

Revisiting Protein Reversed-Phase Chromatography for Bottom-Up Proteomics

Shunsuke Takagi^{1,2}, Nobuyuki Suzuki², Yasushi Ishihama^{1,3*}

¹Department of Molecular Systems BioAnalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan

²Analytical and Quality Evaluation Research Laboratories, Daiichi Sankyo Co., Ltd., Hiratsuka, Kanagawa 254-0014, Japan

³Laboratory of Clinical and Analytical Chemistry, National Institute of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka 567-0085, Japan.

*Correspondence and requests for materials should be addressed to Y.I. (email: yishihama@pharm.kyoto-u.ac.jp, Phone: +81-75-753-4555, FAX: +81-75-753-4601).

Abstract

We revisited protein reversed-phase chromatography (RP), using state-of-the-art RP columns developed for biopharmaceuticals, such as monoclonal antibodies, in order to evaluate the suitability of this methodology as a prefractionation step for bottom-up proteomics. The protein RP prefractionation (Prot-RP) method was compared with two other widely used prefractionation methods, SDS-PAGE and high-pH peptide RP (Pept-RP) by using cell lysates as samples. The overlap between fractions of Prot-RP was comparable to that of SDS-PAGE, and the protein recovery was approximately 2-fold higher. On the other hand, the overlap between fractions of Prot-RP was slightly larger than that of Pept-RP, but Prot-RP was able to identify more protein termini and more isoform-specific peptides than Pept-RP. Our results indicate that the combination of highly efficient protein prefractionation with modern mass spectrometers is particularly effective for proteoform profiling from cellular samples.

Keywords

Protein prefractionation, Reversed-phase chromatography, Proteoform, Biopharmaceuticals

Introduction

Recent developments in mass spectrometers^{1,2}, data acquisition strategies³⁻⁶, and data analysis pipelines^{7,8} have enabled the identification of more than 10,000 human proteins in single-shot bottom-up proteomics experiments⁹. Nevertheless, prefractionation by multidimensional separation is still important to achieve deep proteome coverage¹⁰⁻¹².

While the bottom-up proteomics workflow utilizes proteolytic peptides as samples for liquid chromatography-tandem mass spectrometry (LC/MS/MS) analysis, prefractionation can be applied to both peptides and proteins. Peptide-level prefractionation is performed by subjecting the digested peptide samples to various separation techniques that are orthogonal to the second dimension, such as strong cation exchange chromatography (SCX)¹³, strong anion exchange chromatography (SAX)¹¹, hydrophilic interaction chromatography (HILIC)¹⁴, electrostatic repulsion-hydrophilic Interaction chromatography (ERLIC)¹⁵, and isoelectric focusing (IEF)¹⁶. Among them, reversed-phase chromatography (RP) using a high-pH mobile phase shows excellent fractionation efficiency due to its high separation efficiency and allows extremely deep proteome profiling, although it is not completely orthogonal to the typical second dimension using a low-pH RP^{10,12,17,18}. In contrast, protein-level prefractionation is applied to intact proteins. Several studies have reported that molecular weight information collected during sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)-based prefractionation can enhance proteoform analysis¹⁹⁻²¹. This approach is also very effective for samples whose proteome consists mainly of a few extremely high-abundance proteins, such as serum^{22,23}, plasma^{24,25}, and co-immunoprecipitated samples containing bait proteins²⁶. Thus, protein prefractionation is an important technique, along with peptide prefractionation.

SDS-PAGE is the gold standard method for protein prefractionation, because SDS is a powerful solubilizing agent for sample pretreatment and the method provides high resolution for complex mixtures^{19,20,27-30}. However, the use of SDS-PAGE as a prefractionation technique for bottom-up proteomics generally requires a laborious in-gel digestion process, most of which is manual³¹, though recent efforts to simplify gel handling have improved the situation³². A method has been developed for prefractionation of samples using the principle of SDS-PAGE to recover proteins in solution (Gel-Eluted Liquid Fraction Entrapment Electrophoresis; GELFrEE) and has successfully been applied to proteomics research³³⁻³⁵. However, GELFrEE method

requires a dedicated instrument to conduct experiments which is no longer supplied from the manufacturer³⁶. Protein prefractionation by high-performance liquid chromatography (HPLC) involves a lower experimental burden because proteins can be automatically fractionated by commercially available fraction collectors and proteins can be recovered in solution. Extensive efforts have been devoted to the study of HPLC separation of intact proteins for mass spectrometry^{37,38}. Various separation modes, such as SCX³⁹, SAX²⁴, hydrophobic interaction chromatography (HIC)⁴⁰, size-exclusion chromatography (SEC)^{41,42}, and RP^{22–24,28,43}, have been utilized for protein prefractionation for bottom-up proteomics. But, since proteins are generally less stable in solution than peptides, they can easily precipitate when sample solutions are subjected to harsh conditions to enable retention of proteins on a column (e.g., extreme pH (SCX, SAX), high salt concentration (HIC), etc.). SEC allows direct injection of buffer solutions commonly used for protein sample preparation, but it has limited resolution and is mostly used for the separation of protein complexes^{41,42}. Furthermore, SEC generally cannot trap analytes on the column head and leads to severe dilution of precious samples. Despite these limitations of protein HPLC prefractionation, RP was once thought to be a promising mode because of its capability to retain and concentrate proteins from most aqueous buffers, along with the potential for high resolution. However, although protein RP was utilized in the early days of proteomics⁴³, it did not meet the proteomics community's expectations, and applications of protein RP to bottom-up proteomics have been quite limited since the 2000s. Furthermore, protein RP has rarely been included in comparative studies of sample prefractionation methods^{29,30}.

RP columns packed with silica macroporous particles or poly(styrene-divinylbenzene) copolymer (SDB) monoliths have been developed to improve recovery and resolution, thereby enhancing RP prefractionation of proteins^{22,44}. However, silica macroporous columns require high column temperatures to achieve sufficient separation and recovery²², even though it is well known that proteins can be degraded and/or lost inside columns under such conditions⁴⁵. On the other hand, SDB-based columns can be operated at relatively low temperatures but cannot achieve similar performance to silica-based columns⁴⁶. Therefore, to make protein RP a practical tool, it is necessary to develop a method that can achieve high separation performance under milder conditions.

With the recent advances in biopharmaceuticals, which exhibit high drug efficacy while being less likely to show side effects^{47,48}, the demand for detailed structural characterization of therapeutic proteins has increased, and many manufacturers have

launched new-generation RP columns optimized for such applications⁴⁹. Many of these columns are characterized by short alkyl chains, wide pore size (≥ 45 nm), and superficially porous silica particles, and they perform better at lower temperatures than older-generation silica-based columns⁴⁹. In this study, we revisited protein RP prefractionation using new-generation protein RP columns in order to evaluate its applicability for bottom-up proteomics, and we compared our developed method with two widely used alternative prefractionation methods (SDS-PAGE and high-pH peptide RP). Our results highlight the potential of state-of-the-art protein RP columns for protein prefractionation, especially for proteoform analysis,⁵⁰ when employed in combination with modern mass spectrometers.

Materials and Methods

Materials

Triethylammonium bicarbonate buffer pH 8.5 (TEAB), bovine serum albumin, bovine β -lactoglobulin A, and bovine carbonic anhydrase were from Merck KGaA (Darmstadt, Germany). K562 cell lysate and modified trypsin were obtained from Promega (Madison, WI). Phosphate-buffered saline (PBS) was purchased from Takara Bio (Shiga, Japan). Humanized IgG1k monoclonal antibody (NIST mAb) was from National Institute of Standards and Technology (Gaithersburg, MD). Other reagents were obtained from Fujifilm Wako (Osaka, Japan).

Column screening and bridging study

Column screenings were performed with an ACQUITY UPLC H-Class (Waters, Milford, MA). The columns summarized in Table S1 were employed for the investigation. The column diameter and length were 2.1 mm and 50 mm, respectively, in all cases. Nine standard proteins, listed in Table S2, were dissolved in PBS at 0.5 mg/mL each and the resulting solution was used as a sample (2 μ L/injection). Water and acetonitrile (MeCN) both containing 0.1% trifluoroacetic acid (TFA) were used as mobile phases A and B, respectively. The gradient was 20% B to 50% B in 10 min, with a constant flow rate of 0.4 mL/min. The column temperature was maintained at 60 °C during the analysis and proteins were detected by measuring the ultraviolet (UV) absorbance at 214 nm. Triplicate analyses were conducted for each column.

Bridging study between 2.1 and 4.6 mm i.d. BioResolve RP mAb columns (Waters) was performed with an Agilent 1260 Infinity (Agilent Technologies, Santa Clara, CA). The length of the column was 50 mm for both diameter columns. Eight standard proteins, listed in Table S2 except for bovine serum albumin (BSA), were dissolved in PBS at 0.125 mg/mL each and 16 μ L and 80 μ L were injected to 2.1 and 4.6 mm i.d. columns, respectively (16 and 80 μ g/injection in total). Operating conditions were kept same with the column screening experiments except for the flow rate (2.1 mm i.d.: 0.4 mL/min, 4.6 mm i.d.: 2.0 mL/min). Triplicate analyses were conducted for each internal diameter.

Evaluation of loading capacity

Loading capacity of the BioResolve RP mAb column was evaluated using an Agilent 1260 Infinity. The diameter and length of the column were 4.6 mm and 50 mm, respectively. Eight standard proteins, listed in Table S2 except for bovine serum albumin (BSA), were dissolved in PBS at various concentration and injected on column to achieve 20 to 160 μ g/injection (80 μ L/injection). Operating conditions were kept same with the column screening experiments except for the flow rate (2.0 mL/min). Triplicate analyses were conducted for each injection amount.

Protein prefractionation by RP and subsequent digestion

An Agilent 1260 Infinity equipped with an automated fraction collector was employed for protein RP prefractionation (Prot-RP). Cell lysate was reduced (10 mmol/L dithiothreitol, 37 °C, 30 min) and alkylated (50 mmol/L iodoacetamide, 37 °C, 30 min in the dark) prior to the prefractionation. The treated cell lysate was diluted to 1 mg/mL with 100 mmol/L TEAB containing 8 mol/L urea and an 80 μ g aliquot of proteins was injected into the column. The sample compartment was kept at 5 °C to minimize side reactions while waiting for injections. Separation was conducted on a BioResolve RP mAb column (4.6 $\times$ 150 mm, Waters, catalog number: 186008955). The BioResolve RP mAb column is packed with wide-pore superficially porous particles that have been extensively coated with a polyphenyl group on the silica base material to suppress secondary interactions caused by silanol groups (2.7 μ m particle size, 45 nm pore size)⁴⁹. The same mobile phases as used for the column screening were employed. The gradient was 10% B to 30% B in 4 min, 30% B to 55% B in 18 min, and then 55% B to 95% B in 3 min, with a constant flow rate of 2.0 mL/min. The column temperature and the detection setting were kept the same as those employed in column screening. The fractions were collected

every 0.9 min for 27 min, resulting in 30 aliquots. Each three adjacent fractions were combined to give 10 fractions, which were subjected to the following protein digestion procedures.

The collected proteins were vacuum-evaporated and reconstituted in 200 μ L of 100 mmol/L TEAB containing 8 mol/L urea. The reconstituted proteins were incubated with Lys-C for 3 hours (0.32 μ g/fraction), diluted 5-fold with 50 mmol/L NH_4HCO_3 , and further digested with trypsin overnight (37 $^\circ\text{C}$, 0.32 μ g/fraction). After digestion, the solutions were acidified with TFA and the peptides were desalted on a StageTip packed with poly(styrene-divinylbenzene) copolymer (GL Sciences, Tokyo, Japan)⁵¹. Overall experiments were conducted in triplicate.

Protein prefractionation by SDS-PAGE and in-gel digestion

The cell lysate was reduced and alkylated prior to SDS-PAGE separation in the same manner as described for the sample for Prot-RP. Treated cell lysate was mixed with NuPAGE LDS sample buffer (4x) (Thermo Fisher Scientific, Waltham, MA) and water to achieve 1x sample buffer with 0.5 mg/mL protein, and then heated at 70 $^\circ\text{C}$ for 10 min. SDS-PAGE was performed according to the manufacturer's instructions. Briefly, an 80 μ g aliquot of proteins was divided and loaded onto 8 wells of NuPAGE Bis-Tris mini gels (4 to 12%, 1 mm thickness, 10 wells, Thermo Fisher Scientific) and then separated at 200 V for 30 min. The running buffer was 1x NuPAGE MES SDS running buffer (Thermo Fisher Scientific) and gels were stained with Coomassie G-250 (SimplyBlue SafeStain, Thermo Fisher Scientific). The 8 gel lanes containing proteins were sliced horizontally into 10 equal sized pieces (approximately 8 mm height) to ensure the use of all the loaded proteins (80 μ g), and each piece was further chopped for the following in-gel digestion procedure.

In-gel digestion was carried out as previously reported with slight modifications³¹. Chopped gels were de-stained in 50% 25 mmol/L NH_4HCO_3 /50% MeCN until the gel became completely clear. Gel pieces were then dehydrated with pure MeCN and air-dried for 10 to 15 min. Lys-C and trypsin were added and the gels were left to swell in the enzyme solution (0.5 μ g/fraction each). Digestion was carried out overnight at 37 $^\circ\text{C}$. Solutions remaining in the tube were collected prior to peptide extraction. Peptides were first extracted from the gels into 1% formic acid (FA), followed by 75% MeCN/1% FA. The extracts were combined, completely dried in vacuo, and reconstituted with 4%

MeCN/0.1%TFA. The reconstituted peptide solutions were desalted by StageTips as described in the Prot-RP section. Overall experiments were conducted in triplicate.

Peptide prefractionation by high-pH RP

The cell lysate was digested to generate peptides prior to the high-pH peptide RP prefractionation (Pept-RP). Sample pretreatment (reduction and alkylation) and digestion was performed in the same manner as described in the Prot-RP section, while the enzyme-to-substrate ratio was 1:50. Separation was conducted by an ACQUITY UPLC BEH C18 column (1.7 μ m, 2.1 $\times$ 150 mm, Waters) on the same instrument used for Prot-RP, with 80 μ g of peptides per injection. The amount of peptides were calculated based on the quantity of proteins in order to align the amount of starting material (cell lysate) used in the protein prefractionation methods. The mobile phases were, A: 2.5 mmol/L NH_4HCO_3 and B: 90% MeCN/10% 25 mmol/L NH_4HCO_3 . The gradient was 5% B to 39% B in 37 min, 39% B to 67% B in 3 min, and then 67% B to 78% B in 2 min, with a constant flow rate of 0.2 mL/min. The column temperature was 40 $^\circ\text{C}$ and peptides were detected by measuring the UV absorbance at 214 nm. Fractions were collected for every 4.5 min for 45 min, resulting in 10 fractions, and no further desalting was performed. The digestion of proteins was conducted once and prefractionation of digests was conducted in triplicate.

LC/MS analysis

Fractions collected from the three methods (Prot-RP, SDS-PAGE, and Pept-RP) were completely dried in a centrifugal evaporator and reconstituted with 160 μ L of 4% MeCN/0.1%TFA without adjusting the peptide concentration to prepare the sample solutions.

LC/MS analyses were performed by an Orbitrap Fusion Lumos (Thermo Fisher Scientific) equipped with an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). Separation was performed using an EASY-Spray C18 column (2 μ m, 0.075 $\times$ 150 mm, Thermo Fisher Scientific) by directly injecting the sample solutions (5 μ L). The column temperature was maintained at 50 $^\circ\text{C}$ during the analysis. The mobile phases were, solution A: 0.1% FA, solution B: 80% MeCN/0.1% FA. The gradient was 5% B to 40% B in 60 min, with a constant flow rate of 500 nL/min. The voltage applied to the spray emitter was 2.2 kV.

