Published in final edited form as:

Clin Chim Acta. 2010 July 4; 411(13-14): 972–979. doi:10.1016/j.cca.2010.03.023.

HDL in humans with cardiovascular disease exhibits a proteomic signature

Tomáš Vaisar^{a,*}, Philip Mayer^a, Erik Nilsson^{#,b}, Xue-Qiao Zhao^a, Robert Knopp^a, and Bryan J. Prazien^{#,b}

^aDepartment of Medicine, School of Medicine, University of Washington, Seattle, Washington 98195

^bInsilicos, Seattle WA 98109

Abstract

Background—Alterations in protein composition and oxidative damage of high-density lipoprotein (HDL) have been proposed to impair the cardioprotective properties of HDL. We tested whether relative levels of proteins in HDL₂ could be used as biomarkers for coronary artery disease (CAD).

Methods—Twenty control and eighteen CAD subjects matched for HDL-cholesterol, age, and sex were studied. HDL₂ isolated from plasma was digested with trypsin and analyzed by high-resolution matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and pattern recognition analysis.

Results—Partial least squares discriminant analysis (PLS-DA) of mass spectra clearly differentiated CAD from control subjects with area under the Receiver operating characteristic curve (ROC_{AUC}) 0.94. Targeted tandem mass spectrometric analysis of the model's significant features revealed that HDL₂ of CAD subjects contained oxidized methionine residues of apolipoprotein A-I and elevated levels of apolipoprotein C-III. A proteomic signature composed of MALDI-MS signals from apoA-I, apoC-III, Lp(a) and apoC-I accurately classified CAD and control subjects (ROC_{AUC} = 0.82).

Conclusions—HDL₂ of CAD subjects carries a distinct protein cargo and that protein oxidation helps generate dysfunctional HDL. Moreover, models based on selected identified peptides in MALDI-TOF mass spectra of the HDL may have diagnostic potential.

Keywords

Cardiovascular risk score; inflammation; mass spectrometry; oxidized HDL; partial least squares discriminant analysis

© 2009 Elsevier B.V. All rights reserved.

***Address correspondence to:** Tomáš Vaisar, Division of Metabolism, Endocrinology, and Nutrition, University of Washington, 815 Mercer Street, Seattle WA 98109-4714. Fax: (206) 685-3781, tvaisar@u.washington.edu.. #Mailing address: 111 Queen Anne Avenue N #500, Seattle WA 98109..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Coronary artery disease (CAD) is the leading cause of morbidity and mortality worldwide. Clinical, epidemiological, and genetic studies demonstrate that low levels of high density lipoprotein (HDL) increase the risk for CAD [1]. HDL is a complex, bioactive particle, containing multiple acute phase response proteins, protease inhibitors, and complement regulatory proteins [2-5]. HDL protects against development of atherosclerosis by multiple mechanisms including reverse cholesterol transport [1,6,7], anti-inflammatory, antiapoptotic and anti-oxidant properties [5,8,9].

While low levels of HDL cholesterol (HDL-C) are associated with increased CAD risk [1], high HDL-C levels are not uniformly atheroprotective as indicated by a recent failed clinical trial [10]. This data together with data from animal studies [11] suggests that cardioprotective effects of HDL in humans also depend on the types of particles generated *in vivo* and that HDL in humans with cardiovascular disease can become dysfunctional [5,8]. Both oxidative modifications and alterations in the protein cargo of HDL may alter its biological activity. Indeed, oxidation of HDL has been shown to impair reverse cholesterol transport mediated by HDL [5,12-17]. Furthermore, circulating levels of inflammatory proteins in HDL predict the risk of heart disease in humans [1,4,18] and alterations in the balance between pro- and anti-oxidative enzymes in HDL appear to play a critical role in converting the lipoprotein to a pro-atherogenic form [8,18].

We hypothesized that proteomic fingerprinting of HDL by matrix-assisted laser desorption ionization time-of-flight mass spectrometry could quickly assess HDL's protein cargo and thereby provide indication of HDL disease status. Indeed, SELDI was previously used to detect protein changes in HDL during sepsis [19]. In contrast to SELDI analysis, MALDI-TOF MS provides reproducible, high-resolution spectra [20,21]. High resolution MALDI-TOF MS of peptides from biological samples (such as HDL) leads to data that is intrinsically information rich [22]. Pattern recognition techniques, like partial least squares discriminant analysis (PLS-DA), are useful for extracting relevant features from such large, complicated data sets [23, 24]. PLS-DA is a widely accepted, powerful linear technique for classifying samples using complex data sets, where the number of variables often exceeds the number of samples.

Materials and Methods

Subject selection

The Human Studies Committee at the University of Washington approved all protocols involving human material. Blood anticoagulated with EDTA was collected after an overnight fast from 18 men with established CAD and from 20 apparently healthy men. All CAD subjects were recently diagnosed with symptoms consistent with angina and had abnormal Q waves on their EKG or at least one stenotic lesion (>50% occlusion) on coronary angiography. The CAD subjects were clinically stable, and at least 3 months had elapsed since their acute coronary syndrome. The control subjects had no known history of CAD, were not hyperlipidemic, and had no family history of premature CAD. Subjects smoking, with liver or renal disease, and diabetes were excluded. None of the CAD subjects received lipid-lowering medications for at least 6 weeks before blood collection. All other medications, which are not known to affect lipid metabolism, were continued in the CAD subjects. Levels of plasma LDL and triglycerides were higher in the CAD subjects than in the control subjects, but the 2 groups were otherwise well-matched for known risk factors for vascular disease (Table 1).

