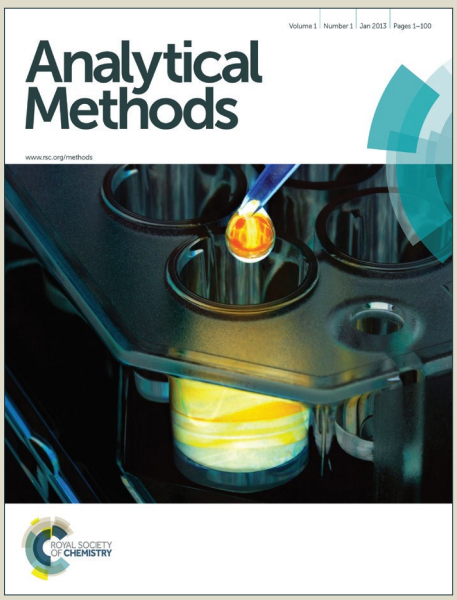
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

ARTICLE

An Optimization of LC-MS/MS Workflow for Deep Proteome Profiling on Orbitrap Fusion

Litong Nie,[‡,a] Mingrui Zhu,[‡,a] Shengnan Sun,[‡,a] Linhui Zhai,[a] Zhixiang Wu,[b] Lili Qian[a] and Minjia Tan*[a]

The development of high-resolution mass spectrometer (MS) has greatly advanced the system-wide proteomic profiling and protein post-translational modification (PTM) studies. However, in contrast to current genomic sequencing technologies, huge time cost and laborious workload are major bottlenecks of current MS-based proteomic approach for large-scale in-depth proteome sequencing of biological samples. Here we present a stepwise optimization of MS parameters and off-line reverse phase HPLC fractionation method in the first tribrid MS platform—Orbitrap Fusion, which integrates quadrupole, ultra high field Orbitrap and linear ion trap mass analyzers. With off-line high pH separation, we identified more than 5,000 proteins using a regular short reverse phase C18 column (10 cm×75 μm, 3 μm particle size) in a single one-hour LC-MS run and 8,493 proteins with 6 orders of magnitude of dynamic range in only 10-hour MS running time. Our study provided a fast, cost-efficient and amenable method for deep proteomic analysis and quantification of large-scale biological samples. Significantly, this strategy would facilitate the proteomic disease biomarker discovery.

Introduction

Proteomics aims to define the complete set of proteins expressed in cells, tissues and organs.[1] One of the great challenges for proteomic analysis is the high complexity and high dynamic range of proteins in cells and tissues. With the rapid advancement in mass resolution, accuracy and dynamic range over the past decade, mass spectrometry has ever-increasingly been the pivotal technology for systematic characterization and quantification of proteins in biological samples.[2, 3] Several different types of MS instruments and their combinations are currently used in proteomic experiments, including quadrupole (Q)/ time of flight (TOF),[4] Fourier transform ion cyclotron resonance (FT-ICR) [5] and Orbitrap mass spectrometers.[6] Q-TOF hybrid mass spectrometer used a quadrupole as mass filter for precursor ion isolation and TOF for high-resolution (40,000) and high-accuracy (<5 ppm) mass analysis. In FT-ICR MS, ions are analyzed in superconducting magnetic fields, which can achieve ultra-high resolution (1,000,000) and mass accuracy (<1 ppm). However, its vast size, formidable cost, high magnetic field and inconvenient operation largely limit its application in regular proteomic experiment. In contrast to TOF and FT-ICR, Orbitrap is a relatively new type of high-resolution bench-top mass spectrometer, in which ions are trapped in electrostatic fields and the mass-to-charge of ions can then be measured from the frequency of ion oscillations.[6, 7] Orbitrap analyzer can achieve high resolving power (up to 450,000 resolution at *m/z* 200) and excellent mass accuracy (1~5 ppm) without magnet field. In 2005, the hybrid MS which first combined an ion trap and an Orbitrap became a powerful tool in proteomic analysis.[8] This combination allows the coupling of fast scan rate and sensitivity of an ion trap for full MS/MS analysis with high resolution and high mass accuracy of Orbitrap for survey full MS scan. Since the introduction of Orbitrap, this technology has greatly advanced the system-wide proteomic profiling and protein post-translational modification (PTM) studies. [9]

With the advancements in MS and efforts from the proteomics community, the first draft of human proteome has been recently completed.[10, 11] In these studies, 17,294 protein-coding genes have been identified which covered approximately 84% of total protein-coding genes in humans. This is an important beginning step toward deciphering and understanding the complete human proteome. Nevertheless, large-scale human proteomic studies are still far behind the human genomic studies, such as the human genome project

*Corresponding authors.
[a] The Chemical Proteomics Center and State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, P.R. China.
E-mail: mjtan@simm.ac.cn
Tel: +86-21-50800172
[b] Department of Pediatric Surgery, Xinhua Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, P.R. China
‡These authors contributed equally to this work.
† (ESI) available: [details of any supplementary information available should be included here]. See DOI: xxxxxxx

(completed in 2003) and the cancer-centric TCGA studies.[12, 13] The main reason is that proteome is much more complex and dynamic than genome in both time and space. One bottleneck for in-depth proteome analysis of the biological samples is the huge time and labor cost in a typical MS-based proteomic experiment, which greatly hinders the large-scale and deep proteome analysis of human biological samples. Currently reported MS-based methods for deep proteome analysis (>8,000 human proteins identified in a given sample) usually need long nano-spray HPLC columns (>50 cm) and long MS machine running time (several days or more for one sample run).[14-20] By using a 50-cm nano-HPLC column coupled to LTQ-Orbitrap Velos MS, Mann et al. identified 5,376 proteins in HEK293 cell lysate in a triplicate run in about one day of MS measure time without prefractionation.[17] In a cell line based deep proteome profiling study, Mann and his coworkers identified 10,255 proteins in HeLa cells taking up to 12 days using the same type of MS instrument.[15] By taking advantage of the high speed and high sensitivity of the quadrupole Orbitrap hybrid MS (Q Exactive), Qin and his coworkers identified more than 8,000 proteins in 12 hours of MS running time in combination with high pH HPLC prefractionation.[14] In a very recent study, 10,544 proteins were identified in a post-mortem brain sample of Alzheimer's disease with about 4 days of MS running time using a 150-cm reverse phase column coupled to Q Exactive MS.[16] Though deep proteome coverage has been achieved with these methods, long instrument running time, laborious workload and complicated instrumentation (such as the expensive ultra-high pressure LC system) limit their application in large-scale proteome analysis. Therefore, a faster and more convenient method for deep proteome analysis of large-scale samples is urgently needed.

