Supporting information

Water droplet-in-oil digestion method for single-cell proteomics

Takeshi Masuda\textsuperscript{1,2*}, Yuma Inamori\textsuperscript{2}, Arisu Furukawa\textsuperscript{2}, Maki Yamahiro\textsuperscript{3}, Kazuki Momosaki\textsuperscript{3}, Chih-Hsiang Chang\textsuperscript{4}, Daiki Kobayashi\textsuperscript{4,5}, Hiroto Ohguchi\textsuperscript{6}, Yawara Kawano\textsuperscript{7}, Shingo Ito\textsuperscript{1,2}, Norie Araki\textsuperscript{4}, Shao-En Ong\textsuperscript{8}, Sumio Ohtsuki\textsuperscript{1,2}

Affiliations
\textsuperscript{1} Department of Pharmaceutical Microbiology, Faculty of Life Sciences, Kumamoto University, Kumamoto 862-0973, Japan
\textsuperscript{2} Department of Pharmaceutical Microbiology, Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto 862-0973, Japan
\textsuperscript{3} Department of Pharmaceutical Microbiology, School of Pharmacy, Kumamoto University, Kumamoto 862-0973, Japan
\textsuperscript{4} Department of Tumor Genetics and Biology, Graduate School of Medical Sciences, Kumamoto University, Kumamoto 860-8556, Japan
\textsuperscript{5} Department of Omics Biology, Graduate School of Medical and Dental Sciences, Niigata University, Niigata 951-8510, Japan
\textsuperscript{6} Division of Disease Epigenetics, Institute of Resource Development and Analysis, Kumamoto University, Kumamoto 860-0811, Japan
\textsuperscript{7} Department of Hematology, Rheumatology, and Infectious Diseases, Faculty of Life Sciences, Kumamoto University, Kumamoto 860-8556, Japan
\textsuperscript{8} University of Washington, Department of Pharmacology, Seattle, WA 98195 USA

*Corresponding author: Takeshi Masuda, Ph.D.
Department of Pharmaceutical Microbiology, Faculty of Life Sciences, Kumamoto University, 5-1 Oe-honmachi, Chuo-ku, Kumamoto 862-0973, Japan
TEL: +81-96-371-4329; FAX: +81-96-371-4329; E-mail: tmasuda@kumamoto-u.ac.jp
Table of Content

Supplementary Experimental Section

Figure S1 Comparison of the extraction efficiency with and without sonication and heating

Figure S2 Recovery of proteins and peptides in a water droplet in ethyl acetate

Figure S3 Comparison of the ISD and WinO methods

Figure S4 Comparison of the addition of six types of magnetic beads on the recovery of peptides using the WinO method

Figure S5 Effect of carboxyl-coated magnetic beads on the WinO method efficiency

Figure S6. Workflow of 100-cell and 10000-cell proteomic analysis using 15 multiple myeloma cell lines

Figure S7. Number of peptides and proteins quantified using the ISD or WinO methods

Figure S8 Distribution of %CV of the ISD and WinO methods

Figure S9 Gene ontology analysis using proteins identified by the WinO method

Figure S10 Application of the WinO method to single-cell proteomics

Figure S11 Distribution of the signal-to-noise ratio of the reporter ions

References
Supplementary Experimental Section

Materials
Sodium deoxycholate (SDC), sodium lauryl sarcosinate (SLS), ammonium bicarbonate (AmBic), dithiothreitol (DTT), iodoacetamide (IAA), mass spectrometry grade lysyl endopeptidase (Lys-C), ethyl acetate, acetonitrile, acetic acid, trifluoroacetic acid (TFA), Dulbecco’s modified Eagle medium (DMEM), negative staining kit, and RPMI-1640 medium were purchased from Fujifilm Wako (Osaka, Japan). Triethylammonium bicarbonate (TEAB), SOURCE 30S beads, and SP Sepharose High Performance beads were from Sigma-Aldrich (St. Louis, MO, USA). Modified trypsin was obtained from Promega (Madison, WI, USA). SDC-XC StageTip was purchased from GL Sciences (Tokyo, Japan). Carboxyl-coated Magnosphere beads were obtained from JSR Life Sciences (Tsukuba, Japan). Benzonase nuclease was purchased from Merck Millipore (Burlington, MA, USA). The lithium dodecyl sulfate (LDS) sample buffer, BCA Assay Kit, TMT reagents, High pH Reversed-Phase Peptide Fractionation Kit, and Dynabeads MyOne Carboxylic acid were procured from Thermo Fisher Scientific (San Jose, CA, USA). FG COOH and NH₂ beads were from TAMAGAWASEIKI (Nagano, Japan).