HDL isolation and digestion

HDL₂ (d = 1.063-1.125 g/ml) was isolated from plasma by sequential density ultracentrifugation. Prior to isolation plasma samples were supplemented with BHT (an inhibitor of lipid peroxidation) and DTPA (a potent inhibitor of metal catalyzed oxidation chemistry). All samples were stored in at -80°C, which in our experience blocks protein oxidation *ex vivo*. To prevent systematic errors during the analysis, all samples from control and CAD subjects were handled randomized in parallel manner. Trypsin digest (50 ng) without further cleanup was then applied to a sample plate for MALDI-TOF analysis (0.5 µl of 100 ng/µl in matrix solvent (70% acetonitrile, 0.1% TFA) and overlaid with 0.5 µl of MALDI matrix (5 mg/ml α -cyano-4-hydroxy-cinnamic acid, CHCA) in matrix solvent).

Mass spectrometric analysis

Mass spectra were acquired on a MALDI tandem mass spectrometer (Applied Biosystems 4700 Proteomics Analyzer, Foster City CA) operated in the reflectron mode using very stringent criteria for spectra acquisition [25]. Spectra were acquired with extensive sampling across the entire area of the sample spot on the MALDI target. Furthermore, only sub-spectra with base peak ion current between 30×10^4 and 80×10^4 cps were accepted to eliminate both weak and saturated spectra. Finally, the spectra acquisition was configured such that a total of 80 accepted “sub-spectra” (total of 2000 laser shots) were accumulated to produce one MALDI spectrum. Internal calibration with 5 peaks of apoA-I peptides (the major protein of HDL) afforded mass accuracy better than 5 ppm across the acquisition mass range.

Analytical precision of the various steps of data acquisition was evaluated (Supplemental Fig. S1) and indicated that classification precision was improved by averaging several spectra from the same spot. We therefore averaged 4 mass spectra from the same spot to generate a master mass spectrum that was used for further analysis.

We determined precision of these spectra across multiple digestions of the same sample. For signals in the selected features model, precision was 12.8% for a CAD subject and 14.4% for a Control subject (Supplemental material Fig. S1, Table S1). Precision of the full spectrum model classification yielded the ProtCAD score standard deviations 0.08 and 0.12 for CAD and Control spectra, respectively. This is less than 6% of the model range indicating high precision of model predictions with intra-sample variability significantly lower than the class range of ProtCAD score (Supplementary Fig. S2).

To minimize bias in data acquisition, the samples were blinded and randomized prior to mass spectrometric analysis. Because the leave-one-out approach requires separation of the data into disease and control groups, the person who performed leave-one-out data analysis was unblinded to the disease status of the samples during the leave-one-out analysis.

MALDI-TOF reproducibility

MALDI-TOF of complex samples can be susceptible to ion suppression. We therefore tested the influence of HDL variability on reproducibility of signals by spiking four synthetic peptides into trypsin digests of HDL isolated from 5 subjects. The peptides were spiked at different levels to yield signals with high, medium and low intensity. For each of the 5 spiked HDL tryptic digests we acquired triplicate MALDI-TOF spectra in exactly the same way as the samples in the main study. In this experiment a high variation in the signal of spiked peptides across the 5 HDL samples would indicate a large ion suppression effects. To test the potential contribution of MALDI-TOF to differential methionine oxidation we also measured the reproducibility of artifactual methionine oxidation by measuring the signal of the oxidized form of the 2 spiked peptides that contained methionine residues across the 5 HDL samples as described above.

Preprocessing of mass spectra

Raw mass spectra were baseline-corrected and centroided using software provided with the MS instrument (ABI 4700 Explorer software, version 3.5) and analyzed with Matlab (ver. 7.0, Mathworks Inc., Natick MA). To ensure that all spectra had the same mass channels, each mass spectrum was transformed to a vector format by placing the signals in bins (40 ppm wide) in the m/z 800 to 5,000 range, for a total of 45,920 bins (or “channels”) per spectrum. Although the accuracy of the instrument was 5 ppm, 40 ppm bins were used to reduce signal shifts and data vectors were aligned. A threshold of 1/10,000 of the spectrum's total signal was used to remove baseline noise.

The data were separated into calibration and test sets prior to preprocessing to avoid data overfitting, because alignment on m/z axis requires determination of the signal channels that were common in the calibration spectra. Natural signals in this data were typically 10 data points apart, and thus signals at adjacent points were considered artifacts of the data binning process. Channels that contained signal in most of the calibration spectra were determined and signals adjacent to these were shifted to align the data. After alignment and filtering, 2,338 channels contained signals. The data was not de-isotoped, because a single error in the de-isotoping can severely mislead pattern recognition analysis. Pattern recognition techniques essentially combine the isotope signals.

