Published in final edited form as:

J Proteome Res. 2018 June 01; 17(6): 2156–2164. doi:10.1021/acs.jproteome.8b00100.

A Targeted, Differential Top-Down Proteomic Methodology for Comparison of ApoA-I Proteoforms in Individuals with High and Low HDL Efflux Capacity

Henrique dos Santos Seckler^{†,‡}, Luca Fornelli^{†,‡}, R. Kannan Mutharasan[§], C. Shad Thaxton^{†,‡,||}, Ryan Fellers[†], Martha Daviglus^{#,‡}, Allan Sniderman[◆], Daniel Rader^{¶,▽,○}, Neil L. Kelleher^{†,‡}, Donald M. Lloyd-Jones^{#,§}, Philip D. Compton^{†,‡}, and John T. Wilkins^{#,§,*}

[†]Departments of Chemistry and Molecular Biosciences and the Proteomics Center of Excellence, Northwestern University, Evanston, IL, USA

[‡]Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, USA

[§]Bluhm Cardiovascular Institute, Northwestern Memorial Hospital, Chicago, IL, USA; The Department of Medicine (Cardiology), Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

[⊥]International Institute for Nanotechnology (IIN), Northwestern University, Evanston, IL, USA

^{||} Feinberg School of Medicine, Department of Urology, Northwestern University, Chicago, IL, USA

[#]Northwestern University Feinberg School of Medicine, Department of Preventive Medicine, Chicago, IL, USA

⁺University of Illinois at Chicago, Institute for Minority Health Research, Chicago, IL, USA

[◆]Royal Victoria Hospital–McGill University Health Centre, Montreal, QC, Canada

[¶]Department of Medicine and Cardiovascular Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[▽]Institute for Translational Medicine and Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[○]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Abstract

Top-down proteomics (TDP) allows precise determination/characterization of the different proteoforms derived from the expression of a single gene. In this study, we targeted apolipoprotein A-I (ApoA-I), a mediator of high-density-lipoprotein cholesterol efflux (HDL-E), which is inversely associated with coronary heart disease risk. Absolute ApoA-I concentration and allelic

*Corresponding Author: John T. Wilkins, Suite 1400, 680 N. Lakeshore Dr. Chicago, IL 60611, tel: +13125031065, j-wilkins@northwestern.edu.

Supporting Information. Further discussion on participant characteristics.

Conflict of Interest statement:

The authors report no conflict of interest.

variation only partially explain inter-individual HDL-E variation. Therefore, we hypothesize that differences in HDL-E are associated with the abundances of different ApoA-I proteoforms. Here, we present a targeted TDP methodology to characterize ApoA-I proteoforms in serum samples and compare their abundances between individuals. We characterized eighteen ApoA-I proteoforms using selected-ion monitoring coupled to electron-transfer dissociation mass spectrometry. We then compared the abundances of these proteoforms between two groups of four participants, representing the individuals with highest and lowest HDL-E values within the Chicago Healthy Aging Study (n=420). Six proteoforms showed significantly ($p<0.0005$) higher intensity in high HDL-E individuals: canonical ApoA-I [fold difference (fd)=1.17], carboxymethylated ApoA-I (fd=1.24) and, with highest difference, four fatty acylated forms: palmitoylated (fd=2.16), oleoylated (fd=2.08), arachidonoylated (fd=2.31) and one bearing two modifications: palmitoylation and truncation (fd=2.13). These results demonstrate translational potential for targeted TDP in revealing, with high sensitivity, associations between inter-individual proteoform variation and physiological differences underlying disease risk.

Keywords

HDL Efflux; Cholesterol; Atherosclerosis; Apolipoproteins; Proteoforms; Top-down Proteomics; Palmitoylation; ApoA-I; Acylations

Introduction

Protein products from a given gene often have a high degree of chemical variation.^{1–2} This variation occurs through combinations of allelic variation, alternative splicing, co- and post-translational modifications.^{1–2} These variants are collectively known as proteoforms¹. Proteoforms are the molecular actuators of gene function and thus proteoform-level information is hypothesized to be more closely related to phenotypic variation than genomic or transcriptomic data.^{2–3} However, acquisition of proteoform-level data is challenging.^{2, 4–5} The classical approach to proteomics, bottom-up proteomics, is used routinely for the identification of thousands of protein groups in a single proteome,⁵ but fails to reconstruct the proteoform distribution present in a given sample.^{5–6} This shortcoming, known as the protein inference problem,⁶ is due to the pre-analytical proteolytic digestion necessary for bottom-up proteomics.^{2, 6} Often, the combination of chemical and genetic modifications creates a highly complex proteoform mixture in biological samples. Individual proteoform structure and their relative abundances cannot be reconstructed through the analysis of proteolytic peptide data because the precise combinations of peptides (including their modified and polymorphic states) that occur in intact forms cannot be inferred from the admixture of proteolytic digestion products.^{2, 5–6} Conversely, top-down proteomics, which analyzes the intact mass of protein molecules in combination with ion fragmentation (MS/MS) data, is capable of full chemical characterization of proteoforms. Furthermore, recent developments in high-resolution mass spectrometry and MS/MS techniques have allowed for the development of robust, label-free quantitative and differential top-down proteomics and this technique has since been adopted by several groups.^{7–11}

While high-throughput differential top-down proteomics can be used for the characterization and quantitative comparison of multiple proteoforms across different biological states, more targeted approaches can discover, characterize, and differentially compare proteoforms of a small number of candidate genes with increased sensitivity. Here, we report on the targeted study of the proteoforms of apolipoprotein A-I (ApoA-I), a predictive marker and possible causal factor for the development of cardiovascular disease (CVD).