Cell culture and sorting
In this study, 15 multiple myeloma cell lines (H929, KMM-1, KMS-11, KMS-12BM, KMS-12PE, KMS-20, KMS-27, KMS-28BM, KMS-28PE, L363, MM.1S, MOLP8, OPM1, RPMI8226, and U266) were cultured in RPMI-1640 medium supplemented with 10% FCS until they reached 80% confluence. Cells were washed three times with phosphate-buffered saline (PBS), and then 1 or 100 cells were sorted into 96-well plates. For 10000 cells, cells were sorted into 1.5 mL tubes. An SH800S Cell Sorter (Sony, Tokyo, Japan) using a 100-μm chip was used for sorting after the dead and doublet cells were removed. HEK293 cells were cultured in DMEM supplemented with 10% FCS until they reached 80% confluence. Cells were collected and washed three times with PBS.

Examination of protein and peptide retention in water droplet in ethyl acetate
Proteins were extracted from HEK293 cells using an extraction buffer (100 mM Tris-HCl (pH 9.0) containing 12 mM SDC and 12 mM SLS), and protein concentration was quantified using the BCA Assay Kit. To examine protein retention in water droplets in ethyl acetate, 50 μL of solution containing 10 μg of HEK293 proteins was
dropped into 500 μL of ethyl acetate in a 1.5 mL tube, and incubated for 24 h at 25 ºC. The ethyl acetate and water droplets were transferred by pipette tip into a new 1.5 mL tube and evaporated with a centrifuge concentrator. Each fraction was reconstituted with 20 μL of LDS sample buffer and separated by 5-20% SDS-PAGE. To prepare the negative control sample, we performed the same process as described above using the extraction buffer, but without the protein. As the positive control, 10 μg of HEK293 proteins were separated by SDS-PAGE together. The protein bands in the gel were detected by the CBB and negative staining.

To examine the peptide retention in the water droplet in ethyl acetate, the HEK293 peptides prepared by the ISD method were used. To prepare the peptide solution, 10 μg of protein was reduced and alkylated with 10 mM DTT and 50 mM IAA, respectively. The protein solution was then diluted 4-fold with 50 mM AmBic prior to enzymatic digestion. Proteins were digested with 0.5 μg of Lys-C followed by 0.5 μg of trypsin overnight at 37 ºC. For the evaluation, 20 μL of the peptide solution was dropped into 200 μL of ethyl acetate and incubated at 25 ºC for 24 h. The ethyl acetate and water droplets were transferred by pipette tip into a new 1.5 mL tube and dried with the centrifuge concentrator. Each fraction was reconstituted in 50 μL of 50 mM AmBic and subjected to the phase transfer method to remove SDC and SLS. The peptides purified with SDB-XC StageTip were analyzed by nanoLC-MS/MS using TripleTOF 5600 (Sciex, Framingham, MA). As the positive control, 10 μg of HEK293 peptides purified by SDB-XC StageTip were used for nanoLC-MS/MS. To prepare the negative control sample, the same process was performed using the buffer containing 3 mM SDC, 3 mM SLS, 37.5 mM AmBic in 25 mM Tris-HCl (pH 9.0), but without the peptides.

**nanoLC-MS/MS analysis**

Three different MS systems were used in this study. To determine the optimum magnetic beads for the WinO method, a TripleTOF 5600 (Sciex, Framingham, MA) and an UltiMate 3000 RSLCnano (Thermo Fisher Scientific) were used. In this system, 5 μL was injected into the LC system. The LC was performed using an Acclaim PepMap RSLC (75 μm × 25 cm, C18, 2 μm, Thermo Fisher Scientific) at 300 nL/min. The mobile phase consisted of (A) 0.1% formic acid and (B) 0.1% formic acid in acetonitrile. Linear gradients of 2–25% B in 60 min, 50–90% B in 15 min, and 90% B for 5 min were applied, and the spray voltage was 2300 V. For information-dependent acquisition (IDA), the precursor scan range was m/z 300–1250 in 250 ms. The top 20 precursor ions with a charge of +2 to +5 were selected, and the product
ion scan was performed for 50 ms over the range of $m/z$ 100–1600. Sequential window acquisition of all theoretical fragment-ion spectra (SWATH) was performed with the LC method described for IDA using 73 variable windows with 40 ms scan times (Table S2). The product ions were collected in the range of $m/z$ 100–1600 in high-sensitivity mode.

