MS method described here uses a combination of high pH reversed phase LC in the first dimension with low pH reversed phase LC in the trap columns and second dimension column followed by high-resolution mass spectrometry. Offline, automated, and online 2D-LC–MS using high and low pH reversed phase LC has been previously demonstrated for proteomics analyses,^{22,23,27,30,35} but to our knowledge, it has not been automated with comprehensive active modulation two-dimensional liquid

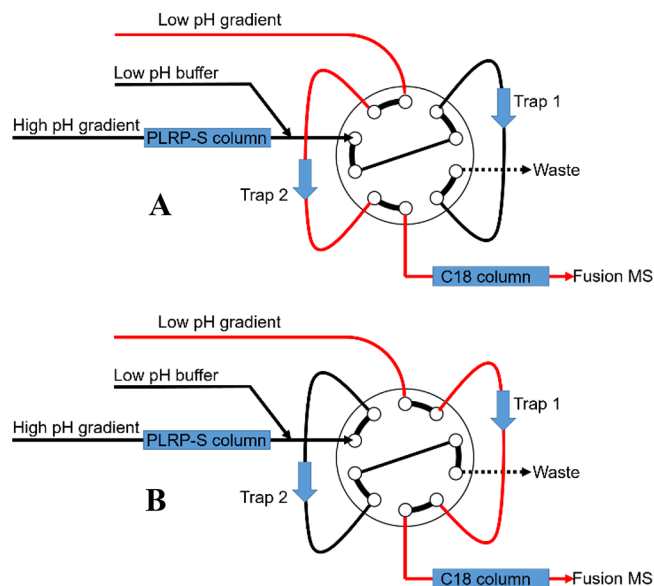


Figure 1. Schematic of the LC $\times$ LC–MS system, showing positions A and B of the 10-port valve.

chromatography mass spectrometry (e.g., with sample collection after the first dimension onto alternating dual trap columns) and applied to metaproteomics previously. This study used environmental samples from the Central Pacific Ocean to compare LC–MS and LC $\times$ LC–MS metaproteomics analysis, and it was successfully applied to produce high-throughput analysis with greatly improved proteomic depth.

MATERIALS AND METHODS

Sample Collection and Preparation

Seawater samples were collected on filters using large-volume in situ pumps (McLane Research Laboratories, Inc.) during a research expedition to study proteomics in an oxygen minimum zone in the Central Pacific Ocean from January to February 2016 on the RV Falkor (FK160115 station 10, 8° N 140° W; chief scientist, M. Saito). Fraction filtration was done using 142 mm diameter 51 μ m pore size acid-cleaned (10% hydrochloric, Seastar) nylon mesh filters positioned before 142 mm 0.2 μ m pore size Supor polyethersulfone membrane filters (Pall Corporation). Pumps were deployed simultaneously at 20, 80, 150, 200, 300, and 400 m and programmed to pump at 4 L/min for 3 h. Filters were then cut and frozen at -80° C. Proteins were extracted from filter quarter sections by a modified magnetic bead extraction procedure as previously described,¹⁸ adapted from the SP3 method for larger volumes needed for metaproteome samples.^{36,37} Briefly, proteins were extracted from the filters with an extraction buffer (50 mM HEPES at pH = 8.5 (Boston BioProducts), 1% SDS in HPLC-grade water), and the isolated protein was reduced and alkylated (dithiothreitol and iodoacetamide) and then purified with magnetic beads (GE Healthcare SpeedBeads). The purified protein was cleaned and reconstituted from the beads, and the protein concentration recovery was calculated using a bicinchoninic acid assay (Pierce) and then trypsin digested (Promega). HeLa peptides were also used for method development (Pierce HeLa Protein Digest Standard 88329).

Liquid Chromatography and Mass Spectrometry Sample Analysis

All liquid chromatographic experiments were performed using a Dionex UltiMate3000 RSLCnano system with a loading pump, an additional RSLCnano pump, a 10-port two-position switching valve (VICI Cheminert C72X-6670SHYD), and a Dionex UltiMate3000 autosampler. All components were controlled with Chromeleon 7 software linked to Thermo Scientific Xcalibur 4.2. Columns were interfaced with a Thermo Fusion mass spectrometer using a Thermo Flex ion source.

To perform LC $\times$ LC–MS, 5 μ g of each sample was directly injected onto a PLRP-S column (200 μ m $\times$ 150 mm, 3 μ m bead size, 300 Å pore size, NanoLCMS Solutions) followed by a nonlinear 8 h gradient with a flow rate of 1 μ L/min from 5 to 95% buffer B using the first RSLCnano pump, where buffer A was 10 mM ammonium formate in water and buffer B was 10 mM ammonium formate in 90% acetonitrile, both adjusted to pH = 10. The PLRP-S column eluent was diluted in-line (10 μ L/min 0.1% formic acid using the loading pump and added to the eluent using a Valco tee, P/N ZT1XCS6) and then trapped and eluted every 30 min on alternating dual traps (300 μ m $\times$ 5 mm, 5 μ m bead size, 100 Å pore size, C18 PepMap100, in PepMap trap holders, P/N 164649, Thermo Scientific). The alternating traps were eluted at 500 nL/min onto a C18 column (100 μ m $\times$ 150 mm, 3 μ m particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH packed in a New Objective PicoFrit column) with repeated 30 min nonlinear gradients of 2–95% buffer D using the second RSLCnano pump, where buffer C was 0.1% formic acid in water and buffer D was 0.1% formic acid in 80% acetonitrile. All solvents were Fisher Optima-grade. A diagram of the flow path is shown in Figure 1, and flow gradients for both dimensions are shown in Figure 2.