The MS was operated in the data-dependent acquisition (DDA) mode with a cycle time of 2 s at “top speed” mode. Survey scans were acquired by the Orbitrap with the following parameters: m/z range of 375 to 1500 Th, resolution of 120,000, and automatic gain control (AGC) of 4.0×10^5 . For MS², ions isolated by quadrupole (1.6 Th window) were fragmented by higher-energy collisional dissociation (collision energy: 27%) and detected by the Orbitrap. Parameters for MS² were set as follows: AGC of 5.0×10^4 , maximum injection time of 54 ms, and resolution of 30,000. The dynamic exclusion time was set to 60 s. All sample solutions were measured once, resulting in 30 raw data files per prefractionation method (10 fractions, $n = 3$ preparation).

Raw data processing

The acquired raw data were grouped for each fractionation method (30 raw data files per method; 10 fractions, $n=3$), and simultaneously searched against the UniProtKB release 2023_03 human database (20,349 sequences) and the common contaminants database by the Sequest HT engine implemented in Proteome Discoverer ver. 2.2.0.388 (Thermo Fisher Scientific). The canonical sequence database was used unless otherwise noted. Proteins identified from the common contaminants database, such as keratins, were excluded from the analysis. Search parameters were set as: precursor mass tolerance of 5 ppm, product ion mass tolerance of 0.05 Th, trypsin enzyme, minimum peptide length of 6, allowing up to 2 missed cleavages. Carbamidomethylation of Cys was set as a static modification. Oxidation of Met and acetylation, Met-loss, and Met-loss + acetylation of protein N-terminus were set as variable modifications. The false discovery rate (FDR) of the peptide spectral match (PSM) was determined by a target-decoy strategy using a reversed-sequence decoy database and controlled by Percolator. The threshold for the FDR to filter PSM was set at a q -value of 0.01. In this study, rather than increasing the accuracy of identifying individual proteins, it was considered important to be able to compare results impartially regardless of the target of prefractionation (proteins or peptides). Therefore, to ensure this strategy, the minimum number of peptides required for protein identification was set to 1. Label-free quantification (LFQ) using the peak area of the extracted ion chromatogram was performed by Proteome Discoverer, enabling the match between runs function. Protein abundances were obtained by summing the quantitative values for unique and razor peptides belonging to the corresponding proteins. The match between runs were only performed to obtain the quantitative values of unidentified peptides, and peptides (and

269 proteins identified from those peptides) detected only in the match between runs were
270 not included in the number of identifications.

Results and Discussion

Column screening for protein RP prefractionation

In this study, we focused on new-generation RP columns targeting biopharmaceuticals, with the aim of developing a highly efficient protein prefractionation method which is expected to show different characteristics with other methods. A total of 12 columns were screened, based on the peak widths at half height and the peak areas of the standard proteins (Table S1). The molecular weight range of the standard proteins used in this study covers 94% of the human canonical proteins registered in UniProt (Fig. S1 and Table S2). First, columns of the same brand were used to identify the modification groups and pore sizes suitable for protein prefractionation. Since the proteome to be fractionated has a wide molecular weight distribution, a high performance is required over the corresponding range. We found that columns with shorter alkyl chains and larger pore size tended to show higher performance over the evaluated molecular weight range (Fig. S2). Therefore, wide pore columns with short alkyl chains, which were recently developed for biopharmaceutical analysis, were considered likely to be applicable to protein prefractionation. A side-by-side comparison of five biologics-focused RP columns showed that they all performed well, however, SunShell C4-100 and BioResolve RP mAb performed best among them (Fig. S3). SDB and non-porous silica particle columns were also evaluated, but BioResolve RP mAb clearly performed better (Fig. S4, S5). After the column screening, we conducted a pilot experiment by loading 80 ug of cell lysate onto a BioResolve RP mAb column with 2.1 mm internal diameter. However, the performance was clearly poor compared to that of 4.6 mm i.d. column, and it was suspected that 80 ug was an overload for a 2.1 mm i.d. column (Fig. S6). Although the screening study was conducted using 2.1 mm i.d. columns, we considered that the screening results could be applied to columns of different internal diameters by matching the linear velocity. In order to demonstrate this idea, we measured standard proteins with columns of different internal diameters (2.1 and 4.6 mm) while matching the linear velocity and confirmed that equivalent profiles and performance could be obtained (Fig. S7). Therefore, it was concluded that the results of the column screening conducted using 2.1 mm i.d. columns could be applied to 4.6 mm i.d. columns by matching the linear velocity. Based on these results, we decided to use a 4.6 mm i.d. BioResolve RP mAb column for the comparative experiment on prefractionation methods to load sufficient amount of proteins for following experimental procedures.

A mixture of 8 standard proteins (Table S1, except for BSA) was injected into a 4.6 mm i.d. BioResolve RP mAb column with total amounts varying from 20 to 160 μg to find an acceptable loading amount. Mean and individual peak widths at half height were employed as indicators of separation performance for each loading amount. BSA was not used in the evaluation because it exhibits complex peak shapes and is known to be unsuitable for evaluating separation performance. As expected, there was a trade-off between sample loading amount and separation performance. However, the performance reduction was limited in the range of loading amounts evaluated in this study (Fig. S8). For example, a 400% increase in loading amount from 20 μg to 80 μg resulted in only a 19% increase in relative mean peak width at half height. Since the balance between separation performance and usability was considered to be acceptable, the loading amount for subsequent fractionation experiments was set at 80 μg .

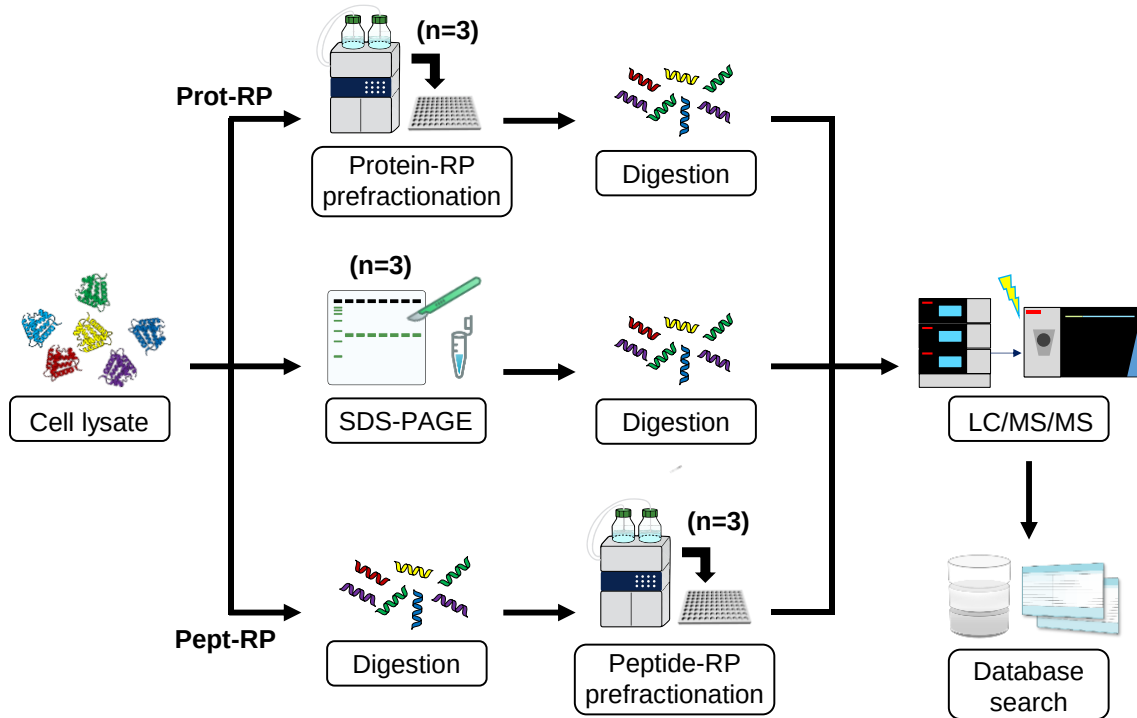


Figure 1. Workflow of this study.

The cell lysates were fractionated by three methods (Prot-RP, SDS-PAGE, and Pept-RP) and finally analyzed by LC/MS/MS. Prefractionations were performed in triplicate for all methods. Obtained raw data were searched against the Uniprot human database. The images are from TogoTV (2016 DBCLS TogoTV, CC-BY-4.0 <https://creativecommons.org/licenses/by/4.0/>).

Prefractionation workflow

Protein prefractionation was performed using the BioResolve RP mAb column, and the most widely used protein and peptide prefractionation methods (SDS-PAGE and Pept-RP) were employed as benchmarks to compare performance. The workflow of the entire experiment is shown in Fig. 1. For each prefractionation method, 80 µg of proteins or digested peptides was used as starting material and fractionated into 10 fractions (n = 3). Chromatograms or gel images obtained by Prot-RP, SDS-PAGE, and Pept-RP are shown in Fig. S9. The identification numbers of proteins and peptides for each method are summarized in Fig. 2a, 2b, and Fig. S10.

The degree of enrichment of individual molecular species into a single fraction is an important metric for prefractionation methods (hereafter referred to as "fractionation efficiency"). To analyze fractionation efficiency, the overlap of identified peptides between adjacent fractions and the enrichment ratio of each molecule were evaluated. Enrichment ratio is defined as the maximum value of distribution ratio among the fractions. The distribution ratio of "peptide X" in "fraction Y" is defined by the following formula.

$$\text{Distribution ratio of } X \text{ in } Y (\%) = \frac{\text{Abundance of peptide } X \text{ in fraction } Y}{\text{Total abundance of peptide } X} \times 100$$

Regarding the definition, enrichment ratio is a quantitative indicator representing how efficiently the molecules are isolated into a single fraction. Prot-RP and SDS-PAGE showed similar levels of overlap ratio and enrichment ratio distribution (Fig. 2c, 2d). On the other hand, Pept-RP showed a slightly lower overlap ratio at the peptide level compared to Prot-RP and SDS-PAGE (Fig. 2c). Comparison of quantitative precision and recovered amount between prefractionation methods revealed slightly better reproducibility at the protein level for Prot-RP and a significantly lower median recovered amount for SDS-PAGE, approximately 50% lower than the other methods (Fig. 3).

As expected, Pept-RP had the highest fractionation efficiency among the prefractionation methods evaluated (Fig. 2c, d). These results indicate that separation of smaller molecules, such as peptides, is more efficient than separation of much larger proteins. Comparison of protein prefractionation methods showed that the fractionation efficiency of Prot-RP is comparable to that of SDS-PAGE, the gold standard for protein prefractionation. Moreover, Prot-RP recovered approximately 2-fold greater amount of peptides than SDS-PAGE and afforded higher protein and peptide identification numbers.

While in-gel digestion is generally considered to limit protease access to proteins and cause lower digestion efficiency and thereby lower recovery than in-solution digestion^{52–54}, Prot-RP digests proteins in the liquid phase after prefractionation, which may account for the difference in recovered amount. Prot-RP is a user-friendly method since most of the steps can be performed in the liquid phase, in contrast to SDS-PAGE, which requires a tedious manual in-gel digestion process. In summary, Prot-RP is a superior protein prefractionation method that is more user-friendly than SDS-PAGE and also has higher performance.

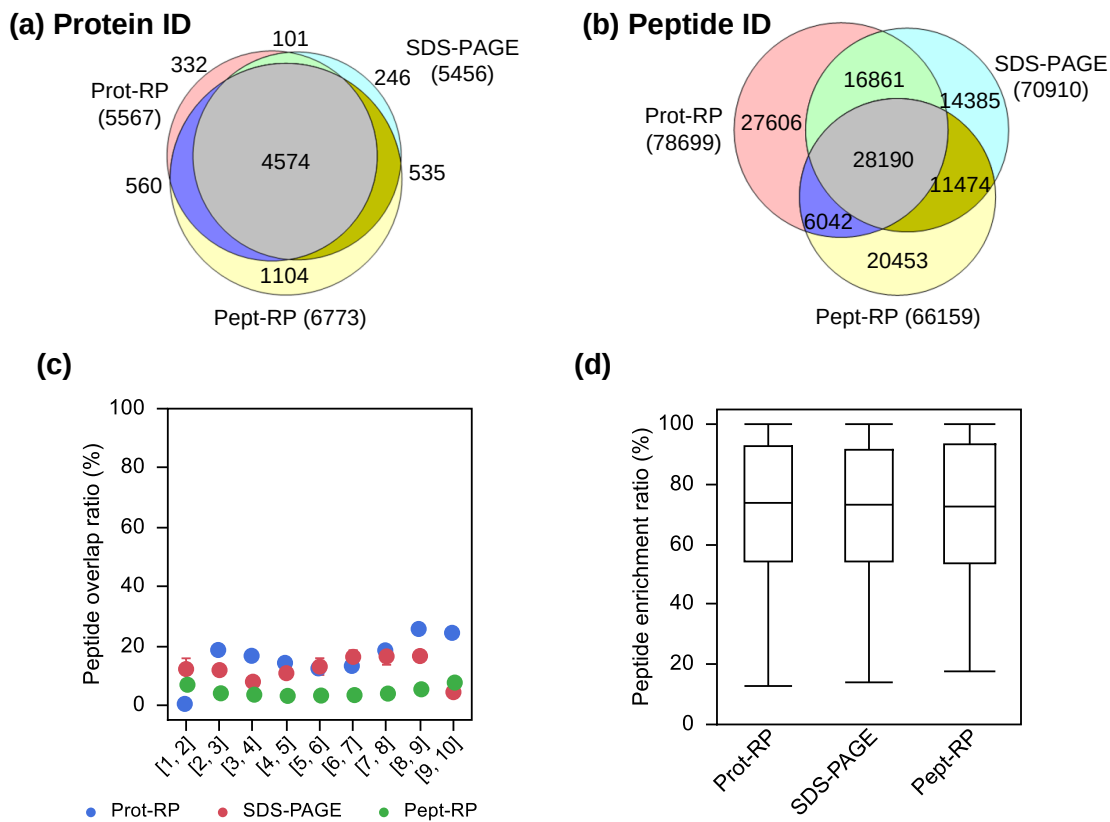


Figure 2. Performance characteristics of the three prefractionation methods.

(a, b) Venn diagrams of the identified proteins (a) and peptides (b) among the three prefractionation methods. (c) Overlap of adjacent fractions for peptides. The overlap ratio is defined as, No. of commonly identified peptide / No. of peptides identified from 2 fractions. Error bars represent standard deviation of $n = 3$ experiments. (d) Distribution of enrichment ratio for peptides. Average of $n = 3$ experiments was employed for the analysis. The enrichment ratio is defined in the main text.