The 100-cell proteomic analysis utilizing the TMT reagents was performed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) and an EASY-nLC™ 1200 system (Thermo Fisher Scientific). Peptides were first loaded onto an Acclaim™ PepMap™ 100 C18 (3 μm, 75 μm ID×20 mm length, P/N 164946 Thermo Fisher Scientific), and separated on a C18 packed emitter column (3 μm, 75 μm I.D. × 150 mm length, Nikkyo Technos, Tokyo, Japan). The injection volume was 5 μL and the flow rate was 300 nL/min. The mobile phase consisted of (A) 0.1% formic acid and (B) 0.1% formic acid in 80% acetonitrile. A multiple-linear gradient elution was performed as follows: 5–35% B in 60 min, 35–100% B in 5 min, and 90% B for 10 min. Multiplex analysis using TMT reagents was performed using a synchronous precursor selection (SPS) MS3 scan in top-speed mode (cycle time = 3 sec). The parameters were as follows: spray voltage, 2300 V; temperature of the ion transfer tube, 250 °C; Orbitrap scan range for precursor ion ($m/z$), 350–1500; resolution for precursor scan, 120,000; ion trap scan range ($m/z$), auto; collision energy for MS2, 35%; collision mode for MS2, CID; collision energy for MS3, 65%; collision mode for MS3, HCD; Orbitrap scan range for MS3 ($m/z$), 100–500; resolution for reporter ion detection, 50,000; maximum injection time, 50 ms for MS1 and MS2 scans and 105 ms for MS3 scan; AGC target, 400,000 for MS1 Orbitrap, 10,000 for MS2 ion trap, and 100,000 for MS3 Orbitrap.

For single-cell proteomics utilizing the TMT reagents, an SPS MS3 scan with a real-time search was performed on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) with the FAIMS Pro interface (Thermo Fisher Scientific) coupled to an EASY-nLC™ 1200 system equipped with an Aurora column (75 μm I.D., 15 cm length, 1.6 μm beads, IonOpticks, Fitzroy, Australia). The mobile phase consisted of (A) 0.1% formic acid and (B) 0.1% formic acid in 80% acetonitrile. LC gradients of 6–20% B in 37 min, 20–30% B in 14 min, 30–40% B in 9 min, 30–90% B in 3 min, and 90% B for 7 min were applied. For the SPS MS3 scan, the following parameters were used: flow rate, 300 nL/min; spray voltage, 2000 V; temperature of ion transfer tube, 275 °C; Orbitrap scan range for precursor ion ($m/z$), 375–1500; resolution for precursor scan, 60,000; FAIMS CV, -50 and -70; ion trap scan range ($m/z$), 200–1200; collision energy for MS2, 30%; collision mode for MS2, CID:
injection time for MS2 (ms), 500; collision energy for MS3, 65%; collision mode for MS3, HCD; Orbitrap scan range for MS3 (m/z), 100–500; reporter ion detection resolution, 50,000; injection time for MS3 (ms), 500; AGC target, 200% (1E5). The parameters for the real-time search were as follows: enzyme, trypsin; maximum search time (ms), 100; SPS mode, true; Xcorr, 1.4; sCn, 0.1; precursor, 10 ppm; precursor range (m/z), 400–1200; precursor exclusion, low 25 ppm and high 25 ppm.