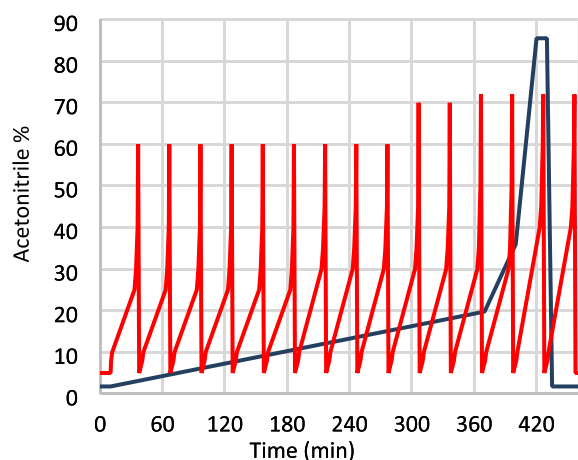


Figure 2. Elution gradients used for LC $\times$ LC active modulation over a 460 min total gradient. First dimension column PLRP-S gradient at 1 μ L/min (blue) and second dimension C18 gradient at 0.5 μ L/min (red) using the two RSLCnano pumps.

For comparison, 1D LC–MS was performed using the same equipment as for LC $\times$ LC–MS but without the additional RSLCnano pump. Then, 5 μ g of each sample was first loaded onto a trap column (300 μ m $\times$ 5 mm, 5 μ m bead size, 100 Å pore size, C18 PepMap100, Thermo Scientific) and washed with 0.1 formic acid and 2% acetonitrile for 10 min at 10 μ L/min. The valve was then switched to include the trap in line

with a C18 column (100 μ m $\times$ 450 mm, 3 μ m particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH packed in a New Objective PicoFrit column) and eluted with a 4 or 8 h nonlinear gradient of 5–95% buffer B, where buffer A was 0.1% formic acid in water and buffer B was 0.1% formic acid in 80% acetonitrile.

During 1D LC–MS and 2D LC $\times$ LC–MS, the Thermo Fusion mass spectrometer collected MS1 spectra at 240 K resolution in the Orbitrap between 380 and 1580 m/z with an AGC target of 4.0×10^5 . MS2 spectra were collected on the ion trap in data-dependent mode with a cycle time of 2 s between master scans. The maximum injection time was 50 ms with the option to inject ions for all available parallelizable time. The CID collision energy was set at 30% and a quadrupole isolation window of 1.6 m/z . Monoisotopic peak determination was set to peptide and to include charge states from 2 to 10 with a mass tolerance of ± 3 ppm. Dynamic exclusion time was 20 s for LC–MS and 5 s for LC $\times$ LC–MS to best match the peak widths obtained for both systems.

Data Analysis and Statistics

The raw mass spectra files were searched using SEQUEST HT within Thermo Proteome Discoverer 2.2 software using a parent ion tolerance of 10 ppm and a fragment tolerance of 0.6 Da and allowing up to 1 missed cleavage site. Oxidation and acetyl dynamic modifications and carbamidomethyl static modifications were included. Percolator peptide spectral matching was used within Proteome Discoverer with a maximum Delta Cn of 0.05 and a decoy search validation based on posterior error probabilities (PEP).³⁸ A custom metagenomic database was used for peptide-to-spectrum matching that was based on six deep metagenomic samples, which were co-collected from this research expedition as described previously.^{17,18} A HeLa reference proteome from UniProt was used to match HeLa digest samples. Processed files were then loaded into Scaffold 5.0 (Proteome Software Inc.) using prefiltered mode with a protein threshold of 1.0% false discovery rate (FDR), a peptide threshold of 0.1% FDR, and a minimum of 1 peptide for analysis. MS1 filtering using Skyline (Skyline 20.2)³⁹ was used to compare MS1 precursor and MS2 fragmentation intensity. RawBeans was used to evaluate peak intensity and full width at half maximum (FWHM).⁴⁰

Data Availability in Repositories

Data used in this study have been deposited at the PRIDE repository (dataset identifier: PXD027351 and [10.6019/PXD027351](https://www.ebi.ac.uk/pride/projects/PXD027351)) and Biological and Chemical Oceanography Data Management Office (BCO-DMO). Sampling metadata, processed data, and co-located environmental data from the ProteOMZ expedition are available at BCO-DMO under the project number 685696 (<https://www.bco-dmo.org/project/685696>) and datasets for proteins, peptides, and FASTA of identified sequences (<https://www.bco-dmo.org/dataset/737620>, <https://www.bco-dmo.org/dataset/737596>, and <https://www.bco-dmo.org/dataset/737611>).

RESULTS AND DISCUSSION

This study demonstrates the development of an active modulation LC $\times$ LC–MS metaproteomics method and compares its application to ocean microbial and human cell line samples. Attributes of the method described include the column and trap chemistry, LC $\times$ LC performance and peak

widths, and instrument settings. Finally, applications of the method to example ocean samples are described.

Selection of PLRP-S and C18 for LC $\times$ LC–MS

Polymeric reversed phase (PLRP-S) stationary phase was chosen for the first LC dimension. High pH separation columns are often used for the first dimension of 2D analysis, which typically has better separation and solvent compatibility than strong cation exchange and size exclusion chromatography columns. A larger-diameter first dimension column (0.2 mm ID) was chosen to increase the total protein capacity and minimize sample loss during direct injection of the sample. C18 was selected for the two active modulation traps and second dimension column due to its high separation capacity for peptides and MS compatibility. Both of these phases combined in 2D-LC–MS exhibited high resolution and solvent compatibility and orthogonal selectivity as previously shown in past experiments.¹⁶ All solvents are highly compatible with mass spectrometry, and the two column phases are stable with the solvents used when the 10 μ L/min 0.1% formic acid in water is added in-line to decrease the pH of the PLRP-S column eluent (Figure 1).