Partial Least Squares-Discriminant Analysis

PLS-DA is a supervised pattern recognition technique that uses a calibration set of samples (cases and controls) for “supervised” creation of a pattern recognition model [23], which is then applied to an independent set of “unknown” samples. PLS-DA yields a single discriminant score [24] that quantifies similarity of the tested spectrum with the model and can be used to predict the class (e.g. CAD or control) of individual samples. PLS-DA models were built with a dummy response matrix containing discrete numerical values for each class (1 for CAD and -1 for control) [23,24]. For each sample being classified, the PLS-DA model then produced a discriminant, which we termed the Proteomics CAD risk score (ProtCAD risk score).

Leave-one-out double cross-validation PLS-DA models for ProtCAD risk score prediction

Both the full spectrum and the selected features predictions were tested using the leave-one-out approach. This approach, especially suitable for studies with limited number of subjects, uses the maximum number of available subjects to build models, and therefore builds the most powerful pattern recognition model possible [24]. The ProtCAD risk score for each subject was determined using a model built from the remaining subjects (e.g., for a CAD subject, the remaining 17 CAD and 20 control subjects). To predict a subject's disease status from the ProtCAD risk score, we compared the value of the ProtCAD risk score to a threshold value corresponding to a selected sensitivity and specificity on a ROC curve.

Double cross-validation was used to avoid overfitting arising from use of all samples for preprocessing or model optimization. Double cross-validation consists of 2 nested cross-validation procedures [26]. The inner loop determines the optimized number of latent variables and the outer loop determines classification of each subject.

Class permutation test

When there are many more m/z values than samples, pattern recognition models may overfit the data. To test that our models did not overfit the data we used a class permutation test [26]. The class labels of the samples (case or control) were randomly permuted and then classification proceeded the same way as when using the correct class labels. The AUC for the ROC curve of each class permutation was recorded. Overfitting would be indicated if the mean

of the permutation results diverted significantly from a random value (0.5) or if the distribution was noticeably skewed.

Identification of informative regression vector features by LC-MALDI-TOF/TOF

We identified peptides corresponding to features in the regression vector of the full spectrum model that were enriched or depleted in the mass spectra of HDL isolated from CAD subjects. We subjected HDL tryptic digests to targeted LC-MALDI-TOF/TOF analysis with internal calibrant included with the matrix. Peptides were identified from tandem mass spectra using a Mascot database search (v2.0, Matrix Science) against the human SwissProt protein database (10/25/2004 ver.) with following parameters: trypsin with up to 2 missed cleavages, methionine oxidation variable modification, precursor tolerance 15 ppm, and fragment ion tolerance 0.2 Da.

Receiver operating characteristic (ROC) curves

Nonparametric ROC curves were constructed from the ProtCAD risk score. Sensitivity and specificity were calculated from the known class identity of each subject in the validation set. Area under the curve (AUC) was determined using the trapezoidal rule [27].

Results

Approach

Our overall goal was to test the hypothesis that MALDI-TOF mass spectra of HDL tryptic digests contain a proteomics signature which can be extracted using pattern recognition techniques and which may allow distinction of normal and disease-modified HDL. Our approach was to isolate HDL₂ from control and CAD subjects, analyze a tryptic digest of HDL₂ proteins by MALDI-TOF-MS, and use pattern recognition analysis of the full-scan mass spectra to select features for further identification by MS/MS (Fig. 1). Identified features were then used to classify disease status of the subjects. We focused our studies on HDL₂ because levels of this buoyant fraction of HDL are strongly associated with CAD status in epidemiological and clinical studies [1,10].

We used MALDI-TOF-MS because it is a sensitive, rapid, and high-resolution mass spectrometric technique that is well-suited for the high-throughput analyses of large-scale clinical studies [28]. Preliminary experiments demonstrated that MALDI-TOF-MS yields reproducible mass spectra from the same spot, from multiple spots, and from parallel HDL digestions (Supplemental Material; Figure S1). Furthermore, four peptides spiked into 5 different HDL samples at high, medium and low levels, exhibited excellent signal reproducibility with average coefficient of variation 13.6% (range 9.2-20.1%)(Supplementary Material, Table S1). Lastly, in a preliminary experiment, MALDI-TOF MS was able to distinguish mixtures with various proportions of CAD and normal HDL (Supplemental Material, Fig. S2).

Full spectrum pattern recognition model

We first built a PLS-DA model using the leave-one-out double cross validation method with all features in the aligned mass spectra. PLS-DA is well suited for analyzing mass spectra, which contain multiple independent signals as well as signals with significant redundancy and/or incomplete selectivity like proteomic mass spectra [24,29]. For each sample being classified, the PLS-DA model produced a discriminant termed the ProtCAD (Proteomics CAD) risk score that was used to predict the clinical status of each subject. To avoid issue of overfitting due to relatively small number of subjects and high number of variables we used a double cross-validation method in which no part of the preprocessing, variable selection or model building

is contaminated with data from the predicted subjects [30,31]. Furthermore, we used a class permutation test to check for bias in the data analysis method.