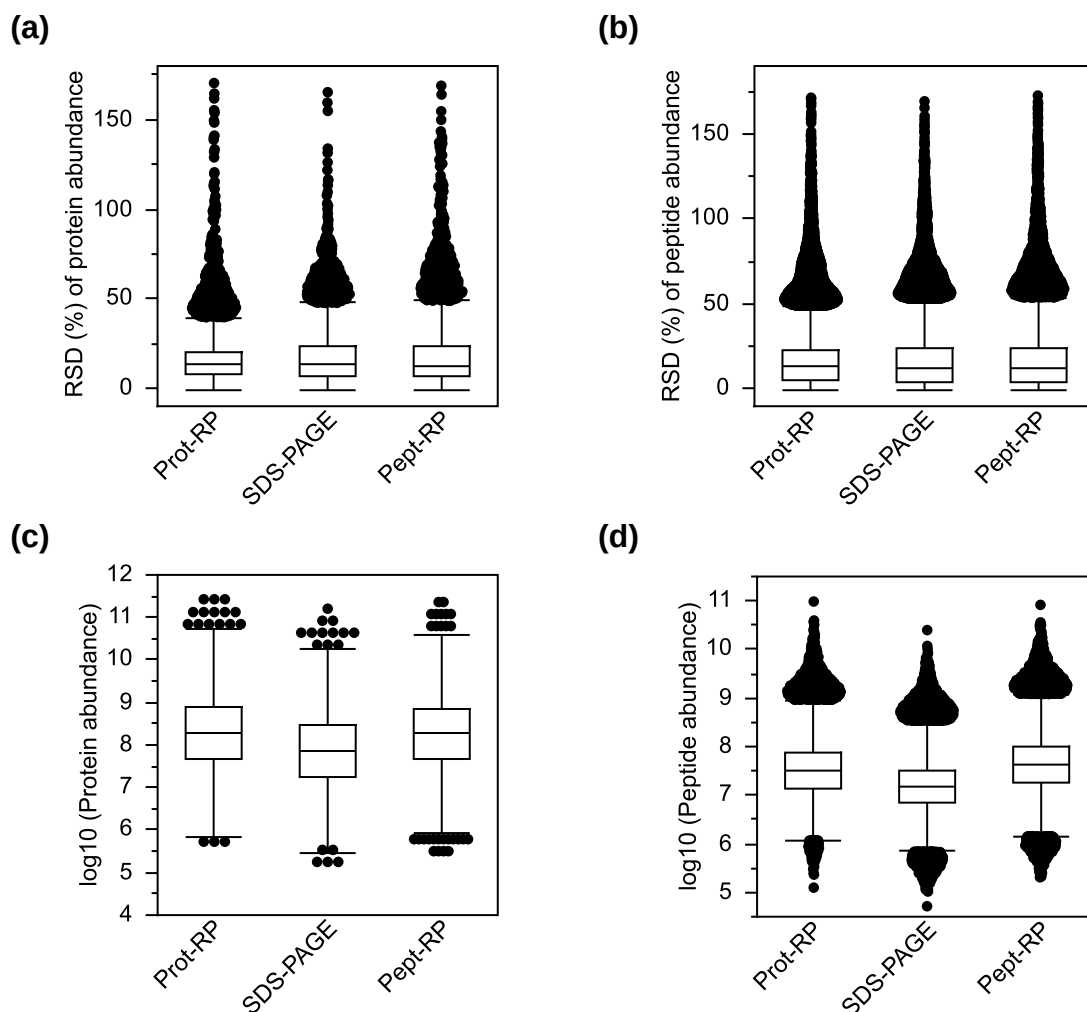


Figure 3. Quantitative precision and recovered analyte amount among the prefractionation methods.

(a, b) Reproducibility distribution of quantitative values for each prefractionation method, shown at the protein (a) and peptide (b) levels. (c, d) Distribution of protein (c) and peptide (d) abundances by LFQ for each prefractionation method. Mean value of triplicated experiments were used.

Applicability to proteoform analysis

The median of identified peptides per protein was 8, 7, and 6 for Prot-RP, SDS-PAGE, and Pept-RP, respectively, and was higher for protein-level prefractionation than for peptide-level prefractionation. Protein prefractionation showed higher frequency in the high-sequence-coverage region ($\geq 60\%$) than peptide prefractionation, and Pept-RP clearly had a greater population in the low-coverage region ($\leq 10\%$) compared to the

other methods (Fig. 4a). The median sequence coverage was 26%, 24%, and 19% for Prot-RP, SDS-PAGE, and Pept-RP, respectively. While the overlap ratio (overlap/union) of proteins among the three methods was 61%, the overlap ratio of peptides was 23%, suggesting that different peptides were identified from the same protein by each method (Fig. 2a, b). These results suggest that some of the peptides specifically identified by protein prefractionation methods do not contribute to the identification of new proteins, but do improve sequence coverage. Since protein prefractionation methods are expected to provide more detailed information on individual proteins than peptide prefractionation, we next focused on the applicability of protein prefractionation to proteoform analysis.

The number and ratio of proteins for which canonical N- and C-termini could be identified are shown in Fig. 4b and c, respectively. Prot-RP showed the highest number and ratio of proteins identified at both termini. In particular, the identification rate of the protein C-terminus of Prot-RP was about twice that of Pept-RP, and thus Prot-RP showed superior performance (Fig. 4c). We compared the ability to identify isoforms among the prefractionation methods by searching raw data against a database containing isoform sequences in addition to canonical sequences (UniProtKB release 2023_03, 42,369 sequences). The number of all proteins identified using the “canonical” and “canonical + isoform” databases are shown in Fig. S11. As in the case of protein termini, Prot-RP also provided the highest identification number and ratio for isoform-specific peptides (Fig. 4d, e). In this study, we defined isoform-specific peptides as peptides that are generated exclusively from isoforms. These results suggest that among the protein prefractionation methods, Prot-RP has the best performance for proteoform analysis. Since complementarity between prefractionation methods was observed in the identification of both protein termini and isoforms, it is considered that the depth of analysis can be further increased by combining multiple prefractionation methods (Fig. S12).

Membrane proteins play critical biological roles and are important targets for proteomic analysis. However, due to their high hydrophobicity, they may be lost in sample prefractionation step due to adsorption to equipment, such as chromatography columns. We evaluated whether membrane proteins could be recovered when proteins are fractionated by reversed-phase chromatography by comparing the results of Prot-RP with SDS-PAGE and Pept-RP. In this analysis, proteins that corresponded to the UniProt Keyword "KW-0472; Membrane" were considered as membrane proteins. The

abundance distributions of all and membrane proteins are shown in Fig. S13a. The abundance distributions of membrane proteins showed the same trend as that of all proteins in all prefractionation methods, indicating that membrane proteins can be recovered in Prot-RP as well as other methods. When the number of identified membrane proteins was compared, Pept-RP showed the highest number, while SDS-PAGE and Prot-RP had similar levels (Fig. S13b). SDS-PAGE gave a slightly higher number of identified membrane proteins (approximately 5%) than Prot-RP, however, the difference was not as great as we expected and the trend was similar to the totally identified proteins. Next, we focused on mitochondrion as an example of a membrane-rich organelle and evaluated the extent to which proteins localized to the mitochondrial membrane were covered. Proteins that corresponded to both "KW-0472; Membrane" and "KW-0496; Mitochondrion" in the UniProt Keyword were considered as known mitochondrial membrane proteins (661 proteins in UniProt KB (Human, Reviewed)). Fig. S13c and d show the numbers and coverages of mitochondrial membrane proteins identified using each fractionation method, respectively. These results indicate that all fractionation methods evaluated in this study were able to identify proteins from mitochondrial membranes at similar coverage levels (54% to 62%). Although Prot-RP is slightly inferior to SDS-PAGE in terms of the number of membrane protein identifications, it has been shown that it is possible to recover even highly hydrophobic proteins to a level similar to that of other methods. Therefore, Prot-RP is considered to be a protein-level prefractionation method that can be applied to various proteoform analysis studies.

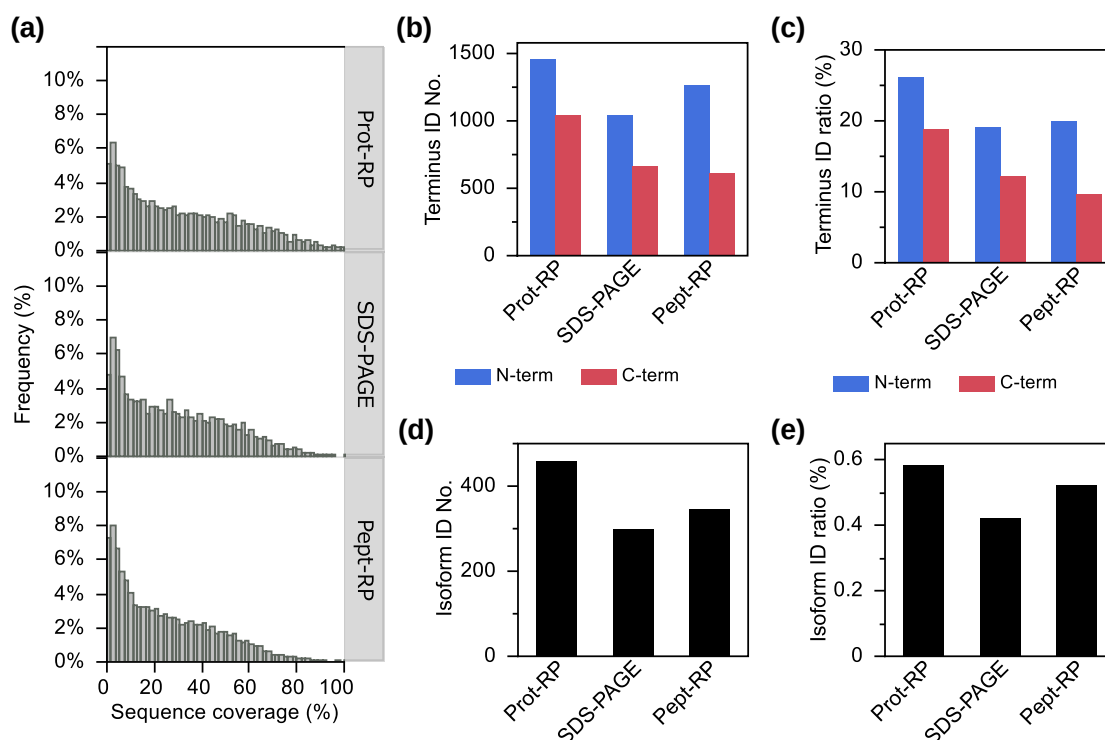


Figure 4. Method performance for proteoform analysis.

Database search results obtained from all the fractions and repetitions were used for the analysis. (a) Distribution of amino acid sequence coverage for the three prefractionation methods. (b) Number of proteins whose termini were identified using each prefractionation method. (c) Ratio of proteins whose termini were identified to total identified proteins for each prefractionation method. (d) Number of isoform-specific peptides identified using each prefractionation method. (e) Ratio of isoform-specific peptides to total identified peptides for each prefractionation method.

Effect of MS performance

The performance of peptide-level and protein-level prefractionation has been compared many times, with differences in the conclusions depending on the generation of instruments which were used in the study^{28–30}. We focused on the effect of MS performance, especially sensitivity, on the conclusions and conducted a simulation analysis to raise the detection limit of MS. In this simulation analysis, detection limit is defined as the lowest precursor ion intensity that can produce a MS/MS spectrum of sufficient quality which could be identified as peptides. The simulation analysis was performed by raising the threshold of precursor ion intensity which could be identified as

peptides. Specifically, the list of precursor ions identified as peptides were first sorted in descending order by their intensity and only certain proportions of the highest intensities were retained (lower sides were removed). The number of identified proteins was calculated with several intensity thresholds ranging from the top 1% to 100% (= no precursor ions removed) (Fig. 5a). This operation was performed for each LC/MS run, and all the identified peptides were merged to evaluate the identified protein numbers for each prefractionation method. The step-by-step workflow of the simulation analysis is shown in Fig. S14. Interestingly, while the highest number of proteins was identified in Pept-RP when all identified precursor ions were used, the order of protein identification numbers for Prot-RP and Pept-RP was reversed when the identified precursor ions were reduced to the top 10% (Fig. 5a). This result suggests that Prot-RP is more advantageous for protein identification than Pept-RP when the MS sensitivity is lower. This finding is in good agreement with the results of Dowell et al. in 2008. They compared protein RP prefractionation with high-pH RP peptide prefractionation using the Q-TOF Micro (Waters, launched in 2001) and reported that protein RP prefractionation identified more proteins when Mascot was used as the search engine²⁸.

To explore the mechanism behind these results, the distribution of $\log_{10}$ (Precursor ion intensity) was compared for all (100%) and the top 10% high intensity precursor ions identified as peptides. When all identified precursor ions were plotted, there was no significant difference in distribution between prefractionation methods (Fig. 5b). However, when identified precursor ions were restricted to the top 10%, the peptide prefractionation method (Pept-RP) showed a narrower and steeper distribution compared to the protein prefractionation methods (Prot-RP and SDS-PAGE) (Fig. 5c). Interquartile ranges (IQR) shown in Table S3 serve as quantitative indices to support these observations. The wider IQR of protein prefractionation methods compared to the peptide fractionation method suggest that there would be a difference in dynamic range among the prefractionation methods when the threshold was set to top 10%. When the MS sensitivity was lower, the protein-level prefractionation was more effective in extending the dynamic range than the peptide-level prefractionation, which is expected to be advantageous for protein identification. When prefractionation is performed at the peptide level, peptides from high-abundance proteins are dispersed in every fraction, which interferes with the detection of low-abundance peptides in all fractions. On the other hand, when prefractionation is performed at the protein level, peptides derived from high-abundance proteins remain within a single fraction, and dispersion of high-

abundance peptides does not occur. Older-generation instruments with limited performance (sensitivity, scan speed, and dynamic range) are expected to benefit strongly from this effect, and in this case protein prefractionation methods enable identification of peptides over a wider dynamic range. It is presumed that this effect led to the results of previous studies, in which more proteins were identified using the protein fractionation method in instruments with lower performance^{28,30}. However, in recent years, MS instruments have improved dramatically in sensitivity, achieving sufficient scan speed and dynamic range to identify low-abundance peptides in the presence of high-abundance peptides. As a result of this improvement in instrument performance, the high fractionation efficiency of peptide prefractionation has become a more significant contributor to the number of protein identifications than the dynamic range extension effect of protein prefractionation. Therefore, peptide prefractionation with high fractionation efficiency, such as Pept-RP, is the best option when the goal is to identify more proteins from samples such as cell lysates using modern MS instruments.

As discussed in the previous section, Prot-RP gave the best results for amino acid sequence coverage, protein terminus identification, and isoform analysis (Fig. 4). Thus, even with the increased performance of instrumentation, the ability to obtain detailed information on individual proteins is still better with protein-level prefractionation than with peptide-level prefractionation. In addition, protein prefractionation provides information on the separation of proteoforms in the first dimension, which is not available with peptide prefractionation. Thus, further improvement of protein prefractionation techniques, such as Prot-RP, is expected to prove especially useful for proteoform analysis in the future.

