Molecular Dynamics Simulation-assisted Ionic Liquid Screening for Deep Coverage Proteome Analysis

Fei Fang,¹‡ Qun Zhao,¹‡ Huiying Chu,²‡ Mingwei Liu,³ Baofeng Zhao,¹ Zhen Liang,¹ Lihua Zhang,¹,* Guohui Li,²,* Liming Wang,⁴ Jun Qin,³,⁵ and Yukui Zhang¹

¹CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Science, National Chromatographic Research and Analysis Center, 457 Zhongshan Road, Dalian 116023, China.

²Laboratory of Molecular Modeling and Design, State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Science, 457 Zhongshan Road, Dalian 116023, China

³State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine; National Center for Protein Sciences (The PHOENIX Center, Beijing), Beijing 102206, China

⁴Division of Hepatobiliary and Pancreatic Surgery, Department of General Surgery, The Second Affiliated Hospital of Dalian Medical University, Dalian 116023, China

⁵Alkek Center for Molecular Discovery, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

* To whom correspondence should be addressed. E-mail: lihuazhang@dicp.ac.cn and ghli@dicp.ac.cn

‡ These authors contributed this work equally

Running Title: Ionic liquid-based deep coverage proteome analysis

Key words: molecular dynamics simulation, ionic liquid, deep coverage proteome analysis
Ionic liquid-based deep coverage proteome analysis

**Abbreviations**

BR - bacteriorhodopsin

C12Im-C1 - 1-dodecyl-3-methylimidazolium chloride

DAVID - Database for Annotation, Visualization and Integrated Discovery

DE - differentially expressed

FDR - false discovery rate

GO - Gene Ontology

GRAVY - grand average of hydropathicity

IAA - iodoacetamide

i-FASP - ionic liquid-based filter-aided sample preparation

IL - ionic liquid

IPA - Ingenuity Pathway Analysis

LINCS - linear constraint solver

MD - molecular dynamics

MEPs - membrane embedded peptides

NH₄HCO₃ - ammonium bicarbonate

PME – particle-mesh Ewald

PSMs - peptide-spectrum matches

RMSD - root-mean-square deviations

RMSF - root mean square fluctuations

TCEP - tris (2-carboxyethyl) phosphine

TMDs - transmembrane domains
Ionic liquid-based deep coverage proteome analysis

ABSTRACT

In-depth coverage of proteomic analysis could enhance our understanding to the mechanism of the protein functions. Unfortunately, many highly hydrophobic proteins and low-abundance proteins, which play critical roles in signaling networks, are easily lost during sample preparation, mainly attributed to the fact that very few extractants can simultaneously satisfy the requirements on strong solubilizing ability to membrane proteins and good enzyme compatibility. Thus, it is urgent to screen out ideal extractant from the huge compound libraries in a fast and effective way. Herein, by investigating the interior mechanism of extractants on the membrane proteins solubilization and trypsin compatibility, a molecular dynamics simulation system was established as complement to the experimental procedure to narrow down the scope of candidates for proteomics analysis. The simulation data shows that the van der Waals interaction between cation group of ionic liquid and membrane protein is the dominant factor in determining protein solubilization. In combination with the experimental data, 1-dodecyl-3-methylimidazolium chloride (C12Im-Cl) is on the shortlist for the suitable candidates from comprehensive aspects. Inspired by the advantages of C12Im-Cl, an ionic liquid-based filter-aided sample preparation (i-FASP) method was developed. Using this strategy, over 3,300 proteins were confidently identified from $10^3$ HeLa cells (~100 ng proteins) in a single run, an improvement of 53% over the conventional FASP method. Then the i-FASP method was further successfully applied to the label-free relative quantitation of human liver cancer and para-carcinoma tissues with obviously improved accuracy, reproducibility and coverage than the commonly used urea-based FASP method. The above results demonstrated that the i-FASP method could be performed as a versatile tool for the in-depth coverage proteomic analysis of biological samples.
INTRODUCTION

With the rapid development of mass spectrometers and techniques, over 16,500 proteins encoded by genes have been elucidated with reliable experimental evidence (1-4). However, many highly hydrophobic and low-abundance proteins are still easily lost during sample preparation, which impedes in-depth proteome analysis (5-7). Therefore, the development of sample preparation methods that are unbiased to such proteins is of great significance to achieve the deep coverage proteome analysis.

As the prerequisite and indispensable steps in ‘shotgun’ proteomics analysis, proteins solubilization and digestion were needed to be both high efficiency and good compatibility. Detergents, such as sodium dodecyl sulfonate (SDS), are commonly used for protein extraction due to their strong solubilizing ability (8). To avoid interference in the subsequent tryptic digestion and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analysis, a filter-aided sample preparation (FASP) method had been developed and successfully applied to the study of various biological issues (9-12). In this strategy, SDS could be removed by a multistep washing procedure involving urea and ammonium bicarbonate (NH₄HCO₃) buffers in the filter, which is not only time consuming but also result in sample loss (13). Additionally, it was found that the dissolving capability of SDS to proteins with extremely high hydrophobicity is still limited.

Recently, to realize the in-depth proteome profiling of low-nanogram samples, Dou et al. (14) developed a microfluidic nanodroplet-based sample processing platform. With this device, the biological samples containing sub-nanogram amounts of protein could be efficiently converted into ready-to-analyze peptides. However, this work is highly professional and special equipment is needed. Furthermore, the commonly used detergents with strong solubilizing ability couldn’t be applied in their device considering the poor compatibility of these detergents with proteases.
Ionic liquid-based deep coverage proteome analysis

In our previous studies, several ionic liquids (ILs) were found to have great performance for membrane proteins analysis (15-20), showing the potential of this underestimated reagents to be extractant candidates. However, testing the limited number of the commercially available ILs with experimental tools is not enough for extractant screening, while it’s unrealistic to synthesize and purify all potential ILs. In the present work, without worrying about much extra time and money for those unavailable extractants, a molecular dynamics simulation-based system, with which the interaction force of any ionic liquids and protein could be calculated, was established. This system could help to narrow down the candidate screening pool of ionic liquids, as well as elucidate the interior mechanism of extractant on membrane proteins solubilization and trypsin compatibility, which is difficult to be solved by experimental method alone. The simulation result showed that the methylimidazolium-based ILs containing alkyl chain length longer than C12 (CnIm-Cl, n =12, 14, 16 …) having stronger interaction force with membrane proteins than other ILs as well as the commonly used extractant, and the van der Waals interaction between cation group of ionic liquid and membrane protein played the dominant role in determining protein solubilization. In combination with the experimental result (19), 1-dodecyl-3-methylimidazolium chloride (C12Im-Cl) was screened out as the ideal extractant. Moreover, we presented a robust and easy-to-use sample preparation method, ionic liquid-based FASP (i-FASP), and successfully applied our method in the qualitative analysis of trace HeLa cells and the quantitative analysis of human liver cancer and para-carcinoma tissues. The experimental results demonstrated the high feasibility by using the established simulation system to design the best extractant in various proteomics fields, and the i-FASP could provide high accuracy, precision and throughput for the deep-coverage proteome analysis, especially for those highly hydrophobic and low-abundance proteins.
Ionic liquid-based deep coverage proteome analysis

EXPERIMENTAL PROCEDURES

Reagents and Materials - The ionic liquids, 1-butyl-3-methylimidazolium thiocyanate (C4Im-SCN) and 1-dodecyl-3-methylimidazolium chloride (C12Im-Cl), were obtained from Shanghai Cheng Jie Chemical (Shanghai, China). Sodium dodecyl sulfonate (SDS), bovine serum albumin (BSA), tris(hydroxymethyl)aminomethane (Tris), tris (2-carboxyethyl) phosphine (TCEP), iodoacetamide (IAA), ethylene diamine tetraacetic acid (EDTA), ethylene glycol tetraacetic acid (EGTA), ammonium bicarbonate (NH₄HCO₃) and protease inhibitor cocktail were purchased from Sigma (St. Louis, MO, USA). Urea was bought from Fluka (Buchs, Germany), and sequencing-grade modified trypsin was from Promega (Madison, WI, USA). Acetonitrile (ACN) and methanol were purchased from Merck (Darmstadt, Germany). The Microcon filtration device with a relative molecular mass cut-off of 10 kDa was from Sartorius AG (Goettingen, Germany). Deionized water was purified using a Milli-Q system from Millipore (Milford, MA, USA). Other chemicals were of analytical grade. Durashell C18 particles (5 μm, 150 Å pore) were obtained from Bona-Agela Technologies (Tianjin, China). Reprosil-Pur C18-AQ particles (3μm, 120 Å pore) were obtained from Dr. Maisch GmbH (Ammerbuch-Entringen, Germany). Fused-silica capillaries (100 μm i.d. / 360 μm o.d.) were purchased from Sino Sumtech (Handan, Hebei, China).

