Chapter 8

Low-Molecular-Weight Plasma Proteome Analysis Using Top-Down Mass Spectrometry

Dong Huey Cheon, Eun Gyeong Yang, Cheolju Lee, and Ji Eun Lee

Abstract

While human plasma has a wealth of diagnostic information regarding the state of the human body in heath and disease, low molecular weight (LMW) proteome (<30 kDa) has been shown to contain a rich source of diagnostic biomarkers. Here we describe a protocol for top-down proteomic analysis to identify and characterize the LMW proteoforms present in four types of human plasma samples without immuno-affinity depletion and with depletion of the top two, six, and seven high-abundance proteins. Each type of plasma sample was first fractionated based on molecular weight using gel-eluted liquid fraction entrapment electrophoresis (GELFrEE). Then, the GELFrEE fractions containing up to 30 kDa were subjected to nanocapillary–LC–MS/MS, and the high-resolution MS and MS/MS data were processed using ProSightPC software. As a result, a total of 442 LMW proteins and cleaved products, including those with posttranslational modifications (PTMs) and single amino acid variations (SAAVs), were identified with a threshold E-value of 1×10^{-4} from the four types of plasma samples.

Key words Human plasma proteome, Low molecular weight, Top-down mass spectrometry, Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), Posttranslational modification (PTM), Single amino acid variation (SAAV), Proteoforms

1 Introduction

While human plasma serves as an invaluable source for disease diagnosis, low molecular weight (LMW) plasma proteome (<30 kDa), which is composed of either small proteins such as hormones, cytokines, and growth factors or peptides derived from the proteolytic degradation of larger proteins, has attracted attention in the field of biomarker discovery [1–3]. The possibility of the LMW components as diagnostic biomarkers was initially observed from peak profiling experiments using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) or surface-enhanced laser desorption/ionization (SELDI)-TOF MS [4, 5]. Although the profiling platform based on MALDI– or SELDI–TOF MS has been combined with tandem

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_8, © Springer Science+Business Media LLC 2017

mass spectrometry for identification of the LMW species [6, 7], it usually focuses on identifying the LMW components showing the differential changes of peak profiles between the control and disease states, which does not result in identifying a great number of LMW components present in the plasma or serum samples.

The enrichment strategies such as centrifugal ultrafiltration followed by bottom-up mass spectrometric analyses have shown to effectively identify a large number of LMW proteins from the plasma or serum samples [8–10]. Hundreds of proteins belonging to LMW plasma or serum proteome were successfully identified from the bottom-up proteomic analyses of the LMW fraction. However, there is a limit in discovering LMW proteins that undergo posttranslational modifications (PTMs) and endogenous proteolytic cleavages associated with disease states because the enzymatic digestion used for bottom-up analysis eliminates the information of intact proteins that naturally occur in plasma or serum.

Top-down mass spectrometric analysis in which intact proteins are directly ionized and fragmented in a mass spectrometer enables a full characterization of the primary structure of a protein and therefore can differentiate diverse protein isoforms called proteoforms, arising from genetic variations, alternative splicing, endogenous proteolysis, and PTMs [11]. While there have been great advances in top-down proteomics for analyzing complex protein mixtures [12, 13], the technologies have not yet been widely applied to clinical samples such as plasma or serum. Here, we report a protocol of top-down mass spectrometric analysis of LMW proteome (<30 kDa) present in four types of human plasma samples without immunoaffinity depletion and with depletion of the top two, six, and seven high-abundance proteins based on our recent work [14]. Prior to top-down MS, the four types of plasma samples were fractionated using continuous tube gel electrophoresis, known as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), in which proteins are constantly eluted from a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) tube gel column based on molecular weight (MW) and collected in liquid form [15, 16]. Then, the GELFrEE fractions containing up to 30 kDa were subjected to nanocapillary–LC–MS/MS, and the high-resolution MS and MS/MS data were processed using ProSightPC software, resulting in identification of 442 LMW proteoforms with molecular weight ranges of 1.2–28 kDa.

2 Materials

All solutions were prepared using HPLC grade water.

2.1 Plasma Preparation

1. Protease and phosphatase inhibitor cocktail tablets.
2. Bicinchoninic acid (BCA) assay kit.

2.2 Depletion of High-Abundance Proteins from Plasma Samples

1. ProteoExtract™ Albumin/IgG removal kit (CALBIOCHEM): albumin removal column, immunoglobulin G (IgG) removal column, 10× binding buffer (250 mM sodium phosphate, pH 7.4), elution buffer for albumin (25 mM sodium phosphate, pH 8.0, 2 M NaCl), and elution buffer for IgG (250 mM citric acid).
2. MARS-6 and MARS-7 columns (4.6 × 50 mm, Agilent Technologies).
3. Buffer A (salt-containing neutral buffer, pH 7.4, Agilent Technologies) and Buffer B (urea buffer, pH 2.2, Agilent Technologies) used for the removal of the top six and seven high-abundance plasma proteins.
4. HPLC system: Agilent 1100 series (Agilent Technologies).
5. Amicon Ultracel-3 centrifugal filter: 3 kDa cutoff.
6. Reduction buffer (1 M dithiothreitol (DTT) stock solution): 0.1542 g of DTT was dissolved in 1 mL of water.
7. Alkylation buffer (0.5 M iodoacetamide (IAA) stock solution): 0.0925 g of IAA was dissolved in 1 mL of water.
8. 0.22 μm polyvinylidene fluoride (PVDF) centrifugal filter.