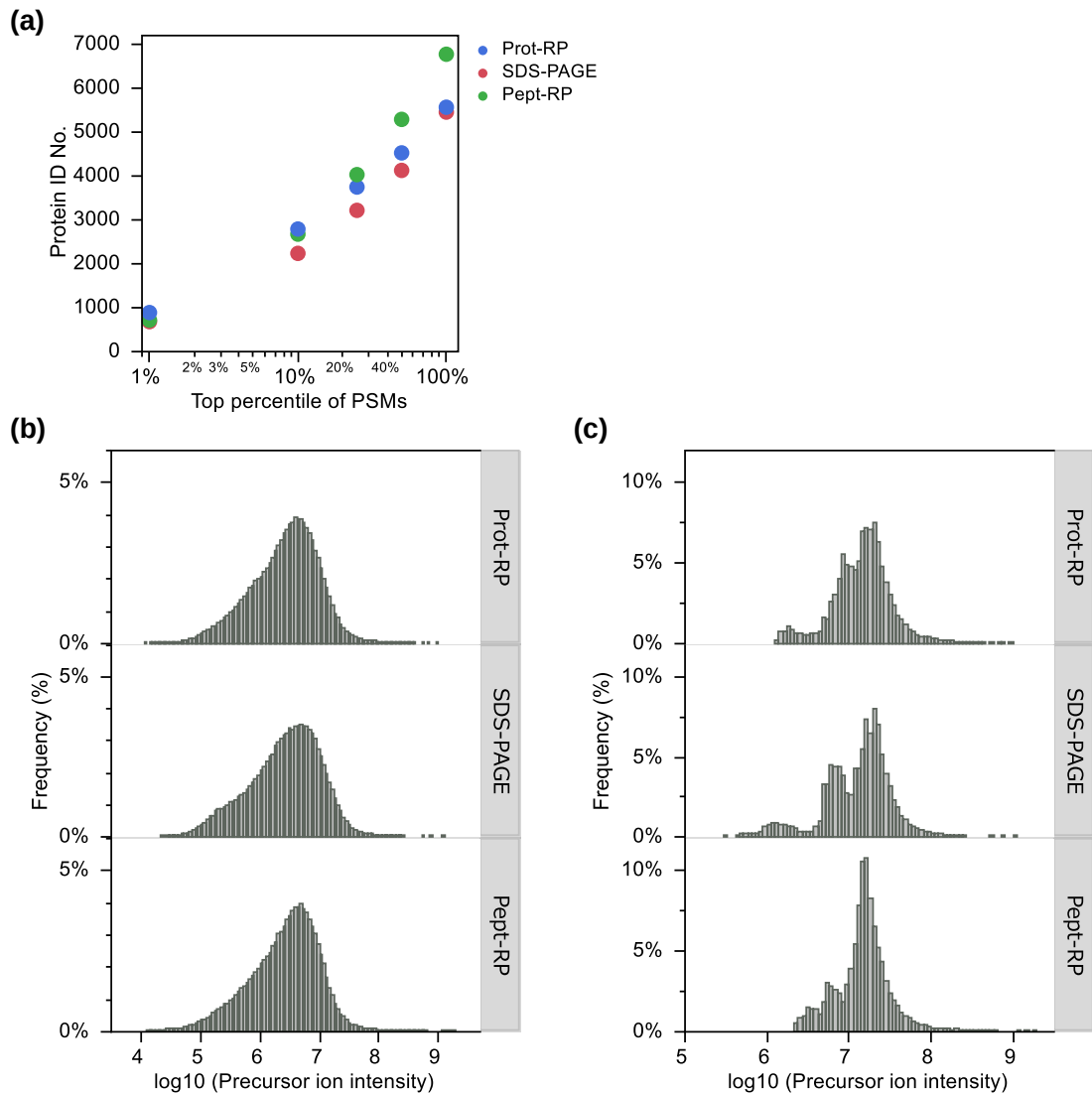


Figure 5. Simulation analysis of earlier-generation mass spectrometers.

(a) Number of proteins identified from precursor ions selected with various cut-off values. Trimming of identified precursor ion list was performed for each distinct LC/MS run and then all results were merged. (b, c) Distribution of $\log_{10}$ (Precursor ion intensity) for all (b) and for the top 10% identified precursor ions (c). $\log_{10}$ (Precursor ion intensity) was normalized among the prefractionation methods by aligning their medians.

Conclusion

In this study, we revisited protein RP prefractionation by using state-of-the-art RP columns focused on biopharmaceuticals and evaluating their applicability for bottom-up proteomics. The performance of our method was compared with that of Pept-RP and

SDS-PAGE. Prot-RP was found to have similar fractionation efficiency to SDS-PAGE while providing 2-fold higher protein recovery. Although the fractionation efficiency of Prot-RP was slightly lower than that of Pept-RP, Prot-RP showed higher performance for proteoform analysis than Pept-RP since it identified more protein-terminal peptides and isoform-specific peptides. The ultimate goal of proteomics is a comprehensive understanding of proteoforms, and for this purpose, top-down proteomics would be preferable. However, top-down proteomics is still a developing technology, and the bottom-up approach is expected to continue to play a key role in proteomics. As protein prefractionation allows top-down and bottom-up proteome analyses from the same fraction and also provides rich proteoform information in bottom-up proteomics experiments, there is a great incentive to achieve higher performance. Protein prefractionation was once a preferable choice in terms of maximizing protein and peptide identification with earlier-generation MS, but has recently been displaced by peptide prefractionation due to the tedious post fractionation procedure while gain in protein identification is relatively lower, especially in SDS-PAGE. However, given the increasing interest of the proteomics community in proteoforms, it seems reasonable to expect a return of protein prefractionation to the mainstream. We believe the protein prefractionation method we have demonstrated here opens the door for highly efficient proteoform profiling.

Associated Information

Data Availability

The raw MS files and the result files (identified PSMS, peptides and proteins) from the Proteome Discoverer have been deposited to the jPOST repository (<https://repository.jpostdb.org/>)⁵⁵, a member of ProteomeXchange Consortium, with the data set identifier PXD049312.

Supporting Information

The Supporting Information is available free of charge at doi. XXXXXXX.

Supplemental Table S1. Columns employed in the screening study; Supplemental Table S2. Details of the standard proteins; Supplemental Table S3. Quartiles in log₁₀ (Precursor ion intensity) distribution shown in Figure 5; Supplemental Figure S1. Materials used for column screening; Supplemental Figure S2. Screening results for SunShell columns with various modifications and pore sizes; Supplemental Figure S3. Screening results for wide pore and short alkyl chain columns; Supplemental Figure S4. Comparison between silica and SDB based columns; Supplemental Figure S5. Comparison between superficially porous and non-porous particle columns.; Supplemental Figure S6. Fractionation efficiency observed in the pilot study utilizing columns with different internal diameters; Supplemental Figure S7. Comparison of chromatographic profiles between columns with different internal diameters; Supplemental Figure S8. Separation performance at various loading amount; Supplemental Figure S9. Chromatograms and a gel image from each prefractionation method; Supplemental Figure S10. Protein and peptide identification numbers for each prefractionation method; Supplemental Figure S11. Number of proteins identified from “canonical” and “canonical + isoform” databases for each prefractionation method; Supplemental Figure S12. Protein terminus and isoforms identified using each prefractionation method; Supplemental Figure S13. Analysis results focused on membrane proteins obtained from each prefractionation method; Supplemental Figure S14. Workflow of the simulation analysis (PDF)

Author Information

Author Contributions

Shunsuke Takagi: Conceptualization, Investigation, Writing –original draft, Visualization.

Nobuyuki Suzuki: Writing – review & editing, Supervision. **Yasushi Ishihama:**

Conceptualization, Writing – review & editing, Supervision, Funding acquisition, Project administration.

Declaration of Competing Interest

Shunsuke Takagi and Nobuyuki Suzuki are employees of Daiichi Sankyo Co., Ltd.

Acknowledgements

This work has been funded by the JST Strategic Basic Research Program CREST (No. JPMJCR1862), AMED-CREST program (No. JP18gm1010010), JSPS Grants-in-Aid for Scientific Research 21H02459, 23H04924 and 23K18185 to Y.I.

References

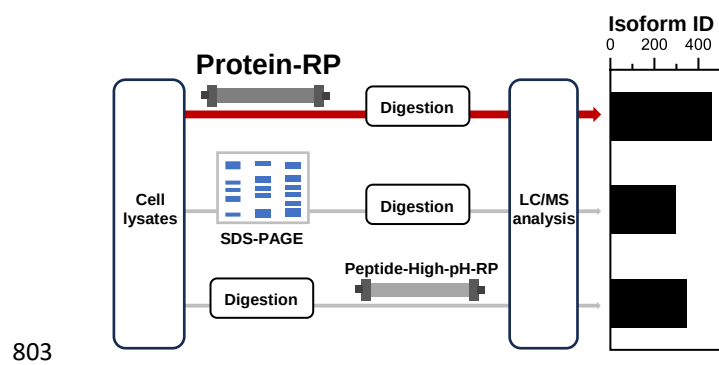
- (1) Meier, F.; Brunner, A.-D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Räther, O.; Mann, M. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **2018**, *17* (12), 2534–2545.
- (2) Heil, L. R.; Damoc, E.; Arrey, T. N.; Pashkova, A.; Denisov, E.; Petzoldt, J.; Peterson, A. C.; Hsu, C.; Searle, B. C.; Shulman, N.; Riffle, M.; Connolly, B.; MacLean, B. X.; Remes, P. M.; Senko, M. W.; Stewart, H. I.; Hock, C.; Makarov, A.; Hermanson, D.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J. Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **2023**, *22* (10), 3290–3300.
- (3) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.
- (4) Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M. Parallel Accumulation-Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14* (12), 5378–5387.
- (5) Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar Acquisition Method Enables Single-Shot Proteomics at a Depth of 10,000 Proteins in 100 Minutes. *Nat. Methods* **2018**, *15* (6), 440–448.
- (6) Meier, F.; Brunner, A.-D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M. diaPASEF: Parallel Accumulation-Serial Fragmentation Combined with Data-Independent Acquisition. *Nat. Methods* **2020**, *17* (12), 1229–1236.
- (7) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17* (1), 41–44.
- (8) Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yilmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J. MaxDIA Enables Library-Based and Library-Free Data-Independent Acquisition Proteomics. *Nat. Biotechnol.* **2021**, *39* (12), 1563–1573.
- (9) Kawashima, Y.; Nagai, H.; Konno, R.; Ishikawa, M.; Nakajima, D.; Sato, H.; Nakamura, R.; Furuyashiki, T.; Ohara, O. Single-Shot 10K Proteome Approach: Over 10,000 Protein Identifications by Data-Independent Acquisition-Based Single-Shot Proteomics with Ion Mobility Spectrometry. *J. Proteome Res.* **2022**, *21*, 1418–1427.
- (10) 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.; Olsen, J. V. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* **2017**, *4* (6), 587–599.e4.
- (11) Ruprecht, B.; Wang, D.; Chiozzi, R. Z.; Li, L.-H.; Hahne, H.; Kuster, B. Hydrophilic Strong Anion Exchange (hSAX) Chromatography Enables Deep Fractionation of Tissue Proteomes. *Methods Mol. Biol.* **2017**, *1550*, 69–82.
- (12) Sinitcyn, P.; Richards, A. L.; Weatheritt, R. J.; Brademan, D. R.; Marx, H.;

- Shishkova, E.; Meyer, J. G.; Hebert, A. S.; Westphall, M. S.; Blencowe, B. J.; Cox, J.; Coon, J. J. Global Detection of Human Variants and Isoforms by Deep Proteome Sequencing. *Nat. Biotechnol.* **2023**, 1–11.
- (13) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, 73 (23), 5683–5690.
- (14) Di Palma, S.; Boersema, P. J.; Heck, A. J. R.; Mohammed, S. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. *Anal. Chem.* **2011**, 83 (9), 3440–3447.
- (15) Hao, P.; Guo, T.; Li, X.; Adav, S. S.; Yang, J.; Wei, M.; Sze, S. K. Novel Application of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) in Shotgun Proteomics: Comprehensive Profiling of Rat Kidney Proteome. *J. Proteome Res.* **2010**, 9 (7), 3520–3526.
- (16) Cargile, B. J.; Sevinisky, J. R.; Essader, A. S.; Stephenson, J. L., Jr; Bundy, J. L. Immobilized pH Gradient Isoelectric Focusing as a First-Dimension Separation in Shotgun Proteomics. *J. Biomol. Tech.* **2005**, 16 (3), 181–189.
- (17) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Two-Dimensional Separation of Peptides Using RP-RP-HPLC System with Different pH in First and Second Separation Dimensions. *J. Sep. Sci.* **2005**, 28 (14), 1694–1703.
- (18) Batth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-pH Reversed-Phase Fractionation for in-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, 13 (12), 6176–6186.
- (19) Lasonder, E.; Ishihama, Y.; Andersen, J. S.; Vermunt, A. M. W.; Pain, A.; Sauerwein, R. W.; Eling, W. M. C.; Hall, N.; Waters, A. P.; Stunnenberg, H. G.; Mann, M. Analysis of the Plasmodium Falciparum Proteome by High-Accuracy Mass Spectrometry. *Nature* **2002**, 419 (6906), 537–542.
- (20) Chen, C.; Wen, M.; Jin, Y. 1DE-MS Profiling for Proteoform-Correlated Proteomic Analysis, by Combining SDS-PAGE, Whole-Gel Slicing, Quantitative LC-MS/MS, and Reconstruction of Gel Distributions of Several Thousands of Proteins. *J. Proteome Res.* **2022**, 21 (10), 2311–2330.
- (21) Takemori, A.; Kaulich, P. T.; Konno, R.; Kawashima, Y.; Hamazaki, Y.; Hoshino, A.; Tholey, A.; Takemori, N. GeLC-FAIMS-MS Workflow for in-Depth Middle-down Proteomics. *Proteomics* **2023**, e2200431.
- (22) Martosella, J.; Zolotarjova, N.; Liu, H.; Nicol, G.; Boyes, B. E. Reversed-Phase High-Performance Liquid Chromatographic Prefractionation of Immunodepleted Human Serum Proteins to Enhance Mass Spectrometry Identification of Lower-Abundant Proteins. *J. Proteome Res.* **2005**, 4 (5), 1522–1537.
- (23) Patil, L. M.; Parkinson, D. H.; Zuniga, N. R.; Lin, H.-J. L.; Naylor, B. C.; Price, J. C. Combining Offline High Performance Liquid Chromatography Fractionation of Peptides and Intact Proteins to Enhance Proteome Coverage in Bottom-up Proteomics. *J. Chromatogr. A* **2023**, 1701, 464044.
- (24) Faca, V.; Pitteri, S. J.; Newcomb, L.; Glukhova, V.; Phanstiel, D.; Krasnoselsky, A.; Zhang, Q.; Struthers, J.; Wang, H.; Eng, J.; Fitzgibbon, M.; McIntosh, M.; Hanash, S. Contribution of Protein Fractionation to Depth of Analysis of the Serum and Plasma Proteomes. *J. Proteome Res.* **2007**, 6 (9), 3558–3565.
- (25) Kaur, G.; Poljak, A.; Ali, S. A.; Zhong, L.; Raftery, M. J.; Sachdev, P. Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *J. Proteome Res.* **2021**, 20 (2), 1261–1279.
- (26) Kerner, M. J.; Naylor, D. J.; Ishihama, Y.; Maier, T.; Chang, H.-C.; Stines, A. P.; Georgopoulos, C.; Frishman, D.; Hayer-Hartl, M.; Mann, M.; Hartl, F. U. Proteome-

- Wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia Coli. *Cell* **2005**, 122 (2), 209–220.
- (27) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996**, 68 (5), 850–858.
- (28) Dowell, J. A.; Frost, D. C.; Zhang, J.; Li, L. Comparison of Two-Dimensional Fractionation Techniques for Shotgun Proteomics. *Anal. Chem.* **2008**, 80 (17), 6715–6723.
- (29) Mostovenko, E.; Hassan, C.; Rattke, J.; Deelder, A. M.; van Veelen, P. A.; Palmblad, M. Comparison of Peptide and Protein Fractionation Methods in Proteomics. *EuPA Open Proteomics* **2013**, 1, 30–37.
- (30) Deng, L.; Handler, D. C. L.; Multari, D. H.; Haynes, P. A. Comparison of Protein and Peptide Fractionation Approaches in Protein Identification and Quantification from *Saccharomyces Cerevisiae*. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2021**, 1162, 122453.
- (31) Paulo, J. A. Sample Preparation for Proteomic Analysis Using a GeLC-MS/MS Strategy. *J Biol Methods* **2016**, 3 (3). <https://doi.org/10.14440/jbm.2016.106>.
- (32) Nishida, H.; Kanao, E.; Ishihama, Y. Centrifugal Gel Crushing Tips for Gel-Based Proteome Analysis. *Anal. Chem.* **2023**. <https://doi.org/10.1021/acs.analchem.3c02527>.
- (33) Tran, J. C.; Doucette, A. A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem.* **2008**, 80 (5), 1568–1573.
- (34) Lee, J. E.; Kellie, J. F.; Tran, J. C.; Tipton, J. D.; Catherman, A. D.; Thomas, H. M.; Ahlf, D. R.; Durbin, K. R.; Vellaichamy, A.; Ntai, I.; Marshall, A. G.; Kelleher, N. L. A Robust Two-Dimensional Separation for Top-down Tandem Mass Spectrometry of the Low-Mass Proteome. *J. Am. Soc. Mass Spectrom.* **2009**, 20 (12), 2183–2191.
- (35) Kachuk, C.; Doucette, A. A. The Benefits (and Misfortunes) of SDS in Top-down Proteomics. *J. Proteomics* **2018**, 175, 75–86.
- (36) Challal, D.; Colin, J.; Villa, T.; Libri, D. A Modified Cross-Linking Analysis of cDNAs (CRAC) Protocol for Detecting RNA–protein Interactions and Transcription at Single-Nucleotide Resolution. In *Methods in Molecular Biology*; Methods in molecular biology (Clifton, N.J.); Springer US: New York, NY, 2022; pp 35–55.
- (37) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. Comprehensive on-Line LC/LC/MS of Proteins. *Anal. Chem.* **1997**, 69 (8), 1518–1524.
- (38) Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. Isoelectric Focusing Nonporous RP HPLC: A Two-Dimensional Liquid-Phase Separation Method for Mapping of Cellular Proteins with Identification Using MALDI-TOF Mass Spectrometry. *Anal. Chem.* **2000**, 72 (6), 1099–1111.
- (39) Gao, M.; Deng, C.; Yu, W.; Zhang, Y.; Yang, P.; Zhang, X. Large Scale Depletion of the High-Abundance Proteins and Analysis of Middle- and Low-Abundance Proteins in Human Liver Proteome by Multidimensional Liquid Chromatography. *Proteomics* **2008**, 8 (5), 939–947.
- (40) Marangon, M.; Van Sluyter, S. C.; Haynes, P. A.; Waters, E. J. Grape and Wine Proteins: Their Fractionation by Hydrophobic Interaction Chromatography and Identification by Chromatographic and Proteomic Analysis. *J. Agric. Food Chem.* **2009**, 57 (10), 4415–4425.
- (41) Kristensen, A. R.; Gsponer, J.; Foster, L. J. A High-Throughput Approach for Measuring Temporal Changes in the Interactome. *Nat. Methods* **2012**, 9 (9), 907–909.