High-density lipoprotein (HDL) particles mediate reverse cholesterol transport, the transfer of cholesterol from tissues such as the vascular intima to the liver for metabolism and excretion, which is one of the primary atheroprotective functions of HDL.¹² HDL efflux capacity is an ex-vivo tissue culture assay that quantifies the rate-limiting step in reverse cholesterol transport. HDL efflux capacity has independent, inverse associations with CVD risk in humans.^{13–14} The molecular reasons for inter-individual differences in HDL efflux are not currently known.

HDL efflux is mediated by interactions between ApoA-I, the major structural protein found on HDL particles, and the ABCA1 transporter.¹⁵ Since the ABCA1 transporter is kept constant in the HDL efflux assay and allelic variation in ApoA-I is rare, we hypothesize that inter-individual differences in HDL efflux are due to inter-individual differences in ApoA-I proteoform distribution. Since CVD is the leading cause of death world-wide,¹⁶ identifying molecular underpinnings to CVD risk has the potential be of significant clinical utility.

In this study, we report the development, from sample preparation to MS analysis, of a targeted, differential top-down proteomic method to discover, characterize and compare abundances of ApoA-I proteoform in samples of human serum. This methodology can serve as basis for larger studies to understand the relationship between an individual's apolipoproteoform distribution and their HDL efflux capacity, as well as other targeted studies that further the goal of proteoform-based personalized medicine.

Experimental Section

Proteomic Sample Preparation

For the initial qualitative analysis and all quantitative analyses, frozen aliquots from banked serum were gently thawed and subjected to ApoB-depletion and subsequent protein extraction. To control for pre-analytical variation all samples were stored under the same conditions and preparation was performed with the same lot for each reagent, under controlled temperature and at the same time. For ApoB-depletion, two 10 μ L aliquots of each serum sample were each added to 10 μ L of LipoSep IP (Sun Diagnostics, New Gloucester, ME) resin, an anti-ApoB immunoprecipitation reagent. The ensuing centrifugation depleted the serum of ApoB-bound particles such as LDL and VLDL. The then HDL-enriched supernatant was collected and a 1 μ L aliquot was submitted to protein extraction, performed via methanol/chloroform/water precipitation as described by Wessel et al¹⁷ and modified to reduce oxidation derived from sample preparation. Modifications to the procedure included extraction and resuspension performed at 4°C and all vortexing and sonication steps substituted for soft mixing. The extracted proteins were then resuspended in

60 µl of 5% acetonitrile and 0.2% formic acid in water. Each set of two aliquots from the same serum sample was analyzed as a preparation replicate.

For ApoA-I proteoform characterization, a 500 µL serum sample from one individual was used to generate an ApoA-I-enriched solution. Ten 50 µl aliquots were prepared and subjected to ProteoPrep Blue (Sigma-Aldrich, St Louis, MI) albumin and IgG depletion, following manufacturer instructions. The lipid fraction of each depleted serum aliquot was then extracted via lipophilic absorption with PHM-L LIPOSORB (Merk Millipore, Billerica, MA) and subsequent boiling in aqueous 100µM ammonium acetate and 10% SDS, pH 7.2. Protein precipitation was then carried out, using the aforementioned modified methanol/chloroform/water extraction protocol. Each aliquot was re-suspended in 20 µL of GELFrEE (Expedeon, San Diego, CA) Sample Buffer and all aliquots were pooled and loaded to an 8%T GELFrEE cartridge for size-based fractionation. The time fraction corresponding to proteins from 0-30 kDa was collected and subjected to methanol/chloroform/water extraction. Protein precipitates were re-suspended in 40 µl of 5% acetonitrile and 0.2% formic acid in water.

Anti-ApoB Western Blot

Prior to methanol/chloroform/water precipitation a 1 µl aliquot of serum, 2 µl of ApoB-depleted serum and a 5 µl aliquot of a 0.2% formic acid extraction of the used anti-ApoB beads were run in an “Any KD” Mini-Protean TGX SDS-PAGE (Bio-Rad Laboratories Inc., Hercules, CA, USA) and transferred to a nitrocellulose membrane in an iBlot 2 Dry Blotting System (Thermo Fisher Scientific, Waltham, MA, USA), using the manufacturer’s protocol. The blotted membrane was blocked with 3% milk TBS-T for 30 mins, incubated with a 1:10000 dilution of a mouse anti-human-ApoB primary antibody (ab39560, Abcam plc, Cambridge, UK) overnight at 4 °C, washed with three 10 min cycles of TBS-T, incubated with a 1:20000 dilution of a goat anti mouse HRP-conjugated secondary antibody (sc-2005, Santa Cruz Biotechnology, Inc.), washed again with three 10 min cycles of TBS-T, developed with 1 ml of Super Signal West Pico Luminol Enhancer Solution and 1 ml of Super Signal West Pico Stable Peroxidase Solution (Thermo Fisher Scientific, Waltham, MA, USA) for 5 min, and recorded in a ChemiDoc XRS+ System (Bio-Rad Laboratories Inc., Hercules, CA, USA) at a 50 s exposure.

Liquid Chromatography – Top-Down Mass Spectrometry

Resuspended protein samples were subjected to reversed-phase liquid chromatography (RPLC) using an Ultimate 3000 LC system (Thermo Scientific, San Jose, CA). Around 1 µg of sample was loaded onto a trap column (20 mm, 150 µm inner diameter, i.d.) packed with PLRP-S resin (Agilent, Santa Clara, CA) for initial wash. For separation of proteoforms, an in-house packed capillary PLRP-S column (200 mm, 75 µm i.d.) was used. Both columns were heated at 55°C. The gradient consisted of a ramp of solvent B from 15 to 50% in 30 min., with a total run time of 60 min. including column wash (at 90% solvent B) and re-equilibration. Solvent A consisted of 5% acetonitrile and 0.2% formic acid in water, while solvent B was composed of 5% H₂O and 0.2% formic acid in acetonitrile. The outlet of the column was on-line coupled to a nanoelectrospray ionization source, to which a ~2 kV potential was applied for ionizing proteoforms for mass spectrometry analysis.