The ProtCAD risk scores from double cross-validation full spectrum model distinguished the CAD and control subjects with high selectivity ($p < 0.0001$, Mann-Whitney non-parametric test) (Fig. 2A). We then used the calculated ProtCAD risk scores to construct an ROC curve. The ROC curve from this model yielded an AUC of 0.94 and a maximum odds ratio of 68 (Figure 2B). From the full scan model ProtCAD risk score ROC curve, we determined a threshold corresponding to 90% sensitivity (ProtCAD threshold = -0.06). At this threshold, the model correctly classified 16 of 18 CAD subjects and 19 of 20 control subjects.

To test for overfitting, we performed class label permutation analysis. The AUC values for 1000 class label permutations were normally distributed, with a mean of 0.46, a value close to that expected for data with no discrimination power (0.5) (Supplemental Materials Fig. S3). This strongly suggests that we did not overfit the data and there is no bias in our model.

Informative features of the PLS-DA regression vector are linked to specific proteins and post-translational modifications in HDL. PLS-DA models are characterized by a regression vector. The dot product of the unknown sample spectra and regression vector create the protCAD score. The regression vector indicates channels on the m/z axis of a mass spectrum that differentiate the two sample classes (Fig. 2C) [32]. The channels in the regression vector with positive values can be used to identify peptides (and indirectly proteins) that are more abundant in CAD samples than in control samples, while the channels with negative values in the regression vector can be used to identify peptides that are less abundant in the CAD samples. We use the term informative feature to describe mass channels in the regression vector representing significant differences between CAD and control samples.

We then used targeted tandem MS analysis to identify peptides corresponding to informative features that classified CAD subjects. The high mass accuracy of the MALDI-TOF spectra and relatively low complexity of HDL proteome (40-60 proteins typically identified in shotgun proteomics analysis)[5] allowed us to target these features in the LC-MALDI-MS/MS experiment. This approach identified 10 of the 13 strongest features that contributed to the PLS-DA model as well as a number of lower intensity features (Supplemental Material, Table S2).

One class of informative features originated from proteins in the HDL₂ fraction that differed in abundance between CAD and control subjects. In the CAD samples, levels of 2 peptides derived from apolipoprotein(a) (Lp(a); Fig. 3A,B) and 2 peptides from apoC-III were elevated (Fig. 3C,D), while two peptides from apoC-I were decreased (Fig. 3E,F).

A second class of informative features in the pattern recognition model centered on post-translationally modified peptides derived from apoA-I, the major protein in HDL. Interestingly, this class included both native peptides containing methionine 112 (Met112) (KWQEEMELYR, m/z 1411.7077 and VQPYLDDFQKKWQEEMELYR, m/z 2645.4139) and their corresponding oxidized peptides (peptide + 16 amu; m/z 1427.6644 and m/z 2661.3337, respectively) (Fig. 4A,C). MS/MS analysis confirmed the sequences of each peptide and demonstrated that the methionine residue had been oxidized to methionine sulfoxide [Met(O), Met + 16 amu](Fig. 4B). Strikingly, the signals for the oxidized peptides increased in CAD subjects, while those for the native Met112 peptides decreased concomitantly (Fig. 4). Signals for other native and oxidized peptides of apoA-I containing methionine did not follow this trend (Supplemental material Fig. S4) suggesting that the difference in levels of oxygenated Met112 did not result from *ex vivo* oxidation. Furthermore, oxidation of two synthetic peptides containing methionine spiked into the HDL digests from 5 different subjects was highly reproducible (CV of 17 and 22% comparable with CV of native

forms). This indicates that although oxidation occurs during the MALDI process it does not affect the measurement precision and will not mask differential oxidation occurring *in vivo* (Supplementary Materials Table S2). Collectively, these observations support the proposal that HDL₂ from control and CAD subjects differ in their protein cargoes and levels of oxidized methionine residues and that MALDI-TOF MS in combination with PLS-DA and LC-MALDI-TOF/TOF analysis is a powerful technique for identification of these features.

Classification using features identified by LC-MS/MS

The full spectrum model used above to identify informative features contained >2000 signals. For practical diagnostic purposes, it is preferable to use a small number of known features. To test the hypothesis that a limited number of selected features can predict disease status we built a new “selected features” model using peptide signals that were both highly informative in the full spectrum regression vector and were identifiable by MS/MS. A total of 24 signals from 7 peptides were used to build this model, including peptide signals from Lp(a), apoC-III, Apo-CI, and ApoA-I peptides carrying oxidized Met112. The selected features model had high discrimination power, yielding ProtCAD risk scores which separated the 2 classes with high statistical significance characterized by the ROC curve with $ROC_{AUC} = 0.82$ and very similar shape to the full-spectrum ROC curve (Fig. 5). Class permutation confirmed that the selected features model does not overfit the data (mean ROC_{AUC} of 0.43) (Supplemental Figure S5). The intensity differences of the individual ions range from a few percent to more than 90% with highly correlated variation in the signals (up to 0.94). The moderately higher discrimination power of the full spectrum model compared to the selected features model is probably caused by a large number of small signals with modest predictive power in the full spectrum model that were not included in the selected features model. Additional patient samples were not available to independently verify the importance of the selected signal and thus selection bias could contribute to this model. This is not an issue in the full spectrum model.