Solubility Measurement for Bacteriorhodopsin – Equal aliquots of bacteriorhodopsin (100 μg) were solubilized in 25 μL of each IL at 1% (m/v), including C4Im-SCN and C12Im-Cl in 50 mM NH₄HCO₃ buffer, as well as C4Im-SCN mixed with 0.5 M NaOH. The samples were respectively sonicated for 10 min in a water bath at room temperature. Afterward, the samples were centrifuged to remove insoluble materials, and the supernatants were collected. The samples were then quantified by the BCA method at 562 nm with BSA as the standard protein. All the measurements were repeated three times in parallel.
Ionic liquid-based deep coverage proteome analysis

Cell Culture - HeLa cells were cultured in MEM-containing 10% fetal bovine serum (v/v) and maintained in a humidified atmosphere of 95% air and 5% CO2 at 37 °C. The adherent cell layer was washed with PBS and then trypsinized with 0.05% trypsin-EDTA solution for 5 min at 37 °C. Then, the cells were centrifuged at 250×g for 5 min to remove trypsin, followed by washing with PBS for three times.

Optimization of Extraction Buffer – For protein extraction, frozen aliquots of 1×10^6 HeLa cells were lysed respectively using 100 μL 0.1 M Tris buffer (pH 7.6) with 1%, 5%, 10% and 15% C12Im-Cl (w/v) in triplicate, to optimize the ionic liquid concentration. For comparison, 100 μL 8 M urea in 0.1 M Tris buffer (pH 8.5) and 4% SDS in 0.1 M Tris buffer (pH 7.6) were used for protein extraction and solubilization, respectively.

For the above different detergents, samples were respectively incubated at 95 °C for 3 min except that for urea, incubated at room temperature for 30 min. Each cell suspension was sonicated on ice for 20 s (pulse on time 5 s, pulse off time 15 s). Subsequently, the cell debris was removed by centrifugation at 16,000×g at 4 °C for 5 min. Finally, BCA protein assay kits from Bio-Rad (Hercules, CA, USA) were used to measure the protein concentration of each supernatant.

Computational Details and Models - To perform molecular dynamics (MD) simulation, the bacteriorhodopsin (PDB code 1M0K (21)) was used to measure the interaction energies of the membrane protein with different chemical reagents, and the protein trypsin (PDB code 4AN7 (22)) was used to test the enzyme activity in different extractants. All molecular dynamics studies were carried out using the GROMACS 5.0.4 software package (23, 24). The all-atom force field expanded by Lopes et al. based on the OPLS-AA frame work was applied to CnIm-Cl, CnIm-Br, CnIm-BF₄, NH₂-CnIm-Cl, OH-CnIm-Cl, CH₂=CH-CnIm-Cl, CN-CnIm-Cl (n=2, 4, 6, 8, 10, 12, 14, 16) (25), while the GROMOS 53a6 force field parameters for SDS were adapted from the work of Tang et
Ionic liquid-based deep coverage proteome analysis

al. (26). The OPLS-AA parameters of methyltributylmethylammonium chloride, 1-dodecyl-3-methyl-pyridinium chloride, and methyltributylmethylphosphonium chloride were obtained from GMXTOP web server (https://Jerkwin.github.io/prog/gmxtop.html). Bacteriorhodopsin was respectively inserted in an 8.5 nm*8.5 nm*11 nm simulation box filled with 1% ionic liquid and 1% SDS (w/v). The systems contained 14 SDS molecules and 11 ionic liquid with TIP3P waters, respectively. The trypsin was inserted in an 8 nm*8 nm*8 nm cubic simulation box filled with two different systems respectively containing 11 SDS molecules and 10 ionic liquid molecules with TIP3P waters. These systems were first constructed using the Packmol package (27).

A periodic boundary condition was applied in all directions of the simulation box. The energy minimization without constraining any solvent atoms was followed by the minimization of the whole system over a few thousand steps to remove conflicting contacts. Subsequently, equilibrations of the systems were performed with gradual release of the position restraints on the protein (1 ns). Finally, for each system, MD simulations were carried out for 50 ns. All the simulations were performed under constant pressure (NPT) conditions. The linear constraint solver (LINCS) method (28) was used to constrain the bonds associated with H atoms, and the integration time step was set to 2 fs. Electrostatic interactions were calculated using the particle-mesh Ewald (PME) algorithm (29, 30). The cut-off distance of the neighbor searching was set to 1.2 nm. To maintain a constant temperature of 300 K, the Nose-Hoover thermostat (31) was applied with a coupling time of 0.1 ps. Therefore, when the NPT conditions were used, the Parinello-Rahman pressure coupling method, which works in conjunction with the Langevin thermostat, was applied at 1.0 bar.

*Sample Preparation by i-FASP* - The protocol for the sample preparation of cells by i-FASP (10^6 cells/experiment) is shown in Figure 3a. For the HeLa cells, an aliquot of 10^6 cells were solubilized
Ionic liquid-based deep coverage proteome analysis

by adding 0.2 mL of lysis buffer (10% C12Im-Cl and 100 mM TCEP dissolved in 100 mM Tris, pH 7.6, 1% protease inhibitor cocktail (v/v)) followed by incubation at 95 °C for 3 min and sonication in a water-ice bath for further solubilization. The cell debris was removed by centrifugation at 16,000×g for 5 min, and 40 μL of the clarified protein extract was transferred to a 10 kDa filter. After centrifugation on the filter at 14,000×g at 20 °C for 15 min, the extracted proteins were retained, diluted with 200 μL of 50 mM NH₄HCO₃ and centrifuged again at 14,000×g at 20 °C for 15 min. Subsequently, 200 μL of 50 mM IAA dissolved in 50 mM NH₄HCO₃ buffer was added, and the samples were incubated in darkness for 20 min. After the filters were washed three times with 200 μL of 50 mM NH₄HCO₃, 100 μL of 10 mM NH₄HCO₃ containing modified trypsin (5 μg) was added, and the sample was digested for overnight at 37 °C. Finally, the tryptic peptides were collected by centrifugation, and 100 μL of water was added to elute the peptide-rich solution. The detailed information for the sample procession of 1,000 HeLa cells and human liver tissue with i-FASP method was included in Supporting Information.