Separation orthogonality of multiple LC–MS combinations has been studied extensively and has shown that high pH and low pH reversed phase columns are complementary for peptide separations.^{41,42} The orthogonality of the two specific LC phases used in this method (PLRP-S 3 μ m bead size, 300 Å pore size, NanoLCMS Solutions and C18 3 μ m bead size, 120 Å pore size, Reprosil-Gold, Dr. Maisch GmbH) was examined by analyzing them individually in 1D mode using 5 μ g of HeLa peptide standard. The same method for LC–MS as previously described was used for the C18 analysis (loading on trap column and eluting with an 8 h gradient). PLRP-S LC–MS 1D data was obtained by loading 5 μ g of HeLa digest directly onto the column followed by the same 8 h gradient of ammonium formate and acetonitrile used during the first dimension of the LC $\times$ LC–MS method, with the eluent delivered to the mass spectrometer with a silica nanospray tip. The peptide retention time between the two columns demonstrated sufficient orthogonality between the materials used in this experiment when comparing each column separately in 1D mode (Figure S1).

Trapping Efficiency

The two C18 traps are the same in design to trap columns commonly used in 1D analysis to remove salts and other impurities in 1D LC–MS, though in this case, used for a longer trapping time (30 min compared to 5–10 min at 10 μ L/min). The additional buffer added at the tee to the PLRP-S eluent reduces the pH and acetonitrile concentration to within the range usually used for 1D trapping. The efficiency of trapping the sample eluting from the first dimension high pH column onto the two trap columns was monitored by attaching the waste from the 10-port valve (Figure 1) to a heated electrospray ionization source attached to the mass spectrometer. HeLa peptides were injected by the normal LC $\times$ LC–MS method, and the waste eluent was monitored for 8 h. Only 25 proteins were identified in this 8 h analysis, indicating low sample loss through the two C18 trap columns (data not shown).

Performance of LC $\times$ LC–MS

Thirty-minute gradients in the second dimension were chosen to have enough time for a quality LC gradient and to minimize

the equilibration time (Figure 2). Shorter gradients are possible but would increase the relative amount of time devoted to column equilibration throughout the analysis, thus decreasing the usable MS analysis time and could result in peak widths too narrow for the acquisition frequency of the mass spectrometer used in this study. The typical peak width was 5 s for 30 min gradients, which enabled the Fusion to obtain quality MS2 spectra with an MS1 cycle time of 2 s. Note that the dynamic exclusion time window was decreased in 2D to 5 from 20 s to compensate for the narrower peak widths in 2D. This resulted in 14 separate second dimension LC chromatograms during the 8 h analysis as seen in Figure 3A (the first 30 min gradient is ignored because the first dimension is yet to load on the trap column).

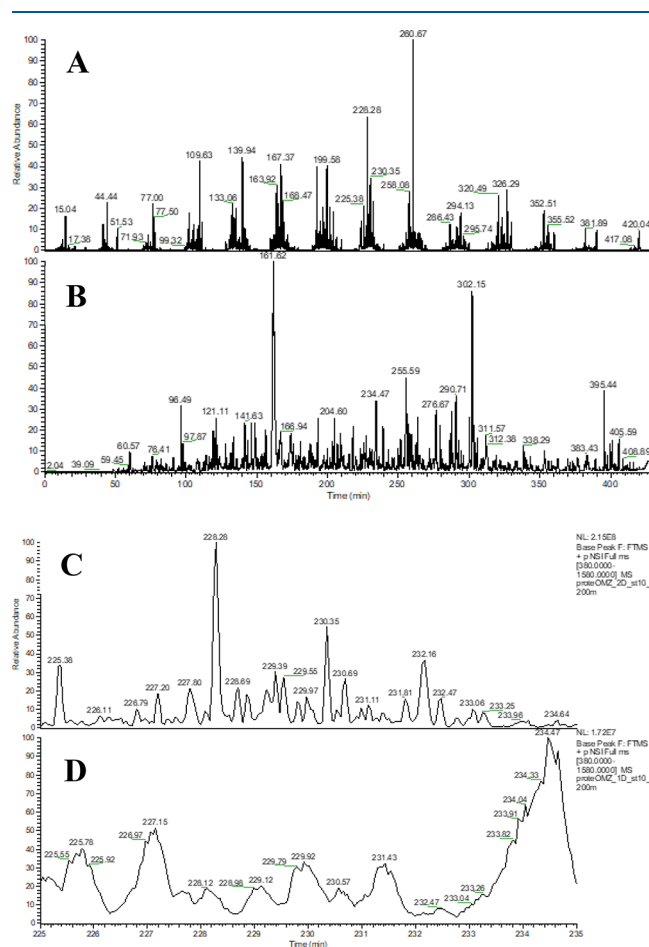


Figure 3. MS1 base peak plots of LC $\times$ LC–MS (A) and LC–MS (B) of a 200 m depth 0.2 μ m water filter metaproteomics sample analyzed for 460 min. Panels (C) and (D) are a 10 min zoom of the same chromatograms, demonstrating the higher peak capacity achieved by LC $\times$ LC–MS (C).

The MS1 chromatograms comparing LC–MS and LC $\times$ LC–MS for one of the sample depths are shown in Figure 3A,B. Narrower peak widths and higher intensities were observed overall throughout the LC $\times$ LC–MS analysis as observed in the 10 min time window comparison in Figure 3C,D. The peak widths versus peak intensity are shown for the two methods in Figure 4A,D (additional depths and HeLa are shown in Figure S2). To demonstrate how this affects data acquisition, an example of a relatively low spectral count protein was chosen (annotated as UPF0365 family protein