2.3 GELFrEE Fractionation

1. Glass tube: 6 mm o.d. × 6.0 cm.
2. Resolving gel buffer (1.5 M Tris–HCl, pH 8.8): 181.71 g of Tris base was dissolved in 900 mL of water and 6 M HCl was added to adjust pH. Then, water was added to make a final volume of 1 L.
3. Stacking gel buffer (0.5 M Tris–HCl, pH 6.8): 60.57 g of Tris base was dissolved in 900 mL of water and 6 M HCl was added to adjust pH. Then, water was added to make a final volume of 1 L.
4. 5× sample buffer: 1 g of sodium dodecyl sulfate (SDS) and 0.05 g of bromophenol blue were added to 5 mL of 0.5 M Tris–HCl (pH 6.8), and then 5 mL of glycerol was finally added to the solution.
5. GELFrEE running buffer: 0.025 M Tris, 0.192 M glycine, and 0.1% SDS.
6. 30% acrylamide/Bis solution.
7. 10% ammonium persulfate (APS): 1 g of APS was dissolved in 10 mL of water.
8. Tetramethylethylenediamine (TEMED).
9. Eight-channel multiplexed device consisting of a cathode chamber, eight gel columns, a collection chamber for each gel column, and an anode chamber (*see* **Note 1**).

2.4 Liquid Chromatography–Mass Spectrometry

1. NanoLC 2D system (Eksigent Technologies).
2. Mobile phase A: 0.2% formic acid and 99.8% water.
3. Mobile phase B: 0.2% formic acid and 99.8% acetonitrile.
4. Orbitrap XL mass spectrometer (Thermo Fisher Scientific).
5. PLRP-S 1000 Å 5 μm resin (Agilent Technologies).
6. PicoTip emitters (New Objective).

2.5 MS Data Analysis

1. ProSightPC 3.0: search engine for protein identification and characterization (Thermo Fisher Scientific).

3 Methods

The workflow for top-down mass spectrometric analysis of the LMW proteome present in human plasma samples is seen in Fig. 1. The current protocol includes plasma sample preparation without immunoaffinity depletion and with depletion of the top two, six, and seven high-abundance proteins, molecular weight-based separation, nanocapillary–LC–MS/MS, and high-resolution MS and MS/MS data processing for identification of proteoforms.

3.1 Human Plasma Sample Preparation

1. Protease and phosphatase inhibitor cocktails were added to individual plasma sample (*see* **Note 2**).
2. Protein concentration of the plasma sample was determined using the bicinchoninic acid (BCA) method according to the manufacturer's instructions (*see* **Note 3**).

3.2 Depletion of High-Abundance Proteins

For top-down proteomic analysis of plasma samples with the removal of the top two high-abundance proteins (albumin and IgG), the top six high-abundance proteins (albumin, IgG, immunoglobulin A (IgA), serotransferrin (TRFE), haptoglobin (HPT), and alpha-1 antitrypsin (A1AT), and the top seven high-abundance proteins (albumin, IgG, IgA, TRFE, HPT, A1AT, and fibrinogen) high-abundance proteins, the human plasma sample was depleted of its high-abundance proteins using three different immunoaffinity columns.

3.2.1 Depletion of Top Two High-Abundance Proteins

1. 180 μL of plasma was diluted with 180 μL of 10× binding buffer, which was provided by the manufacturer, and 1440 μL of water.
2. After a syringe was filled with 6 mL of 1× binding buffer, which was diluted from 10× binding buffer using water, without introducing air bubble, it was connected to an albumin removal column. Then, gentle pressure was applied so that 1× binding buffer was passed through each column for column equilibration at a flow rate of 0.25 mL/min. The process was repeated for the other albumin removal column and an IgG removal column. Two albumin and one IgG removal columns were connected together prior to sample loading. The flow-through fraction was discarded.