Recently, the first Orbitrap tribrid mass spectrometer model Orbitrap Fusion was marketed, which combines quadrupole, ultra high field Orbitrap and linear ion trap mass analyzers, and is capable of versatile modes for sample analyses.[21] In this study, we reported a fast and optimized approach for deep proteome analysis using Orbitrap Fusion by a stepwise optimization of its instrument parameters and off-line reverse phase (RP) HPLC fractionation method. With the optimized method and off-line high pH separation, we identified more than 5,000 proteins using an in-house packed 10-cm C18 RP column in a single one-hour LC-MS run and 8,493 proteins with 6 orders of magnitude of dynamic range in only 10-hour MS running time in HeLa cells. We also identified more than 4,500 phosphopeptides in one-hour LC-MS run after optimization. Our study provided a fast, convenient and cost-efficient method for deep proteomic analysis and quantification of large-scale biological samples. Therefore, our study may be highly instructive for proteomic sample analysis using Orbitrap Fusion, particularly for large-scale sample analysis. Significantly, this strategy would be promising for the identification of proteomic disease biomarkers.

Materials and Methods

Chemicals and materials

BCA Protein assay kit was obtained from Beyotime (Shanghai, China). Modified sequencing-grade trypsin was purchased from Promega (Fitchburg, MI, USA). MS grade water, acetonitrile and other chemicals were from Thermo Fisher Scientific (Waltham, MA, USA). DMEM medium and FBS were from Life technology (San Diego, CA, USA). SILAC-DMEM medium was from Silantes (München, Germany). All-in-one phosphoprotease inhibitor was from Solarbio (Beijing, China).

Ethics Statement

All the animal experiments were carried out in strict conformity with the guidelines approved by the Institutional Animal Care and Use Committee of the Shanghai Institute of Materia Medica, Chinese Academy of Sciences.

Cell culture

HeLa cells were cultured in high glucose DMEM medium supplemented with 10% FBS and 1% Penicillin. For SILAC (stable isotope labeling by amino acids in cell culture) labeling, cells were cultured in SILAC-DMEM medium. Light labeled cells were cultured in DMEM medium containing $^{12}C_6$-lysine and $^{12}C_6{}^{14}N_4$-argine and heavy labeled cells were in medium containing $^{13}C_6$-lysine and $^{13}C_6{}^{15}N_4$-argine. Cells were harvested for further analysis when labeling efficiency was higher than 95%.[22]

Protein extraction and digestion

HeLa cells were harvested by centrifugation at 1,500 rpm for 5 min. The cells were washed twice with ice-cold PBS, then resuspended in lysis buffer (8 M urea, 50 mM NH_4HCO_3) and lysed on ice for 30 min. Protein concentration was measured by BCA assay. For SILAC labeled cells, proteins from heavy and light labeled cells were mixed equally and digested in solution as described previously.[23] Briefly, the protein mixture was digested by trypsin at an enzyme to protein mass ratio of 1:50 (w/w) overnight at 37 °C. The peptides were reduced with 5 mM dithiothreitol (DTT) for 30 min at 56 °C and then alkylated with 15 mM iodoacetamide (IAA) for 30 min at room temperature in darkness. The reaction was quenched with 30 mM cysteine. Additional trypsin at an enzyme to protein mass ratio of 1:100 (w/w) was added to digest the sample completely. The resulting peptides were dried under vacuum and desalted by C18 Sep-Pak column (Waters Corp., Milford, MA).

For the mouse liver tissue, it was quickly dissected into pieces on ice and homogenized gently. The homogenate was filtered through 70-μM cell strainer and then centrifuged at 1,000 g for 5 min. The cells were washed twice with ice-cold PBS, then resuspended in lysis buffer (8 M urea, 50 mM NH_4HCO_3, and 100 U/mL phosphoprotease) and lysed on ice for 30 min. Then the dissolved proteins were digested with trypsin via the filter-aided sample preparation (FASP) procedure using 30K microcon centrifugal filter units (Millipore Corp., Billerica, MA).[24] Briefly, the protein solution was added into filter units and centrifuged at 16,000 g for 15 min at room temperature. Filters were washed twice with 100 μL 50 mM NH_4HCO_3 buffer.

Finally, trypsin was added to each filter. The ratio of protein to enzyme was 100:1. Samples were incubated and digested overnight at 37 °C. After digestion, the peptides were collected by centrifugation and then dried.

First dimension high pH RP HPLC fractionation

Peptide separation was performed on an Agilent 1100 Series HPLC system (Agilent Technologies Inc., Folsom, CA) with a BEH-130 column (5 μm, 250 mm×4.6 mm I.D., Waters Corp., Milford, MA) as previously described.[25] The peptides were separated by using a 100-min gradient from 0% to 100% buffer B (98% acetonitrile, 2% H_2O, adjust to pH 10.0 with $NH_3{\cdot}H_2O$) in buffer A (2% acetonitrile, 98% H_2O, adjust to pH 10.0 with $NH_3{\cdot}H_2O$) at a flow rate of 0.7 mL/min. The elution was collected every one minute from 20 min to 100 min (fraction 1 to fraction 80). Then the 80 fractions were dried for further analyses.

Enrichment of phosphopeptides by TiO_2 micro-columns

For phosphoproteomic experiment, TiO_2 micro-column were manually packed as the Larsen group described.[26] Two milligrams of mouse liver peptides were resolved with binding buffer (1M lactic acid in 70% acetonitrile and 5% trifluoroacetic acid (TFA) (vol/vol)) and loaded onto the TiO_2 micro-column. The TiO_2 micro-column was washed six times with binding buffer, once with washing buffer 1 (30% acetonitrile and 0.5% TFA) and twice with washing buffer 2 (80% acetonitrile and 0.5% TFA). After washing, the phosphopeptides were eluted by using 150 μL elution buffer (40% acetonitrile with pH 10.5 adjusted by ammonia solution) via slow centrifugation at ~300 g for 10 min at room temperature. The eluted phosphopeptides were dried for further analysis.

Nano-HPLC-MS/MS analysis

The peptides were dissolved in solvent A (0.1% formic acid in 2% acetonitrile and 98% H_2O), then loaded onto a manually packed reverse phase C18 column (10 cm× 75 μm I.D., packed with C18 resin, 3 μm particle size, 90 Å, Dikma Technologies Inc., Lake Forest, CA) coupled to EASY-nLC 1000 system (Thermo Fisher Scientific, Waltham, MA). Peptides were eluted from 5% to 80% solvent B (0.1% formic acid in 90% acetonitrile and 10% H_2O) in solvent A at a flow rate of 300 nL/min. For 60 min gradient, the condition were as follows: 5-32% B over 50 min, 32-80% over 7 min, and then held at 80% B for 3 min. For 90 min gradient, the condition were as follows: 5-32% B over 80 min, 32-80% over 6 min, and then held at 80% B for 4 min. For 120 min gradient, the condition were set as follows: 5-32% B over 110 min, 32-80% over 6 min, and then held at 80% B for 4 min. The eluted peptides were analyzed by Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Parameters were as follows: For full MS spectra, the scan range was *m/z* 350 ~ 1,300 with a resolution of 120,000 at *m/z* 200. MS/MS acquisition was performed in top speed mode with 3 s cycle time. The resolution was 15,000 at m/z 200.Intensity threshold was 5,000, and maximum injection time was 35 ms. AGC target was set to 7,000, and the isolation window was 1 *m/z*. Ions with charge states 2+, 3+ and 4+ were sequentially fragmented by higher energy collisional dissociation (HCD) with normalized collision energy (NCE) of 32%. The dynamic exclusion duration was set as 60 s.