All mass spectrometry measurements were performed using an Orbitrap Fusion Lumos (Thermo Scientific, San Jose, CA), operating in “intact protein mode” (N_2 pressure in the HCD cell of 2 mTorr). For the initial qualitative experiment and all quantitative MS experiments, the instrument was operated in MS1-mode only (i.e., no fragmentation was applied). The instrument was cycling between a full scan (over a 500-2000 m/z window) and a selected ion monitoring (SIM) scan (over a 780.70-803.3 m/z window, to target the 35+, 36+ and 37+ charged states of ApoA-I proteoforms). The applied resolving power was 120,000 (at 200 m/z), while the automatic gain control (AGC) target was set at $2e5$ and $5e4$ for broadband MS1 and SIM, respectively. For qualitative experiments, aimed at mapping intact proteoforms and characterizing each form with complete molecular specificity, targeted MS/MS experiments were performed during the elution time window of ApoA-I via electron transfer dissociation (ETD), with or without supplemental activation using collision-induced dissociation (ETciD). ETD duration was set at 5-10 ms, with fluoranthene AGC target of $7e5$ charges. For ETciD a 12% NCE was applied. For fragmentation technology comparison, higher-energy collisional dissociation (HCD) was applied at 15% NCE. MS/MS scans were recorded over a 400-2000 m/z window at 60,000 resolving power (at 200 m/z). The AGC target for each selected precursor was $5e5$, with isolation window of 1 m/z to ensure selectivity of fragmentation by excluding adjacent proteoforms. To confirm specific proteoform isolation, each MS/MS scan was accompanied by a SIM scan within the same precursor isolation window and at an AGC target of $1e4$. All Full MS, SIM and MS/MS scans were obtained by averaging 4 microscans.

Proteoform Characterization from Top-Down MS Data

For ApoA-I proteoform characterization and post-translational modification (PTM) mapping, SIM scans of the initial qualitative LC-MS runs were manually analyzed to search for MS peaks of similar charge distribution, mass and retention time of ApoA-I. All peaks that fit these parameters were further analyzed by fragmentation. Intact mass spectra were recorded and fragmentation data was acquired as previously described. Then, ion masses deconvoluted from the fragmentation and intact mass spectra were analyzed with the freeware ProSight Lite (<http://prosightlite.northwestern.edu/>)¹⁸ against ApoA-I sequence and modification variants found in the UniProt database (UniProt Accession #P02647) and the biochemical and clinical literature on ApoA-I. All MS peaks that were identified as species containing an ApoA-I backbone but for which the mass did not match any previously known proteoforms were characterized with a classical top-down proteomics approach for proteoform discovery.⁴ Briefly, the highly resolved intact mass difference (m) of a proteoform to the canonical form was used to narrow the list of possible chemical modifications or genetic variations to the proteoform and the presence and location of each modification was tested by fragmentation coverage of the modification site. The generated fragmentation maps were assigned P-scores for proteoform identification confidence. A list of the proteoforms identified in this study is reported in Table S-1, along with identification confidence parameters and permanent proteoform identifiers. All proteoforms of ApoA-I detectable in this study were chemically characterized. All graphical fragmentation maps used in this analysis as well as the respective RAW files have been deposited to the MassIVE database (<https://massive.ucsd.edu/>), dataset identifier: MSV000082051. Proteoforms were

also uploaded to the top-down proteomics proteoform repository (<http://repository.topdownproteomics.org/>).

Participant Sample Selection

Serum samples were collected from the Chicago Healthy Aging Study (CHAS) repository. CHAS design has been described previously.¹⁹ CHAS is a community-based cohort study of 1,395 asymptomatic, older US adults. From 2007-2010 CHAS participants underwent detailed in-person interview to assess health histories, examination to assess CVD risk factor measurement, venipuncture to obtain blood samples, and imaging for sub-clinical atherosclerosis. CHAS was approved by the Northwestern IRB. Serum samples obtained at the initial examination were stored at -80C for future analysis. In 2013, a sample of 420 CHAS participant serum samples were selected for HDL efflux capacity measurement.²⁰ From this subgroup of 420 participants, we chose 4 samples with the highest and 4 with the lowest measured HDL efflux capacity, to cover the range of efflux values. To enable data acquisition to occur over one contiguous block of liquid-chromatography tandem mass spectrometry (LC-MS/MS) instrument time (5 days), and therefore diminish intensity variation due to instrumental drift, we chose a total sample size of 8 participant samples.

Demographics, smoking status and current medications were obtained via participant self-report. Height and weight were measured using MESA and CARDIA protocols and equipment.²¹ Height and weight measurements were made with participants wearing light clothing and no shoes. Height was measured by standard stadiometer. After resting for 5 min BP was measured three times in the right arm with the participants seated. The reported BP is an average of the last two seated measurements. Blood samples were analyzed for glucose, hs-CRP, and standard lipid fractions.

HDL Efflux Measurement

Frozen aliquots from banked serum were gently thawed and apolipoprotein B (ApoB)-depleted serum was prepared using the polyethylene glycol precipitation method. The cholesterol efflux assay used has been described previously.^{15, 20} Briefly, J774 mouse macrophages radiolabeled with 3H-cholesterol at 2 microCi/ml and treated with 2 microg/ml ACAT inhibitor (Sandoz 58-035; Sigma-Aldrich, St. Louis, MO) and 30 microM of cAMP analog [8-(4-chlorophenylthio)adenosine 3',5'-cyclic monophosphate sodium salt; Sigma-Aldrich] were incubated with a 2.0% final concentration of ApoB-depleted serum for 4 h. After 4 h, radioactivity in both the cell culture media and the cells was assessed and percent efflux was calculated as follows: (radioactivity in the medium after efflux – radioactivity in blank medium lacking serum)/total radioactivity in labeled cells. The percent efflux for technical triplicates was averaged together. To normalize across assay runs, a serum sample pool drawn from healthy volunteers was run with each sample. Normalized efflux was calculated as percent efflux for the sample divided by percent efflux for the sample pool.