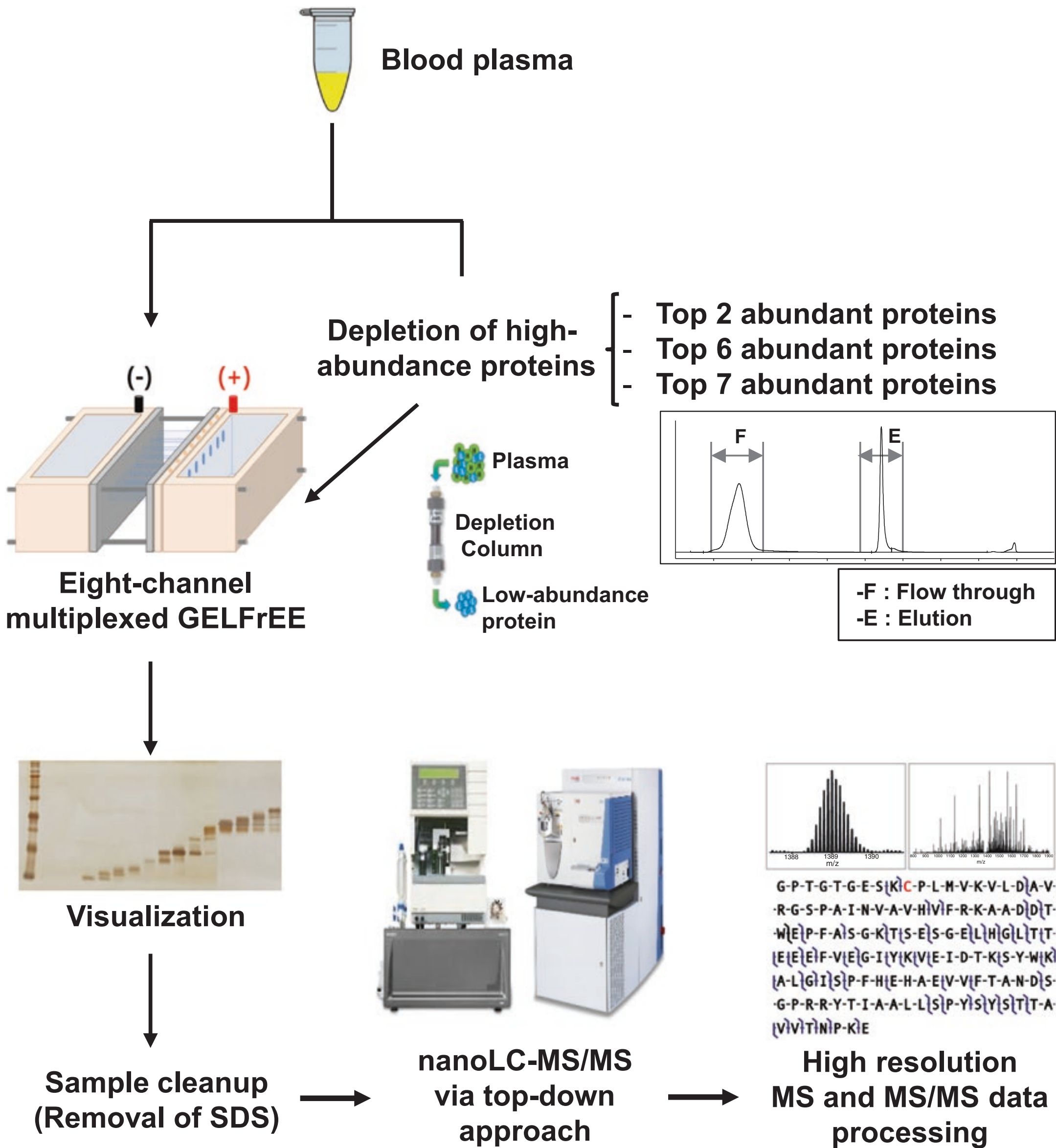

Fig. 1 Workflow for top-down mass spectrometric analysis of four types of human plasma samples. Four types of human plasma without depletion of high-abundance proteins and with depletion of the top two, six, and seven abundant proteins were fractionated using eight-channel multiplexed GELFrEE. The GELFrEE fractions containing up to 30 kDa were subjected to nanoLC-MS/MS, and the resulting high resolution MS and MS/MS data are processed using ProSightPC software tailored for top-down analysis. Reproduced from [14] with permission from the publisher

3. A new syringe was filled with the diluted sample and connected to the removal columns. Then, gentle pressure was applied so that the diluted sample was slowly loaded to the removal columns at a flow rate of 0.1 mL/min. The flow-through fraction was collected.
4. After a syringe filled with 6 mL of 1× binding buffer was connected to the removal columns, gentle pressure was applied so

that the binding buffer was passed through the removal columns at a flow rate of 0.25 mL/min. The flow-through fraction was also collected and combined with the flow-through fraction that was previously collected.

5. The albumin and IgG removal columns were disconnected prior to elution of bound proteins. After a new syringe filled with 6 mL of albumin elution buffer was connected to the albumin removal columns, gentle pressure was applied so that the albumin elution buffer was passed through them at a flow rate of 0.25 mL/min. When the bound proteins need to be analyzed, the eluted fraction can be collected. This process was also repeated for the IgG removal column using 3 mL of IgG elution buffer. Then, the individual removal columns were equilibrated with 2 mL of 1× binding buffer and stored at 4 °C.

3.2.2 Depletion of Top Six and Seven High-Abundance Proteins

1. Buffers A and B, which were provided by the manufacturer, were filtered using 0.45 μm regenerated cellulose membrane and then degassed for 10 min.
2. 20 μL of plasma was diluted with 80 μL of buffer A and then filtered using 0.22 μm polyvinylidene fluoride (PVDF) centrifugal filter at 16,000 × *g* for 1 min.
3. MARS-6 or MARS-7 column was connected to Agilent 1100 HPLC system and equilibrated with buffer A for 20 min at a flow rate of 1 mL/min.
4. 100 μL of the diluted plasma sample was injected onto a MARS-6 or MARS-7 column at a flow rate of 0.25 mL/min and then separated using the following gradient conditions: 0 min 100% buffer A (0.25 mL/min), 0–9 min 100% buffer A (0.25 mL/min), 9–9.01 min 100% buffer B (1 mL/min), 9.01–12.5 min 100% buffer B (1 mL/min), 12.5–12.6 min 100% buffer A (1 mL/min), and 12.6–20 min 100% buffer A (1 mL/min).
5. The flow-through fraction, which was not bound to the MARS-6 or MARS-7 column, was collected (*see* **Note 4**).
6. Another 100 μL of the diluted plasma sample was injected again onto a MARS-6 or MARS-7 column and repeated for the removal of the top six or seven high-abundance proteins.
7. When the depletion process was done, MARS-6 or MARS-7 column was equilibrated with buffer A for 7.4 min at a flow rate of 1 mL/min and kept at 4 °C.