Analysis of MS data

Raw files for human protein identification were analyzed by Proteome Discoverer (version 1.3) using Mascot search engine (version 2.3.01) with percolator (strict FDR of 0.01 and a relaxed FDR of 0.05) against UniProt Human database (88,817 sequences, updated on 07/12/2014).[27] Methionine oxidation and protein N-terminal acetylation were chosen as variable modifications, and cysteine alkylation by iodoacetamide was chosen as a fixed modification. Mass error for parent ion mass was ±10 ppm with fragment ion as ±0.5 Da. The protease was specified as trypsin with 2 maximum missing cleavages. Peptide identifications were accepted with high confidence, corresponding to less than 1% false discovery rate (FDR).[28] Peptides assigned with a Mascot score lower than 20 were further discarded.[28] On the protein level, PSMs with at least high confidence and strict maximum parsimony principle were applied.

Besides, to stringently analyze the data of our optimized method from off-line two dimensional HPLC separation (10 fractions×60 min), we used PepDistiller[29] to calculate false discovery rate (FDR) on protein level. FDR for protein with target-decoy based strategy was set to be 0.01.

Raw files of phosphorylation data were analyzed by Proteome Discoverer (version 1.3) using Mascot search engine (version 2.3.01) with percolator against UniProt Mouse database (50,128 entries). In addition to the above modifications, phosphorylation of serine, threonine and tyrosine residues (STY) was set as variable modifications. The other parameters were the same as the human protein identification analysis.

Raw files of SILAC samples were analyzed by MaxQuant (version 1.4.1.2)[30] against UniProt Human database (88,817 sequences). Methionine oxidation, protein N-terminal acetylation, $^{13}C_6$-lysine and $^{13}C_6{}^{15}N_4$-argine were chosen as variable modifications and cysteine alkylation by iodoacetamide was chosen as a fixed modification. 20 ppm of mass tolerance was set for the first search. Main mass tolerance of precursor ion was 4.5 ppm, and 20 ppm was allowed for MS/MS search. A minimum peptide length of 7 amino acids was considered for peptide identification. One ratio count was also set for SILAC quantification. False discovery rate thresholds for protein, peptide and modification site were set to 0.01. The reversed versions and common contaminants were filtered for further data analysis.

Bioinformatics analysis

For human protein identification and mouse phosphorylation data, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis[31] were performed using DAVID 6.7 (The Database for Annotation, Visualization and Integrated Discovery) bioinformatics tool[32] with background of the total

Homo sapiens and Mus musculus genome information, respectively.

Result and Discussion

Optimization of instrument parameters on Orbitrap Fusion

Orbitrap Fusion was the first commercialized tribrid mass spectrometer, which combined quadrupole, ultra high field Orbitrap and linear ion trap mass analyzers together. In Orbitrap Fusion, the precursor ions can be isolated in either quadrupole (Q) or ion trap (IT), and the fragment product ions can be analyzed in Orbitrap (OT) or IT. For the first time, the HCD (Higher-energy collisional dissociation) fragment ions can be analyzed in both low-resolution ion trap and high-resolution Orbitrap analyzers.[33] Based on the new structure design, many different analysis modes can be used, such as Q-OT-HCD-IT, Q-OT-CID-IT, Q-OT-HCD-OT and Q-OT-CID-OT. In order to establish the best mode for protein identification, we compared these modes for protein profiling analysis. In a typical proteomic experiment, full MS1 analysis in the high-resolution Orbitrap mass analyzer can greatly benefit the confident identification of proteins in complex samples. Therefore, we chose OT as the mass analyzer for precursor ion analysis. As shown in Table 1, more than 2,800 and 2,600 proteins were identified in Q-OT-HCD-IT and Q-OT-CID-IT modes, respectively. The numbers of peptides and proteins identified under Q-OT-HCD-IT mode were higher than those under CID-IT and the HCD-OT modes (Table 1 and Fig. S1). The highest number of MS/MS spectra was identified in HCD-IT mode, which was most likely due to more fragment ions generated with HCD fragmentation. Consistent with our anticipation, the numbers of identified peptides and proteins decreased about 20% in Q-OT-HCD-OT mode as compared to Q-OT-HCD-IT mode (Fig. S1). This was probably due to the significantly slower acquisition rate by sequential analysis of precursor and fragment ions in only one high resolution mass analyzer (Orbitrap) in contrast to the parallelized MS1 and MS2 data acquisition in different mass analyzers in Q-OT-HCD-IT mode (Table 1). Therefore, we used Q-OT-HCD-IT mode for protein identification.

Next, we optimized the key parameters for MS analysis in Q-OT-HCD-IT mode, including mass resolution, maximum inject time, isolation window, automatic gain control (AGC) target of MS2 and normalized collision energy (Fig. 1A and Fig. S2). All the optimizations were performed in top speed mode, which could maximize the number of high-quality MSn spectrum acquisition within each duty cycle. Orbitrap mass analyzer allows six different resolutions from 15,000 to 450,000 at *m/z* 200 in Orbitrap Fusion. Higher resolution can increase mass accuracy and the sensitivity of low abundance precursor ion detection, but decrease the data acquisition rate. Therefore, we selected the resolution at 60,000, 120,000 and 240,000 at *m/z* 200 to test the best condition for protein identification. The highest number of proteins was identified at the resolution of 120,000 (Fig. 1A), suggesting that it was a good compromise between the detection sensitivity and acquisition speed.

In Q-OT-HCD-IT mode, precursor ions are isolated by the quadrupole mass filter for subsequent MS2 fragmentation. Bigger isolation window will benefit isotopic ion detection but also increase the detection possibility of undesirable contaminated ions from co-eluted species with similar *m/z*. To optimize the isolation widow, we compared the quadrupole isolation widths at 0.7, 1 and 1.5 *m/z*. As shown in Figure 1A, the maximum protein number was identified at 1 *m/z*. In the ion trap type MS analyzer, better ion statistics can be achieved when reasonably more ions are trapped. However, excessive ions can lead to the space charge effects, which will compromise ion trap performance.[34] To overcome this issue, maximum injection time and automatic gain control (AGC) were introduced to maintain the optimum quantity of data acquisition, which are important MS parameters for protein identification. We therefore further evaluated maximum injection time and AGC in proteome analysis. We compared the maximum injection time from 30 to 60 ms and the AGC target from 5,000 to 10,000. The results showed that the highest number of proteins was identified at 35 ms of maximum injection time and 7,000 of AGC (Fig. 1A).