Sample Preparation by FASP - The cells (10⁶ cells/experiment) was prepared using the FASP method following the procedures described previously with minor modifications (9). Briefly, each aliquot of 1×10⁶ cells was solubilized by 0.2 mL of lysis buffer (4% SDS and 100 mM TCEP in 100 mM Tris, pH 7.6, 1% protease inhibitor cocktail (v/v)). Then, the samples were incubated at 95 °C for 3 min with ultrasonication for 20 s (pulse on time 5 s, pulse off time 15 s) to reduce the viscosity of the lysate and sonicated in a water-ice bath for further solubilization. The cell debris was removed by centrifugation at 16,000×g for 5 min, and 40 μL of the clarified protein extract was transferred to a 10 kDa filter. After centrifugation at 14,000×g at 20 °C for 15 min, the protein sample was retained on the filter. The concentrate was diluted in the device with 200 μL of UA buffer (8 M urea dissolved in 50 mM NH₄HCO₃) and centrifuged at 14,000×g at 20 °C for 15 min.
Ionic liquid-based deep coverage proteome analysis

Subsequently, 200 µL of 50 mM IAA dissolved in UA buffer was added to the filter, and the samples were incubated in darkness for 20 min. Next, the filters were washed three times with 200 µL of UA buffer and 200 µL 50 mM NH₄HCO₃, respectively. After that, 5 µg of trypsin in 100 µL of 10 mM NH₄HCO₃ was added to the sample, and the samples were then digested overnight at 37 °C. Finally, the tryptic peptides were collected by centrifugation, and 100 µL of water was added to elute the peptide-rich solution.

The detailed information for the sample processing of 1,000 HeLa cells and human liver tissue with FASP method is included in Supporting Information.

Sample Preparation by in-solution digestion method - The cells (10⁶ cells/experiment) was prepared using the in-solution digestion method following the procedures described previously (15).

In brief, three aliquots containing 1×10⁶ cells were solubilized in 4% C12Im-Cl (w/v, in 50 mM NH₄HCO₃, pH 8.0). Each cell suspension was sonicated on ice for 60 s (pulse on time 10 s, pulse off time 10 s). Subsequently, the cell debris was removed by centrifugation at 16,000 ×g at 4 °C for 5 min, and the supernatant were maintained at 90 °C for 20 min for thermal denaturation.

Subsequently, the samples were cooled, followed by reduction in 10 mM TCEP at 56 °C for 1 h. Afterwards, the cysteines were alkylated in darkness in 25 mM IAA for 30 min at room temperature. The tryptic digestion was diluted by addition of 50 mM NH₄HCO₃ to 5 times and performed with a trypsin/protein ratio (m/m) of 1:25 at 37 °C for 12 h. Tryptic digests were divided into two equal parts for C12Im-Cl removal by precipitation with bis(trifluoromethane)sulfonimide lithium salt and strong cation exchange trap column, respectively, followed by desalting with a homemade C18 trap column.

Nano RPLC-ESI-MS/MS Analysis - The detailed information is included in Supporting Information.
Ionic liquid-based deep coverage proteome analysis

**Database Searching** - The nano-RPLC-ESI-MS/MS raw files obtained with the Q-Exactive mass spectrometer were searched against the Human UniProtKB database (release 2019_8, 20,434 protein entries) using Proteome Discoverer 2.2.0.388 or MaxQuant v1.6.7.0. Peptides were searched using fully tryptic cleavage constraints, and up to two cleavage sites were allowed for tryptic digestion. Cysteine carbamidomethylation was set as a static modification, while N-terminal acetylation and methionine oxidation were set as variable modifications. Peptide identification was based on a search with an initial mass deviation of up to 6 ppm for the precursor ions and an allowed fragment mass deviation of 20 ppm. A false discovery rate (FDR) of 0.01 for both proteins and peptides were required to filter the results.

The nano-RPLC-ESI-MS/MS raw files obtained by the Orbitrap Fusion Lumos mass spectrometer were searched against the Human UniProtKB database (release 2016_7, 42,164 protein entries) with the Proteome Discoverer 2.1.0.81. Peptides were searched using fully tryptic cleavage constraints, and up to two cleavage sites were allowed for tryptic digestion. Cysteine carbamidomethylation was set as a static modification, while N-terminal acetylation and methionine oxidation were set as variable modifications. The mass tolerances for MS and MS/MS were set to 20 ppm and 0.5 Da, respectively.

For the identification quality control, the target-decoy based strategy was applied to keep both peptide and protein level false discovery rates (FDRs) lower than 1%. PepDistiller was used to calculate the probability value (q-value) that keep the FDR measured by the decoy hits lower than 1% for every peptide-spectrum matches (PSMs). The peptides with length longer than 6 amino acids and shorter than 60 amino acids were kept. The q-values of both target and decoy peptide sequences were dynamically increased until the corresponding protein FDR was less than 1% employing the parsimony principle. FDR estimation on the level of peptide spectrum matches was
Ionic liquid-based deep coverage proteome analysis performed using the peptide validator node with filtering for 1% FDR (high confidence filter). For the fractionated sample, all the PSMs in all fractions were combined to achieve a more stringent quality control.

Experimental Design and Statistical Rationale - In each experiment, three biological replicates were compared. For the quantification analysis, a home-made program was used to carry out statistics. Briefly, the proteins quantified in triplicates were remained for followed analysis, and a two-tailed t-test applied with correction for multiple testing (Benjamini-Hochberg). Volcano plots were constructed using the permutation-based approach of Tusher and coworkers (32), to implement an FDR of 0.01. All statistical analysis was performed using GraphPad software (Graphpad Prism, RRID:SCR_002798). Statistical significance between datasets was determined by performing two-tailed, Wilcoxin test. In statistical analysis, \( p > 0.05 \) is indicated as not significant (n.s.), whereas statistically significant values are indicated by asterisks as follows: \( **p < 0.001 \), \( ****p < 0.0001 \).

Bioinformatics analysis - The assignment of protein cellular localization and molecular function was according to the Gene Ontology (GO)-annotation, the UniProtKB database (http://www.uniprot.org/), and the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) (33, 34). TMHMM algorithm (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict transmembrane domains (TMDs) of identified proteins and the transmembrane peptides. The topology of protein was visualized with the Protter software (http://wlab.ethz.ch/protter/). The protein interaction network analysis was performed by STRING software v10.0 (http://www.string-db.org/). Disease network and regulator effect analysis were performed by using Ingenuity Pathway Analysis (IPA, Ingenuity Systems Inc., Redwood City, CA, USA).
Ionic liquid-based deep coverage proteome analysis

RESULTS

Theoretical Simulation of Ionic Liquid-based Membrane Protein Solubilization and Trypsin Biocompatibility – To screen out an ideal extractant for in-depth proteome analysis from a myriad of chemical reagents, it’s crucial to develop an efficient platform to narrow down the scope of candidates before experimental procedure. Molecular dynamics (MD) (35, 36), an overview system of the dynamic evolution, was employed to simulate the interaction between the small molecules and the proteins, as well as elucidate the interior mechanism of extractants on the membrane proteins solubilization and trypsin biocompatibility. As several ionic liquids (ILs) were found to have great performance in our previous proteomics research (15-19), various ILs fine-tuned by the independent selection of cation and anion groups were applied to the simulation system as a pool to prescreen out the extractant candidates. Firstly, an equation was needed to calculate the interaction energy between the small molecules and protein. In simulations system, the forces between atoms and the potential energy of the system are defined by molecular mechanics biomolecular force fields, which are parameterized to fit quantum-mechanical calculations and experimental spectroscopic data. Parameterization involves definition of chemical bonding, atomic angles and dihedral angles, as well as determination of partial atomic charges for calculation of the electrostatic-interaction energies, identification of appropriate van der Waals atomic radii, etc (37). The detailed modeling and parameters extraction performed in our simulation system were illustrated in the Methods Section. When equilibrium is reached, the interaction energy between the protein and solvent is calculated based on the following equation in triplicate

\[ \Delta E_{ele+vdW} = E_{ele+vdW}^{protein+solvent} - E_{ele+vdW}^{protein} - E_{ele+vdW}^{solvent} \]