3.2.3 Sample Enrichment

1. The flow-through fractions obtained from the depletion of top two, six, and seven high-abundance proteins were diluted with threefold with 10 mM Tris–HCl (pH 7.5) and concentrated using Amicon Ultracel-3 centrifugal filters (3 kDa cutoff). Prior to concentration of each depleted plasma sample, a centrifugal filter device was first rinsed two times with 450 μL of

water 14,000 × *g* for 10 min and conditioned two times with 450 μL of 10 mM Tris–HCl (pH 7.5) at 14,000 × *g* for 10 min. The concentrated samples having approximately 100 μL were recovered by inverting the filter device and centrifuged at 2000 × *g* for 2 min.

2. The protein concentration of the concentrated samples was determined using the BCA assay kit.

3.3 GELFrEE Fractionation of Four Types of Plasma Samples

3.3.1 GELFrEE Fractionation

1. One end of a glass tube was tightly covered with Parafilm (2 cm × 2 cm).
2. In order to cast 17.5% T for the resolving gel, 1.7 mL of water, 2.5 mL of 1.5 M Tris–HCl (pH 8.8), and 5.8 mL of 30% acrylamide/Bis (37.5:1) solution were first mixed in a 15 mL falcon tube, and then 50 μL of 10% APS solution and 5 μL of TEMED were finally added. Then, 849 μL of the resolving gel buffer was slowly added to the glass tube in order to make a 3 cm length of resolving gel, and 100 μL of 2-methyl-2-butanol was immediately added to the top layer of the resolving gel.
3. After the resolving gel was polymerized, 1.5 cm-long stacking gel was cast to 4% T. 6.1 mL of water, 2.5 mL of 0.5 M Tris–HCl (pH 6.8), and 1.3 mL of 30% acrylamide/Bis solution (37.5:1) were first mixed in a 15 mL falcon tube, and then 50 μL of 10% APS solution and 10 μL of TEMED were finally added. After the residual 2-methyl-2-butanol over the resolving gel was removed by inverting the glass tube on clean absorbent paper, 300 μL of the stacking gel buffer was slowly added to the top of the resolving gel, and 100 μL of 2-methyl-2-butanol was immediately added to the top layer of the stacking gel. After the stacking gel was polymerized, the residual 2-methyl-2-butanol was also removed as described above (*see* **Note 5**).
4. 350 μg of the depleted plasma sample was diluted to 77 μL using Tris–HCl (pH 7.5). Then, 0.5 μL of 1 M DTT solution was added to the depleted plasma sample followed by incubation for 35 min at 56 °C to reduce cysteine residues. Then, 2.5 μL of 500 mM IAA solution was added to the depleted plasma sample followed by incubation for 30 min at room temperature in the dark for alkylation (*see* **Note 6**).
5. 80 μL of the plasma sample mixed with 5× sample buffer and heated for 10 min at 95 °C. Then, 100 μL of each sample was loaded onto a SDS-polyacrylamide tube gel column (*see* **Note 7**).
6. The cathode and anode chambers of eight-channel multiplexed GELFrEE device were filled with fresh running buffer. 150 μL of running buffer was also added to each collection chamber of the GELFrEE device. Then, the eight-channel multiplexed GELFrEE device was operated with a constant application of 240 V in a stop and go cycle, collecting fractions from each gel columns at defined time points by transferring the solution in

each collection chamber to a siliconized microcentrifuge tube (*see* **Note 8**). After collecting the fractions at each time point, 150 μL of fresh running buffer was introduced into each collection chamber, and the power supply was resumed to continue separation. 16 to 18 GELFrEE fractions containing up to 30 kDa were collected for each type of plasma sample.

7. Each fractionation was visualized by silver staining of an SDS-PAGE slab gel with 8 μL of each 150 μL GELFrEE fraction (Fig. 2).