Discussion

We tested the hypothesis that HDL₂ of CAD subjects carries a unique protein cargo that might serve as a signature for cardiovascular disease. A PLS-DA model based on full MALDI-TOF spectra of tryptic digests of HDL₂ (full spectrum model) distinguished CAD from control subjects with high accuracy. Targeted LC-MS/MS analysis revealed two classes of informative spectral features in the regression vector of this model. The first class consisted of peptides that were differentially detected in tryptic digests of HDL₂. These peptides were derived from proteins apoC-III, apoC-I and Lp(a). The second class contained post-translationally modified peptides with oxidized methionine residues derived from apoA-I. A refined PLS-DA model based only on these identified features distinguished CAD from control subjects with nearly the same sensitivity and specificity as the model based on all the features in the full spectrum. These observations are in agreement with the proposal that oxidative damage and alterations in protein composition impair the cardioprotective properties of HDL in human CAD [5,8, 10] and that the signature of the dysfunctional HDL may serve as additional test for risk of cardiovascular disease.

Proteomic fingerprinting of HDL by MALDI-TOF-MS offers a number of important advantages for building classification models that identify subjects at risk for CAD. First, HDL is causally linked to CAD pathogenesis, which increases the likelihood that informative features in the model will be causally involved in the disease. Second, the HDL proteome is much simpler than the plasma proteome (which has been estimated to contain $>10^4$ different proteins and peptides with relative concentrations ranging over 12 orders of magnitude), which

greatly facilitates MS analysis. Third, our approach interrogates tryptic digests, thereby significantly enhancing the accuracy and precision of the mass spectrometric analysis.

Our initial full spectrum PLS-DA model in conjunction with targeted tandem MS analysis of informative features in the regression vector revealed that number of peptides, including those derived from apoC-III, apoC-I, and Lp(a), were differentially detected in tryptic digests of HDL₂. Interestingly, Lp(a) is a well-established risk factor for CAD [33], supporting the proposal that pattern recognition analysis of the HDL₂ fraction can identify subjects with established CAD. Although Lp(a) is not classically associated with HDL, the elevated levels of Lp(a) peptides in HDL₂ of CAD subjects are likely due to overlap of buoyant density of Lp(a) (1.05-1.2 g/ml) with HDL₂ (1.063- 1.12 g/ml). Thus, Lp(a) contributes to the protein composition of HDL₂ lipoprotein fraction. Levels of apoC-III peptides also appeared higher in the CAD subjects than in the healthy controls. It is noteworthy that apoC-III inhibits lipoprotein lipase and the hepatic uptake of triglyceride-rich lipoproteins, which might increase levels of atherogenic triglyceride-rich lipoproteins [34,35]. In contrast apoC-I, which abundance was decreased in CAD subjects, inhibits cholesteryl ester transfer protein (CETP), a protein known to increase levels of HDL [35]. Thus, alterations in apoC-I and C-III levels might contribute to lipid remodeling and the formation of pro-atherogenic HDL particles.

ApoA-I peptides containing oxidized methionine residues constituted a second class of informative signals in the PLS-DA classification model indicating increased levels of oxidation in CAD subjects. Thus, the level of peptides containing Met(O)112 was elevated in HDL of CAD subjects, whereas the level of peptides containing Met112 was lower. Interestingly, the apoA-I peptides directly adjacent to the peptides containing Met112 were also significantly changed in CAD subjects, and previous studies have demonstrated that oxidation of methionine residues alters the susceptibility of apoA-I to proteolytic digestion [36]. These observations suggest that oxidation of methionine residues in apoA-I increases in CAD subjects and that such oxidation may lead to local changes in the protein's conformation. Although methionine oxidation is facile and often is an artifact of sample preparation and MALDI analysis, several other identified Met containing peptides did not show any difference between control and CAD subjects (Supplementary Material Figure 5), indicating that the observed Met112 oxidation is likely a product of *in vivo* oxidative events. *In vitro* studies demonstrated that lipid hydroperoxides and reactive intermediates derived from myeloperoxidase oxidize methionine residues in apoA-I. Furthermore, oxidation of methionine residues impairs ability of apoA-I to remove cholesterol from lipid-laden macrophages [13,37]. Moreover, recent studies demonstrate that humans with type 2 diabetes have elevated levels of Met(O) in the apoA-I of circulating HDL [38]. Collectively, this data indicate that the informative features which distinguished the CAD and control subjects are directly related to the cardiovascular disease.

We previously used shotgun proteomics to investigate the protein composition of HDL₃. Those studies suggested that levels of apoC-IV, PON-1, complement factor C3, and apoE were higher in HDL₃ of CAD subjects than in that of control subjects [5]. Those earlier studies used an approach specifically designed to maximize number of detected proteins and used liquid chromatography in concert with electrospray ionization (ESI) to introduce peptides into the mass spectrometer. In the current studies we used a different approach which established a signature of HDL from a single MALDI-MS spectrum without specifically attempting to maximize number of identified proteins. Furthermore, we ionized peptides with MALDI rather than with ESI. It is well established that ESI and MALDI ionize different classes of peptides with different efficiencies [39]. For example, arginine containing peptides are much more readily ionized by MALDI than lysine containing peptides. Thus, the differences in identified proteins and protein expression levels observed in the two studies probably reflect the different methods used to ionize peptides. It is also likely that HDL₂ and HDL₃ carry different protein cargoes.