Nested Sample Replication for Reproducible Quantitative Comparison of Proteoforms

Nested data acquisition was set up as previously described for a quantitative top-down proteomics workflow.^{5, 7} Briefly, in order to mitigate the effects of measurement errors arising from sample preparation and instrumental performance drift, each preparation

replicate was injected 4 times through LC-MS. Samples were injected in a randomized fashion while avoiding clusters of samples from patients of the same efflux group. Proteoform intensity variation between injection replicates and preparation replicates was quantified to parse out the effects of pre-analytical variation and increase statistical confidence in the variation observed between individuals and efflux groups.

MS Data Analysis for Quantitative Comparison of Proteoforms and Validation

For proteoform abundance comparison, the generated SIM scans (780.7-803.3 m/z window) were searched for isotopic patterns of target proteoforms by a previously described fitter algorithm.⁷ Briefly, a list of chemical formulas for the characterized ApoA-I proteoforms was used to calculate isotopic patterns for the 35+, 36+ or 37+ charge state of each form. Then, the centroided peaks of the observed spectrum above a minimum intensity (NL=100) were searched against the expected m/z of the two most abundant isotopomers calculated for the given chemical formulas. A window of 10 ppm was allowed for matches in order to account for mass accuracy drift. The full isotopic distribution for matching formulas was then adapted to the observed mass error and the spectra was searched for the remaining isotopomers with a 2 ppm allowance, to account for charge-space effects such as coalescence.²²⁻²³ The relative intensities of the matched isotopic peaks were compared to the expected distribution in order to generate a fitting score. Fitting scores ranged from 0-1 and were calculated by a least squares model weighed positively towards higher intensities. Proteoform intensities were quantified for each SIM scan in which the relative chemical formula passed a fitting score threshold of 0.3. The total intensity per LC-MS run was calculated for each proteoform. The total intensity of each form was normalized, per LC-MS run, by the total ion intensity (or total ion current, TIC) in the SIM scans, i.e., by the total ApoA-I intensity within the specified charge states.

To validate the association between ApoA-I proteoform intensity as identified by the aforementioned method and abundance in the initial sample, 5 samples of known ApoA-I concentration (0.25 $\mu\text{g}/\mu\text{L}$, 0.18625 $\mu\text{g}/\mu\text{L}$, 0.125 $\mu\text{g}/\mu\text{L}$, 0.093125 $\mu\text{g}/\mu\text{L}$, 0.06125 $\mu\text{g}/\mu\text{L}$) were prepared from a pure ApoA-I standard (MyBioSource, San Diego, CA, USA) and run with the same LC-MS method as all previous samples. The proteoforms present in the sample were analyzed similarly to all quantitative samples in this study and the R^2 correlation between MS intensity and injected mass was described.

Results and Discussion

Sample Preparation and Proteoform Characterization

In order to increase relative concentration of ApoA-I and reduce non-protein impurities, serum samples were depleted of ApoB-bound particles, such as LDL and VLDL. ApoB depletion efficiency was analyzed by western blot (Figure S-1).

Protein content of the ApoB-depleted protein extracts was analyzed by LC-MS/MS and found to contain ApoA-I (Figure S-2). Following ApoA-I observation, an LC-Full MS/SIM method was developed to increase sensitivity in the m/z range between charged states of the highly abundant canonical proteoform of ApoA-I. Proteoforms observed in the SIM scans

were fragmented for characterization. Intact mass and fragmentation data from candidate species allowed for identification of 18 distinct ApoA-I proteoforms. A diagram depicting the top-down proteomics methodology for proteoform characterization from SIM data is shown in Figure 1.

Proteoform characterization in this study was permitted by recent advances in mass spectrometry. For instance, several of the charge states of the different proteoforms of ApoA-I were often observed less than 1 Th apart and thus required highly selective quadrupolar isolation in order to produce proteoform-specific characterization (Figure S-3). Equally narrow precursor isolation windows have been used before for top-down characterization of the different proteoforms of histone H3.²⁴ However, while that analysis was performed in a direct-injection regime, in the present study we were able to characterize ApoA-I proteoforms on the LC time scale. To do so, we utilized – with and without supplemental activation – high-capacity ETD, which has been described to reduce the need for scan averaging and increase fragmentation coverage in TDP.²⁵ Indeed, the difference in ApoA-I fragmentation coverage generated by ETD as compared to HCD allowed for unequivocal chemical characterization of ApoA-I proteoforms, especially those bearing PTMs on internal regions of the primary sequence. A comparison of the fragmentation coverages achieved with each ion fragmentation technique applied to canonical ApoA-I is shown in Figure S-4.

In total, 11 PTMs were observed on the primary sequence of ApoA-I, including: oxidation, phosphorylation, glycation by a hexose, carboxymethylation and acylation by palmitic, oleic and arachidonic acids. Importantly, six of the 18 proteoforms included multiple PTMs within their sequences (e.g., glutamine truncation + palmitoylation on the same ApoA-I primary sequence). Figure 2 shows all the modification sites confidently localized by fragmentation, mapped onto the previously described secondary and tertiary structures of ApoA1.

Several PTMs of ApoA-I observed in this study had been previously described in the literature. Glycated forms of ApoA-I have been observed endogenously^{26–27} including forms bearing glycation by a hexose and advanced glycation end products such as carboxymethylation on a lysine.²⁸ Notably, both glycation modifications were localized to the same residue of ApoA-I, suggesting that we observed early- and late-stage glycation products.²⁸ Moreover, we observed acylations of 4 different types on ApoA-I. While ApoA-I had only been described previously to be modified by palmitic acylation,²⁹ palmitoylated proteins are known to be modified by other long fatty acids as a result of acyltransferase's lack of substrate specificity.^{30–31} This may explain the observation of acylations by longer fatty acids, such as oleic, arachidonic and docohexaenoic acids.