**Data analysis**

The IDA and SWATH data from TripleTOF 5600 were analyzed using ProteinPilot version 4.5 (Sciex) and PeakView version 2.1 (Sciex). The false discovery rate was estimated by searching against a decoy database generated by randomization of the UniProt human reference database. The data obtained from the Orbitrap Fusion Tribrid mass spectrometer were analyzed using MaxQuant 1 ver.1.6.17.0 and the UniProt human reference database was used as the reference. The following parameters were used: type of parameter section, reporter ion MS3; isobaric labels, 10plex TMT or 11plex TMT; enzyme, trypsin/P; variable modifications, oxidation (M) and acetyl (protein N-term); and fixed modifications, carbamidomethyl (C). For the data obtained from the Orbitrap Eclipse Tribrid mass spectrometer, Sequest HT in Proteome Discoverer v.2.5 (Thermo Fisher Scientific) was used with the following parameters: enzyme, trypsin (full); maximum missed cleavage site, 2; dynamic modification, oxidation (M), acetyl (N-terminal of protein), Met-loss (N-terminal of protein), and Met-loss+ acetyl (N-terminal of protein); static modification, carbamidomethyl (C). Peptides meeting the following criteria were used for reporter ion quantification: minimum average reporter S/N, 10; maximum isolation interference, 50; minimum SPS mass matches, 65. The sum of the S/N ratio of reporter ions was used as the protein quantification value. The grand average of hydropathy (GRAVY) scores of proteins and peptides was calculated in accordance with a published procedure. 2. For gene ontology (GO) analysis, DAVID version 6.7 (https://david.ncifcrf.gov/summary.jsp) was used. Uniform manifold approximation and projection (UMAP) was generated using the R script. 3. The t-test was performed using GraphPad Prism 8.4.3 (GraphPad Software, CA) or Excel (Microsoft, MA). One-way ANOVA Tukey’s multiple comparison test was performed using GraphPad Prism 8.4.3. In proteomic data, fold changes ≥ 2 or ≤ 0.5 with p-value < 0.05 were considered significant.

**Data availability**
The MS raw data and result files have been deposited in the ProteomeXchange Consortium (http://www.proteomexchange.org/, PXD029814) via the jPOST partner repository (https://jpostdb.org, JPST001390)\(^4\).
Figure S1 Comparison of the extraction efficiency with and without sonication and heating

The RPMI8226 cell pellet containing 5E5 cells were dissolved in 100 mL of 100 mM Tris-HCl (pH 9.0) containing 125 unit of benzonase, 12 mM SDC, and 12 mM SLS. In one group, proteins were extracted by sonication for 20 min followed by heating at 95 °C for 5 min. In another group, proteins were extracted by leaving samples on the benchtop at 25 °C for 30 min. The extracted protein amounts were quantified by the BCA assay. We used 20 mL of the sample for protein digestion. These processes were performed in triplicate. The digested peptides were analyzed by nanoLC-MS/MS using TripleTOF 5600. Triplicate data were averaged, and error bars show standard deviation. Unfixed cells were used in (A). The protein concentrations extracted from unfixed and fixed cells were compared (B). Fixation of cells were performed by incubation in 4% formaldehyde for 20 min at RT. After fixation, cells were washed twice with PBS before extraction. The extraction was performed by leaving samples on the benchtop at 25 °C for 30 min in 100 mL of 100 mM Tris-HCl (pH 9.0) containing 125 unit of benzonase, 12 mM SDC, and 12 mM SLS. Unpaired t-test was performed using GraphPad Prism 8.4.3.
To evaluate the protein and peptide recovery in a water droplet in ethyl acetate (EtAc), 10 μg HEK293 whole cell lysate or 10 μg digested peptides dissolved in 3 mM SDC, 3 mM SLS, 37.5 mM AmBic in 25 mM Tris-HCl (pH 9.0) were added into EtAc. After 24 h incubation, EtAc and water droplets were collected. The distributions of proteins or peptides in each fraction were confirmed by performing 5-20% SDS-PAGE followed by CBB or negative staining (A) and nanoLC-MS/MS (B), respectively. The peptide peak area was quantified on TripleTOF 5600 using DIA. As a control, protein or peptide solution used for this examination were also analyzed. Unpaired t-test was performed using GraphPad Prism 8.4.3.