3.3.2 Sample Processing for Top-Down Mass Spectrometric Analysis

1. The GELFrEE fractions with similar molecular weight ranges collected from eight to 24 channel replicates of GELFrEE were typically combined and concentrated using an Amicon Ultracel-3 centrifugal filter (*see* **Note 9**). Prior to concentra-

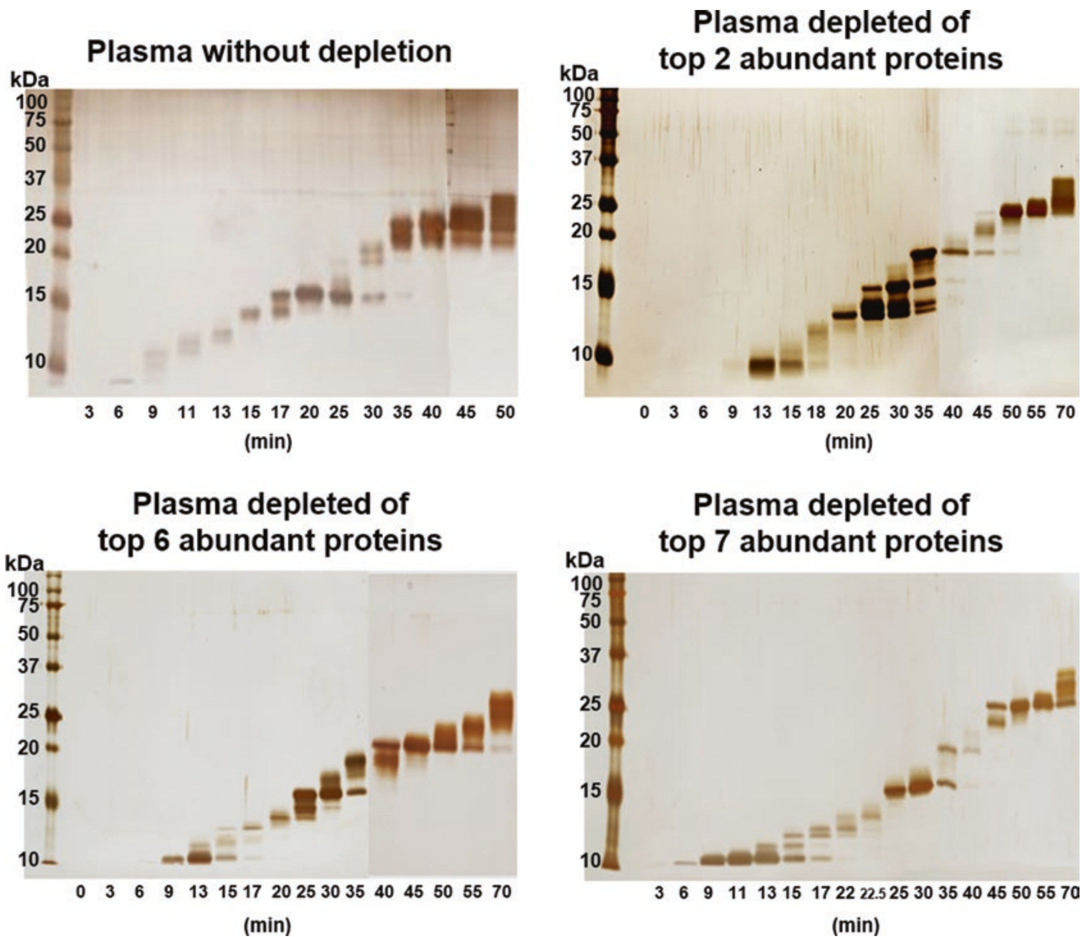

Fig. 2 Slab gel visualizations of GELFrEE fractionation for LMW proteome (<30 kDa) present in four types of human plasma samples. Plasma samples without depletion of high-abundance proteins and with depletion of the top two, six, and seven abundant proteins were fractionated using 17.5% tris-glycine GELFrEE. Reproduced from [14] with permission from the publisher

tion of the fractionated samples, the centrifugal filter devices were rinsed two times with 450 μL of water at 14,000 × *g* for 10 min and conditioned two times with 450 μL of 10 mM Tris–HCl (pH 7.5) at 14,000 × *g* for 10 min. GELFrEE fractions were diluted threefold with 10 mM Tris–HCl (pH 7.5) and centrifuged at 14,000 × *g* for 10 min. The concentrated solute having approximately 100 μL was recovered by inverting the filter device and centrifuged at 2000 × *g* for 2 min.

2. For the removal of SDS from the fractionated samples, the concentrated samples were precipitated using chloroform/methanol/water precipitation [17]. 400 μL of methanol was added to each sample and vortexed vigorously for 1 min. 100 μL of chloroform was then added and vortexed vigorously again. 300 μL of water was added and vortexed again. The samples were centrifuged at 16,000 × *g* for 20 min. After centrifugation, the top aqueous/methanol layer was carefully pipetted off and discarded, while the protein pellet over chloroform layer was not touched. 400 μL of methanol was added to the protein pellet over chloroform layer and gently mixed with the protein pellet. Then, the samples were centrifuged again at 16,000 × *g* for 20 min. The supernatant was carefully removed, while the protein pellet was not disturbed. Then, 400 μL of methanol was added again to the protein pellet and gently mixed with the protein pellet. Then, the samples were centrifuged again. After the supernatant was removed, residual solvent was allowed to dry in a fume hood. Then, the protein pellets were resuspended with 20 μL of solution consisting of solution 0.2% formic acid, 94.8% water, and 5% acetonitrile.