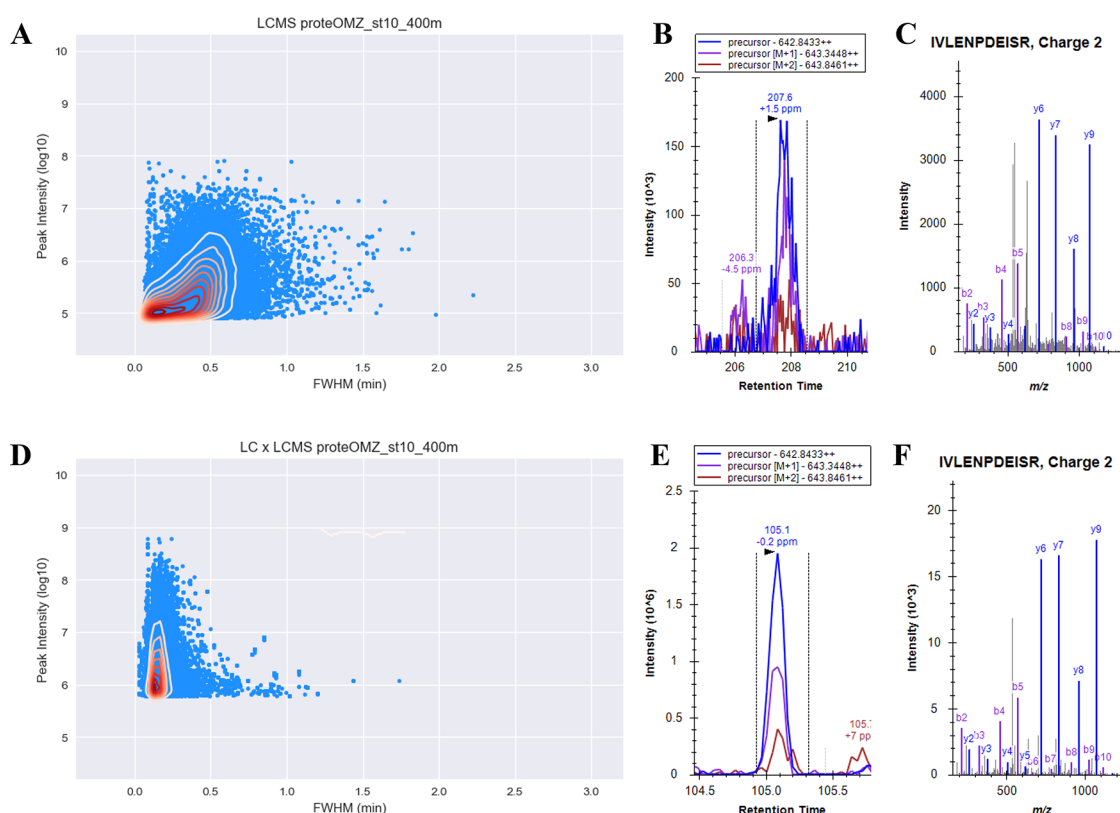


Figure 4. Comparison of full width at half maximum (FWHM) to peak intensity for LC–MS (A) and LC × LC–MS (D) of a 400 m depth 0.2 μ m water filter metaproteomics sample, showing a large increase in both FWHM and peak intensity for LC × LC–MS. MS1 precursor peak intensity and MS2 fragmentation intensity comparison between LC–MS (B, C) and LC × LC–MS (E, F) of a relatively low intensity peptide (Candidatus Marinimicrobia bacterium, sequence identifier: NODE_160939_length_858_cov_11.7995_1_858_-).

Table 1. Comparison between LC–MS and LC × LC–MS Data from Seawater Filters

depth (m)	ProteOMZ protein and peptide matches									
	LC–MS					LC × LC–MS				
	proteins	peptides	unique peptides	unique MS/MS spectra	total MS/MS	proteins	peptides	unique peptides	unique MS/MS spectra	total MS/MS
20	5024	21,492	8733	10,912	141,095	8268	37,811	13,926	17,988	230,809
20 4 h	4413	15,201	7620	9376	85,496					
80	3165	15,015	5119	6278	135,668	6140	25,361	9718	12,691	223,438
80 4 h	2693	11,114	4204	5028	79,509					
150	5032	23,325	10,749	12,991	140,813	8925	42,262	17,598	22,463	229,439
200	6769	30,706	15,650	18,768	144,007	13,192	57,030	26,757	33,158	231,696
300	6984	31,642	16,818	20,022	152,732	13,508	60,085	28,266	35,070	233,835
400	6625	30,164	15,416	18,572	154,306	15,584	68,509	31,618	40,011	244,689

from the Candidatus Marinimicrobia bacterium) and MS1 filtering was performed in Skyline to plot the corresponding MS1 peak chromatography and matching MS2 fragment spectra for the detected peptides and is shown in Figure 4 (additional peptides in Figure S3). Increasing peak separation simplifies MS1 spectra, which increases the number of ions for a particular mass when injecting ions into the Orbitrap mass analyzer to reach the AGC target. Figure 4 shows the increased peak intensity and lower interference of other ions when using LC × LC–MS compared to LC–MS, resulting in a larger number of unique peptide identifications and percentage of matched MS2 spectra for LC × LC–MS as observed in Table 1. This is particularly important in metaproteomic and microbiome samples where there are many low-abundance species producing a large amount of low-concentration

peptides. For example, a prior analysis of ocean metaproteomics demonstrated far more detectable peaks (~50%) with a much lower average intensity (>10-fold less) relative to a human cell line with similar amounts of total protein injected (Saito et al.; their Figure 5).²⁰ This appears to be less of a problem with single organism proteomics as observed in Table S1 when comparing HeLa digest analysis using both methods, where the numbers of proteins and unique peptides are much closer between methods compared to the ocean metaproteomics samples (Figure 6).