where \( E_{ele+vdW}^{protein+solvent} \) is the sum of the electrostatic and van der Waals interaction energy of
Ionic liquid-based deep coverage proteome analysis

protein dissolving in solvent, while $E_{\text{ele+vdW}}^{\text{protein}}$ and $E_{\text{ele+vdW}}^{\text{solvent}}$ respectively denote the energy of protein and the solvent. In this work, the systems of bacteriorhodopsin (BR), an integral membrane protein with 7 transmembrane helices, was selected as the model and dissolved in different ILs with four commonly used cation groups, including two branched ILs of methyltributylammonium chloride and methyltributylphosphonium chloride, as well as two linear ILs of 1-dodecyl-3-methylimidazolium chloride (C12Im-Cl) and 1-dodecyl-3-methylpyridinium chloride (C12Py-Cl), were under simulation. As shown in Table S1, the imidazolium and pyridinium-based ILs had the comparable interaction energy when binding with BR, with at least 15% increase obtained in comparison with that between the branched ILs and BR (856.85 vs 734.33 KJ/mol). While the Cl$^-$ contributed negligible portions to the total interaction force, the van der Waals energy between cation groups of ILs and BR accounted for a large proportion in determining its solubilizing ability to membrane protein.

Furthermore, various methylimidazolium chlorides with different lengths of alkyl side chains (CnIm-Cl, n = 2, 4, 6, 8, 10, 12, 14 and 16) were respectively modeled. The result illustrated that the total interaction energy between BR and CnIm-Cl was strengthened as the length of the alkyl side chain increased; however, it reached a plateau at the C12Im-Cl (Table S2). The rule was further proved by evaluating the CnIm-based ILs with other commonly used anion groups like Br$^-$ and BF$_4^-$ (Table S3 and S4). The simulation data showed that there was no obvious difference among the interaction forces of BR and C12Im-based ILs with different anion groups. However, since the water solubility of the CnIm-Br and CnIm-BF$_4$ became dramatically worse as the length of the alkyl side chain increased, the ILs containing Cl$^-$ were preferred to be extractant candidates for proteome analysis.

The ILs bearing functional groups were also subjected to the simulation system. As shown in Table
Ionic liquid-based deep coverage proteome analysis

S5-S8, both the $\pi-\pi$ conjugate interactions of cations with the unsaturated alkyl substituent (CH$_2$=CH-C10Im-Cl and CN-C12Im-Cl) and the hydrogen bond interaction of the cations with -NH$_2$ and -OH substituent contributed little to the ILs on the interaction energy. Interestingly, the rule of the total interaction energy between BR and CnIm-Cl was strengthened as the length of the alkyl side chain increased and reached a plateau when $n=12$ still works for these specific ILs. In combination with the above result and our previous experimental result (19), as well as the poor water solubility of the CnIm-Cl ($n>12$), the C12Im-Cl was on the shortlist for the ideal extractants with strong solubilization capability to membrane proteins. Sodium dodecyl sulfate (SDS), which has been widely considered as one of the strongest detergents with high solubilizing ability, was taken as the reference for the simulation system. It was found that, compared to the tight interaction between C12Im-Cl and BR (Figure 1a), SDS and BR were much farther apart (Figure 1b), which was in accordance with our previous result that SDS provided a relatively weaker dissolving capacity than C12Im-Cl.

To further investigate the performance of the MD simulation for pre-screen of protein extractant. The 1-butyl-3-methylimidazolium thiocyanate (C4Im-SCN), an ionic liquid was found that experimentally work to extract proteins very efficiently (38), was subjected to our MD simulation. The result showed that the interaction force between C4Im-SCN and BR is much weaker than that of C12Im-Cl (Table S9). We also performed experimental procedure to validate our simulation result. As shown in the Figure S1, with bacteriorhodopsin as protein sample, the C4Im-SCN in 50 mM NH$_4$HCO$_3$ almost extracted nothing; even when it was mixed with the alkali solvent (0.5 M NaOH), very few amounts of membrane protein were extracted. Oppositely, a large amount of membrane protein could be extracted and dissolved by C12Im-Cl.

To reveal the mechanism of the high efficiency of C12Im-Cl on membrane proteins solubilization,
Ionic liquid-based deep coverage proteome analysis

the weights of different interaction forces devoted to the total energy were compared. Notably, the van der Waals energy between [CnIm]⁺ and BR was the dominant factor in the total interaction energy for each IL (Table S2-S8). Furthermore, as shown in Figure 1a, the conformation for the BR-C12Im-Cl system demonstrated that it was the alkyl chain, rather than the cation ring of C12Im-Cl, adsorbed onto the surface of BR, and the C12Im-Cl was mainly bound to the exposed hydrophobic amino acid residues of the membrane protein (Figure 1c). Hence, it concluded that C12Im-Cl could efficiently solubilize membrane proteins because of strong hydrophobic interactions.

To evaluate the compatibility of extractants with trypsin activity, the conformations of trypsin dissolved in water, C12Im-Cl and SDS were simulated separately. First, to determine the overall stability of the developed system, the root-mean-square deviations (RMSD) from the starting structure were calculated for all Cα of trypsin in water, C12Im-Cl and SDS systems against simulation time (Figure S2a). As shown in Figure S1a, with the equilibrated protein structures obtained, trypsin dissolved in C12Im-Cl displayed the least motion in all the MD simulations; it rapidly equilibrated within ~1 ns, and the RMSD remained constant throughout the rest of the simulation, with an average value of 1 Å, similar to that of trypsin dissolved in water. By contrast, the equilibration of trypsin in SDS took approximately 3 ns, with the average RMSD of 1.6 Å. Similar to that observed in water (Figure S2b), there was little change in the conformation of trypsin dissolved in C12Im-Cl compared to the native crystal structure (Figure S2c). However, when trypsin was dissolved in SDS, parts of the twisted β-sheets became flexible, and the positions of α helices changed substantially (Figure S2d). The phenomenon was also reflected in the more obvious fluctuation (shown as larger root-mean-square fluctuation value) in the positions of each amino acid of trypsin in SDS system (Figure S2e). As the center of the active site of protease, the change occurred in the
Ionic liquid-based deep coverage proteome analysis

conformation of the catalytic triad would dramatically attenuate the enzyme activity of trypsin. As shown in Figure S1f, unlike the obvious change observed in SDS system, the position and orientation of the catalytic triad in C12Im-Cl system were consistent with those of the native crystal structure, crucial for maintaining the enzyme activity of trypsin. The above results demonstrated the excellent trypsin biocompatibility of C12Im-Cl. This is the first time to elucidate the interior interaction mechanism between ILs and proteins at molecular level. As two independent events, the simulation result was in perfect accord with the experimental data, demonstrating that the molecular dynamics simulation system could be an important complement to experimental procedures for screening out the optimal extractant for proteome analysis.

*Establishment of i-FASP Method* - Based on the superiority of C12Im-Cl regarding membrane protein solubilization and trypsin biocompatibility, C12Im-Cl has been utilized in the analysis of membrane proteins from large amounts of samples in our previous work. However, the C12Im-Cl were removed by using strong cation exchange, inducing the sample loss, which limited its applications in the trace amount samples, as well as the in-depth coverage analysis of proteomics. With an entrenchment at every step, we established an ionic liquid-based filter-aided sample preparation (i-FASP) method to achieve the in-depth proteome analysis in this work (Figure 2a). With $1 \times 10^6$ HeLa cells as the sample, the lysis buffer containing C12Im-Cl was used to disrupt the cells and extract proteins at 95 °C, and we observed that the concentration of extracted protein increased with that of C12Im-Cl (Figure 2b). C12Im-Cl with 10% (w/v) was chosen as the optimal system because it was more efficient than the 4% SDS and 8 M urea in proteins extraction (Figure 2c). Subsequently, *in situ* alkylation, desalting and digestion were integrated on a filter with a cut-off molecular weight of 10 kDa. Unlike the conventional FASP method using SDS as the extractant,
Ionic liquid-based deep coverage proteome analysis

repeated washing steps were not needed to remove the detergent from the C12Im-Cl/trypsin solution because of their excellent compatibility. After on-filter digestion with trypsin in 10 mM NH₄HCO₃, an equal volume of water was used to elute the residual peptides without additional desalting steps for peptide purification. Therefore, the sample preparation time could be shortened by ca. 4h relative to FASP (Figure 2d).