3.4 Liquid Chromatography–Mass Spectrometry

1. Trap (150 μm i.d. × 3 cm) and analytical (75 μm i.d. × 10 cm) columns were packed with PLRP-S media (1000 Å, 5 μm).
2. Typically, 5–10 μL of sample was injected onto a trap column using an autosampler and separated on an analytical column (1000 Å, 5 μm) with 350 nL/min. The typical gradient conditions: 0 min 95% buffer A (100% water with 0.2% formic acid) and 5% buffer B (100% acetonitrile with 0.2% formic acid), 0–5 min 5–20% B, 5–10 min 20–21% B, 10–55 min 21–30% B, 55–70 min 30–40% B, 70–78 min 40–52% B, 78–83 min 52–85% B, 83–88 min 85–5% B, and 88–100 min at 5% B.
3. Data were collected on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) using the Orbitrap mass analyzer with AGC targets of 1×10^6 for MS (4–16 microscans, 60,000 or 100,000 resolving power at *m/z* 400, typical scan range of 800–1600 *m/z*) and MS/MS (4–16 microscans, 60,000 or 100,000 resolving power at *m/z* 400).

4. The spray voltage was set to 2.5 kV, and the temperature of the heated capillary was held at 250 °C.
5. Fragmentation was achieved using data-dependent collision-induced dissociation (CID) or source-induced dissociation (SID). In most cases, GELFrEE fractions with molecular ranges up to 13 kDa were subjected to CID fragmentation, and the ones with molecular ranges from 13 to 30 kDa were subjected to SID fragmentation. CID was pursued with a 15 or 25 *m/z* isolation window for either the most and fourth intense ions or the third and sixth intense ions from the previous full MS scans to decrease the chances of fragmenting the different charge states originated from the same proteins. MS/MS settings for the CID were as follows, minimum signal threshold = 1000 counts, normalized collision energy = 41%, activation Q = 0.4, and activation time = 100 ms. Dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 480 s, and a repeat duration of 120 s. SID utilized a potential of 75 V, and data were collected with a scan range of 400–1800 *m/z*.

3.5 Top-Down MS Data Analysis

1. Each LC-MS/MS file was analyzed using ProSightPC 3.0 program.
2. Intact precursor and fragment masses from LC-MS/MS files were determined using Xtract algorithm within ProSightHT of ProSightPC software to determine monoisotopic neutral masses from high-resolution precursor and fragment ion spectra and compiled into a ProSight upload file (.puf). From precursor selection criterion within ProSightHT, multiplexing mode was also selected (*see* **Note 10**).
3. Each .puf file was searched in absolute mass mode via an iterative search tree method. The first absolute mass search was initiated against a shotgun-annotated human proteome database containing PTMs, known alternative splice forms, coding single nucleotide polymorphisms (cSNPs), and peptide cleavage events (UniProt release 2012_06, 10,535,964 proteoforms) with 10,000 Da precursor window and 10 ppm fragment mass tolerance (*see* **Note 11**).
4. For initial searches that did not identify a protein below an E-value cutoff of 1×10^{-4}, a second absolute mass search took place against a simplified database including N-terminal acetylation and initial methionine cleavage (UniProt release 2012_06, 472,735 proteoforms) with 100,000 Da precursor window and 10 ppm fragment mass tolerance.
5. As for LC–MS/MS files generated from GELFrEE fractions containing up to 15 kDa, the .puf files were additionally searched in biomarker search mode against a simplified database (UniProt release 2012_06, 237,388 proteoforms) with

2.2 Da precursor window and 10 ppm fragment tolerance (*see* **Note 12**). At least four matched fragment ions and an E-value lower than 1×10^{-4} were required for protein identification [12, 13] (Fig. 3).

6. A Sequence Gazer tool in ProSightPC software was used to manually determine PTMs or single amino acid variations (SAAVs) (Fig. 4).