- (42) Kirkwood, K. J.; Ahmad, Y.; Larance, M.; Lamond, A. I. Characterization of Native Protein Complexes and Protein Isoform Variation Using Size-Fractionation-Based Quantitative Proteomics. *Mol. Cell. Proteomics* **2013**, *12* (12), 3851–3873.
- (43) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate Quantitation of Protein Expression and Site-Specific Phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (12), 6591–6596.
- (44) Melchior, K.; Tholey, A.; Heisel, S.; Keller, A.; Lenhof, H.-P.; Meese, E.; Huber, C. G. Protein- versus Peptide Fractionation in the First Dimension of Two-Dimensional High-Performance Liquid Chromatography-Matrix-Assisted Laser Desorption/ionization Tandem Mass Spectrometry for Qualitative Proteome Analysis of Tissue Samples. *J. Chromatogr. A* **2010**, *1217* (40), 6159–6168.
- (45) Fekete, S.; Rudaz, S.; Veuthey, J.-L.; Guilleme, D. Impact of Mobile Phase Temperature on Recovery and Stability of Monoclonal Antibodies Using Recent Reversed-Phase Stationary Phases. *J. Sep. Sci.* **2012**, *35* (22), 3113–3123.
- (46) Jaag, S.; Shirokikh, M.; Lämmerhofer, M. Charge Variant Analysis of Protein-Based Biopharmaceuticals Using Two-Dimensional Liquid Chromatography Hyphenated to Mass Spectrometry. *J. Chromatogr. A* **2021**, *1636*, 461786.
- (47) Kesik-Brodacka, M. Progress in Biopharmaceutical Development. *Biotechnol. Appl. Biochem.* **2018**, *65* (3), 306–322.
- (48) Martins, A. C.; Oshiro, M. Y.; Albericio, F.; de la Torre, B. G.; Pereira, G. J. V.; Gonzaga, R. V. Trends and Perspectives of Biological Drug Approvals by the FDA: A Review from 2015 to 2021. *Biomedicines* **2022**, *10* (9). <https://doi.org/10.3390/biomedicines10092325>.
- (49) Bobály, B.; Lauber, M.; Beck, A.; Guilleme, D.; Fekete, S. Utility of a High Coverage Phenyl-Bonding and Wide-Pore Superficially Porous Particle for the Analysis of Monoclonal Antibodies and Related Products. *J. Chromatogr. A* **2018**, *1549*, 63–76.
- (50) Smith, L. M.; Agar, J. N.; Chamot-Rooke, J.; Danis, P. O.; Ge, Y.; Loo, J. A.; Paša-Tolić, L.; Tsybin, Y. O.; Kelleher, N. L.; Consortium for Top-Down Proteomics. The Human Proteoform Project: Defining the Human Proteome. *Sci Adv* **2021**, *7* (46), eabk0734.
- (51) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **2003**, *75* (3), 663–670.
- (52) Speicher, K. D.; Kolbas, O.; Harper, S.; Speicher, D. W. Systematic Analysis of Peptide Recoveries from in-Gel Digestions for Protein Identifications in Proteome Studies. *J. Biomol. Tech.* **2000**, *11* (2), 74–86.
- (53) Rabilloud, T.; Vaezzadeh, A. R.; Potier, N.; Lelong, C.; Leize-Wagner, E.; Chevallet, M. Power and Limitations of Electrophoretic Separations in Proteomics Strategies. *Mass Spectrom. Rev.* **2009**, *28* (5), 816–843.
- (54) Takemori, A.; Takemori, N. Chapter Six - Sample Preparation for Structural Mass Spectrometry via Polyacrylamide Gel Electrophoresis. In *Methods in Enzymology*; Shukla, A. K., Ed.; Academic Press, 2023; Vol. 682, pp 187–210.
- (55) 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* (D1), D1107–D1111.

802 For Table of Contents Only



Supporting Information

Revisiting Protein Reversed-Phase Chromatography for Bottom-Up Proteomics

Shunsuke Takagi^{a,b}, Nobuyuki Suzuki^b, Yasushi Ishihama^{a,c}

- a) Department of Molecular Systems BioAnalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan
- b) Analytical and Quality Evaluation Research Laboratories, Daiichi Sankyo Co., Ltd., Hiratsuka, Kanagawa 254-0014, Japan
- c) Laboratory of Clinical and Analytical Chemistry, National Institute of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka 567-0085, Japan.

*Correspondence and requests for materials should be addressed to Y.I. (email: yishihama@pharm.kyoto-u.ac.jp, Phone: +81-75-753-4555, FAX: +81-75-753-4601).

Contents

Supplemental Table S1.	Columns employed in the screening study
Supplemental Table S2.	Details of the standard proteins
Supplemental Table S3.	Quartiles in log ₁₀ (Precursor ion intensity) distribution shown in Figure 5
Supplemental Figure S1.	Materials used for column screening.
Supplemental Figure S2.	Screening results for SunShell columns with various modifications and pore sizes.
Supplemental Figure S3.	Screening results for wide pore and short alkyl chain columns.
Supplemental Figure S4.	Comparison between silica and SDB based columns.
Supplemental Figure S5.	Comparison between superficially porous and non-porous particle columns.
Supplemental Figure S6.	Fractionation efficiency observed in the pilot study utilizing columns with different internal diameters.
Supplemental Figure S7.	Comparison of chromatographic profiles between columns with different internal diameters.
Supplemental Figure S8.	Separation performance at various loading amount.
Supplemental Figure S9.	Chromatograms and a gel image from each prefractionation method.
Supplemental Figure S10.	Protein and peptide identification numbers for each prefractionation method.
Supplemental Figure S11.	Number of proteins identified from “canonical” and “canonical + isoform” databases for each prefractionation method.
Supplemental Figure S12.	Protein terminus and isoforms identified using each prefractionation method.
Supplemental Figure S13.	Analysis results focused on membrane proteins obtained from each prefractionation method.
Supplemental Figure S14.	Workflow of the simulation analysis.

Supplemental Table S1. Columns employed in the screening study

Manufacture	Column name	Stationary phase	Pore size (nm)	Particle size (μm)
Chromanik Technologies (Osaka, Japan)	SunShell C4-30	C4	30	2.6
	SunShell C4-100	C4	100	2.6
	SunShell C8-30	C8	30	2.6
	SunShell HFC18-30	C18	30	2.6
	SunShell HFC18-16	C18	16	2.6
Waters (Milford, MA)	BioResolve RP mAb	Polyphenyl	45	2.7
Agilent Technologies (Santa Clara, CA)	AdvanceBio RP-mAb	C4	45	3.5
	AdvanceBio RP-mAb	C8 (SB-C8)	45	3.5
	AdvanceBio RP-mAb	Diphenyl	45	3.5
	PLRP-S	SDB	30	5
	PLRP-S	SDB	100	5
Imtakt (Kyoto, Japan)	Presto FF-C18	C18	Non-porous	2

SDB = poly(styrene-divinylbenzene) copolymer

Column dimension were 2.1 × 50 mm for all columns

Supplemental Table S2. Details of the standard proteins

Peak No. ^a	Protein description	Species	Observed mass (Da) ^b
1	Ribonuclease A (RNaseA)	Bovine	13,684
2	Insulin	Human	5,808
3	Lysozyme	<i>Gallus gallus</i>	14,307
4	Transferrin	Human	75,155
5	Bovine serum albumin (BSA)	Bovine	66,437
6	Trypsin inhibitor	Soybean	19,980
7	Humanized IgG1κ monoclonal antibody (NIST mAb)	-	148,229
8	β-Lactoglobulin A (β-LgA)	Bovine	18,365
9	Carbonic anhydrase (CA)	Bovine	29,027

^a Peak numbers refer to Fig. S1.

^b Mass analysis was conducted using a quadrupole-time of flight instrument (Synapt G2-Si, Waters). Values of the dominant peak are shown.

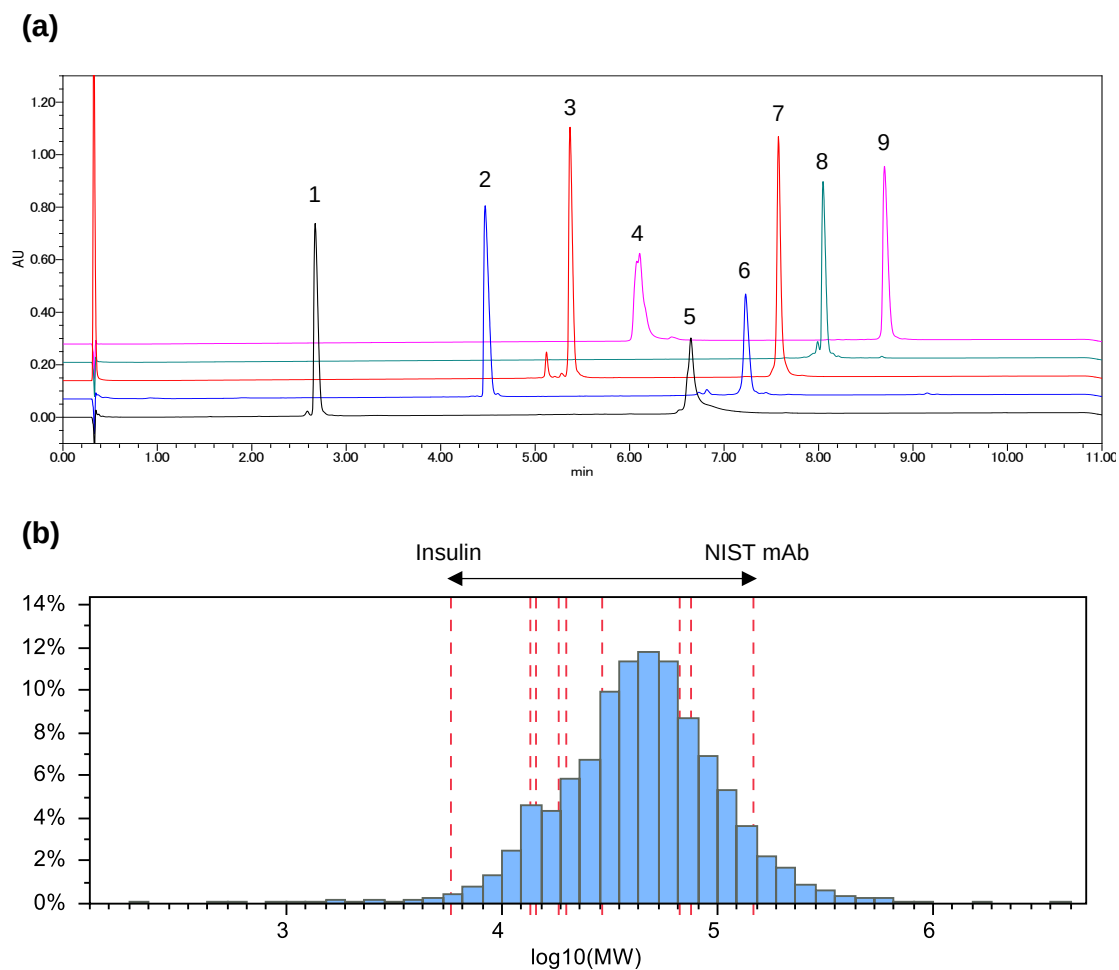
Supplemental Table S3. Quartiles in $\log_{10}$ (Precursor ion intensity) distribution shown in Figure 5

(a) Quartiles in $\log_{10}$ (Precursor ion intensity) distribution for top 100% precursor ions (Figure 5c)

Top 100%	$\log_{10}$ (Precursor ion intensity)		
	Prot-RP	SDS-PAGE	Pept-RP
3rd quartile	6.80	6.84	6.80
1st quartile	6.07	6.05	6.06
Interquartile range (IQR)	0.73	0.79	0.74
Relative IQR to Prot-RP	–	108	101