High collision energy would produce many unpredictable fragment ions, while low collision energy may lead to insufficient fragmentation. Both were unfavorable for peptide identification.[35] To get the most appropriate collision energy for peptide identification, we examined a range of normalized collision energy (NCE) values, 28%, 30%, 32% and 35% for HCD fragmentation (Fig. 1A). Our results showed that the number of identified proteins slightly increased when NCE value was set to 32%. When the NCE value increased to 35%, the number of identified proteins and peptides decreased dramatically. After we manually checked the raw data, we found that a significant portion of unpredictable fragment ions was generated at NCE of 35% (Fig. 1).

Top N mode is prevalent in data dependent acquisition method in previous generation of Orbitrap mass spectrometers in proteome profiling analysis. However, in top N mode, different MS analyzers, such as Orbitrap and ion trap, cannot maximize their performance in parallelized ion detection. Top speed mode was therefore introduced in Orbitrap Fusion. With the specific hardware configuration and sophisticated control software improvement, top speed mode can acquire maximum high quality MS data from each cycle. Our data showed that, in top speed mode, the number of acquired MS/MS scans ranged from 1 to 35 in each duty cycle (Fig. 1B), demonstrating the advantage of top speed mode over top N mode for data dependent acquisition. We next optimized the cycle time of top speed analysis for our one-hour HPLC gradient analysis. The highest number of proteins and percentage of identified MS/MS spectra were achieved when top speed was set at 3 s (Fig. 1A).

After instrument optimization, we next evaluated the sensitivity of Orbitrap Fusion for complex sample analysis. Six orders of magnitude of sample loading (from 0.02 to 2000 ng) derived from HeLa whole cell lysate were subjected to

Orbitrap Fusion analysis using a 1 h HPLC gradient, respectively. As shown in Fig. 2, the number of acquired MS/MS spectra and identified proteins increased with higher sample amount injection. We also found that peptide-spectrum matches (PSMs) and protein numbers of 2 ng peptides was 8 folds and 5 folds increased than those of 0.02 ng peptides, respectively. For 2,000 ng sample injection, numbers of PSMs and proteins were increased by 300 and 34 folds than those of 0.02 ng peptides, respectively. Meanwhile, the numbers of acquired MS/MS spectra of 2,000 ng and 2 ng were 400% and 50% more than those of 0.02 ng peptides, respectively. We identified more than 2,500 proteins from 200 ng whole cell lysate and 90% peptides were identified within a mass accuracy of 1 ppm (Fig. S3). This data suggested that our LC-MS/MS method was good for global protein identification for as low as 200 ng whole cell lysate sample.

Evaluation of protein quantification based on SILAC method by Orbitrap Fusion

System-wide quantification of protein abundance change in a given biological or pathological condition plays a critical role in characterization of protein functions. The SILAC (stable isotope labeling by amino acids in cell culture) approach based on high-resolution MS is a powerful method for proteome wide relative quantification.[36] Orbitrap Fusion can reach the resolving power as high as of 450,000 at *m/z* 200, which could benefit the accurate isotopic peak recognition and quantification of SILAC pair ions. However, higher resolution will compromise the data acquisition speed. Therefore, there is a balance between protein identification and quantification accuracy for MS resolution selection in SILAC sample analysis. To optimize the Orbitrap resolution for SILAC experiment, we mixed equal amount of proteins from heavy and light labeled HeLa cells for LC-MS/MS analysis. In a triplicate measurement experiment, our results showed that the SILAC ratio distribution of heavy to light labeled proteins at the resolutions of 120,000 and 240,000 were more centered at the 1:1 ratio. Less than 3% SILAC ratio of the proteins were over 1.5 folds at the resolution of 120,000 and 240,000, meanwhile more than 5% SILAC ratio of the proteins were over 1.5 folds at the resolution of 15,000, 30,000, 60,000, and 45,000, which suggested resolution of 120,000 and 240,000 could achieve better quantitative accuracy (Fig. 3B and Table S1). Our results showed the highest numbers of proteins in both SILAC and non-SILAC labeled samples were identified at the resolution of 120,000 (Fig. 1, 3A and Table S1). These results suggested that the resolution of 120,000 for MS1 analysis was an optimal setting for protein identification and quantification.

Two dimensional HPLC-MS/MS analysis for deep proteome profiling

Mounting evidence showed that the off-line two dimensional HPLC separation was a superior approach in deep proteome mapping due to its simple handling and high reproducibility.[25, 37] To further evaluate our approach for global protein profiling by Orbitrap Fusion, we employed off-line basic RP chromatography fractionation to reduce the complexity of whole-cell tryptic digest (Fig. S4A). After using a concatenation strategy as previously reported,[25] 20 fractions were pooled (fraction 1, 21, 41, 61 were combined, 2, 22, 42 were combined) and subjected to a one-hour gradient nano-HPLC/MS/MS analysis, respectively (Exp 1) (Table 2).

More than 3,900 proteins were identified in each fraction, which were 1,000 proteins more than the number of proteins identified in a previous study with a fast sequencing workflow.[14] In total, 8,922 proteins with about 80,000 peptides were identified within 20 hours of MS running time (Fig. S5, Table S2).

A major bottleneck for deep proteome profiling of large-scale biological samples is that huge amount of MS time is needed for deep proteome coverage. To optimize a fast deep proteome profiling approach using Orbitrap Fusion, we next pooled the off-line HPLC fractionated tryptic peptides into 10 or 5 fractions for LC-MS/MS analysis. We tested 4 methods (Exp 2-5 as shown in Table 2) by using different LC-MS/MS gradients. For 10 fraction samples, 8,229 and 8,632 proteins were identified from the 60 min LC-MS/MS method (~5,100 proteins identified in each fraction) (Exp 2) and the 90 min LC-MS/MS method (~5,400 protein identified in each fraction) (Exp 3), respectively. For the 5 fraction samples, 6,767 proteins were identified from the 120 min LC-MS/MS method (~4800 proteins identified in each fraction) (Exp 4) and 6,599 proteins from the 60 min LC-MS/MS method (~4,600 proteins identified in each fraction) (Exp 5) (Table 2 and Tables S3-S6). Obviously, the total number of identified proteins decreased when the fraction number was smaller (Table 2). Our results showed that 10 fractions with a 60-min HPLC-MS/MS gradient was an optimized approach for fast deep proteome sequencing, which made a tradeoff between operation time and the depth of protein identification. Then PepDistiller[29] were used to control 1% FDR of proteins from 10 fractions with a 60-min HPLC-MS/MS gradient. With this method, 8,493 proteins and 79,134 unique peptides (Table S7) were detected with about 30% of average protein sequence coverage, which was the highest number ever achieved using a 10-cm manually packed RP column in such a short time (Fig. 4A). In our analysis, about 90% proteins were identified by more than one peptide (Fig. S6). Remarkably, the intensity range of identified proteins across all fractions covered nearly 6 orders of magnitude (Fig. 4B). Therefore, our optimized method would benefit deep proteomic study, particularly amenable to large-scale sample sequencing.