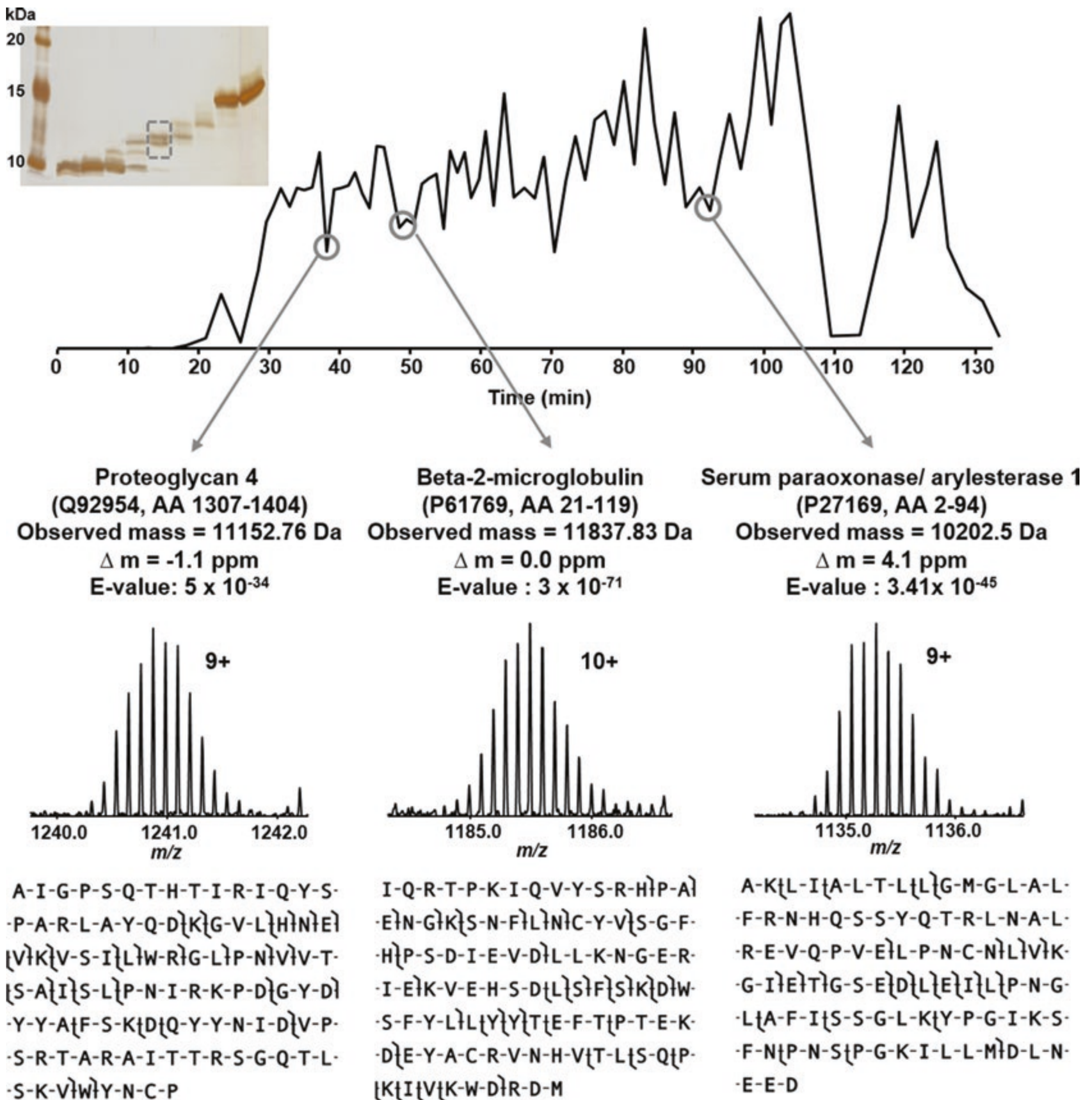

Fig. 3 An example of identification of LMW proteoforms from a GELFrEE fraction of plasma sample depleted of its seven high-abundance proteins via top-down approach. A total ion chromatogram is shown with intact mass spectra and graphical fragmentation maps for beta-2-microglobulin and cleaved products of proteoglycan 4 and serum paraoxonase/arylesterase 1. Reproduced from [14] with permission from the publisher

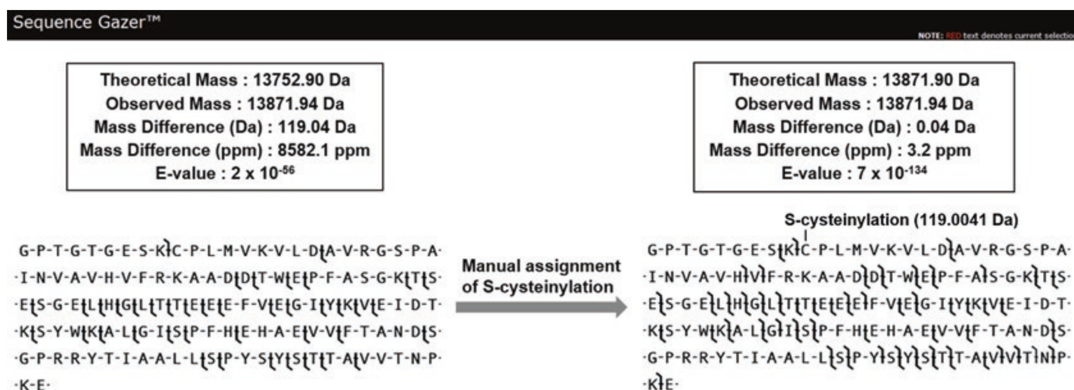

Fig. 4 Graphical fragmentation map of transthyretin (TTR) with S-cysteinylation that was manually assigned from Sequence Gazer tool in ProSightPC software. TTR protein was originally identified with one b-ion and 40 y-ions and with 119.04 Da of intact mass difference. According to UNIMOD (www.unimod.org), the mass difference can result from S-cysteinylation (119.0041 Da). When the mass value was added on the cysteine residue of the protein sequence from Sequence Gazer tool, 36 b-ions and one y-ion were additionally matched with intact mass difference of 3.2 ppm

4 Notes

1. Eight-channel multiplexed device for separating proteins based on molecular weight was fabricated as previously described [16]. GELFREE8100 device (Expedeon) can be used for the same purpose prior to top-down mass spectrometric analysis.
2. The individual human plasma sample was collected as suggested by the HUPO Plasma Proteome Project [18]. It is important to avoid freeze-thaw cycles for plasma samples without addition of protease inhibitors because the degraded products produced from sample processing steps can be identified via top-down approach. Protease and phosphatase inhibitor cocktails can be added immediately after the plasma samples are prepared from blood or added when they are first thawed.
3. When the plasma samples are needed to be pooled for top-down mass spectrometric analysis, they may be pooled with equal amounts prior to analysis.
4. The peak of flow-through fraction was typically seen from 1.5 min to 6 min from the chromatogram of the MARS-6 or MARS-7 column.
5. The resolving and stacking gels were usually polymerized within 1 h.