To further demonstrate the potential of i-FASP, we performed the qualitative proteome analysis of 1×10⁶ HeLa cells. First, the retention behavior of residual C12Im-Cl on LC-MS/MS was tested to evaluate the compatibility of C12Im-Cl with MS analysis. As shown in Figure S3a, the residual C12Im-Cl was eluted until the mobile phase composed of 80% ACN (v/v), with no negative effects on the identification of peptides.

Totally, the i-FASP method identified 3,339 proteins, providing 21.6% improvement in comparison with FASP method (Figure 3a and Supplementary Table 1), and 87% (2385/2747) of the proteins identified by FASP method were also detected with i-FASP method. Additionally, a high repeatability of 81.0% in biological triplicates was obtained by the i-FASP method, comparative to that obtained by FASP method of 79.6% (Figure 3b and Figure S3b). Impressively, the percentage of peptides containing missed cleavage sites identified from i-FASP method was 12.7 ± 0.3% (n=3), twice less than that of FASP method (27.5 ± 0.8%). Afterwards, the hydropathic character of the identified peptides was investigated. Compared with the FASP method, the distributions of peptide grand average of hydropathicity (GRAVY) values were noticeably shifted towards hydrophobicity for the solely identified peptides by using i-FASP method (shown in Figure 3c). The scatter plots shown in Figure 3d illustrated that many larger and more basic peptides were exclusively identified by i-FASP. There might be proton transfer reduction caused by the C12Im-Cl adduction, which would be further investigated in our future work.
Ionic liquid-based deep coverage proteome analysis

Additionally, the protein recoveries of the i-FASP and FASP methods were measured by employing 25 μg of BSA as the standard protein, with the produced tryptic digests were quantified by UV detection (214 nm) according to previous work (19). As a result, the recovery of FASP method was 66 ± 1% (n=4), comparative to the published as 50% (39), much lower than that of the i-FASP method as 90 ± 5% (n=4). Additionally, with different starting sample amounts, the i-FASP method afforded recoveries ranging from 80% to 100% (Figure S3c). It could be explained by that the residual C12Im-Cl favors the reduction of the nonspecific adsorption of proteins and peptides. To further investigate the performance of the developed sample preparation method, equal amount of proteins extracted from HeLa cells with C12Im-Cl were treated with i-FASP and in-solution digestion methods, respectively. Thanks to the low sample loss attributed to the in-situ processing, 35% improvement in protein identification was obtained by i-FASP method in comparison with in-solution digestion (Figure S4a and Supplementary Table 2), and 82% (1788/2186) of the proteins identified by in-solution method were also detected with i-FASP method. In comparison with the in-solution method, a higher repeatability in biological triplicates was obtained by the i-FASP method (83.0% vs 78.6%). Expectedly, due to the impurities removal before protease digestion with the ultrafiltration unit, as few as 6.1% of the peptides identified with i-FASP method were predicted with missed cleavage sites, while only 68.3% of the peptides from in-solution digestion method were completely digested (Figure S4b). Additionally, in comparison with high-abundance proteins (>100,000 copies per cell), much fewer low- (500-5,000 copies per cell) and middle-abundance (5,000-100,000 copies per cell) proteins were detected by in-solution method than i-FASP method (Figure S4c).

Furthermore, the long-term compatibility with LC-MS/MS system by the residual C12Im-Cl from the i-FASP method was investigated. As shown in Figure S5, after subjecting the digested peptides
Ionic liquid-based deep coverage proteome analysis pretreated by i-FASP strategy into the same column for more than one month, there’s almost no change for both peak intensity and peptide retention time. After running the samples for two months, the chromatographic peak intensity is almost as high as the 1st run and the retention time shifts only 4-7 min for a 60 min LC gradient. Additionally, there was no significant peak broadening even after three months running (Figure S5b). The above result showed a long-term stability of LC-MS system with the C12Im-Cl resided in i-FASP products.

The above results demonstrated that the i-FASP method allowed unbiased proteome analysis of both hydrophobic and hydrophilic proteins with high recovery contributed from two aspects. The first is the excellent solubilizing ability to hydrophobic proteins of C12Im-Cl at high concentration; the second is that with the introduction of ultrafiltration unit, most of the C12Im-Cl was depleted with NH₄HCO₃ washing, minimizing the sample loss caused by the harsh washing; the third is that since the small amount of C12Im-Cl was enzyme-friendly and MS-compatible, the residual extractant in the sample could help to enable the hydrophobic proteins remaining well spread over the surface of the spin column, ensuring their exposure to enzymatic digestion.

Qualitative Proteome Analysis of Trace Amount Cell Samples - Cell proteomics with relatively few cells is of great interest to the biological and clinical fields (40, 41). Due to tedious sample preparation, it is still difficult to complete deep coverage proteome analysis for trace amounts of samples. Based on its good performance, the i-FASP method was utilized to the analysis of proteins from 1,000 HeLa cells. As shown in Supplementary Table 3, 20,528 ± 226 peptides, corresponding to 3,329 ± 41 proteins (n=3) could be identified within 5 h of sample preparation and 1 h RPLC-MS/MS analysis, obviously improved compared to the number identified by FASP (10,613 ± 310 peptides, corresponding to 2,173 ± 26 proteins, n=3). The abundance of the proteins
Ionic liquid-based deep coverage proteome analysis

identified from $10^3$ HeLa cell sample prepared by $i$-FASP spanned approximately 5.9 orders of magnitude, broader than that by FASP with 4.8 orders (Supplementary Table 4). Additionally, nearly 90% (2410/2683) of the proteins identified with FASP method were covered by the $i$-FASP method. For those proteins both identified by the two methods, 93% of them had higher intensity (Figure 4a) and sequence coverage (Figure 4b) were achieved by $i$-FASP method. The distribution of the cellular components and biological functions for the identified proteins were investigated by using DAVID software. For the proteins identified with $i$-FASP method, 963 of them were assigned as membrane proteins, including 32.1% of them with more than one transmembrane domain (TMD) (Supplementary Table 3). While for the FASP method, only 741 of the identified proteins were assigned as membrane proteins, among which 26.7% were predicted to have ≥2 TMDs (Figure 4c). Due to their high hydrophobicity, it is difficult to extract and identify the membrane embedded peptides (MEPs) by the FASP methods, while many more MEPs were efficiently identified by using $i$-FASP method (Figure 4d). For example, transmembrane protein 14C (GN: TMEM14C), a 4-TM protein, couldn’t be identified by FASP method considering its high percentage of MEPs, while the sequence coverage could be achieved as high as 67% by using $i$-FASP method (Figure 4d).