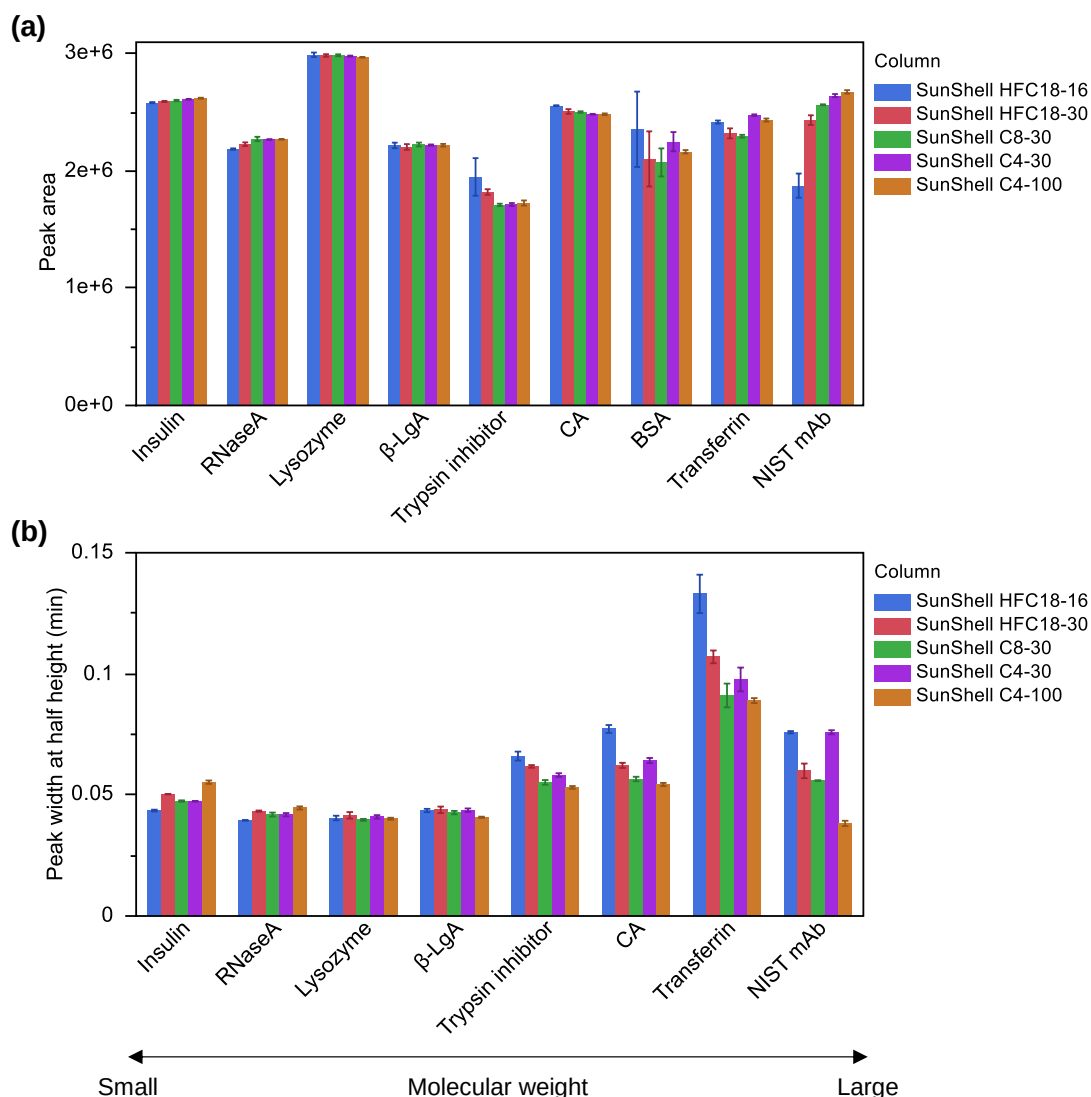
(b) Quartiles in $\log_{10}$ (Precursor ion intensity) distribution for top 10% precursor ions (Figure 5d)

Top 10%	$\log_{10}$ (Precursor ion intensity)		
	Prot-RP	SDS-PAGE	Pept-RP
3rd quartile	7.35	7.35	7.32
1st quartile	6.95	6.87	7.02
Interquartile range (IQR)	0.40	0.48	0.30
Relative IQR to Prot-RP	–	120	75



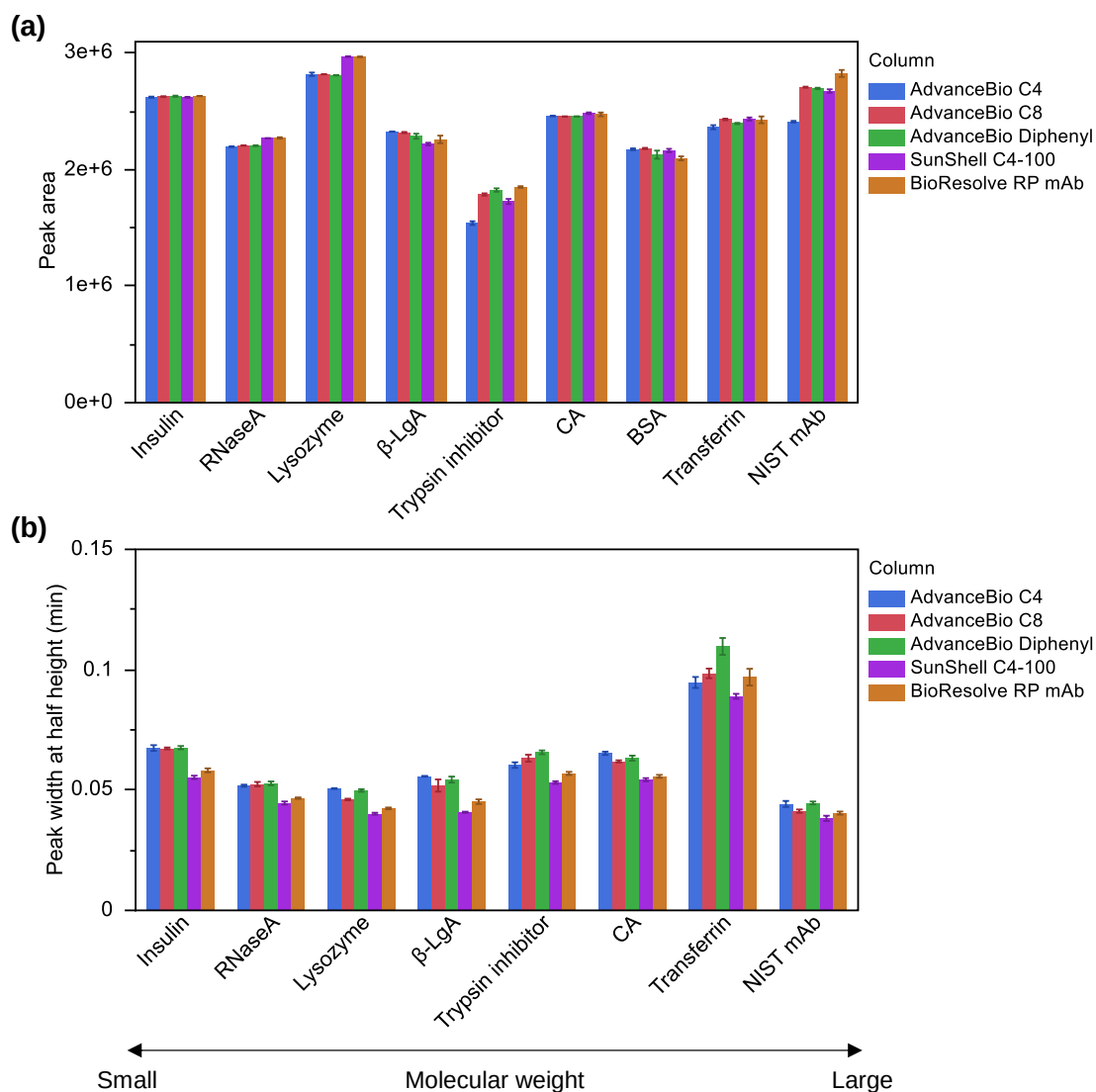
Supplemental Figure S1. Materials used for column screening.

(a) Typical chromatogram of the standard proteins obtained using BioResolve RP mAb. Descriptions of the peak numbers are given in Table S2. (b) Molecular weight distribution of human canonical proteins registered in Uniprot. Red dashed lines indicate the molecular weights of standard proteins employed for the column screening study.



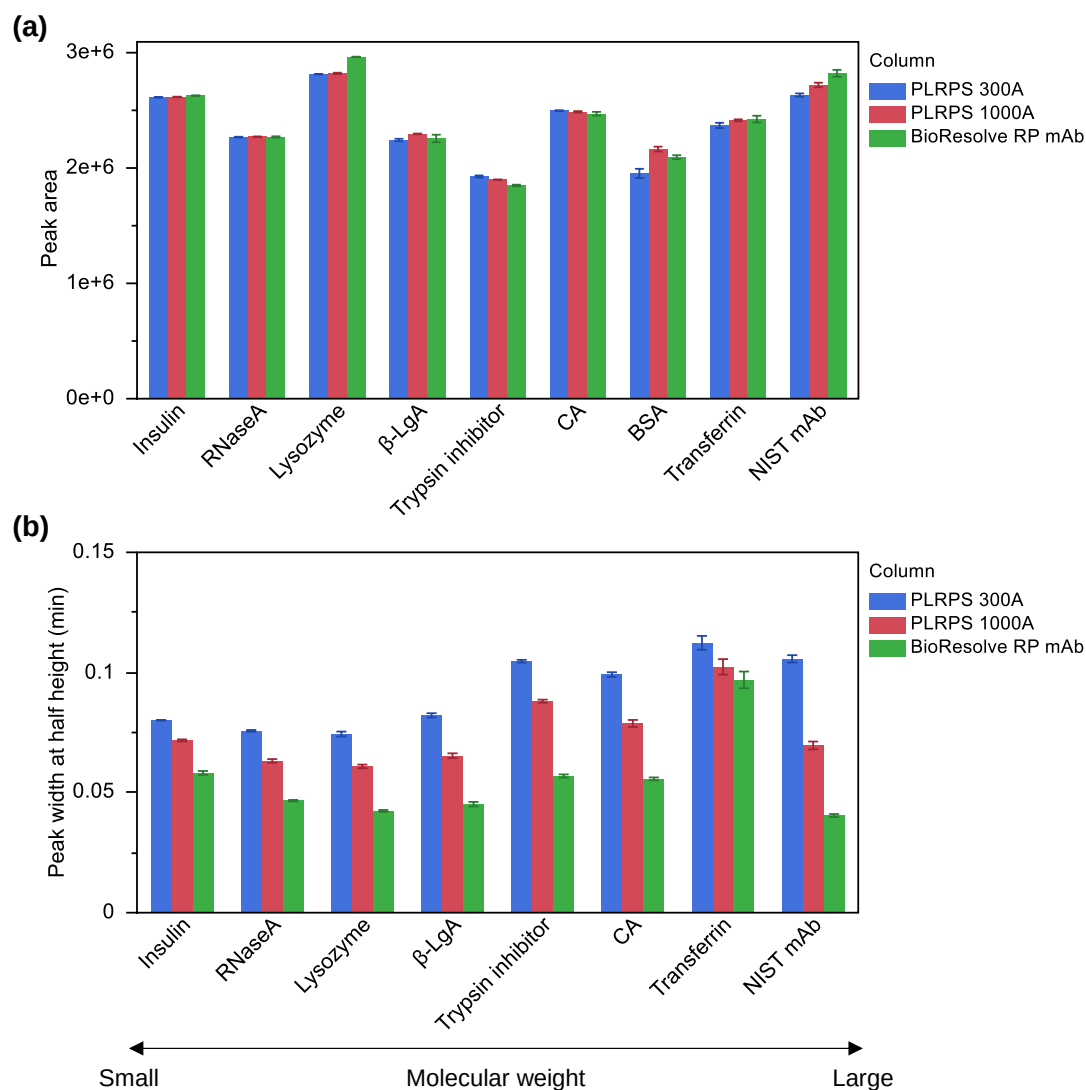
Supplemental Figure S2. Screening results for SunShell columns with various modifications and pore sizes.

Peak area (a) and peak width at half height (b) of standard proteins obtained from SunShell brand columns. Details of columns are provided in Table S1. Error bars represent standard deviation of $n = 3$ measurements. Peak width at half height of BSA was not calculated since it exhibited a complex peak shape (multiple peak tops and shoulders) and was considered inappropriate for inclusion in the evaluation of separation efficiency. Dunnett's test employing SunShell C4-100 as the control detected the following statistically significant differences (family wise error rate (FWER) < 0.05). NISTmAb (peak width at half height): Between all columns, NISTmAb (peak area): Between SunShell HFC18-16 and SunShell HFC18-30, Transferrin (peak width at half height): Between SunShell HFC18-16 and SunShell HFC18-30. The statistical analysis was performed using JMP ver. 17.0.0 (SAS Institute).



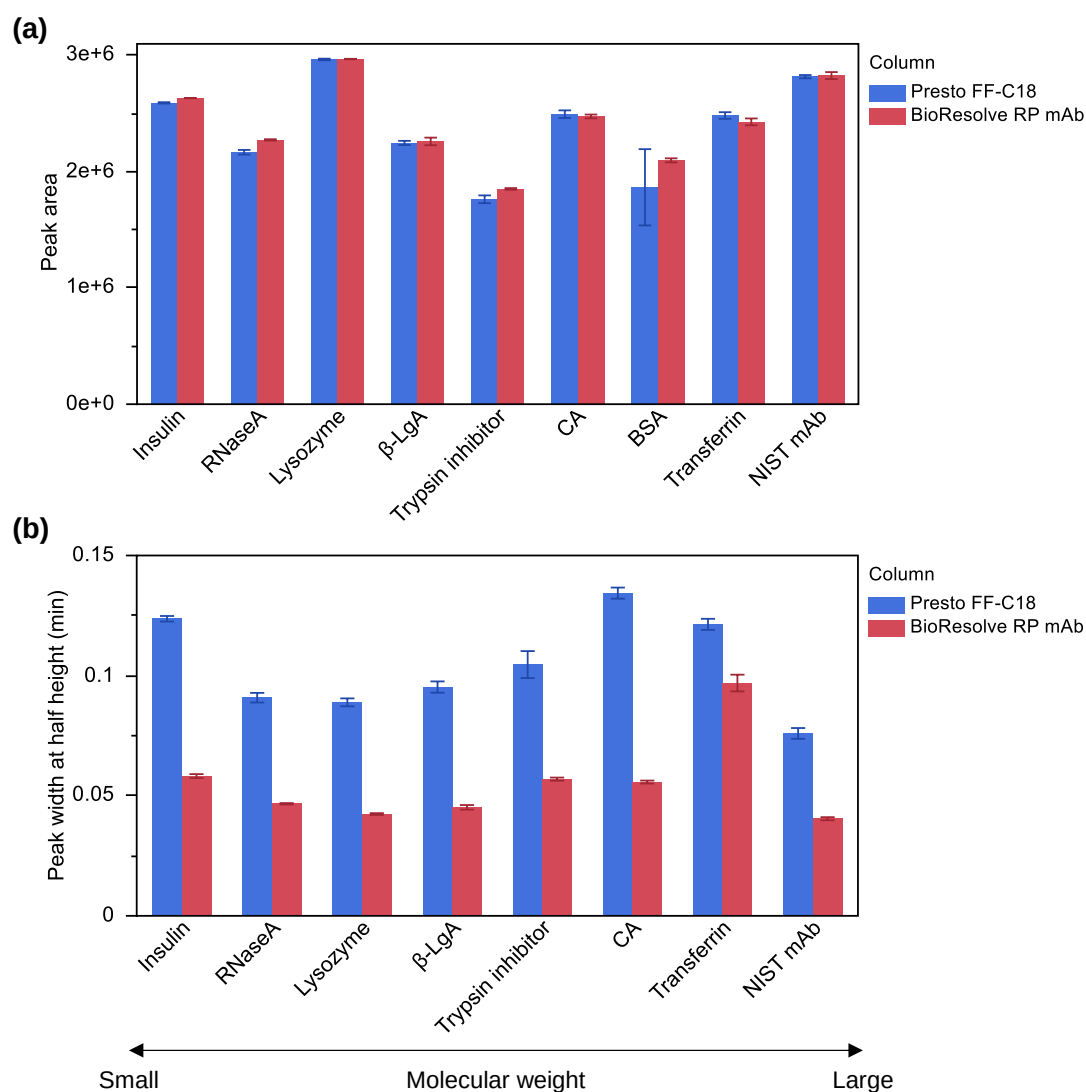
Supplemental Figure S3. Screening results for wide pore and short alkyl chain columns.

Peak area (a) and peak width at half height (b) of standard proteins obtained from SunShell C4-100 and RP columns designed for biopharmaceuticals. Error bars represent standard deviation of $n = 3$ measurements. BSA was excluded from the evaluation of separation efficiency for the reason stated in the legend of Fig. S2. Dunnett's test employing BioResolve RP mAb as the control detected the following statistically significant differences (FWER < 0.05). NISTmAb (peak area): Between all columns. The statistical analysis was performed using JMP ver. 17.0.0 (SAS Institute).