However, while several of the aforementioned PTMs had been previously described, few attempts had been made to discover their endogenous modification sites and thus fully chemically characterize these ApoA-I species. Moreover, this is, to our knowledge, the first attempt to characterize the full distribution of proteoforms of ApoA-I. Finally, these data constitute empirical evidence that the distribution of proteoforms within a given sample cannot be reconstructed by the analysis of proteolytic peptides. For instance, the inference

that acylated proteoforms happen concomitantly with truncation but not – at least to the degree of sensitivity present in this study – with glycations or before propeptide cleavage can only be made from intact mass data and proteoform-specific fragmentation.

Comparison of Proteoform Abundances between Efflux Levels

To examine the linearity of the correlation between mass spectrometric intensity (as detected by our in-house fitting algorithm) and the abundance of ApoA-I proteoforms, a standard correlation curve was generated, using a commercial ApoA-I standard (Figure S-5). The curve fit a linear distribution with $R^2 = 0.9977$ within the range of concentrations of ApoA-I normally found in human serum.³²

Using the same label-free top-down MS analysis, extracted serum samples from 8 selected Chicago Healthy Aging Study (CHAS) participants were compared. The relative abundance of each of the previously characterized ApoA-I proteoforms was determined for the samples. Substantial difference in ApoA-I proteoform distribution was observed between individuals (Figure 3).

Furthermore, to assess the power of the methodology to compare proteoform abundances across groups with different HDL effluxes, participants were separated into two groups: 4 individuals with high efflux and 4 with low efflux (representing the high and low extremes found in 420 CHAS participants). Characteristics of the 8 participants are shown in Table 1. The mean normalized HDL efflux was 0.65 (± 0.05) in the low HDL efflux group and 1.7 (± 0.05) in the high efflux group. Mean abundances for each ApoA-I proteoform were compared between groups. Figure 4 displays a volcano plot (reporting quantitation confidence versus fold difference between groups for each characterized proteoforms) of comparative TDP results. Overall, 6 ApoA-I proteoforms had significantly higher relative abundance (normalized by total ApoA-I signal) in individuals in the high efflux group: the canonical form, a carboxymethylated form and 4 acylated forms: palmitic acylation, oleic acylation, arachidonic acylation and a palmitoylated and truncated form. Intensity of the canonical form was on average 17% higher in individuals with high efflux. Acylated ApoA-I proteoforms showed an approximate 2-fold higher intensity in individuals with high HDL efflux capacity (palmitic acylation (fd=2.16, $p < 0.0001$), oleic acylation (fd=2.08, $p < 0.0001$), arachidonic acylation (fd=2.31, $p < 0.001$), and the palmitic acylation + truncation form (fd=2.13, $p = 0.003$). The carboxymethylated form was 24% higher in the high HDL efflux group as well. The variation of all other proteoforms was not significant between efflux groups.

The estimated effects of each source of proteoform intensity variation, including injection, preparation, individual, group-based and residual variation are depicted in Figure S-6. Proteoform intensity variation between injection and preparation replicates accounted for a relatively small fraction of the overall variation and – aside from residual, unidentified variation – most of the variation was seen between individuals and the two efflux groups. These results underline the power of targeted TDP in analyzing ApoA-I proteoform abundance variation across individuals and casts it as a useful tool for understanding the correlation between modifications to the ApoA-I backbone and HDL efflux.

While sample size forbids extrapolation of results to the whole population, the results are nonetheless interesting from a biological standpoint. Previous studies had demonstrated modest associations between ApoA-I concentration and HDL efflux capacity, suggesting that concentration of ApoA-I explains some differences in HDL efflux.¹⁴ Our data, if replicated in larger samples, could add to this finding, as they raise the possibility that it is the concentration of certain proteoforms of ApoA-I, such as acylated proteoforms, that explains inter individual difference in HDL efflux.

Notably, most data on modifications of ApoA-I suggest that non-canonical chemical variants are associated with lower efflux capacity or a null effect on efflux.^{33–34} For instance, post-translationally modified variants such as W72-oxidized ApoA-I and a truncated form at S228 were shown to poorly mediate HDL efflux *in vitro*.^{35–36} Interestingly, in the case of W72-oxidized ApoA-I, higher concentration of this particular proteoform were associated with CVD events in a small clinical sample. Our study, however, is the first to suggest that certain non-canonical, endogenous proteoforms of ApoA-I are associated with higher efflux values. Of the modified forms found to be higher in our high efflux group, the ones bearing acylations are especially interesting, because these modifications are commonly observed in membrane-bound proteins, in which they function to facilitate the association between protein and membrane, by embedding of the non-polar region of the fatty acid into the phospholipid layer.³¹ We posit that acylations of ApoA-I may increase the stability of the HDL particle and allow for more efflux per particle or faster lipidation of nascent HDL. Likewise, acylations could prolong the interaction between ApoA-I and ABCA1, thus increasing cholesterol transport.

However, the present study is a cross-sectional analysis; thus, causality cannot be inferred. Future mechanistic studies, especially of the effect of acylation of ApoA-I on efflux capacity, will increase our understanding of the relationship observed in our findings. Moreover, while this is, to our knowledge, the most direct attempt to characterize the endogenous ApoA-I proteoform distribution, all data presented here were gathered from extracts and are subject to preparation bias. Therefore, PTMs generally regarded as non-enzymatic, such as methionine oxidation, are possibly observed as byproducts of sample preparation and ionization.^{37–38} However, most ApoA-I PTMs observed result from enzymatic activity (acylations, phosphorylation and backbone cleavages) or endogenous reactions (glycations), and thus the relative abundance of the related proteoforms characterized in this study are likely representative of endogenous modification states of ApoA-I.³⁹ Finally, any extraction bias was distributed across samples from all individuals and, while absolute concentration of proteoforms is not inferable from the data, relative variations in proteoforms intensity are representative of biological changes.