Because pattern recognition methods interrogate multiple features in single analysis [24,40], such methods should be able to identify and classify subjects more accurately than methods that use a single protein marker. Importantly, our PLS-DA model built using selected features was effective at predicting CAD status, with an ROC_{AUC} of 0.82. For a CAD diagnostic test, an ROC_{AUC} of 0.7 to 0.8 is generally considered acceptable, and values over 0.8 are considered excellent [41]. It should be noted that the individual ions included in the model would not be able to distinguish CAD and Control subjects by themselves and the discrimination power comes from their combination in the pattern recognition model. Although the full spectrum mass spectrum PLS-DA model outperformed the selected features model built on the 24 identified features, this is likely explained by presence of many minor features in the full spectrum model that contributed small improvements to the model, which combined to produce better overall performance. However, predictive models built using a small set of well characterized proteomic features are much more suited for the rigorous control and reproducibility required for routine clinical analysis [42]. Furthermore, such well defined features allow for synthesis of stable-isotope labeled peptides or proteins, which would further improve precision of the measurement and minimize ion suppression effects. Therefore the set of biologically relevant features we identified in this study could be directly used in the follow up large scale validation studies without further discovery work. Importantly, the selected features model results compare favorably with other single lipoprotein-associated risk factors, such as LDL-C and HDL-C [43-45]. Thus, our approach appears to have diagnostic potential.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge assistance of Chris Fraley with the statistical analysis. This research was supported by grants from the National Institutes of Health (HL083578, HL086798, P30ES07033, P30DK017047, and P01HL030086). TV was supported by a Pilot and Feasibility Award from the Diabetes and Endocrinology Research Center and a Research and Technology Development award from Washington Technology Center. Mass spectrometry experiments were supported by the Mass Spectrometry Resource, Department of Medicine, and the Mass Spectrometry Core, Diabetes and Endocrinology Research Center, University of Washington.

Abbreviations

PLS-DA	partial least squares discriminant analysis
ProtCAD	proteomics CAD risk score

References

1. Movva R, Rader DJ. Laboratory assessment of HDL heterogeneity and function. *Clin Chem* 2008;54:788–800. [PubMed: 18375481]
2. Khovidhunkit W, Kim MS, Memon RA, et al. Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. *J Lipid Res* 2004;45:1169–1196. [PubMed: 15102878]
3. Van Lenten B, Hama S, de Beer F, et al. Anti-inflammatory HDL becomes pro-inflammatory during the acute phase response. Loss of protective effect of HDL against LDL oxidation in aortic wall cell cocultures. *J Clin Invest* 1995;96:2758–2767. [PubMed: 8675645]
4. Getz GS, Reardon CA. SAA, HDL biogenesis, and inflammation. *J Lipid Res* 2008;49:269–270. [PubMed: 18199807]
5. Vaisar T, Pennathur S, Green PS, et al. Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL. *J Clin Invest* 2007;117:746–756. [PubMed: 17332893]