Moreover, to evaluate the performance of $i$-FASP method to the low-abundance proteins, our results were referred to the copy numbers of proteins identified in previous work (42) (Supplementary Table 5). As shown in Figure 4e, most of the high-abundance proteins (>100,000 copies per cell), such as the cytoskeletal protein vimentin (gene name: VIM) with 20 million copies per cell (43), were successfully identified by both the $i$-FASP and FASP methods. However, for the middle- (5,000-100,000 copies per cell) and low-abundance proteins (500-5,000 copies per cell), only few of them were uniquely identified with FASP method, while at least three times more proteins
Ionic liquid-based deep coverage proteome analysis were uniquely identified by using i-FASP method, including the very low abundant proteins (<500 copies per cell). The cellular component of the proteins solely identified by i-FASP method was analyzed with DAVID software. Most of the low abundant proteins were assigned as cytoplasm, nucleus, plasma membrane, membrane and exosome proteins, while the top subcellular organelles for the high abundant proteins belonging to nucleus, cytoplasm, exosome and mitochondrion (Figure 4e and Supplementary Table 5). The above results demonstrated the superiority of i-FASP for pre-treating highly hydrophobic and low-abundance proteins, beneficial to improving the coverage of proteome analysis.

Quantitative Proteome Analysis of Human Hepatoma Tissue by Label-Free Method- The label-free quantitative method has been widely employed for the analysis of different samples with deep coverage and handy operation. However, it is extremely required the high reproducibility and protein identification sensitivity. Based on its excellent performance in qualitative proteome analysis, the i-FASP method was then employed to the label-free quantitative proteome analysis of human hepatoma tissues. In contrast to cell and other tissue samples, the high lipid content (5-10%) in human liver tissues can seriously affect the subsequent tryptic digestion and LC-MS/MS analysis, especially for the tumor tissues (44). In conventional lipid removal methods, the proteins extracted from the lipid-rich tissues were usually precipitated with acetone, resulting in serious sample loss (45). In our i-FASP method, different organic solvents, including methanol, ethanol and isopropanol, mixed with equal volume of NH₄HCO₃, were respectively used as washing buffers to remove the lipids during C12Im-Cl removal process. And methanol mixed with equal volume of 50 mM NH₄HCO₃ was chosen as the optimal washing buffer based on the good performance of protein identification (Table S10). The cancer and para-carcinoma tissues of human liver samples were
Ionic liquid-based deep coverage proteome analysis treated using i-FASP method. As a comparison, the same amounts of samples were back-to-back treated by using FASP method with 8 M urea as the extractant (46, 47). Overall, an average of 7,176 proteins (corresponded to 56,492 peptides) were identified from liver cancer tissue by i-FASP method, 6% and 16% increase in comparison with the protein and peptide number identified from FASP method (shown in Table S11 and Supplementary Table 6). In addition, the higher Pearson’s correlation coefficient between the replicate experiments was achieved by the i-FASP method, demonstrating the good reproducibility of this method (Figure S6). These results can be attributed to the improved digestion efficiency and operational simplicity of the i-FASP method. The broader coverage and better repeatability obtained with the i-FASP method were prerequisites for achieving high accuracy and precision in the MS/MS-based label-free proteome quantitation method.

The proteins quantified in all three replicates by the i-FASP and modified FASP methods were separately subjected to Benjamini-Hochberg (BH) FDR estimation, and those passed the 1% BH-FDR threshold were retained in the volcano plot (Figure 5a). Totally, 4352 and 3946 proteins were confidently quantified with the i-FASP and m-FASP method, respectively (Supplementary Table 7). To the best of our knowledge, this was the largest quantitative proteomic dataset for single paired tumor and non-tumor tissues of hepatocellular carcinoma. Gene ontology (GO) analysis suggests that while the proteins associated with biological processes such as cell-cell adhesion, mRNA splicing and tRNA export from nucleus are up-regulated, the proteins involved in the biological processes associated with mitochondrial functions such as oxidation-reduction process, assembly of respiratory chain complex I, electron transport and the tricarboxylic acid cycle are down-regulated (Figure 5b). The difference between the functions regulated by the nucleus and mito-
Ionic liquid-based deep coverage proteome analysis

Chondria in cancer cells might be attributed to the mitochondria has its own genomic system independent from the nucleus, and they possess their own transcription, translation, and protein assembly machinery. Considering that the mitochondria electron transport system contributes to the metabolic remodeling in cancer cells by rebalancing the tumor bioenergetics towards glycolysis, the results of our GO analysis advances a possibility that the dysfunction in mitochondria electron transport might be partially responsible for the tumorigenesis of liver cancer.

With further control of the two-fold standard deviation in the accuracy evaluation as the cutoff with significance, 125 and 58 proteins were confidently quantified as differentially expressed (DE) proteins by using the i-FASP and FASP method, respectively. Among them, 30 proteins were quantified with the consistent changes by both methods. We further explored the underlying information of the proteins exclusively quantified by two methods, and the other 28 DE proteins quantified using FASP method were all confidently identified by i-FASP method, but without showing significant changes caused by the poor repeatability in low-abundance proteins (Supplementary Table 6). Expressively, among the DE proteins quantified by i-FASP method, the percentage of membrane proteins and low-abundance proteins, including extracellular exosomes and transcription factors, were as high as 50.8% and 40.8%, respectively (Figure 5c). Additionally, among the proteins exclusively quantified with i-FASP method, 28 of them even could not be confidently identified when using the FASP method due to the limitation of this method in detecting proteins with low abundance and high hydrophobicity (Supplementary Table 7).

The 125 DE proteins quantified by i-FASP were then subjected to disease network with Ingenuity Pathway Analysis (IPA) database, and most of the DE proteins were enriched to the liver-associated cancers (Figure 5d). We further referred the biological functions of the DE proteins quantified by i-FASP to the previous reports (details shown in Supplementary Table 7), and as many as 45.6%
Ionic liquid-based deep coverage proteome analysis of the differential expressed proteins were reported to be associated with liver cancer, such as nestin (NES), and adenylyl cyclase-associated protein 1 (CAP1) (48, 49). Furthermore, 34.4% of the proteins were reported to be associated with other cancers, including macrophage-capping protein (CAPG), which is a as breast cancer biomarker (50), and perilipin-2 (PLIN2), which is a renal cell carcinoma biomarker (51). These proteins are potentially involved in the same oncogene pathway in liver cancer and may be potential biomarkers for multiple cancers. Interestingly, most of the DE proteins, not reported to be associated with tumors, were found closely related to cancer-associated proteins by using STRING software (Figure S7a). The DE proteins were also subjected to the regulator effect analysis. The top two scoring networks were arteriosclerosis (Figure S7b) and inflammation of liver (Figure S5c). The predicted regulators, including C-C chemokine receptor (CCR2), peroxisome proliferator-activated receptor gamma coactivator (PPARGC1A), nuclear receptor (NR1I3) and peroxisomal acyl-coenzyme A oxidase (ACOX1), might be biomarker candidates in liver cancer, but needs more biological validation in the future work.

DISCUSSION
Sample preparation for proteome analysis typically involves the extraction of proteins from biological species and protease digestion. Therefore, the capacity for protein solubilization and protease compatibility of the extractant are critical elements for global proteome analysis. In this work, molecular dynamics simulation system was developed to play subsidiary role to experimental methods, facilitating to narrow down the scope of extractant candidate as well as reveal interior mechanism of extractants on both membrane proteins solubilization and trypsin biocompatibility at the molecular level. C12Im-Cl, a methylimidazolium-based ionic liquid, showed strong interactions with membrane proteins via its alkyl chain, while the conformation of trypsin, especially its
Ionic liquid-based deep coverage proteome analysis

catalytic triad, remained stable in this distinctive system.

Contributed by the merits of C12Im-Cl, an ionic liquid-based sample preparation method, \textit{i-FASP}, was proposed for the in-depth coverage analysis of proteomics, especially for the low-abundance and highly hydrophobic proteins from trace amounts of samples. The \textit{i-FASP} method showed superior protein extraction capacities and excellent compatibility with trypsin digestion, as well as high efficiency and throughput during sample preparation. When used in the qualitative analysis of cells and tissues, \textit{i-FASP} showed great promise in deep-coverage proteome analysis. In combination with quantitative proteome strategies and the excellent performance of \textit{i-FASP} in the analysis of membrane and low-abundance proteins, many differentially expressed proteins were confidently identified and quantified in our study with high accuracy and deep coverage. The results demonstrated the great potential of our \textit{i-FASP} method to provide a rich resource for understanding the mechanisms of diseases and discovering novel biomarkers.