Conclusions

We report a targeted top-down proteomics methodology that can be used for the discovery of novel ApoA-I proteoforms and differential comparison of their abundances across serum samples from multiple individuals, including individuals from both extremes of HDL efflux capacity. This methodology can empower future experiments of larger cohorts to generate conclusive biological results on the associations between proteoforms of ApoA-I and reverse

cholesterol metabolism. Similar targeted top-down proteomic methodologies may greatly aid in parsing down the association of the distribution of proteoforms of any gene of interest to individual phenotype, allowing for advances in proteoform-based personalized medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Work performed for this study was funded by The American Heart Association, under grant SDG 27250022, The National Institute of Health, under grants K23 HL133601-01 and RO1 HL081141 and The National Institute of General Medical Sciences, under grant P41 GM108569. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ABBREVIATIONS

TDP	top-down proteomics
ETD	electron-transfer dissociation
ETciD	electron-transfer dissociation with collisionally-induced supplemental activation
HCD	high-energy collision-induced dissociation
SIM	selected ion monitoring

References

- Smith LM, Kelleher NL, Proteomics CTD. Proteoform: a single term describing protein complexity. *Nat Methods*. 2013; 10(3):186–187. [PubMed: 23443629]
- Schluter H, Apweiler R, Holzthutter HG, Jungblut PR. Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J*. 2009; 3(1):11. [PubMed: 19740416]
- Gregorich ZR, Ge Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*. 2014; 14(10):1195–1210. [PubMed: 24723472]
- Toby TK, Fornelli L, Kelleher NL. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem*. 2016; 9:499–519.
- Ntai I, LeDuc RD, Fellers RT, Erdmann-Gilmore P, Davies SR, Rumsey J, Early BP, Thomas PM, Li S, Compton PD, Ellis MJ, Ruggles KV, Fenyo D, Boja ES, Rodriguez H, Townsend RR, Kelleher NL. Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol Cell Proteomics*. 2016; 15(1):45–56. [PubMed: 26503891]
- Nesvizhskii AI, A R. Interpretation of shotgun proteomic data: The protein inference problem. *Mol Cell Proteomics*. 2005; 4:1419–1440. [PubMed: 16009968]
- Ntai I, Kim K, Fellers RT, Skinner OS, Smith ADt, Early BP, Savaryn JP, LeDuc RD, Thomas PM, Kelleher NL. Applying label-free quantitation to top down proteomics. *Anal Chem*. 2014; 86(10):4961–4968. [PubMed: 24807621]
- Mazur MT, Cardasis HL, Spellman DS, Liaw A, Yates NA, Hendrickson RC. Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *P Natl Acad Sci USA*. 2010; 107(17):7728–7733.
- Wildburger NC, Esparza TJ, LeDuc RD, Fellers RT, Thomas PM, Cairns NJ, Kelleher NL, Bateman RJ, Brody DL. Diversity of Amyloid-beta Proteoforms in the Alzheimer's Disease Brain. *Sci Rep-Uk*. 2017; 7(1):9520.

10. Cleland TP, DeHart CJ, Fellers RT, VanNispen AJ, Greer JB, Leduc RD, Parker WR, Thomas PM, Kelleher NL, Brodbelt JS. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res.* 2017; 16(5):2072–2079. [PubMed: 28412815]
11. Zhang J, Guy MJ, Norman HS, Chen YC, Xu QG, Dong XT, Guner H, Wang SJ, Kohmoto T, Young KH, Moss RL, Ge Y. Top-Down Quantitative Proteomics Identified Phosphorylation of Cardiac Troponin I as a Candidate Biomarker for Chronic Heart Failure. *J Proteome Res.* 2011; 10(9):4054–4065. [PubMed: 21751783]
12. Rosenson RS, Brewer HB Jr, Davidson WS, Fayad ZA, Fuster V, Goldstein J, Hellerstein M, Jiang XC, Phillips MC, Rader DJ, Remaley AT, Rothblat GH, Tall AR, Yvan-Charvet L. Cholesterol efflux and atheroprotection: advancing the concept of reverse cholesterol transport. *Circulation.* 2012; 125(15):1905–1919. [PubMed: 22508840]
13. Rohatgi A, Khera A, Berry JD, Givens EG, Ayers CR, Wedin KE, Neeland IJ, Yuhanna IS, Rader DR, de Lemos JA, Shaul PW. HDL cholesterol efflux capacity and incident cardiovascular events. *The New England journal of medicine.* 2014; 371(25):2383–2393. [PubMed: 25404125]
14. Saleheen D, Scott R, Javad S, Zhao W, Rodrigues A, Picataggi A, Lukmanova D, Mucksavage ML, Luben R, Billheimer J, Kastelein JJ, Boekholdt SM, Khaw KT, Wareham N, Rader DJ. Association of HDL cholesterol efflux capacity with incident coronary heart disease events: a prospective case-control study. *The lancet Diabetes & endocrinology.* 2015; 3(7):507–513. [PubMed: 26025389]
15. de la Llera-Moya M, Drazul-Schrader D, Asztalos BF, Cuchel M, Rader DJ, Rothblat GH. The ability to promote efflux via ABCA1 determines the capacity of serum specimens with similar high-density lipoprotein cholesterol to remove cholesterol from macrophages. *Arteriosclerosis, thrombosis, and vascular biology.* 2010; 30(4):796–801.
16. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J, Fornage M, Gillespie C, Isasi CR, Jimenez MC, Jordan LC, Judd SE, Lackland D, Lichtman JH, Lisabeth L, Liu S, Longenecker CT, Mackey RH, Matsushita K, Mozaffarian D, Mussolino ME, Nasir K, Neumar RW, Palaniappan L, Pandey DK, Thiagarajan RR, Reeves MJ, Ritchey M, Rodriguez CJ, Roth GA, Rosamond WD, Sasson C, Towfighi A, Tsao CW, Turner MB, Virani SS, Voeks JH, Willey JZ, Wilkins JT, Wu JH, Alger HM, Wong SS, Muntner P. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation.* 2017; 135(10):e146–e603. [PubMed: 28122885]
17. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry.* 1984; 138(1):141–3. [PubMed: 6731838]
18. Fellers RT, Greer JB, Early BP, Yu X, LeDuc RD, Kelleher NL, Thomas PM. ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics.* 2015; 15(7):1235–1238. [PubMed: 25828799]
19. Pirzada A, Reid K, Kim D, Garside DB, Lu B, Vu TH, Lloyd-Jones DM, Zee P, Liu K, Stamler J, Daviglus ML. Chicago Healthy Aging Study: objectives and design. *American journal of epidemiology.* 2013; 178(4):635–644. [PubMed: 23669655]
20. Mutharasan RK, Thaxton CS, Berry J, Daviglus ML, Yuan C, Sun J, Ayers C, Lloyd-Jones DM, Wilkins JT. HDL efflux capacity, HDL particle size, and high-risk carotid atherosclerosis in a cohort of asymptomatic older adults: the Chicago Healthy Aging Study. *Journal of lipid research.* 2017; 58(3):600–606. [PubMed: 28049656]
21. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-Ethnic Study of Atherosclerosis: objectives and design. *American journal of epidemiology.* 2002; 156(9):871–881. [PubMed: 12397006]
22. Gorshkov MV, Fornelli L, Tsybin YO. Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom.* 2012; 26(15):1711–7. [PubMed: 22730091]
23. Tarasova IA, Surin AK, Fornelli L, Pridatchenko ML, Suvorina MY, Gorshkov MV. Ion coalescence in Fourier transform mass spectrometry: should we worry about this in shotgun proteomics? *Eur J Mass Spectrom.* 2015; 21(3):459–470.