6. Tall AR. Cholesterol efflux pathways and other potential mechanisms involved in the athero-protective effect of high density lipoproteins. *J Intern Med* 2008;263:256–273. [PubMed: 18271871]
7. Oram JF, Heinecke JW. ATP-binding cassette transporter A1: a cell cholesterol exporter that protects against cardiovascular disease. *Physiol Rev* 2005;85:1343–1372. [PubMed: 16183915]
8. Barter PJ, Nicholls S, Rye KA, Anantharamaiah GM, Navab M, Fogelman AM. Antiinflammatory properties of HDL. *Circ Res* 2004;95:764–772. [PubMed: 15486323]
9. Laberge MA, Moore KJ, Freeman MW. Atherosclerosis and innate immune signaling. *Ann Med* 2005;37:130–140. [PubMed: 16026120]
10. Rader DJ. Illuminating HDL--is it still a viable therapeutic target? *N Engl J Med* 2007;357:2180–2183. [PubMed: 17984168]
11. Warden CH, Hedrick CC, Qiao JH, Castellani LW, Lusis AJ. Atherosclerosis in transgenic mice overexpressing apolipoprotein A-II. *Science* 1993;261:469–472. [PubMed: 8332912]
12. Bergt C, Pennathur S, Fu X, et al. The myeloperoxidase product hypochlorous acid oxidizes HDL in the human artery wall and impairs ABCA1-dependent cholesterol transport. *Proc Natl Acad Sci U S A* 2004;101:13032–13037. [PubMed: 15326314]
13. Shao B, Cavigiolio G, Brot N, Oda MN, Heinecke JW. Methionine oxidation impairs reverse cholesterol transport by apolipoprotein A-I. *Proceedings of the National Academy of Sciences* 2008;105:12224–12229.
14. Shao B, Bergt C, Fu X, et al. Tyrosine 192 in apolipoprotein A-I is the major site of nitration and chlorination by myeloperoxidase, but only chlorination markedly impairs ABCA1-dependent cholesterol transport. *J Biol Chem* 2005;280:5983–5993. [PubMed: 15574409]
15. Panzenböck U, Stocker R. Formation of methionine sulfoxide-containing specific forms of oxidized high-density lipoproteins. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 2005;1703:171–181.
16. Zheng L, Nukuna B, Brennan ML, et al. Apolipoprotein A-I is a selective target for myeloperoxidase-catalyzed oxidation and functional impairment in subjects with cardiovascular disease. *J Clin Invest* 2004;114:529–541. [PubMed: 15314690]
17. Francis GA. High density lipoprotein oxidation: in vitro susceptibility and potential in vivo consequences. *Biochim Biophys Acta* 2000;1483:217–235. [PubMed: 10634938]
18. Navab M, Anantharamaiah GM, Reddy ST, Van Lenten BJ, Ansell BJ, Fogelman AM. Mechanisms of disease: proatherogenic HDL--an evolving field. *Nat Clin Pract Endocrinol Metab* 2006;2:504–511. [PubMed: 16957764]
19. Levels JHM, Pajkrt D, Schultz M, et al. Alterations in lipoprotein homeostasis during human experimental endotoxemia and clinical sepsis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 2007;1771:1429–1438.
20. Villanueva J, Philip J, Chaparro CA, et al. Correcting Common Errors in Identifying Cancer-Specific Serum Peptide Signatures. *J Proteome Res* 2005;4:1060–1072. [PubMed: 16083255]
21. Gillette MA, Mani DR, Carr SA. Place of Pattern in Proteomic Biomarker Discovery. *J Proteome Res* 2005;4:1143–1154. [PubMed: 16083265]
22. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006;116:271–284. [PubMed: 16395409]
23. Barker M, Rayens W. Partial Least Squares for Discrimination. *Journal of Chemometrics* 2003;16:166–173.
24. Lee KR, Lin X, Park DC, Eslava S. Megavariate data analysis of mass spectrometric proteomics data using latent variable projection method. *Proteomics* 2003;3:1680–1686. [PubMed: 12973725]
25. Szajli E, Feher T, Medzihradszky KF. Investigating the quantitative nature of MALDITOF MS. *Mol Cell Proteomics* 2008;7:2410–2418. [PubMed: 18653768]
26. Smit S, van Breemen MJ, Hoefsloot HCJ, Smilde AK, Aerts JMFG, de Koster CG. Assessing the statistical validity of proteomics based biomarkers. *Analytica Chimica Acta* 2007;592:210–217. [PubMed: 17512828]
27. Fawcett T. An Introduction to ROC Analysis. *Pattern Recognition Letters* 2006;27:861–874.

28. Callesen AK, Madsen JS, Vach W, Kruse TA, Mogensen O, Jensen ON. Serum protein profiling by solid phase extraction and mass spectrometry: a future diagnostics tool? *Proteomics* 2009;9:1428–1441. [PubMed: 19235169]
29. Michaud FT, Garnier A, Lemieux L, Duchesne C. Multivariate analysis of single quadrupole LC-MS spectra for routine characterization and quantification of intact proteins. *Proteomics* 2009;9:512–520. [PubMed: 19137552]
30. Somorjai RL, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics* 2003;19:1484–1491. [PubMed: 12912828]
31. Hendriks MM, Smit S, Akkermans WL, et al. How to distinguish healthy from diseased? Classification strategy for mass spectrometry-based clinical proteomics. *Proteomics* 2007;7:3672–3680. [PubMed: 17880000]
32. Heidema AG, Thissen U, Boer JM, Bouwman FG, Feskens EJ, Mariman EC. The association of 83 plasma proteins with CHD mortality, BMI, HDL-, and total-cholesterol in men: applying multivariate statistics to identify proteins with prognostic value and biological relevance. *J Proteome Res* 2009;8:2640–2649. [PubMed: 19351182]
33. Suk Danik J, Rifai N, Buring JE, Ridker PM. Lipoprotein(a), measured with an assay independent of apolipoprotein(a) isoform size, and risk of future cardiovascular events among initially healthy women. *JAMA* 2006;296:1363–1370. [PubMed: 16985228]
34. Ooi EM, Barrett PH, Chan DC, Watts GF. Apolipoprotein C-III: understanding an emerging cardiovascular risk factor. *Clin Sci (Lond)* 2008;114:611–624. [PubMed: 18399797]
35. Shachter NS. Apolipoproteins C-I and C-III as important modulators of lipoprotein metabolism. *Curr Opin Lipidol* 2001;12:297–304. [PubMed: 11353333]
36. Roberts LM, Ray MJ, Shih TW, Hayden E, Reader MM, Brouillette CG. Structural analysis of apolipoprotein A-I: limited proteolysis of methionine-reduced and -oxidized lipid-free and lipid-bound human apo A-I. *Biochemistry* 1997;36:7615–7624. [PubMed: 9200714]
37. Shao B, Oda MN, Oram JF, Heinecke JW. Myeloperoxidase: an inflammatory enzyme for generating dysfunctional high density lipoprotein. *Curr Opin Cardiol* 2006;21:322–328. [PubMed: 16755201]
38. Brock JW, Jenkins AJ, Lyons TJ, et al. Increased methionine sulfoxide content of apoA-I in type 1 diabetes. *J Lipid Res* 2008;49:847–855. [PubMed: 18202432]
39. Stapels MD, Barofsky DF. Complementary Use of MALDI and ESI for the HPLCMS/MS Analysis of DNA-Binding Proteins. *Anal Chem* 2004;76:5423–5430. [PubMed: 15362902]
40. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons; New York: 1989.
41. Lloyd-Jones DM, Liu K, Tian L, Greenland P. Narrative review: Assessment of C-reactive protein in risk prediction for cardiovascular disease. *Ann Intern Med* 2006;145:35–42. [PubMed: 16818927]
42. Anderson NL. The roles of multiple proteomic platforms in a pipeline for new diagnostics. *Mol Cell Proteomics* 2005;4:1441–1444. [PubMed: 16020426]
43. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937–952. [PubMed: 15364185]
44. Walldius G, Jungner I. The apoB/apoA-I ratio: a strong, new risk factor for cardiovascular disease and a target for lipid-lowering therapy--a review of the evidence. *J Intern Med* 2006;259:493–519. [PubMed: 16629855]
45. Walldius G, Jungner I. Apolipoprotein A-I versus HDL cholesterol in the prediction of risk for myocardial infarction and stroke. *Curr Opin Cardiol* 2007;22:359–367. [PubMed: 17556890]