An increase in peptide and protein identifications is often observed with an increase in depth in ocean metaproteomics samples (Table 1 and Figure 6). This is likely due to an increase in the diversity of organisms with depth, where the upper water column is often dominated by several abundant

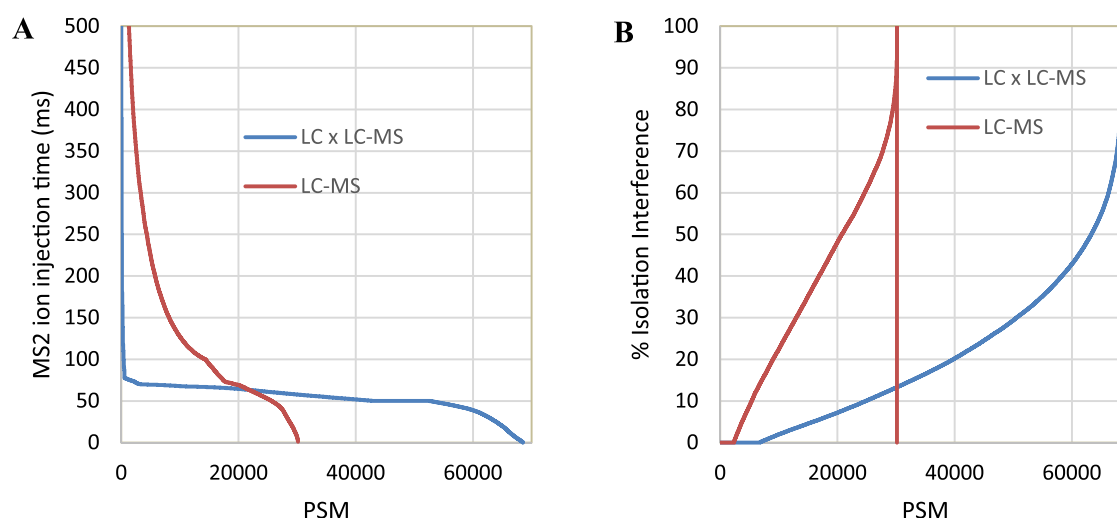


Figure 5. MS2 ion injection time (A) for all peptide spectral matches (PSM) for LC $\times$ LC-MS (blue) and LC-MS (red) of a 400 m depth 0.2 μ m water filter metaproteomics sample. Percent isolation interference for all PSMs of the same sample are shown in (B), where the % isolation interference = $100 \times [1 - (\text{precursor intensity in isolation window} / \text{total intensity in isolation window})]$. A high % isolation interference indicates the presence of co-isolated peptide species.

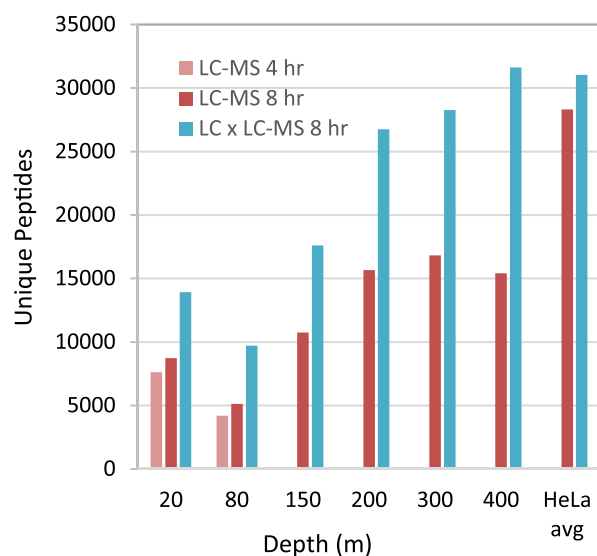


Figure 6. Comparison of unique peptide identification results by LC-MS (red) and LC $\times$ LC-MS (blue) analyzing sea water metaproteomics samples at different depths and HeLa.

phototrophic and lithotrophic species. The LC $\times$ LC-MS method would become particularly useful in the more complex and diverse deep ocean. Similar improvements could occur in soil and sediment metaproteomics and complex microbiomes (e.g., digestive systems) that are highly diverse.¹⁹

Instrument Settings

The Thermo Fusion mass spectrometer has the advantage of performing MS1 scans in the Orbitrap mass analyzer while simultaneously performing MS2 scans in the ion trap. This allows a high-resolution MS1 scan without sacrificing time for the Orbitrap scan to finish before obtaining MS2 spectra. When performing data-dependent scans in cycle time mode (2 s between MS1 master scans as opposed to a set number of MS2 scans per MS1 scan, or top N) and allowing ions to be injected for all available parallelizable time, the MS2 injection time during metaproteomics sample analysis is often above the set maximum injection time of 50 ms, which can be beneficial

for very low concentration peptides. The injection time for identified peptides typically ranges between 25 and 500 ms for LC-MS metaproteomics samples (Figure 5) but is significantly lower for LC $\times$ LC-MS with a majority of injection times less than 75 ms. Previous results using only the Orbitrap for MS1 and MS2 scans resulted in much fewer peptide identifications when analyzing metaproteomics samples.

Different MS database search engines and post-search algorithms can produce varying results on the same samples.⁴³ For metaproteomics samples, more peptide matches were initially observed with CID compared to HCD collision for MS2 analysis using the ion trap on the Fusion MS, and this was even more pronounced in LC-MS than in LC $\times$ LC-MS. When data was analyzed using SEQUEST HT and Target Decoy PSM Validator in Proteome Discoverer followed by loading the resulting files in Scaffold using either PeptideProphet or Local False Discovery Rate (LFDR) algorithms, it was necessary to increase the Xcorr peptide threshold values to get an acceptable FDR of less than 1% for proteins and 0.1% for peptides. However, when SEQUEST HT was used with Percolator PEP scoring followed by prefiltered mode in Scaffold (available in version 4.11 and later), HCD collision obtained slightly more peptide identifications than CID, and both performed better overall. The better Xcorr values obtained using CID may be a result of a decrease in low mass fragments in CID, which may result in more interferences in complex samples. However, the increased speed of HCD compared to CID (10 ms collision time required only for CID) may overcome this difference depending on the search algorithms used.