24. Zheng YP, Fornelli L, Compton PD, Sharma S, Canterbury J, Mullen C, Zabrouskov V, Fellers RT, Thomas PM, Licht JD, Senko MW, Kelleher NL. Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. *Molecular & Cellular Proteomics*. 2016; 15(3):776–790. [PubMed: 26272979]
25. Riley NM, Mullen C, Weisbrod CR, Sharma S, Senko MW, Zabrouskov V, Westphall MS, Syka JE, Coon JJ. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J Am Soc Mass Spectrom*. 2016; 27(3):520–31. [PubMed: 26589699]
26. Curtiss LK, Witztum JL. Plasma apolipoproteins AI, AII, B, CI, and E are glucosylated in hyperglycemic diabetic subjects. *Diabetes*. 1985; 34(5):452–461. [PubMed: 3921419]
27. Nobecourt E, Tabet F, Lambert G, Puranik R, Bao S, Yan L, Davies MJ, Brown BE, Jenkins AJ, Dusting GJ, Bonnet DJ, Curtiss LK, Barter PJ, Rye KA. Nonenzymatic glycation impairs the antiinflammatory properties of apolipoprotein A-I. *Arteriosclerosis, thrombosis, and vascular biology*. 2010; 30(4):766–772.
28. Brown BE, Nobecourt E, Zeng J, Jenkins AJ, Rye K-A, Davies MJ. Apolipoprotein A-I glycation by Glucose and Reactive Aldehydes Alters Phospholipid Affinity but Not Cholesterol Export from Lipid-Laden Macrophages. *PLOS ONE*. 2013; 8(5):e65430. [PubMed: 23741493]
29. Hoeg JM, Meng MS, Ronan R, Fairwell T, Brewer HB Jr. Human apolipoprotein A-I. Post-translational modification by fatty acid acylation. *The Journal of biological chemistry*. 1986; 261(9):3911–3914. [PubMed: 3005308]
30. Kerkhoff C, Habben K, Gehring L, Resch K, Kaefer V. Substrate specificity of acyl-CoA:Lysophospholipid acyltransferase (LAT) from pig spleen. *Archives of biochemistry and biophysics*. 1998; 351(2):220–226. [PubMed: 9514659]
31. Resh MD. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et biophysica acta*. 1999; 1451(1):1–16. [PubMed: 10446384]
32. Mashayekhi NR, Sadrnia S, Chehrei A, Javaheri J. The Correlation between Serum ApoA1 and B and Coronary Artery Disease as Well as Its Severity. *Int Cardiovasc Res J*. 2014; 8(1):1–5. [PubMed: 24757643]
33. Alexander ET, Weibel GL, Joshi MR, Vedhachalam C, de la Llera-Moya M, Rothblat GH, Phillips MC, Rader DJ. Macrophage reverse cholesterol transport in mice expressing ApoA-I Milano. *Arteriosclerosis, thrombosis, and vascular biology*. 2009; 29(10):1496–1501.
34. Anthanont P, Asztalos BF, Polisecki E, Zachariah B, Schaefer EJ. Case report: A novel apolipoprotein A-I missense mutation apoA-I (Arg149Ser)Boston associated with decreased lecithin-cholesterol acyltransferase activation and cellular cholesterol efflux. *Journal of Clinical Lipidology*. 2015; 9(3):390–395. [PubMed: 26073399]
35. Dinnes DL, White MY, Kockx M, Traini M, Hsieh V, Kim MJ, Hou L, Jessup W, Rye KA, Thaysen-Andersen M, Cordwell SJ, Kritharides L. Human macrophage cathepsin B-mediated C-terminal cleavage of apolipoprotein A-I at Ser228 severely impairs antiatherogenic capacity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2016; 30(12):4239–4255. [PubMed: 27630170]
36. Huang Y, DiDonato JA, Levison BS, Schmitt D, Li L, Wu Y, Buffa J, Kim T, Gerstenecker GS, Gu X, Kadiyala CS, Wang Z, Culley MK, Hazen JE, DiDonato AJ, Fu X, Berisha SZ, Peng D, Nguyen TT, Liang S, Chuang C-C, Cho L, Plow EF, Fox PL, Gogonea V, Tang WHW, Parks JS, Fisher EA, Smith JD, Hazen SL. An abundant dysfunctional apolipoprotein A1 in human atheroma. *Nat Med*. 2014; 20(2):193–203. [PubMed: 24464187]
37. Lagerwerf FM, van de Weert M, Heerma W, Haverkamp J. Identification of Oxidized Methionine in Peptides. *Rapid Communications in Mass Spectrometry*. 1996; 10(15):1905–1910. [PubMed: 9004526]
38. Schey KL, Finley EL. Identification of Peptide Oxidation by Tandem Mass Spectrometry. *Accounts of Chemical Research*. 2000; 33(5):299–306. [PubMed: 10813874]
39. Knorre DG, Kudryashova NV, Godovikova TS. Chemical and Functional Aspects of Posttranslational Modification of Proteins. *Acta Naturae*. 2009; 1(3):29–51. [PubMed: 22649613]

40. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nature biotechnology*. 2001; 19(10):952–957.
41. Borhani DW, Rogers DP, Engler JA, Brouillette CG. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. *Proc Natl Acad Sci USA*. 1997; 94(23): 12291–12296. [PubMed: 9356442]

Figure 1. Top-down characterization of two proteoforms of ApoA-I

(A) SIM mass spectrum of intact proteoforms of ApoA-I. Colors represent spectra of two different proteoforms: in pink, canonical ApoA-I, and in blue, a previously uncharacterized proteoform 264.3 Da higher in mass than the canonical. Each proteoform was isolated and fragmented by ETD to create panels B and C. (Tandem mass spectrum not shown). (B) Canonical ApoA-I fragment map. Fragment ions generated by ETD were assigned to specific positions where backbone bonds are cleaved in the fragmentation process; these assignments were then mapped graphically onto the canonical sequence to display fragmentation coverage. A P-score was calculated for confidence in proteoform identification⁴⁰. (C) A fragment map of the higher mass proteoform. Alongside fragmentation covering unmodified residues, 10 fragment ion masses (shown in dark blue) covering bond cleavages past K88 were 264.24 Da higher than expected, confirming the presence of a modification the size of an oleoyl group at the site marked in blue. For reference, the chemical structure of unmodified and oleoylated lysine residues is shown on the right. The fragmentation and intact mass data confirm the characterization of the complete primary sequence and PTM location of this ApoA-I proteoform.

Figure 2. PTM Sites on HDL-bound ApoA-I Structure

(A), Secondary structure of ApoA-I. The colored cylinders represent alpha helices 1 through 11. Sites of observed modifications are marked along the sequence. Location of the dehydration modification was narrowed down to the region between D150 and Y166, but could not be specified further. The red triangle represents the N-terminal 5-residue propeptide cleaved during Proto-ApoA-I activation. (B), 3D structure of 1-43 ApoA-I in reconstructed HDL as described in Borhani et al⁴¹. Side chains are shown for modified residues.

Figure 3. Proteoform abundance per participant
Average percent contribution of each proteoform to total ApoA-I abundance in participant samples.

Figure 4. Proteoform quantification

(A), Volcano plot of differential top-down proteomics results, comparing relative proteoform abundances in the high efflux group to the low efflux group. Color shows statistical significance of the proteoform intensity difference between groups, as measured by one-way ANOVA. In red, proteoforms for which the multiple-test-adjusted P score is at an instantaneous false discovery rate (FDR) $< 5\%$ and thus normalized intensity difference was non-significant; in green and yellow, 6 proteoforms for which the instantaneous FDR is above the significance threshold, numbered in the order they appear in panels B and C. (B), Average intensities of proteoform in each group. Blue: high efflux, red: low efflux. Bars show standard deviation within group. (C), Average intensities of acylated proteoforms in each group. Blue: high efflux, red: low efflux. Bars show standard deviation within group.

Table 1

Individual-Level CHAS Participant Characteristics.

Participant ID	Low Efflux				High Efflux			
	1	2	3	4	5	6	7	8
Efflux	0.67	0.59	0.61	0.72	1.67	1.77	1.7	1.75
Age (y)	70	68	72	65	74	67	65	76
Sex	M	M	M	M	F	F	F	M
Race	B	W	W	W	W	W	W	W
BMI (Kg/m ²)	37	29.6	26.4	30.7	20.7	21.8	16.9	30.1
SBP (mmHg)	123	119	132	118	112	127	118	118
DBP (mmHg)	77	80	75	68	63	76	67	69
Smoker	N	N	N	N	N	N	N	N
TC (mg/dL)	168	147	172	85	228	245	224	153
LDL-C (mg/dL)	101.2	67	108.6	45.8	78.4	124.8	140.8	62.8
HDL-C (mg/dL)	40	50	29	31	130	110	55	39
Triglycerides (mg/dL)	134	151	172	41	98	51	141	256
HDL Particles (μmol/L)	32.2	37.2	26.9	27.9	44.9	49.7	39.3	33.6
LDL Particles (nmol/L)	1084	742	1353	701	834	1187	1007	905
Chol-Meds	N	Y	Y	N	N	N	N	Y
Hs-CRP (mg/L)	5.97	11.75	5.16	1.59	2.2	0.49	0.6	0.79
Glucose (mg/dL)	100	101	113	124	91	104	84	128
CAC (AU)	0	128.3	46.4	64.9	Na	0	0	932.5