6. The plasma sample without depletion of high-abundance proteins was not subjected to reduction and alkylation because immunoglobulin light chains were supposed to be produced under reducing conditions and to be present in the GELFrEE fraction around 25 kDa. Thus, the reduction of disulfide bonds using DTT for the plasma sample without removal of high-abundance proteins was omitted in order to maximize the identification of plasma proteoforms that were endogenously present in the molecular weight region.
7. As for the protein loading amount to an SDS-PAGE tube gel, 850 μg of plasma sample without depletion of high-abundance proteins and 350 μg of the depleted plasma samples were loaded onto each SDS-polyacrylamide gel column because 850 μg of the plasma sample without depletion revealed similar separation efficiency of LMW fractions (<30 kDa) compared to that obtained from 350 μg of the depleted plasma samples. Since the high-abundance proteins including albumin and IgG constitute over 75% of the total proteins of plasma sample without depletion of highly abundant proteins, it is assumed that the amount of LMW proteoforms present in the plasma sample without removal of high-abundance proteins is relatively smaller than that present in the same amount of the depleted plasma samples.
8. After the entire portion of the blue dye had entered the collection chamber, the first fraction was collected.
9. While a GELFrEE fraction can be analyzed by top-down mass spectrometry, we found that a combination of multiple GELFrEE fractions with similar molecular weight ranges resulted in identification of more number of LMW proteoforms due to more intense MS and MS/MS signals.
10. From multiplexing mode, multiple precursor masses were selected within an isolation range as multiple precursors, based on an intensity cutoff (set at 10% here) relative to the base peak of the analysis window. The multiplexing mode allows identification of multiple precursors that are fragmented together in the same isolation window.
11. Absolute mass search involves matching the observed precursor mass to theoretical intact masses from a database within a user-specified tolerance and then comparing the observed fragment masses to those calculated from possible forms within a user-specified tolerance.
12. Biomarker search involves matching an observed mass to the theoretical masses of possible subsequences from the database within a precursor mass tolerance and then comparing the observed fragment masses to those calculated from the candidate subsequences within a user-specified tolerance.

Acknowledgments

The authors gratefully acknowledge the technical assistance from the Kelleher Group at Northwestern University. This work was supported by grants from the Multi-omics Program (2012M3A9B9036679) and the Brain Research Program (2015M3C7A1064795) funded through NRF supported by the Korean Ministry of Science, ICT and Future Planning and a KIST institutional program.

References

1. Liotta LA, Petricoin EF (2006) Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. J Clin Invest 116(1):26–30
2. Petricoin EF, Belluco C, Araujo RP, Liotta LA (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer 6(12):961–967
3. Drake RR, Cazares L, Semmes OJ (2007) Mining the low molecular weight proteome of blood. Proteomics Clin Appl 1(8):758–768
4. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359(9306):572–577
5. de Noo ME, Deelder A, van der Werff M, Ozalp A, Mertens B, Tollenaar R (2006) MALDI-TOF serum protein profiling for the detection of breast cancer. Onkologie 29(11): 501–506
6. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 116(1):271–284
7. Ebert MPA, Niemeyer D, Deininger SO, Wex T, Knippig C, Hoffmann J, Sauer J, Albrecht W, Malfertheiner P, Rocken C (2006) Identification and confirmation of increased fibrinopeptide A serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. J Proteome Res 5(9):2152–2158
8. Greening DW, Simpson RJ (2010) A centrifugal ultrafiltration strategy for isolating the low-molecular weight (<or=25K) component of human plasma proteome. J Proteome 73(3):637–648
9. Harper RG, Workman SR, Schuetzner S, Timperman AT, Sutton JN (2004) Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis 25(9):1299–1306
10. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD (2003) Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2(10):1096–1103
11. Smith LM, Kelleher NL, Consortium for Top Down P (2013) Proteoform: a single term describing protein complexity. Nat Methods 10(3):186–187
12. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 480(7376):254–258
13. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL (2013) Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. Mol Cell Proteomics 12(12):3465–3473
14. Cheon DH, Nam EJ, Park KH, Woo SJ, Lee HJ, Kim HC, Yang EG, Lee C, Lee JE (2016) Comprehensive analysis of low-molecular-weight human plasma proteome using top-down mass spectrometry. J Proteome Res 15(1):229–244
15. Tran JC, Doucette AA (2008) Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. Anal Chem 80(5):1568–1573
16. Tran JC, Doucette AA (2009) Multiplexed size separation of intact proteins in solution phase for mass spectrometry. Anal Chem 81(15):6201–6209
17. Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution

in the presence of detergents and lipids. Anal Biochem 138(1):141–143

18. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, Mehigh RJ, Cockrill SL, Scott GB, Tammen H, Schulz-Knappe P, Speicher DW, Vitzthum F, Haab BB, Siest G, Chan DW (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277