Figure S2 Recovery of proteins and peptides in a water droplet in ethyl acetate
Figure S3 Comparison of the ISD and WinO methods

The relative peptide levels between the WinO and ISD methods are shown as a volcano plot (A). Red dots indicate peptides with a significant change ($p < 0.05$, 2-fold or more). The distribution of %CV from the ISD and WinO methods is presented in (B). The %CV for peptide intensities was calculated from triplicate data generated using each method. The relative peptide levels between the WinO and ISD methods in the fully cleaved peptide and the mis-cleaved peptide (C). The peptide recovery was calculated by dividing the peptide intensity obtained from the WinO method by that from the ISD method. Unpaired $t$-test was performed using GraphPad Prism 8.4.3.
Figure S4 Comparison of the addition of six types of magnetic beads on the recovery of peptides using the WinO method

We evaluated the use of amine-coated Dynabeads, methyl sulfonate-coated SOURCE 30S beads, sulfopropyl-coated SP sepharose beads, carboxyl-coated Dynabeads, carboxyl-coated FG beads, or carboxyl-coated Magnosphere beads on the efficacy of the WinO method for protein and peptide preparation. The beads were equilibrated with 50 mM AmBic containing 12 mM SDC and 12 mM SLS. The WinO method was performed in triplicate using 10 ng HEK293 whole cell lysate in the presence of 1.65 mg beads. The peptide peak area was quantified on TripleTOF 5600 using DIA. Each bar indicates the total intensity of the quantified peptides. Error bars show standard deviations. One-way ANOVA and multiple comparisons were performed using the GraphPad Prism 8.4.3. *p*-values showing the significant differences (*p* < 0.05) between samples are shown.
**Figure S5 Effect of carboxyl-coated magnetic beads on the WinO method efficiency**

The proportion of mis-cleaved peptides in the WinO method with and without beads is shown in (A). These proportions were calculated based on peptide level and averaged across triplicate samples. Error bars indicate standard deviation. The correlation between the GRAVY protein score and relative protein levels from the WinO method with to without beads is shown in (B). Correlation of the recovery rate of peptides in the WinO method with beads and the frequency of acidic or basic amino acids (C).
Figure S6. Workflow of 100-cell and 10000-cell proteomic analysis using 15 multiple myeloma cell lines

As starting materials, 10000 or 100 cells from 15 multiple myeloma cell lines were sorted and digested in triplicate or quadruplicate. For the ISD and WinO methods, 10000 or 100 cells were sorted into 1.5 mL tubes and 96-well plates, respectively. Peptides labeled with TMT reagents were combined. For the ISD method, five TMT batches were prepared (Table S1), and an internal standard was added to all batches. The internal standard was prepared by mixing and digesting protein extracts from 15 cell lines. The peptide mixture was separated into nine fractions using the High-pH Reversed-Phase Fractionation Kit (Thermo Fisher Scientific). Half of the prepared samples were analyzed using nanoLC-MS/MS.
Figure S7. Number of peptides and proteins quantified using the ISD or WinO methods
The numbers of quantified peptides and proteins are shown in (A) and (B), respectively. Each bar indicates the average and standard deviations of replicate data.
Figure S8 Gene ontology analysis using proteins identified by the WinO method

Gene ontology analysis with DAVID version 6.7 (https://david.ncifcrf.gov/summary.jsp) was performed using 798 proteins identified from 15 multiple myeloma cell lines by the WinO method.
Figure S9 Comparison of the proteome profiles obtained with the ISD and WinO methods. Distribution of transmembrane domains (TMDs) in proteins identified with either the ISD or WinO method. We identified 835 and 70 TM proteins from the 15 cell lines using the ISD and WinO methods, respectively. The number of TMDs in each protein was determined from the UniProt database (https://www.uniprot.org/).
Figure S10 Application of the WinO method to single-cell proteomics

Numbers of peptides quantified using the ISD or WinO method (A). The graphs plot the average number of quantified peptides and the standard deviation of quadruplicate data. (B) compares the Log10 protein levels obtained by the WinO method between commonly quantified with the ISD method (239 proteins) and the uniquely quantified in the WinO method (247 proteins). Each bar shows the median.
Figure S11 Distribution of the signal-to-noise ratio of the reporter ions

Distribution of the accepted and rejected spectra based on the signal-to-noise ratio (S/N) of the reporter ions detected in carrier channels (A). Using the Proteome Discoverer tool, spectra with average S/N < 10 were rejected. Distribution of proteins identified using the ISD or WinO method based on protein levels (B). The protein levels were calculated as the total S/N of the reporter ions using Proteome Discoverer.

References
(1) Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008, 26, 1367-72.
(2) Kyte, J.; Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982, 157, 105-32.
(3) McInnes, L.; Healy, J.; Melville, J., UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 2020.
(4) Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.;
Goto, S.; Ishihama, Y., jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res* **2017**, *45*, D1107-D1111.