Applications in Environmental and Microbiome Metaproteomics

The 2D LC-MS method described here is suitable for high-throughput metaproteomic analysis of environmental and microbiome samples. Because both dimensions are fully automated, users only load each sample a single time. For example, three 8 h samples per day can be conducted on 24/7 operations. In contrast, offline 2D operations typically require user intervention to load the second dimension extracts and could be susceptible to handling errors. As a result, the 2D

LC–MS method enables the analysis of large datasets with great depth to provide coherence across large sample sets, such as large-scale clinical cross-sectional or longitudinal microbiome studies or environmental studies with large-scale longitudinal (literally in this case) sampling schemes. In the latter case, we have applied this 2D LC–MS method to two >100 sample datasets from the Central Pacific and North Atlantic Oceans¹⁸ (also Saunders et al., in prep; Saito et al., in prep). Data visualizations of these datasets demonstrate coherence of low-abundance proteins across spatial scales covering thousands of kilometers and a thousand meters deep. For example, a relatively low-abundance sugar transporter from alphaproteobacteria was identified by four unique peptides across 4000 km in the Central Pacific in the upper mesopelagic ocean where sugar release by degradation of sinking organic materials is expected (Figure 7). This specific visualization and

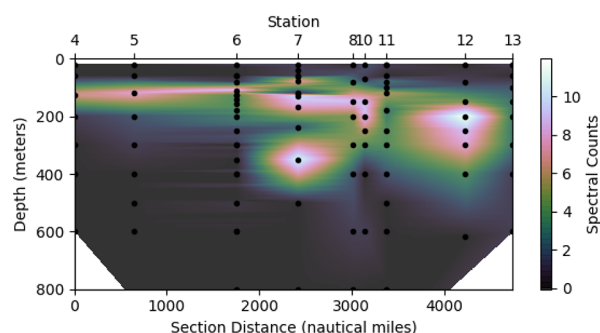


Figure 7. Distribution of a sugar ABC transporter from the FK160115 ProteOMZ expedition in the central Pacific Ocean, revealing a presence in the upper mesopelagic depths (100–400 m) where sugars are expected to be released from the degradation of sinking organic matter. Top taxonomic Blast hits for this protein are associated with the alphaproteobacteria (sequence ID: NODE_13436_length_3609_cov_27.852_2554_3531_ within the Ocean Protein Portal).

tens of thousands more proteins are accessible through a data access and visualization portal, the Ocean Protein Portal, demonstrating the ability of high-throughput 2D LC–MS to generate robust results that have been made accessible for use in research and education.¹⁸

The 2D LC–MS method described here is also suited to the generation of spectral libraries needed for data-independent acquisition analysis. Deep 2D analysis of a subset of samples could be employed for library generation that encompasses much of the diversity present in the dataset. Further library improvement parsing through machine learning-based spectral generation (ProSIT) could avoid instances of false positives.⁴⁴

CONCLUSIONS

This study was a comparison of 1D and 2D active modulation LC methods for metaproteomics of oceanic filter samples and demonstrated that protein identification increased by a factor of 1.5 or more when using 2D active modulation LC while using the same instrument and amount of sample. The system requires only an additional LC pump and columns added to a typical LC–MS and therefore has the potential to be used in many labs. Based on these findings, it is recommended that metaproteomics sample analysis would benefit from using LC × LC–MS. This method has now been routinely used in this lab for over 4 years and has analyzed hundreds of ocean metaproteomics samples.^{18,20,45}

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00588>.

Mass spectrometry proteomics data deposited to the ProteomeXchange Consortium via the PRIDE⁴⁶ partner repository with the dataset identifier PXD027351 and 10.6019/PXD027351; MS spectra and additional experimental results: comparison of LC–MS and LC × LC–MS data from the HeLa standard, orthogonality between the first (PLRP-S) and second (C18) phases, FWHM plot comparisons between LC–MS and LC × LC–MS for several depths and HeLa, and MS1 precursor peak intensity and MS2 fragmentation intensity comparison between LC–MS and LC × LC–MS of several low intensity peptides from a single protein (PDF)

AUTHOR INFORMATION

Corresponding Author

Matthew R. McIlvin – Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02563, United States; orcid.org/0000-0002-5301-8365; Email: mmcilmv@whoi.edu

Author

Mak A. Saito – Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02563, United States; orcid.org/0000-0001-6040-9295

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00588>

Author Contributions

†M.R.M. and M.A.S. contributed equally to this work.

Funding

NIH R01GM135709, NSF 1925445, NSF 1850719, and G.B Moore Foundation were granted to M.A.S.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dawn Moran (Woods Hole Oceanographic Institution) for sample preparation. We thank Andrea Gargano (Universiteit van Amsterdam) for analytical method discussions.

REFERENCES

- (1) Bender, S. J.; Moran, D. M.; McIlvin, M. R.; Zheng, H.; McCrow, J. P.; Badger, J.; DiTullio, G. R.; Allen, A. E.; Saito, M. A. Colony formation in Phaeocystis antarctica: connecting molecular mechanisms with iron biogeochemistry. *Biogeosciences* **2018**, *15*, 4923–4942.
- (2) Bergauer, K.; Fernandez-Guerra, A.; Garcia, J. A. L.; Sprenger, R. R.; Stepanauskas, R.; Pachiadaki, M. G.; Jensen, O. N.; Herndl, G. J. Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E400–E408.
- (3) Blakeley-Ruiz, J. A.; Erickson, A. R.; Cantarel, B. L.; Xiong, W.; Adams, R.; Jansson, J. K.; Fraser, C. M.; Hettich, R. L. Metaproteomics reveals persistent and phylum-redundant metabolic

functional stability in adult human gut microbiomes of Crohn's remission patients despite temporal variations in microbial taxa, genomes, and proteomes. *Microbiome* **2019**, *7*, 1–15.