Figure 1. Scheme of the experimental design

MALDI-TOF spectra of 18 CAD and 20 control subjects were used to generate full spectrum model. The 50 most significant informative features (a) were subjected to targeted LC-MS/MS and then a selected features model was built using 24 identified informative features corresponding to 7 peptides (b).

Figure 2. Full spectrum PLS-DA model

Data from all the subjects in each group and a leave-one-out double cross validation approach were used to build a PLS-DA model to determine a ProtCAD risk score for each subject. (A) The ProtCAD risk score distinguished subjects with high statistical significance (B). An ROC curve constructed using the ProtCAD risk score showed high selectivity and sensitivity (ROC_{AUC} of 0.94). (C) Regression vector of the full spectrum PLS-DA model indicates number of features distinguishing CAD and control subjects. The x axis (m/z) represents mass channels of the MALDI-TOF mass spectrum. *Positive* and *negative* features on the regression vector correspond to *enrichment* and *depletion* of the signals from CAD samples relative to control samples. Features were identified using the targeted LC-MALDI-TOF/TOF approach. (* Feature not identified)

Figure 3.

Tandem mass spectrometric analysis identifies changes in relative peptide abundance as one class of informative features. Mass peaks corresponding to regression vector informative features that distinguished between control and CAD subjects were subjected to targeted LC-MALDI-TOF/TOF MS/MS analysis. (A) Positive informative feature, m/z 1440. (B) MS/MS spectrum of m/z 1440 identifying the feature as Lp(a) peptide NPDVAAPYCYTR. (C) Positive informative feature, m/z 1716. (D) MS/MS spectrum of m/z 1716 identifying the feature as apoC-III peptide DALSSVQESQVAQQAR (E) Negative informative feature, m/z 1488. (F) MS/MS spectrum of m/z 1488 identifying the feature as apoC-I peptide MREWFSETFQK. Note that positive signals indicate an increase in relative abundance of the

feature in CAD subjects, while negative signals indicate a relative decrease in abundance. All peptides exhibited MASCOT database search scores with a confidence interval of 100%.

Figure 4. Tandem mass spectrometric analysis identifies methionine oxidation as a second class of informative features

Ions corresponding to full spectrum PLS-DA model informative features that distinguished between control and CAD subjects were subjected to LC-MALDI-TOF/TOF MS/MS analysis. (A) Informative features m/z 1411 (M) and m/z 1427 (M + 16). Note that the feature at m/z 1411 (M) is negative, while the feature at m/z 1427 (M + 16) is positive, indicating depletion and enrichment of the corresponding peptides in CAD subjects. (B) MS/MS analysis identifying ions of m/z 1411 and m/z 1427 as apoA-I peptide KWQEEMELYR and KWQEEM(O)ELYR, respectively. (C) Informative features m/z 2645 (M) and 2661 (M+16). Note that the feature m/z 2645 is negative, while the feature m/z 2661 (M + 16) is positive in CAD subjects. MS/MS analysis of the ions identified them as apoA-I peptides VQPYLDDFQKKWQEEMELYR and VQPYLDDFQKKWQEEM(O)ELYR, respectively. M, peptide; M + 16, oxygenated peptide (data not shown).

Figure 5. Selected features model performance

PLS-DA model was build using leave-one-out double cross validation method and 24 identified informative features. ROC curve of constructed using the ProtCAD risk scores derived from the selected features model.

Table 1

Clinical characteristics of study subjects.

	Control	CAD
N	20	18
Age (y)	57 (6)	57 (6)
Male (%)	100	100
BMI	25.2 (1.5)	30.3 [#] (4.2)
Aspirin	0%	100%
Antihypertensive	0%	100%
Cholesterol (mg/dl)	197 (13)	223 (27)
Triglycerides (mg/dl)	104 (29)	146 [#] (67)
HDL-C (mg/dl)	42 (8)	41 (8)
LDL-C (mg/dl)	134 (14)	160 [#] (25)

Results represent means (SD).

C, cholesterol.

[#]P<0.02 by the paired 2-tailed Student's *t*-test.