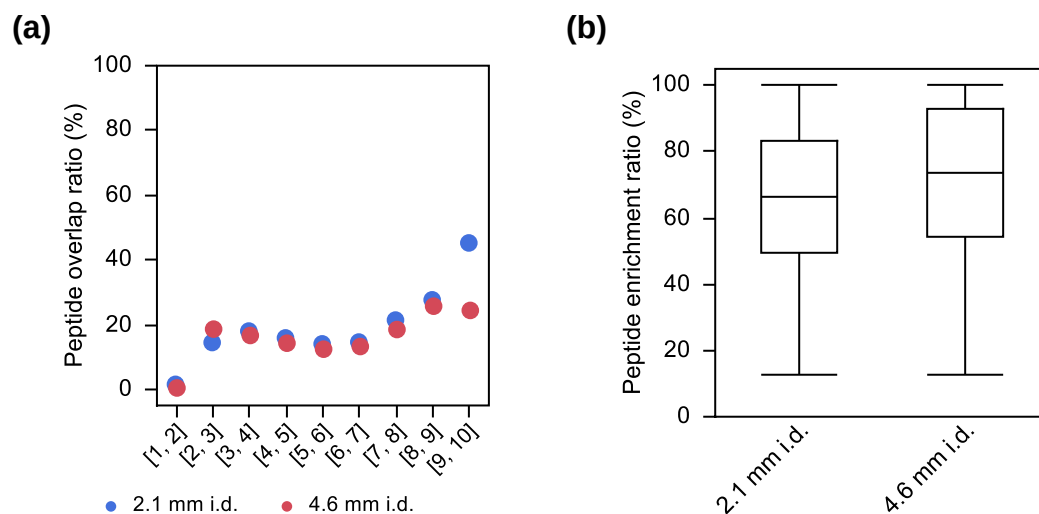
Supplemental Figure S4. Comparison between silica and SDB based columns.

Peak area (a) and peak width at half height (b) of standard proteins obtained from silica (BioResolve RP mAb) and SDB (PLRPS) based columns. Error bars represent standard deviation of $n = 3$ measurements. BSA was excluded from the evaluation of separation efficiency for the reason stated in the legend of Fig. S2.



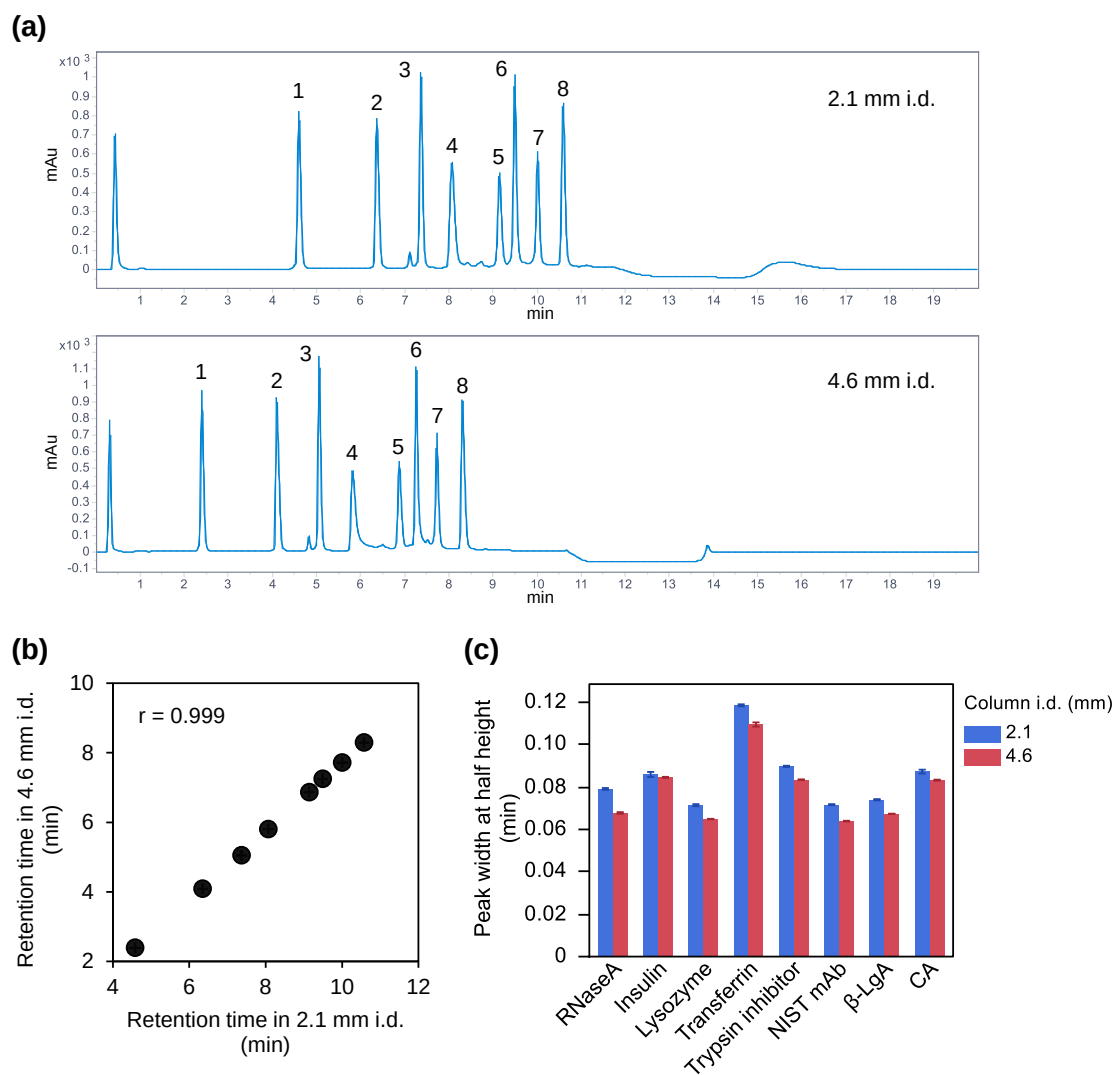
Supplemental Figure S5. Comparison between superficially porous and non-porous particle columns.

Peak area (a) and peak width at half height (b) of standard proteins obtained from columns packed with superficially porous (BioResolve RP mAb) and non-porous (Presto FF-C18) particles. Error bars represent standard deviation of $n = 3$ measurements. BSA was excluded from the evaluation of separation efficiency for the reason stated in the legend of Fig. S2.



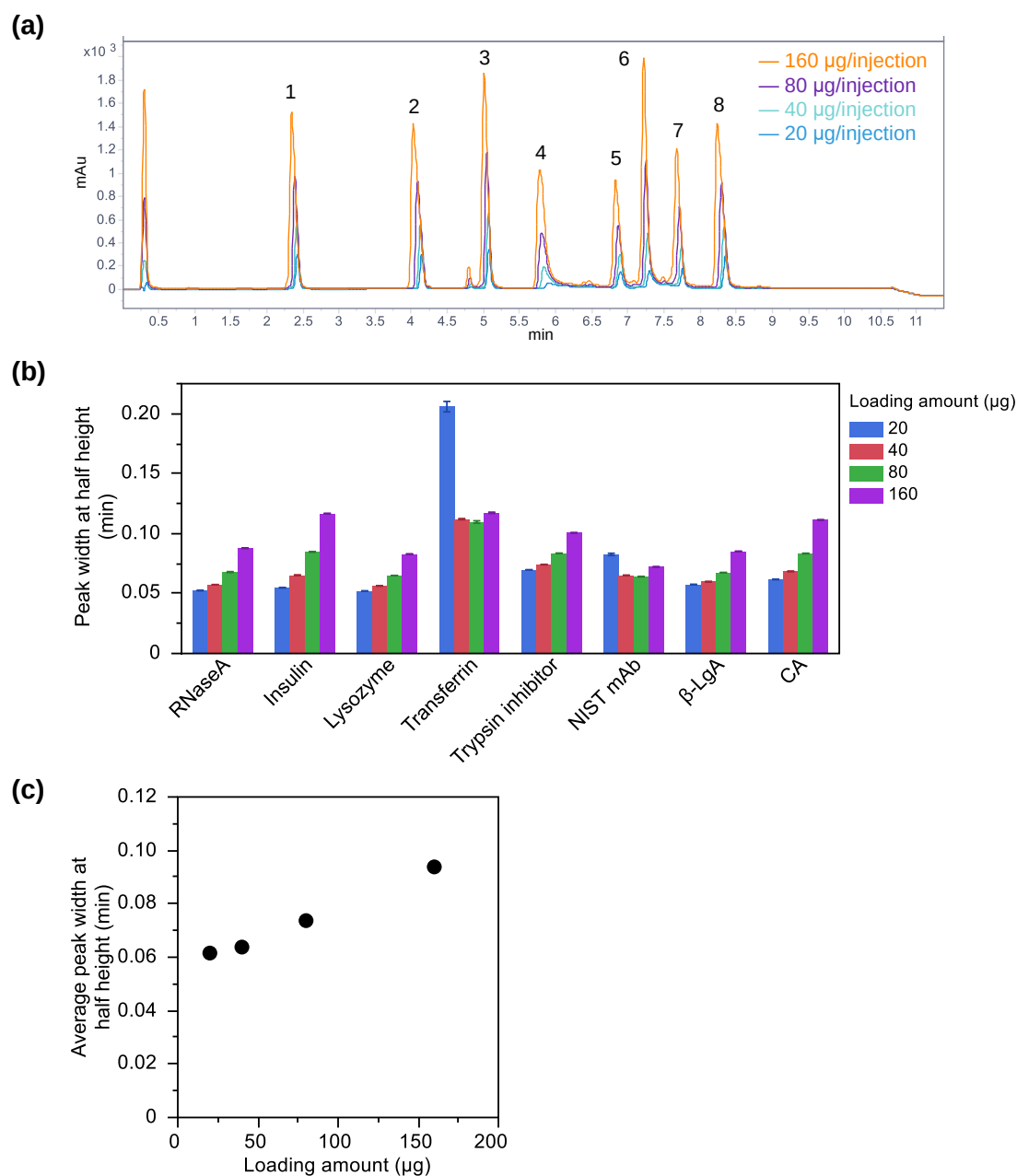
Supplemental Figure S6. Fractionation efficiency observed in the pilot study utilizing columns with different internal diameters.

(a) Overlap of adjacent fractions for peptides. The overlap ratio is defined as No. of commonly identified peptide / No. of peptides identified from 2 fractions. Note that, although the standard deviations of $n = 3$ measurements are shown as error bars, variations are too small to be visible on the figure. (b) Distribution of enrichment ratio for peptides. Average of $n = 3$ experiments was employed for the analysis. The enrichment ratio is defined in the main text, section "Prefractionation workflow". Prefractionation experiments were conducted in a same manner for 2.1 mm and 4.6 mm i.d. columns except for flow rates.



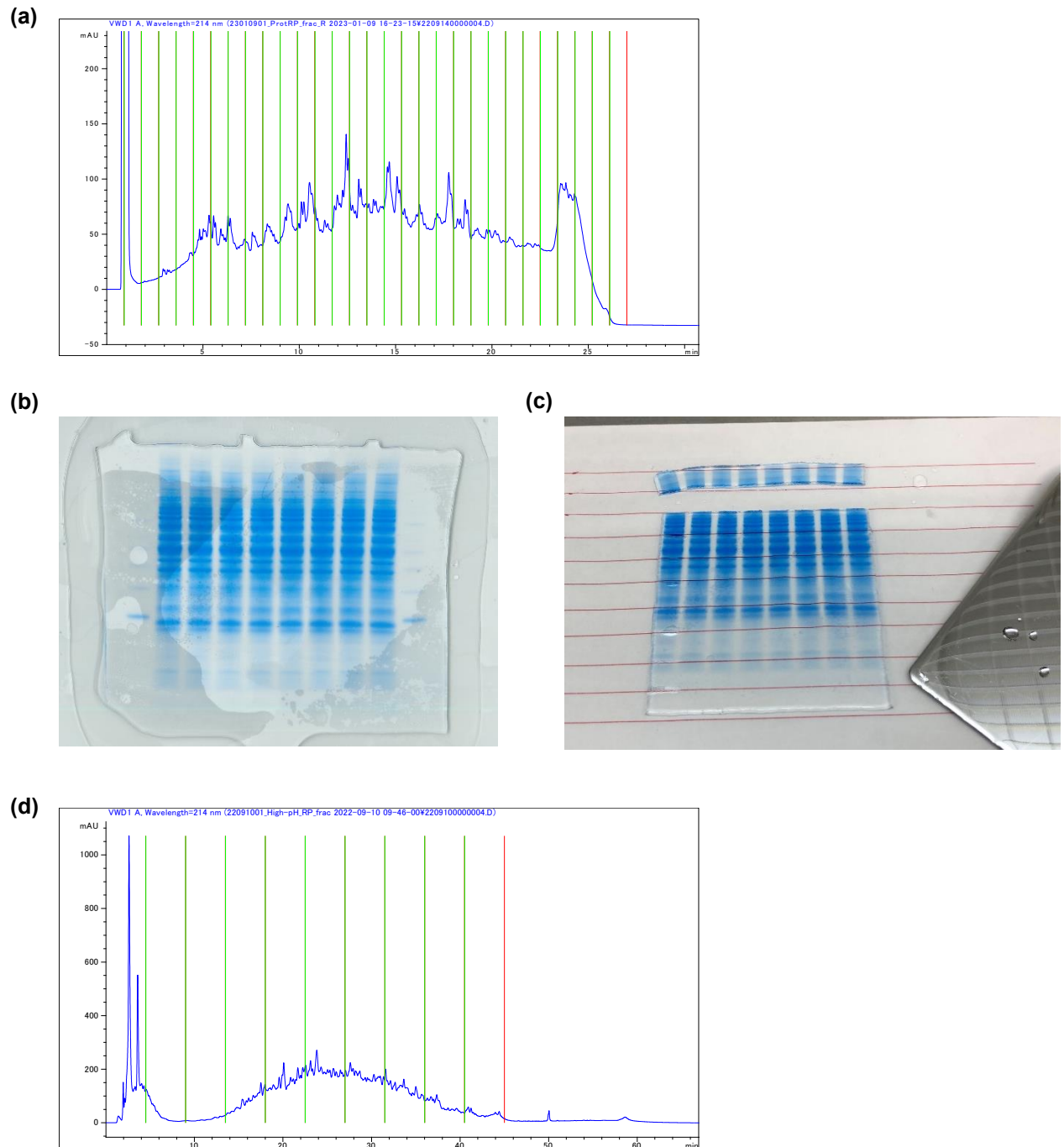
Supplemental Figure S7. Comparison of chromatographic profiles between columns with different internal diameters.

(a) Chromatograms obtained by injecting a standard protein mixture into BioResolve RP mAb columns (2.1 mm i.d. and 4.6 mm i.d.). (Peak No.1: RNaseA, 2: insulin, 3: lysozyme, 4: transferrin, 5: trypsin inhibitor, 6: NIST mAb, 7: β -Lactoglobulin A (β -LgA), 8: carbonic anhydrase (CA)). (b) Retention times for peaks obtained from columns with 2.1 and 4.6 mm i.d. Note that, although the standard deviations (SD) of $n = 3$ measurements are shown as error bars, variations are too small to be visible on the figure. (c) Peak width at half height of each protein for each internal diameter. Error bars represent SD of $n = 3$ measurements.