\textbf{Acknowledgements.} This work was supported by the National Key Research and Development Program of China (2017YFA0505003, 2016YFA0501401), National Natural Science Foundation (21775150, 21834006, 91543201), CAS Key Project in Frontier Science (QYZDY-SSW-SLH017) and Innovation program from DICP, CAS (DICP TMSR201601, DICP ZZBS201712).

\textbf{Author contributions:} F.F., Q.Z., G.H.L., L.H.Z. and Y.K.Z. designed research; L.M.W. provided liver tissue sample; F.F., Q.Z. and M.W.L. performed proteome analysis; J.Q. provided analytical tools; F.F., Q.Z. and H.Y.C. performed data analysis; F.F., Q.Z., H.Y.C., B.F.Z., Z.L and L.H.Z. wrote the manuscript; L.H.Z. and Y.K.Z. supervised research.
Ionic liquid-based deep coverage proteome analysis

Data Availability. All mass spectrometry data and search results have been deposited to the MASSIVE (ftp://MSV000083818@massive.ucsd.edu) and project name ‘Molecular Dynamics Simulation-directed Ionic Liquid Screening for Deep Coverage Proteome Analysis’. Reviewers can access the dataset by using ‘feifang’ as ID and ‘6Nm9z#55’ as the password.
Ionic liquid-based deep coverage proteome analysis

REFERENCES

1. Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T. C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., and Pandey, A. (2014) A draft map of the human proteome. *Nature* **509**, 575-581

2. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587

3. Malmstrom, E., Kilsgard, O., Hauri, S., Smeds, E., Herwald, H., Malmstrom, L., and Malmstrom, J. (2016) Large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics. *Nat Commun* **7**, 10261

4. Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., Sorensen, K. D., Hoyer, S., Orntoft, T. F., Andersen, C. L., Nielsen, M. L., and Olsen, J. V. (2017)
Ionic liquid-based deep coverage proteome analysis

An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes.

Cell Syst 4, 587-599

5. Hughes, C. S., Moggridge, S., Muller, T., Sorensen, P. H., Morin, G. B., and Krijgsveld, J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc 14, 68-85

6. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R., 3rd (2013) Protein analysis by shotgun/bottom-up proteomics. Chem Rev 113, 2343-2394

7. Navarro, P., Kuharev, J., Gillet, L. C., Bernhardt, O. M., MacLean, B., Rost, H. L., Tate, S. A., Tsou, C. C., Reiter, L., Distler, U., Rosenberger, G., Perez-Riverol, Y., Nesvizhskii, A. I., Aebersold, R., and Tenzer, S. (2016) A multicenter study benchmarks software tools for label-free proteome quantification. Nat Biotechnol 34, 1130-1136

8. Howard, R. J., and Barnwell, J. W. (1984) The detergent solubility properties of a malarial (Plasmodium knowlesi) variant antigen expressed on the surface of infected erythrocytes. J Cell Biochem 24, 297-306

9. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. Nat Methods 6, 359-362

10. Zielinska, D. F., Gnadt, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell 141, 897-907

11. Casanovas, A., Pinto-Llorente, R., Carrascal, M., and Abian, J. (2017) Large-Scale Filter-Aided Sample Preparation Method for the Analysis of the Ubiquitinome. Anal Chem 89, 3840-3846

12. Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., Li, C., Zhang, L., Qin, G., Zhang, M., Chen, N., Zhang, M., Huang, Y., Zhou, J., Zhao, Y., Liu,
Ionic liquid-based deep coverage proteome analysis

M., Zhu, X., Qiu, Y., Sun, Y., Huang, C., Yan, M., Wang, M., Liu, W., Tian, F., Xu, H., Zhou, J., Wu, Z., Shi, T., Zhu, W., Qin, J., Xie, L., Fan, J., Qian, X., He, F., and Chinese Human Proteome Project, C. (2019) Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 567, 257-261

13. Berger, S. T., Ahmed, S., Muntel, J., Cuevas Polo, N., Bachur, R., Kentsis, A., Steen, J., and Steen, H. (2015) MStern Blotting-High Throughput Polyvinylidene Fluoride (PVDF) Membrane-Based Proteomic Sample Preparation for 96-Well Plates. *Mol Cell Proteomics* 14, 2814-2823

14. Dou, M., Zhu, Y., Liyu, A., Liang, Y., Chen, J., Piehowski, P. D., Xu, K., Zhao, R., Moore, R. J., Atkinson, M. A., Mathews, C. E., Qian, W. J., and Kelly, R. T. (2018) Nanowell-mediated two-dimensional liquid chromatography enables deep proteome profiling of <1000 mammalian cells. *Chem Sci* 9, 6944-6951

15. Li, Y., F. Fang, M. Sun, Q. Zhao, Y. Hu, Z. Sui, Z. Liang, L. Zhang, and Y. Zhang. (2020) Ionic liquid-assisted protein extraction method for plant phosphoproteome analysis. *Talanta* 213, 120848.

16. Taoka, M., Horita, K., Takekiyo, T., Uekita, T., Yoshimura, Y., and Ichimura, T. (2019) An Ionic Liquid-Based Sample Preparation Method for Next-Stage Aggregate Proteomic Analysis. *Anal Chem* 91, 13494-13500

17. Sun, L., Tao, D., Han, B., Ma, J., Zhu, G., Liang, Z., Shan, Y., Zhang, L., and Zhang, Y. (2011) Ionic liquid 1-butyl-3-methyl imidazolium tetrafluoroborate for shotgun membrane proteomics. *Anal Bioanal Chem* 399, 3387-3397

18. Tao, D., Qiao, X., Sun, L., Hou, C., Gao, L., Zhang, L., Shan, Y., Liang, Z., and Zhang, Y. (2011) Development of a highly efficient 2-D system with a serially coupled long column and its application in identification of rat brain integral membrane proteins with ionic liquids-assisted
Ionic liquid-based deep coverage proteome analysis

solubilization and digestion. *J Proteome Res* 10, 732-73.

19. Zhao, Q., Fang, F., Liang, Y., Yuan, H., Yang, K., Wu, Q., Liang, Z., Zhang, L., and Zhang, Y. (2014) 1-Dodecyl-3-methylimidazolium chloride-assisted sample preparation method for efficient integral membrane proteome analysis. *Anal Chem* 86, 7544-7550

20. Zhao, Q., Fang, F., Shan, Y., Sui, Z., Zhao, B., Liang, Z., Zhang, L., and Zhang, Y. (2017) In-Depth Proteome Coverage by Improving Efficiency for Membrane Proteome Analysis. *Anal Chem* 89, 5179-5185

21. Schobert, B., Cupp-Vickery, J., Hornak, V., Smith, S., and Lanyi, J. (2002) Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. *J Mol Biol* 321, 715-726

22. Patil, D. N., Chaudhary, A., Sharma, A. K., Tomar, S., and Kumar, P. (2012) Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. *FEBS J* 279, 4547-4564

23. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26, 1701-1718

24. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845-854

25. Canongia Lopes, J. N., Deschamps, J., and Pádua, A. A. H. (2004) Modeling Ionic Liquids Using a Systematic All-Atom Force Field. *J Phys Chem B* 108, 2038-2047