(4) Brum, J. R.; Ignacio-Espinoza, J. C.; Kim, E. H.; Trubl, G.; Jones, R. M.; Roux, S.; VerBerkmoes, N. C.; Rich, V. I.; Sullivan, M. B. Illuminating structural proteins in viral "dark matter" with metaproteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2436–2441.

(5) Colatriano, D.; Ramachandran, A.; Yergeau, E.; Maranger, R.; Gélinais, Y.; Walsh, D. A. Metaproteomics of aquatic microbial communities in a deep and stratified estuary. *Proteomics* **2015**, *15*, 3566–3579.

(6) Kleiner, M.; Thorson, E.; Sharp, C. E.; Dong, X.; Liu, D.; Li, C.; Strous, M. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* **2017**, *8*, 1558–1514.

(7) Maier, T. V.; Lucio, M.; Lee, L. H.; VerBerkmoes, N. C.; Brislawn, C. J.; Bernhardt, J.; Lamendella, R.; McDermott, J. E.; Bergeron, N.; Heinzmann, S. S.; Morton, J. T.; Gonzalez, A.; Ackermann, G.; Knight, R.; Riedel, K.; Krauss, R. M.; Schmitt-Koppin, P.; Jansson, J. K. Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *MBio* **2017**, *8*, No. e01343.

(8) Matallana-Surget, S.; Jagtap, P. D.; Griffin, T. J.; Beraud, M.; Wattiez, R. Comparative metaproteomics to study environmental changes. *Metagenomics*; Academic Press: 2018 (pp. 327–363).

(9) Moore, E. K.; Nunn, B. L.; Goodlett, D. R.; Harvey, H. R. 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*, 324–359.

(10) Mikan, M. P.; Harvey, H. R.; Timmins-Schiffman, E.; Riffle, M.; May, D. H.; Salter, I.; Noble, W. S.; Nunn, B. L. Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *The ISME Journal* **2020**, *14*, 39–52.

(11) Saito, M. A.; McIlvin, M. R.; Moran, D. M.; Goepfert, T. J.; DiTullio, G. R.; Post, A. F.; Lamborg, C. H. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **2014**, *345*, 1173–1177.

(12) Saito, M. A.; Breier, C.; Jakuba, M.; McIlvin, M.; Moran, D. Envisioning a chemical metaproteomics capability for biochemical research and diagnosis of global ocean microbiomes. *The Chemistry of Microbiomes: Proceedings of a Seminar Series*; National Academies Press: National Academies of Sciences, Engineering, Medicine: 2017 (pp. 29–36).

(13) Verberkmoes, N. C.; Russel, A. L.; Shah, M.; Godzik, A.; Rosenquist, M.; Halfvarson, J.; Lefsrud, M. G.; Apajalahti, J.; Tysk, C.; Hettich, R. L.; Jansson, J. K. Shotgun Metaproteomics of the Human Distal Gut Microbiota. *ISME J.* **2009**, *3*, 179–189.

(14) Wilmes, P.; Heintz-Buschart, A.; Bond, P. L. A decade of metaproteomics: where we stand and what the future holds. *Proteomics* **2015**, *15*, 3409–3417.

(15) Xiong, W.; Giannone, R. J.; Morowitz, M. J.; Banfield, J. F.; Hettich, R. L. Development of an enhanced metaproteomic approach for deepening the microbiome characterization of the human infant gut. *J. Proteome Res.* **2015**, *14*, 133–141.

(16) Saito, M. A.; Dorsk, A.; Post, A. F.; McIlvin, M. R.; Rappé, M. S.; DiTullio, G. R.; Moran, D. M. Needles in the blue sea: sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics* **2015**, *15*, 3521–3531.

(17) Cohen, N. R.; McIlvin, M. R.; Moran, D. M.; Held, N. A.; Saunders, J. K.; Hawco, N. J.; Brosnahan, G. R.; Camborg, C.; McCrow, J. P.; Dupont, C. L.; Allen, A. E.; Saito, M. A. Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat. Microbiol.* **2021**, *6*, 173–186.

(18) Saito, M. A.; McIlvin, M. R.; Moran, D. M.; Santoro, A. E.; Dupont, C. L.; Rafter, P. A.; Saunders, J. K.; Kaul, D.; Lamborg, C. H.; Westley, M.; Valois, F.; Waterbury, J. B. Abundant nitrite-

oxidizing metalloenzymes in the mesopelagic zone of the tropical Pacific Ocean. *Nat. Geosci.* **2020**, *13*, 355–362.

(19) Blakeley-Ruiz, J. A.; Erickson, A. R.; Cantarel, B. L.; Xiong, W.; Adams, R.; Jansson, J. K.; Fraser, C. M.; Hettich, R. L. Metaproteomics reveals persistent and phylum-redundant metabolic functional stability in adult human gut microbiomes of Crohn's remission patients despite temporal variations in microbial taxa, genomes, and proteomes. *Microbiome* **2019**, *7*, 18–15.

(20) Saito, M. A.; Bertrand, E. M.; Duffy, M. E.; Gaylord, D. A.; Held, N. A.; Hervey, W. J., IV; Hettich, R. L.; Jagtap, P. D.; Janech, M. G.; Kinkade, D. B.; Leary, D. H.; McIlvin, M. R.; Moore, E. K.; Morris, R. M.; Neely, B. A.; Nunn, B. L.; Saunders, J. K.; Shepherd, A. I.; Symmonds, N. I.; Walsh, D. A.; et al. *J. Proteome Res.* **2019**, *18*, 1461–1476.

(21) Hsieh, E. J.; Bereman, M. S.; Durand, S.; Valaskovic, G. A.; MacCoss, M. J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 148–153.

(22) Bath, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13*, 6176–6186.