To evaluate the depth of proteome coverage in our experiment, we then subjected the identified 8,493proteins to pathway database of the Kyoto Encyclopedia of Genes and Genome (KEGG). Our data covered almost all (198 of 200) KEGG pathways and more than 100 of which had at least 50% coverage, such as RNA degradation (55 of 57 proteins) and citrate cycle (TCA cycle) (26 of 31 proteins) (Fig. 4C and Table S8). Furthermore, 350 transcription factors (23% of all the members in the DBD database[38]) and 168 kinases (33% of all the members in the human kinome database[39]) were

identified, respectively. These results indicated the depth of protein identification of the complex samples.

Phosphoproteomic analysis

Protein post-translational modifications (PTMs) are of central importance for all living organisms. Among them, protein phosphorylation plays key roles in diverse cellular processes, such as transcription, homeostasis, cellular signaling and communication.[40, 41] Mass spectrometry has become a powerful tool for global analysis of phosphoproteome.[9] However, the stoichiometries of many phosphorylation substrates in cells are relatively low. Therefore, mass spectrometry performance, together with phosphopeptide enrichment method, plays a key role in deep phosphoproteomic study. Following our optimized proteomic profiling method, we further investigated the optimized MS parameters for phosphopeptide identification after TiO_2 column enrichment of a mouse liver tissue sample (Fig. S4B). An important feature for protein phosphorylation identification is that phosphopeptides show significant neutral loss of 98 Da due to the labile structure of O-P in MS/MS fragmentation. Therefore, we examined two key parameters, normalized collision energy and max injection time, for phosphopeptide identification. We tested NCE values at 25%, 30%, 32% and 35%. A tiny increase in identified phosphopeptide number was observed at the NCE value of 32% (Fig. 5). We further observed that the best result was achieved at the maximum injection time of 35 ms (Fig. 5). Briefly, we reasoned that 32% HCD collision energy with 35 ms max injection time could afford a considerable gain for phosphoproteomic analysis. With this optimized MS parameters, a total of 4,515 phosphopeptides were identified in a one-hour gradient LC/MS analysis (Table S9), which was faster than a recent study without phosphopeptide prefractionation.[42]

We performed KEGG pathway analysis on the phosphopeptide data by using DAVID 6.7 bioinformatics tool to evaluate the quality of our phosphoproteomic data. Most phosphorylation related pathways were covered, such as mTOR, MAPK, and ErbB signaling pathways, and oxidative phosphorylation (Table S10).

Conclusion

With very fast MS/MS scan rate, high resolution and versatile combination of MS analyzers, Orbitrap Fusion offers unprecedented capacity for deep proteome analysis. In our study, we showed that the Q-OT-HCD-IT mode achieved the best performance for proteome sequencing, since it combines the advantages of the rich peptide fragment information by HCD and very fast acquisition rate in IT. Using optimized instrument parameters in Q-OT-HCD-IT mode on Orbitrap Fusion, more than 2,500 proteins were identified from only 200 ng whole-cell lysate, and about 90% peptides were identified within a mass accuracy of 1 ppm. Our results provided a practical basis for the analysis of precious and scarce samples. It is noteworthy that the resolution of 120,000 not only provided high MS2 scan rate but also achieved more accurate protein identification and quantification. In combination with off-line reverse phase (RP) HPLC fractionation method, we achieved highly efficient and fast analysis of the deep proteome in HeLa cell line. We identified 8,493 proteins for human cell line with six orders of magnitude of dynamic range in only 10 hours of MS running time. To our best knowledge, this was the highest protein number ever achieved using a 10-cm manually packed RP column with 3 μm-particle size resin in such a short time. Finally, we investigated the application of the tribrid MS system into phosphoproteomic analysis. More than 4,500 phosphopeptides were identified in a single one-hour LC-MS gradient. Our results suggest that Orbitrap Fusion can reach high proteome coverage in very short time, and is a robust tool for phosphorylation analysis. In addition, Orbitrap Fusion can efficiently achieve higher analysis throughput relative to the previous Orbitrap models in deep quantitative proteome coverage mainly due to its increased scan speed, although the intrinsic analytical sensitivity of Orbitrap Fusion is similar to that of the previous generation of Orbitrap mass spectrometers. The previous Orbitrap models will require more extensive offline peptide fractions, longer chromatographic gradient, and more carefully chosen Orbitrap settings to possibly achieve the same depth of proteome coverage as Orbitrap Fusion does. Therefore, we provided a cost-efficient and amenable method with no need of complicated instrumentation (such as ultra-high pressure LC system) for rapid in-depth proteomic analysis. It is noteworthy that this method can be easily applied into the analysis of almost all kinds of samples, such as cell lines, animal or plant samples, and human pathological tissues. This method is particularly amenable to large-scale sample analysis. With ultra-high dynamic range, our results especially benefit the discovery of predictive and diagnostic biomarkers and the potential therapeutic targets for clinical applications.

Acknowledgments

This work is supported by the Natural Science Foundation of China (No. 31370814), the National Science and Technology Major Project “Key New Drug Creation and Manufacturing Program” of China (2014ZX09507-002), the National Basic Research Program of China (973 Program) (No. 2014CBA02004), Shanghai Municipal Science and Technology Commission (No. 14DZ2261100 and No.15410723100).

References

1. W. P. Blackstock and M. P. Weir, *Trends in biotechnology*, 1999, **17**, 121-127.
2. R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198-207.
3. T. C. Walther and M. Mann, *The Journal of cell biology*, 2010, **190**, 491-500.