26. Tang, X., Koenig, P. H., and Larson, R. G. (2014) Molecular dynamics simulations of sodium dodecyl sulfate micelles in water-the effect of the force field. *J Phys Chem B* 118, 3864-3880
Ionic liquid-based deep coverage proteome analysis

27. Martinez, L., Andrade, R., Birgin, E. G., and Martinez, J. M. (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem 30, 2157-2164

28. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A Linear Constraint Solver for Molecular Simulations. J Comput Chem 18, 1463-1472

29. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 98, 10089-10092

30. Wang, H., Gao, X., and Fang, J. (2016) Multiple Staggered Mesh Ewald: Boosting the Accuracy of the Smooth Particle Mesh Ewald Method. J Chem Theory Comput 12, 5596-5608

31. Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A Gen Phys 31, 1695-1697

32. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98, 5116-5121

33. Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat. Protoc 4, 44-57

34. Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1-13

35. Shukla, D., Peck, A., and Pande, V. S. (2016) Conformational heterogeneity of the calmodulin binding interface. Nat Commun 7, 10910

36. Cao, C., Wang, S., Cui, T., Su, X. C., and Chou, J. J. (2017) Ion and inhibitor binding of the double-ring ion selectivity filter of the mitochondrial calcium uniporter. Proc Natl Acad Sci U S A 114, E2846-E2851

37. Nair, P. C. and Miners, J. O. (2014) Molecular dynamics simulations: from structure function
Ionic liquid-based deep coverage proteome analysis

relationships to drug discovery. *In Silico Pharmacol* 2, 4

38. Taoka, M., Horita, K., Takekiyo, T., Uekita, T., Yoshimura, Y. and Ichimura, T. (2019) An Ionic Liquid-Based Sample Preparation Method for Next-Stage Aggregate Proteomic Analysis. *Anal Chem* 91, 13494-13500

39. Erde, J., Loo, R. R., and Loo, J. A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res* 13, 1885-1895

40. Waanders, L. F., Chwalek, K., Monetti, M., Kumar, C., Lammert, E., and Mann, M. (2009) Quantitative proteomic analysis of single pancreatic islets. *Proc Natl Acad Sci U S A* 106, 18902-18907

41. Nagrath, S., Sequist, L. V., Maheswaran, S., Bell, D. W., Irimia, D., Ulkus, L., Smith, M. R., Kwak, E. L., Digumarthy, S., Muzikansky, A., Ryan, P., Balis, U. J., Tompkins, R. G., Haber, D. A., and Toner, M. (2007) Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* 450, 1235-1239

42. Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., Sørensen, K. D., Høyer, S., Ørntoft, T. F., Andersen, C. L., Nielsen, M. L., and Olsen, J. V. (2017) An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* 4, 587-599.e584

43. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* 11, 319-324

44. Reddy, J. K., and Rao, M. S. (2006) Lipid metabolism and liver inflammation. II. Fatty liver disease and fatty acid oxidation. *Am J Physiol Gastrointest Liver Physiol* 290, G852-858
Ionic liquid-based deep coverage proteome analysis

45. Gorg, A., Drews, O., and Weiss, W. (2006) Extraction and solubilization of mouse liver proteins enriched for alkaline proteins. CSH Protoc 2006

46. Luebker, S. A., and Koepsell, S. A. (2016) Optimization of Urea Based Protein Extraction from Formalin-Fixed Paraffin-Embedded Tissue for Shotgun Proteomics. Int J Proteomics 2016, 4324987

47. Bougnoux, A., Mange, A., and Solassol, J. (2012) Proteomic analysis of differently archived breast cancer tissues. Cancer Genom Proteom 9, 93-99

48. Tschaharganeh, D. F., Xue, W., Calvisi, D. F., Evert, M., Michurina, T. V., Dow, L. E., Banito, A., Katz, S. F., Kastenhuber, E. R., Weissmueller, S., Huang, C. H., Lechel, A., Andersen, J. B., Capper, D., Zender, L., Longerich, T., Enikolopov, G., and Lowe, S. W. (2014) p53-dependent Nestin regulation links tumor suppression to cellular plasticity in liver cancer. Cell 158, 579-592

49. Fan, Y. C., Cui, C. C., Zhu, Y. S., Zhang, L., Shi, M., Yu, J. S., Bai, J., and Zheng, J. N. (2016) Overexpression of CAP1 and its significance in tumor cell proliferation, migration and invasion in glioma. Oncol Rep 36, 1619-1625

50. Westbrook, J. A., Cairns, D. A., Peng, J., Speirs, V., Hanby, A. M., Holen, I., Wood, S. L., Ottewell, P. D., Marshall, H., Banks, R. E., Selby, P. J., Coleman, R. E., and Brown, J. E. (2016) CAPG and GIPC1: Breast Cancer Biomarkers for Bone Metastasis Development and Treatment. J Natl Cancer Inst 108, djv360

51. Morrissey, J. J., Mellnick, V. M., Luo, J. Q., Siegel, M. J., Figenshau, R. S., Bhayani, S., and Kharasch, E. D. (2015) Evaluation of Urine Aquaporin-1 and Perilipin-2 Concentrations as Biomarkers to Screen for Renal Cell Carcinoma A Prospective Cohort Study. Jama Oncology 1, 204-212
Figure 1. Conformation of bacteriorhodopsin-extractants simulation system. Snapshots of direct adsorption between the surface of bacteriorhodopsin and (a) cation group of C12Im-Cl or (b) anion group of SDS; and (c) the binding sites between the C12Im-Cl and the surface of bacteriorhodopsin. The bacteriorhodopsin is shown in violet, C12Im-Cl is shown in cyan, and SDS is shown in grey.
Figure 2. Evaluation on performance of i-FASP method. (a) Flowchart of the i-FASP method. (b) The effects of the C12Im-Cl concentration on the protein extraction. (c) A comparison of the
Ionic liquid-based deep coverage proteome analysis

efficiency on protein extraction using three extractants. (d) A comparison of the sample preparation times for the $i$-FASP (solid line) and FASP (dash line) methods.
Figure 3. Qualitative proteome analysis results with $10^6$ HeLa cells. (a) Overlap of the proteins identified by the i-FASP and FASP methods, and (b) overlap of the proteins identified from biological triplicates by the i-FASP method. The distribution of (c) the GRAVY values and (d) molecular weights (Mw) versus isoelectric points (pI) for the i-FASP-specific peptides (diamond), FASP-specific peptides (round) and shared peptides (circle).
Figure 4. Qualitative proteome analysis results for $10^3$ HeLa cells. A comparison of (a) the abundance and (b) sequence coverage of commonly identified proteins from a 1000-cell sample.
Ionic liquid-based deep coverage proteome analysis

prepared using the i-FASP and FASP methods; ****p<0.0001 between i-FASP and FASP methods with two-tailed, Wilcoxon test. The distribution of (c) transmembrane domains (TMDs), (d) membrane-embedded peptides (MEPs) and (e) copy number of the proteins identified from the 1000-cell sample prepared using the i-FASP and FASP methods. Additionally, (d) the topology of the TMEM14C, including the four MEPs (red labeled amino acids) exclusively identified with i-FASP method, is shown in the inset. (e) The distribution of cellular component of the high-, middle- and low-abundance proteins uniquely identified with i-FASP method is shown in the inset.
Ionic liquid-based deep coverage proteome analysis

![Graphs showing proteome analysis results](image-url)
Ionic liquid-based deep coverage proteome analysis

**Figure 5. Label-Free quantitative proteome analysis of human hepatoma tissue.** (a) Volcano plot and (b) gene ontology analysis of the differentially expressed proteins quantified using the i-FASP method. (c) Subcellular location of the 125 differentially expressed proteins identified by the i-FASP method. (d) Number of molecules associated with diseases generated from Ingenuity Pathway Analysis database and our input data of 125 differentially expressed proteins.