(23) Lee, H.; Mun, D. G.; So, J. E.; Bae, J.; Kim, H.; Masselon, C.; Lee, S.-W. Efficient Exploitation of Separation Space in Two-Dimensional Liquid Chromatography System for Comprehensive and Efficient Proteomic Analyses. *Anal. Chem.* **2016**, *88*, 11734–11741.

(24) Ficarro, S. B.; Zhang, Y.; Carrasco-Alfonso, M. J.; Garg, B.; Adelmant, G.; Webber, J. T.; Kuckey, C. J.; Marto, J. A. Online Nanoflow Multidimensional Fractionation for High Efficiency Phosphopeptide Analysis. *Mol. Cell. Proteomics* **2011**, *10*, O111.011064.

(25) Yuan, H.; Jiang, B.; Zhao, B.; Zhang, L.; Zhang, Y. Recent Advances in Multidimensional Separation for Proteome Analysis. *Anal. Chem.* **2019**, *91*, 264–276.

(26) Hinzke, T.; Kouris, A.; Hughes, R.-A.; Strous, M.; Kleiner, M. More Is Not Always Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics. *Front. Microbiol.* **2019**, *10*, 238.

(27) Lam, M. P. Y.; Law, C. H.; Quan, Q.; Zhao, Y.; Chu, I. K. Fully automatable multidimensional reversed-phase liquid chromatography with online tandem mass spectrometry. *Methods Mol. Biol.* **2014**, *1156*, 39–51.

(28) Stoll, D. R.; Shoykhet, K.; Petersson, P.; Buckenmaier, S. Active Solvent Modulation: A valve-based approach to improve separation compatibility in two-dimensional liquid chromatography. *Anal. Chem.* **2017**, *89*, 9260–9267.

(29) Stoll, D. R.; Lhotka, H. R.; Harnes, D. C.; Madigan, B.; Hsiao, J. J.; Staples, G. O. High resolution two-dimensional liquid chromatography coupled with mass spectrometry for robust and sensitive characterization of therapeutic antibodies at the peptide level. *J. Chromatogr., B* **2019**, *1134–1135*, 121832.

(30) Richard, V. R.; Domanski, D.; Percy, A. J.; Borchers, C. H. An online 2D-reversed-phase – Reversed-phase chromatographic method for sensitive and robust plasma protein quantitation. *J. Proteomics* **2017**, *168*, 28–36.

(31) Pirok, B. W. J.; Gargano, A. F. G.; Schoenmakers, P. J. Optimizing separations in online comprehensive two-dimensional liquid chromatography. *J. Sep. Sci.* **2018**, *41*, 68–98.

(32) Gargano, A. F. G.; Shaw, J. B.; Zhou, M.; Wilkins, C. S.; Fillmore, T. L.; Moore, R. J.; Paša-Tolić, L. Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS. *J. Proteome Res.* **2018**, *17*, 3791–3800.

(33) Baglai, A.; Blokland, M. H.; Mol, H. G.; Gargano, A. F.; van der Wal, S.; Schoenmakers, P. J. Enhancing detectability of anabolic-steroid residues in bovine urine by actively modulated online comprehensive two-dimensional liquid chromatography – high-resolution mass spectrometry. *Anal. Chim. Acta* **2018**, *1013*, 87–97.

- (34) Pirok, B. W. J.; den Uijl, M. J.; Moro, G.; Berbers, S. V. J.; Croes, C. J. M.; van Bommel, M. R.; Schoenmakers, P. J. Characterization of dye extracts from historical cultural-heritage objects using state-of-the-art comprehensive two-dimensional liquid chromatography and mass spectrometry with active modulation and optimized shifting gradients. *Anal. Chem.* **2019**, *91*, 3062–3069.
- (35) Dwivedi, R. C.; Spicer, V.; Harder, M.; Antonovici, M.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhn, O. V. Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal. Chem.* **2008**, *80*, 7036–7042.
- (36) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsvel, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **2014**, *10*, 757.
- (37) Moggridge, S.; Sorensen, P. H.; Morin, G. B.; Hughes, C. S. Extending the compatibility of the SP3 paramagnetic bead processing approach for proteomics. *J. Proteome Res.* **2018**, *17*, 1730–1740.
- (38) Spivak, M.; Weston, J.; Bottou, L.; Käll, L.; Noble, W. S. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. *J. Proteome Res.* **2009**, *8*, 3737–3745.
- (39) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968.
- (40) Morgenstern, D.; Barzilay, R.; Levin, Y. RawBeans: A simple, vendor-independent, raw-data quality-control tool. *J. Proteome Res.* **2021**, *20*, 2098–2104.
- (41) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Anal. Chem.* **2005**, *77*, 6426–6434.
- (42) Yeung, D.; Mizero, B.; Gussakovsky, D.; Klaassen, N.; Lao, Y.; Spicer, V.; Krokhn, O. V. Separation Orthogonality in Liquid Chromatography–Mass Spectrometry for Proteomic Applications: Comparison of 16 Different Two-Dimensional Combinations. *Anal. Chem.* **2020**, *92*, 3904–3912.
- (43) Tu, C.; Li, J.; Shen, S.; Sheng, Q.; Shyr, Y.; Qu, J. Performance investigation of proteomic identification by HCD/CID fragmentations in combination with high/low-resolution detectors on a tribrid, high-field orbitrap instrument. *PLoS One* **2016**, *11*, No. e0160160.
- (44) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548–1510.
- (45) Held, N. A.; Webb, E. A.; McIlvin, M. R.; Hutchins, D. A.; Cohen, N. R.; Moran, D. M.; Kunde, K.; Lohan, M. C.; Mahaffey, C.; Woodward, E. M. S.; Saito, M. A. Co-occurrence of Fe and P stress in natural populations of the marine diazotroph *Trichodesmium*. *Biogeosciences* **2020**, *17*, 2537–2551.
- (46) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. (PubMed ID: 30395289)