4. A. V. Loboda, A. N. Krutchinsky, M. Bromirski, W. Ens and K. G. Standing, *Rapid communications in mass spectrometry : RCM*, 2000, **14**, 1047-1057.
5. W. Haas, B. K. Faherty, S. A. Gerber, J. E. Elias, S. A. Beausoleil, C. E. Bakalarski, X. Li, J. Villen and S. P. Gygi, *Molecular & Cellular Proteomics*, 2006, **5**, 1326-1337.
6. Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman and R. Graham Cooks, *Journal of mass spectrometry : JMS*, 2005, **40**, 430-443.
7. A. Makarov, *Analytical chemistry*, 2000, **72**, 1156-1162.
8. A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning, *Analytical chemistry*, 2006, **78**, 2113-2120.
9. K. Sharma, R. C. D'Souza, S. Tyanova, C. Schaab, J. R. Wisniewski, J. Cox and M. Mann, *Cell reports*, 2014, **8**, 1583-1594.
10. M. S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda and A. Pandey, *Nature*, 2014, **509**, 575-581.
11. N. Cancer Genome Atlas Research, *Nature*, 2014, **513**, 202-209.
12. B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M. C. Chambers, L. J. Zimmerman, K. F. Shaddox, S. Kim, S. R. Davies, S. Wang, P. Wang, C. R. Kinsinger, R. C. Rivers, H. Rodriguez, R. R. Townsend, M. J. Ellis, S. A. Carr, D. L. Tabb, R. J. Coffey, R. J. Slebos, D. C. Liebler and C. Nci, *Nature*, 2014, **513**, 382-387.
13. N. Cancer Genome Atlas Research, *Nature*, 2014, **511**, 543-550.
14. C. Ding, J. Jiang, J. Wei, W. Liu, W. Zhang, M. Liu, T. Fu, T. Lu, L. Song, W. Ying, C. Chang, Y. Zhang, J. Ma, L. Wei, A. Malovannaya, L. Jia, B. Zhen, Y. Wang, F. He, X. Qian and J. Qin, *Molecular & cellular proteomics : MCP*, 2013, **12**, 2370-2380.
15. N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo and M. Mann, *Molecular systems biology*, 2011, **7**, 548.
16. H. Wang, Y. Yang, Y. Li, B. Bai, X. Wang, H. Tan, T. Liu, T. G. Beach, J. Peng and Z. Wu, *Journal of proteome research*, 2015, **14**, 829-838.
17. S. S. Thakur, T. Geiger, B. Chatterjee, P. Bandilla, F. Frohlich, J. Cox and M. Mann, *Molecular & Cellular Proteomics*, 2011, **10**.
18. T. Kocher, P. Pichler, R. Swart and K. Mechtler, *Nature protocols*, 2012, **7**, 882-890.
19. T. Kocher, R. Swart and K. Mechtler, *Analytical chemistry*, 2011, **83**, 2699-2704.
20. X. Yin, X. Liu, Y. Zhang, G. Yan, F. Wang, H. Lu, H. Shen and P. Yang, *Proteomics*, 2014, **14**, 2008-2016.
21. M. W. Senko, P. M. Remes, J. D. Canterbury, R. Mathur, Q. Y. Song, S. M. Eliuk, C. Mullen, L. Earley, M. Hardman, J. D. Blethrow, H. Bui, A. Specht, O. Lange, E. Denisov, A. Makarov, S. Horning and V. Zabrouskov, *Analytical chemistry*, 2013, **85**, 11710-11714.
22. T. Geiger, J. R. Wisniewski, J. Cox, S. Zanivan, M. Kruger, Y. Ishihama and M. Mann, *Nature protocols*, 2011, **6**, 147-157.
23. S. C. Kim, Y. Chen, S. Mirza, Y. D. Xu, J. Lee, P. S. Liu and Y. M. Zhao, *Journal of proteome research*, 2006, **5**, 3446-3452.
24. J. R. Wisniewski, A. Zougman, N. Nagaraj and M. Mann, *Nature methods*, 2009, **6**, 359-362.
25. Y. Wang, F. Yang, M. A. Gritsenko, Y. Wang, T. Clauss, T. Liu, Y. Shen, M. E. Monroe, D. Lopez-Ferrer, T. Reno, R. J. Moore, R. L. Klemke, D. G. Camp, 2nd and R. D. Smith, *Proteomics*, 2011, **11**, 2019-2026.
26. T. E. Thingholm, T. J. Jorgensen, O. N. Jensen and M. R. Larsen, *Nature protocols*, 2006, **1**, 1929-1935.
27. D. N. Perkins, D. J. Pappin, D. M. Creasy and J. S. Cottrell, *Electrophoresis*, 1999, **20**, 3551-3567.
28. B. K. Koo, M. Spit, I. Jordens, T. Y. Low, D. E. Stange, M. van de Wetering, J. H. van Es, S. Mohammed, A. J. Heck, M. M. Maurice and H. Clevers, *Nature*, 2012, **488**, 665-669.
29. N. Li, S. Wu, C. Zhang, C. Chang, J. Zhang, J. Ma, L. Li, X. Qian, P. Xu, Y. Zhu and F. He, *Proteomics*, 2012, **12**, 1720-1725.
30. J. Cox and M. Mann, *Nature biotechnology*, 2008, **26**, 1367-1372.
31. M. Kanehisa and S. Goto, *Nucleic acids research*, 2000, **28**, 27-30.
32. W. Huang da, B. T. Sherman and R. A. Lempicki, *Nucleic acids research*, 2009, **37**, 1-13.
33. A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall and J. J. Coon, *Molecular & cellular proteomics : MCP*, 2014, **13**, 339-347.
34. A. H. Payne and G. L. Glish, *Methods in enzymology*, 2005, **402**, 109-148.
35. J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann, *Nature methods*, 2007, **4**, 709-712.
36. M. Mann, *Nature reviews. Molecular cell biology*,

2006, **7**, 952-958.
37. P. Wiczling, M. J. Markuszewski, M. Kaliszan and R. Kaliszan, *Analytical chemistry*, 2005, **77**, 449-458.
38. D. Wilson, V. Charoensawan, S. K. Kummerfeld and S. A. Teichmann, *Nucleic acids research*, 2008, **36**, D88-92.
39. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, **298**, 1912-1934.
40. C. Chang and R. C. Stewart, *Plant physiology*, 1998, **117**, 723-731.
41. J. B. Stock, A. J. Ninfa and A. M. Stock, *Microbiological reviews*, 1989, **53**, 450-490.
42. H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. Heck and S. Mohammed, *Journal of proteome research*, 2013, **12**, 260-271.

Table 1 Comparison of the performance of Orbitrap Fusion between Q-OT-HCD-IT and Q-OT-CID-IT modes.

	HCD	CID	% Difference
Peptides	22520±300	16600±200	39
Unique peptides	16300±100	12450±150	33
Proteins	2834±6	2600±60	11
MS spectra	6100±200	5200±100	19
MS/MS spectra	57670±2000	51600±100	16
% Identified MS/MS	39±1	31±1	20