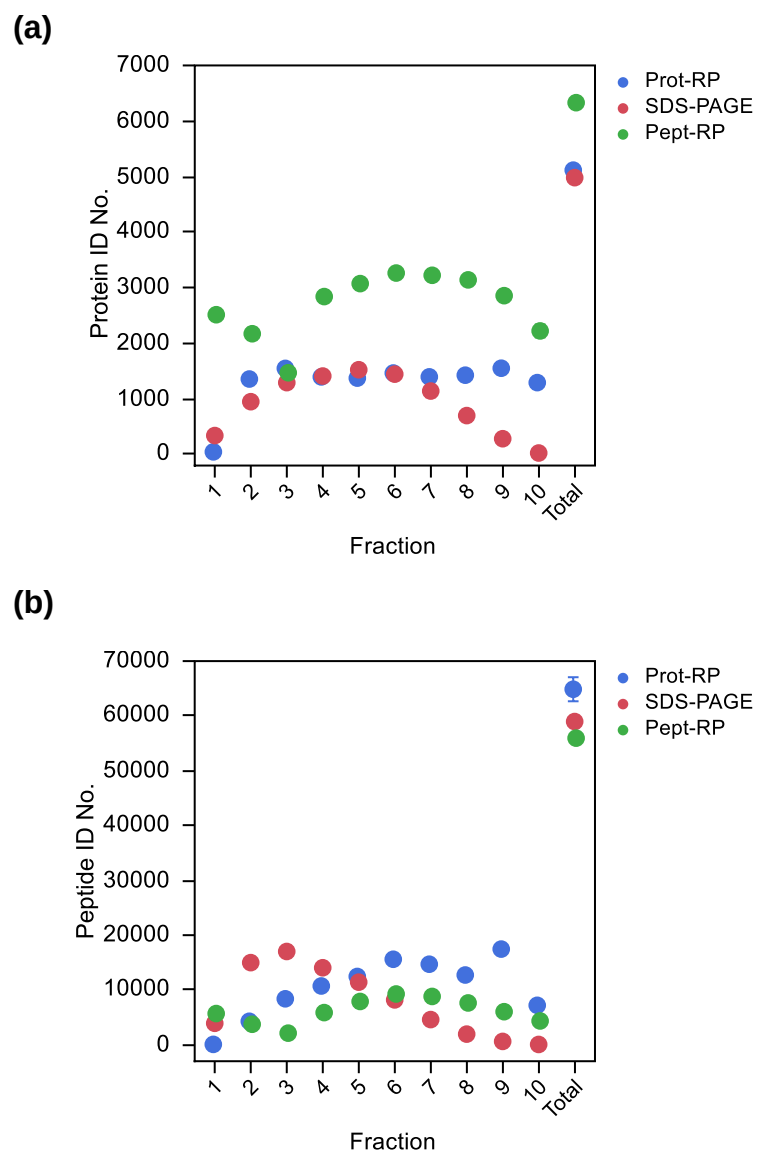
Supplemental Figure S8. Separation performance at various loading amount.

(a) Chromatograms obtained by injecting a standard protein mixture into a BioResolve RP mAb column (4.6 mm i.d.) with loading amounts varying from 20 to 160 µg (Peak No.1: RNaseA, 2: insulin, 3: lysozyme, 4: transferrin, 5: trypsin inhibitor, 6: NIST mAb, 7: β-Lactoglobulin A (β-LgA), 8: carbonic anhydrase (CA)). (b) Peak width at half height of each protein for each injection amount. Error bars represent standard deviation (SD) of $n = 3$ measurements. (c) Mean peak widths at half height were obtained for all proteins except transferrin and plotted against loading amounts (µg). Note that, although the SDs of $n = 3$ measurements are shown as error bars, variations are too small to be visible on the figure.



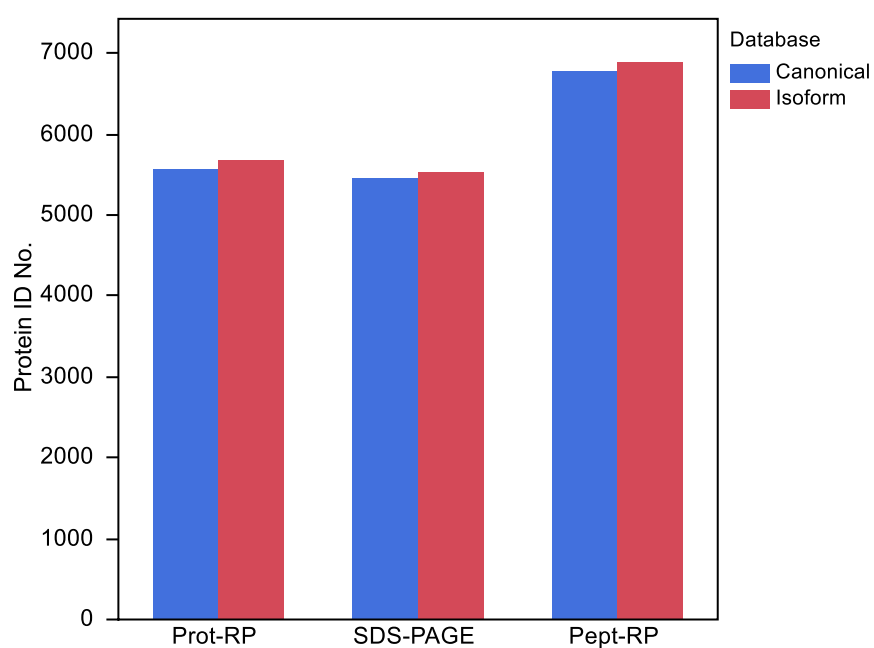
Supplemental Figure S9. Chromatograms and a gel image from each prefractionation method.

(a) Chromatogram of the cell lysates obtained from Prot-RP. (b) Gel images of the cell lysates separated by SDS-PAGE. (c) Gels were first horizontally cut into 10 pieces across 8 lanes to utilize whole proteins loaded on gels. (d) Chromatogram of the cell digests obtained from Pept-RP. Green and red lines in the chromatograms (a, d) represent start and stop times of an automated fraction collector, respectively. “Stop lines” coincide exactly with “Start lines”, except for the final fraction, and thus are not visible for earlier fractions.



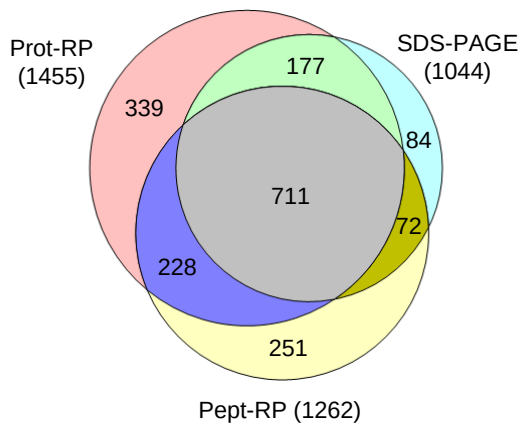
Supplemental Figure S10. Protein and peptide identification numbers for each prefractionation method.

Number of proteins (a) and peptides (b) identified from each fraction, for each of the three prefractionation methods. Error bars represent standard deviation of $n = 3$ experiments.

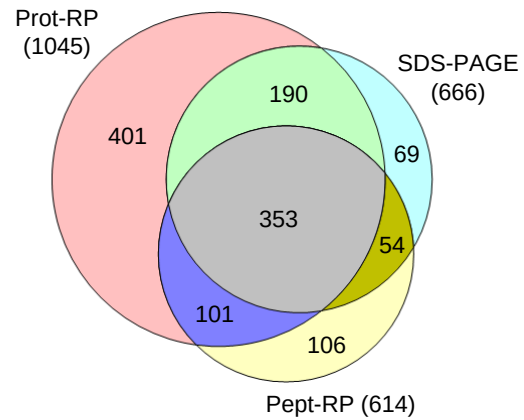


Supplemental Figure S11. Number of proteins identified from “canonical” and “canonical + isoform” databases for each prefractionation method.

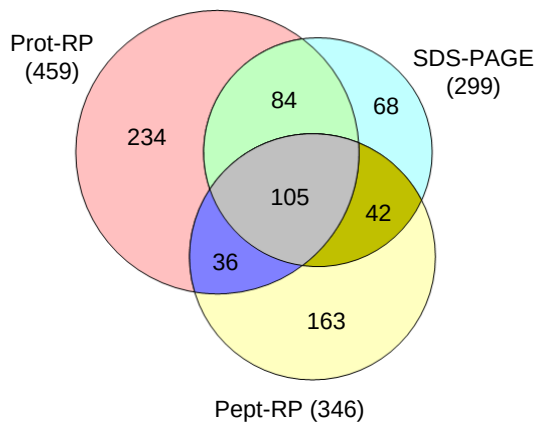
(a) N-terminus



(b) C-terminus

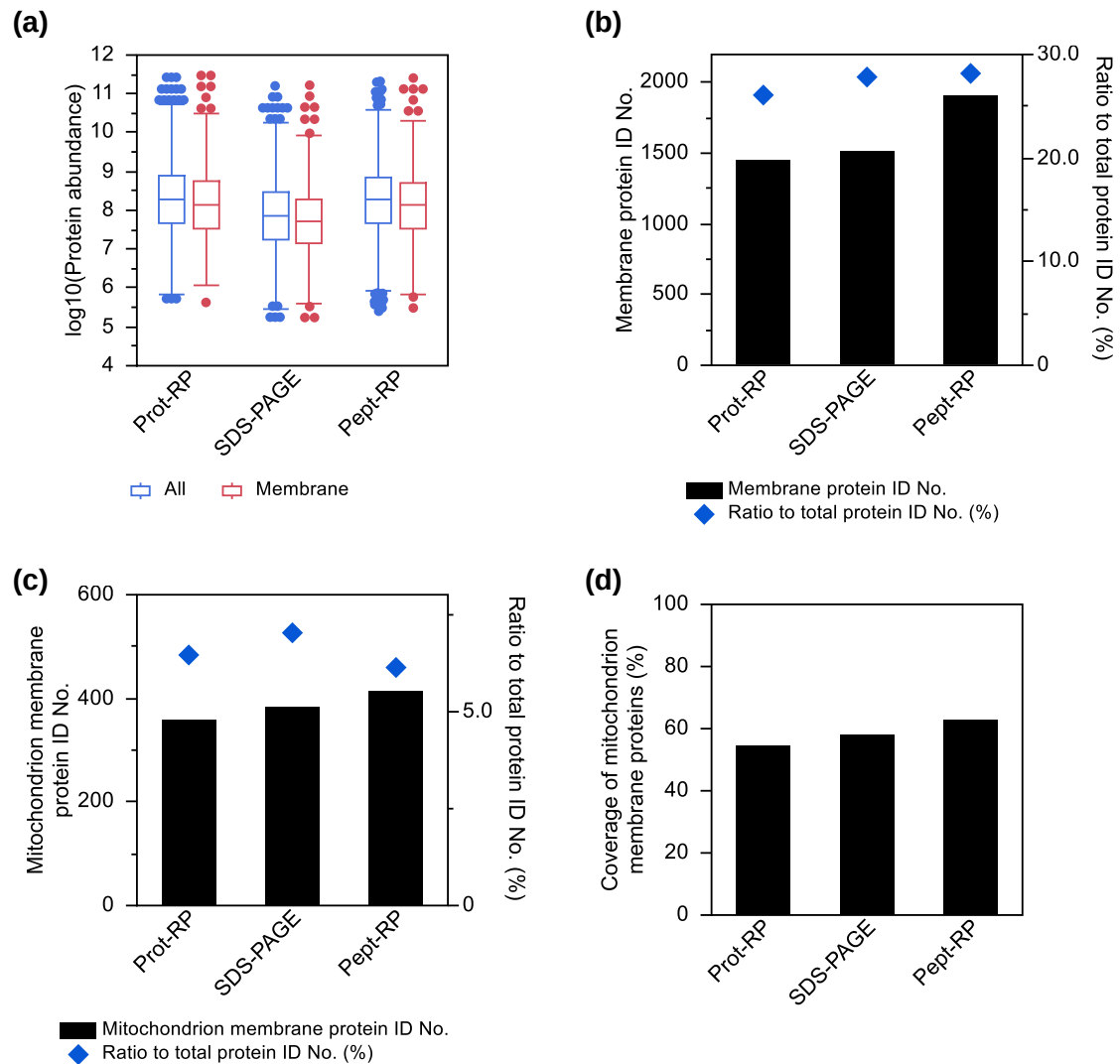


(c) Isoform-specific peptides



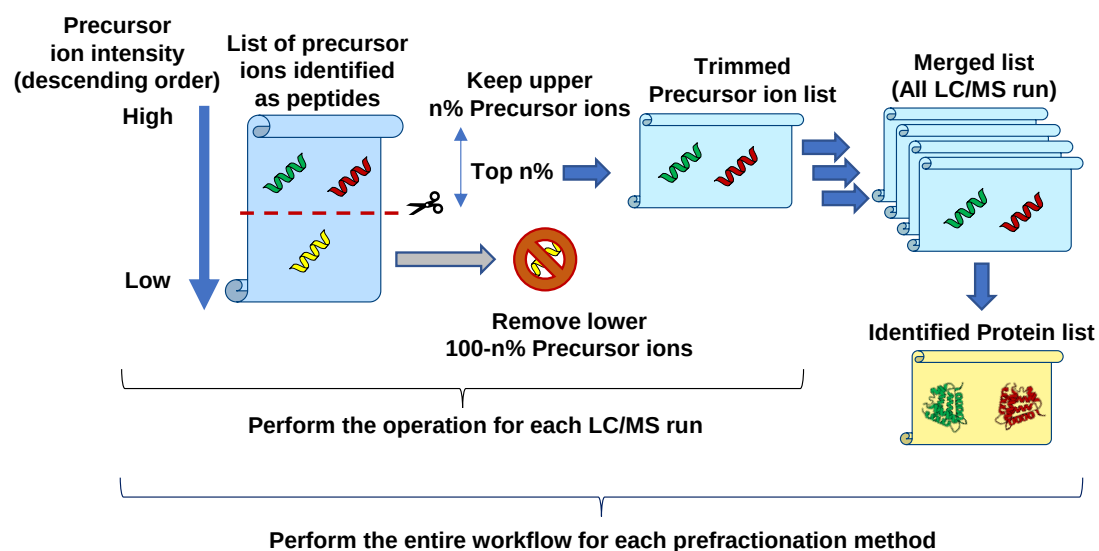
Supplemental Figure S12. Protein terminus and isoforms identified using each prefractionation method.

Venn diagrams for the three prefractionation methods (Prot-RP, SDS-PAGE, and Pept-RP), showing (a) proteins whose N-terminus was identified, (b) proteins whose C-terminus was identified, (c) isoform-specific peptides.



Supplemental Figure S13. Analysis results focused on membrane proteins obtained from each prefractionation method.

(a) Distribution of membrane protein abundances calculated for each prefractionation method. Blue and red color represent all the quantified proteins and membrane proteins, respectively. (b) Number of membrane proteins identified from each prefractionation method (black bar) and ratio to totally identified proteins provided in Fig. 2a (blue marker). (c) Identification number of proteins localized to the mitochondrial membrane for each prefractionation method (black bar) and ratio to totally identified proteins provided in Fig. 2a (blue marker). (d) Coverage (%) of mitochondrion membrane proteins for each prefractionation method. Total number of mitochondrion membrane proteins was 661.



Supplemental Figure S14. Workflow of the simulation analysis.

First, the list of precursor ions identified as peptides for each LC/MS run was sorted in descending order based on their intensity. From the sorted list, precursor ions with the top $n\%$ intensity was retained, and the rest were removed ($n\%$: 1%, 10%, 25%, 50%, 100%). The trimmed lists were merged for each prefractionation method to obtain a list of identified proteins. The images are from TogoTV (2016 DBCLS TogoTV, CC-BY-4.0 <https://creativecommons.org/licenses/by/4.0/>).