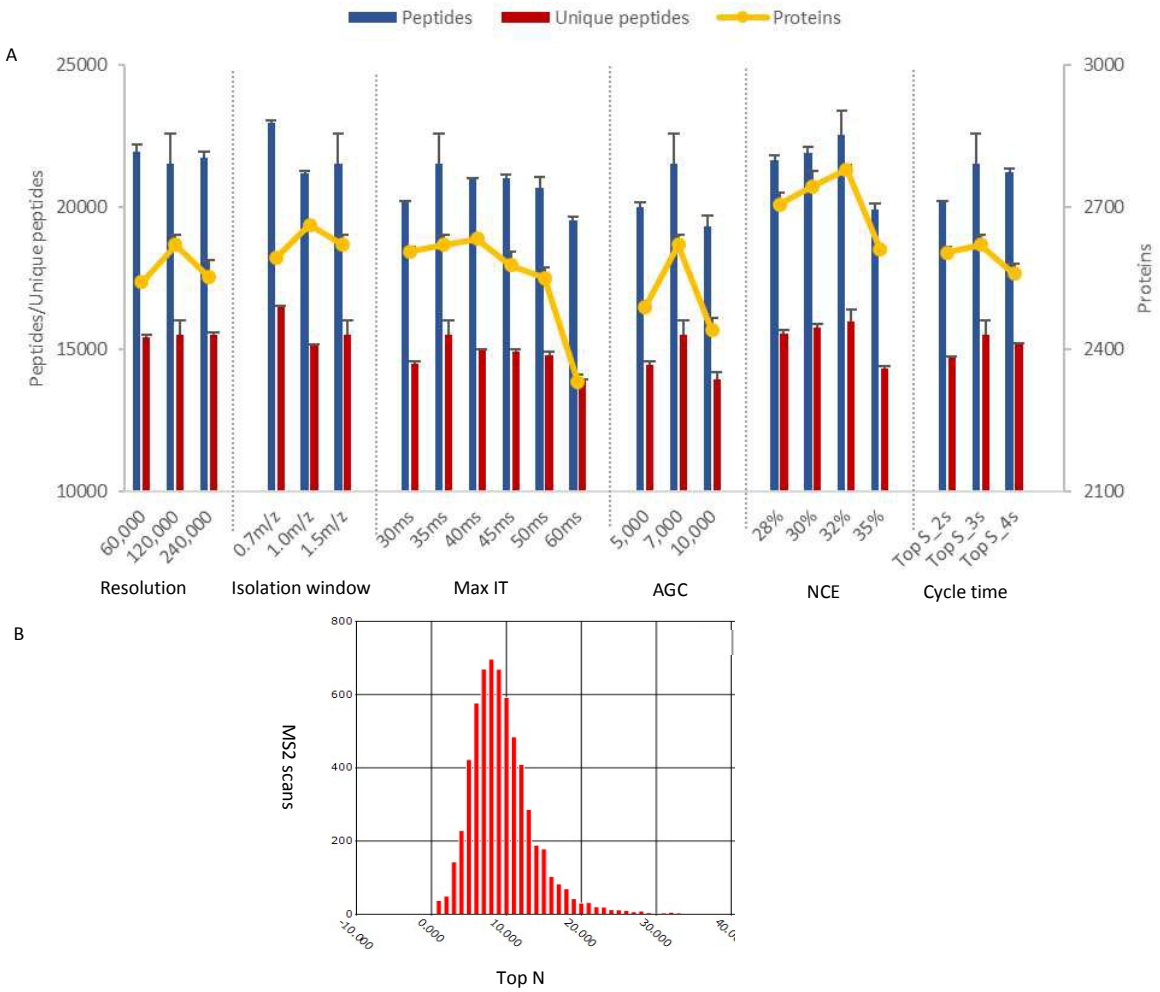


Fig. 1 MS method optimization in Q-OT-HCD-IT mode. (A) Optimization of instrument parameters in Orbitrap-HCD-IT mode. All error bars depict the standard deviation of a replicate measurement. Resolution, isolation window, maximum injection time (Max IT), AGC target, HCD normalized collision energy (NCE) and cycle time of top speed were optimized. General parameters were as described in method of nano-HPLC-MS/MS analysis while optimizing each single parameter. Number of proteins, peptides and unique peptides identified at different parameters were displayed. (B) Top N spacing, number of MS2 scans in between consecutive scan (Visulaized by RawMeat version 2.1).

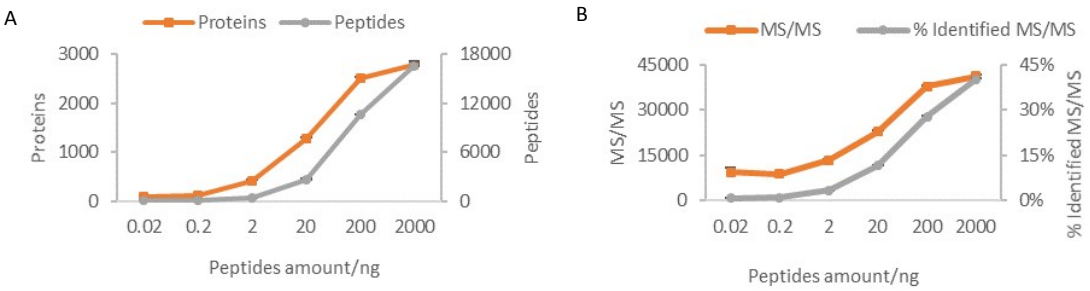


Fig. 2 Evaluation of the sensitivity of Orbitrap Fusion at HCD-IT mode. The sample loading was varied across six orders of magnitude (0.02 ng-2,000 ng). All error bars depict the standard deviation of a triplicate measurement. (A) Number of proteins and peptides identified. (B) Number of MS/MS events and percentage of identified MS/MS.

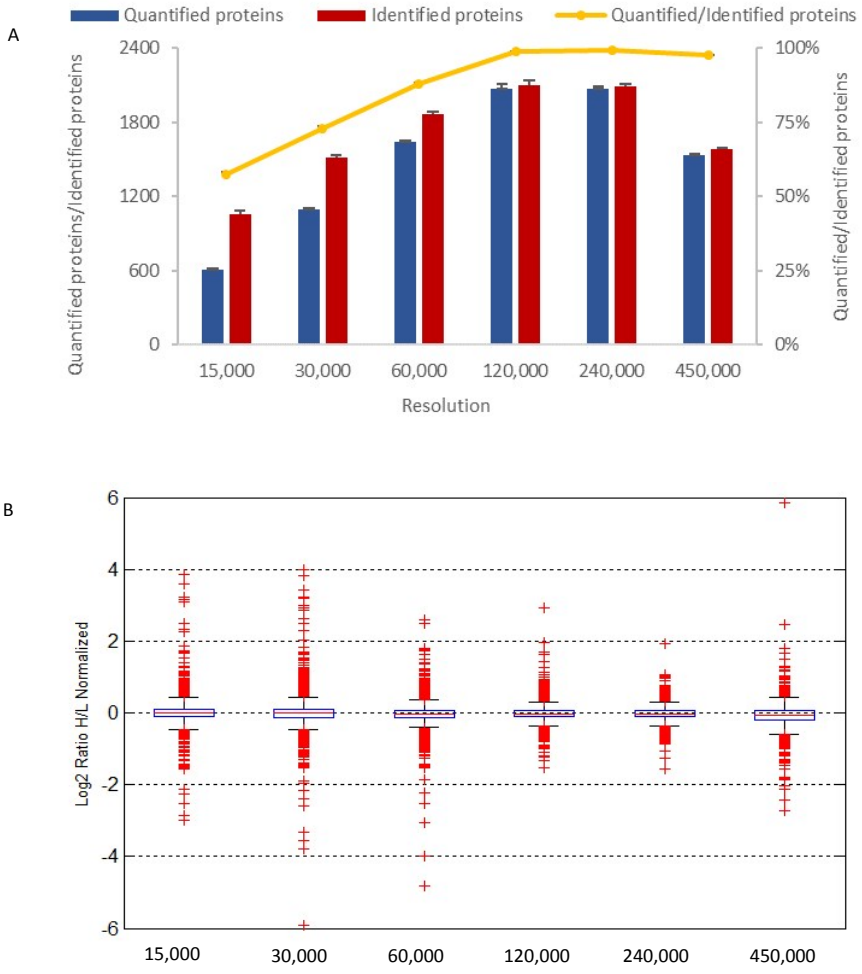


Fig. 3 Evaluation of protein quantification at different resolutions of full MS. All error bars depict the standard deviation of a triplicate measurement. (A) Number of quantitative proteins, identified proteins and the percentage of quantitative proteins identification and (B) distribution of log2 Ratio H/L (heavy to light) at MS1 resolving powers of 15,000, 30,000, 60,000, 120,000, 240,000 and 450,000, at m/z 200.

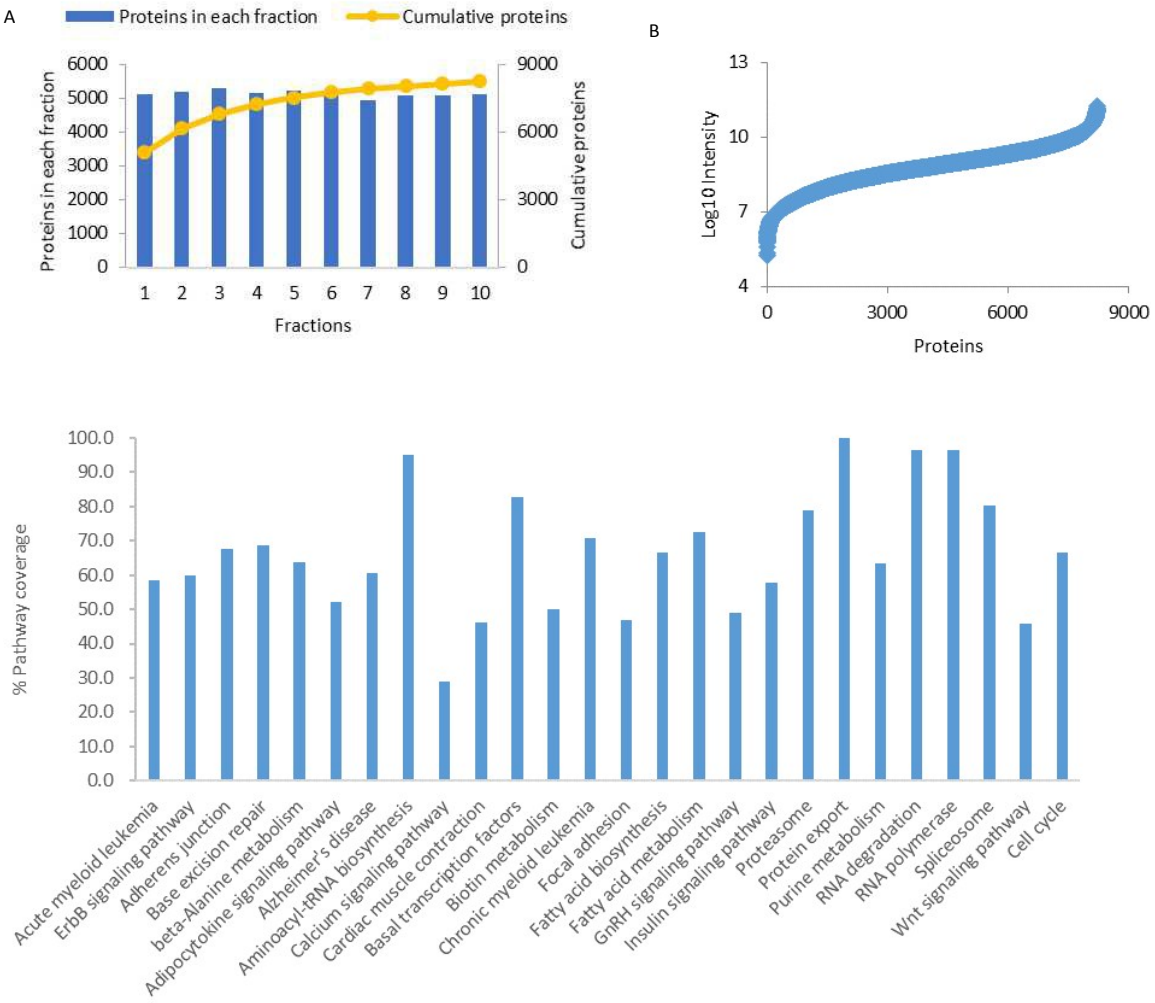


Fig. 4 Deep proteome coverage in two-dimension RP method. (A) Individual protein identified in each fractions and cumulative protein identified across all the 10 fractions over a 60-min LC-MS/MS gradient. (B) Intensity distribution of proteins identified across all the 10 fractions over a 60-min LC-MS/MS gradient. (C) KEGG pathway analysis of the proteins identified across all the 10 fractions over a 60 min gradient.

Table 2 Summary of protein identification results with different LC-MS/MS gradient and combined fractions.

Experiments	Fraction numbers	Time of MS gradient (min)	Total proteins	Proteins in each fraction	MS running time (h)
Exp-1	20	60	8922	3927	20×1
Exp-2	10	60	8229	5120	10×1
Exp-3	10	90	8632	5484	10×1.5
Exp-4	5	120	6767	4866	5×2
Exp-5	5	60	6599	4661	5×1

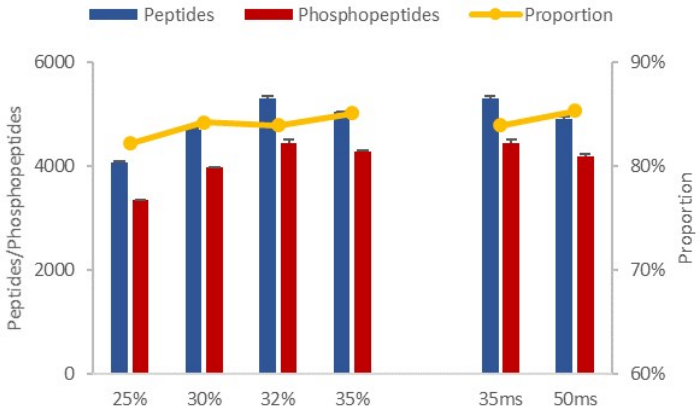


Fig. 5 Method optimization for phosphoproteomic analysis on the Orbitrap Fusion by replicate measurement. Number of peptides, phosphopeptides and proportion of phosphopeptides identified at different values of normalized collision energy and maximum injection time.