Universiteit
Leiden
The Netherlands

Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches

Demus, D.A.

Citation

Demus, D. A. (2024, October 1). *Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches*. Retrieved from https://hdl.handle.net/1887/4093481

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4093481

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 4

4. Large-scale analysis of apolipoprotein CIII glycosylation by ultrahigh resolution mass spectrometry

Daniel Demus[1,2], Annemieke Naber[3], Viktoria Dotz[1†], Bas C. Jansen[1,2], Marco R. Bladergroen[1], Jan Nouta[1], Eric J. G. Sijbrands[3], Mandy van Hoek[3], Simone Nicolardi[1], Manfred Wuhrer[1]

[1] Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, Netherlands

[2] Ludger Ltd., Culham Science Centre, Abingdon, Oxfordshire, United Kingdom

[3] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, Netherlands

Reprinted and adapted: Demus D, Naber A, Dotz V, Jansen BC, Bladergroen MR, Nouta J, Sijbrands EJG, Van Hoek M, Nicolardi S, Wuhrer M. Large-Scale Analysis of Apolipoprotein CIII Glycosylation by Ultrahigh Resolution Mass Spectrometry. Front Chem. 2021 May 7;9:678883. doi: 10.3389/fchem.2021.678883

Copyright © 2021 The Authors

ABSTRACT

Apolipoprotein-CIII (apo-CIII) is a glycoprotein involved in lipid metabolism and its levels are associated with cardiovascular disease risk. Apo-CIII sialylation is associated with improved plasma triglyceride levels and its glycosylation may have an effect on the clearance of triglyceride-rich lipoproteins by directing these particles to different metabolic pathways. Large-scale sample cohort studies are required to fully elucidate the role of apo-CIII glycosylation in lipid metabolism and associated cardiovascular disease. In this study, we revisited a high-throughput workflow for the analysis of intact apo-CIII by ultrahigh-resolution MALDI FT-ICR MS. The workflow includes a chemical oxidation step to reduce methionine oxidation heterogeneity and spectrum complexity. Sinapinic acid matrix was used to minimize the loss of sialic acids upon MALDI. MassyTools software was used to standardize and automate MS data processing and quality control. This method was applied on 771 plasma samples from individuals without diabetes allowing for an evaluation of the expression levels of apo-CIII glycoforms against a panel of lipid biomarkers demonstrating the validity of the method. Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced by apo-CIII sialylation. Interestingly, the association of apo-CIII glycoforms with triglyceride levels was found to be largely independent of body mass index. Due to its precision and throughput, the new workflow will allow studying the role of apo-CIII in the regulation of lipid metabolism in various disease settings.

INTRODUCTION

Lipid metabolism is regulated by complex biological mechanisms in which apolipoproteins – proteins embedded in lipoprotein particles – modulate the transport and availability of blood lipids[117]. Apolipoprotein-CIII (apo-CIII) is a 79 amino acid glycoprotein present on the surface of triglyceride-rich lipoproteins and is an inhibitor of lipoprotein lipase (LPL), an enzyme that hydrolyses triglycerides into fatty acids[118-120]. Apo-CIII has been associated with increased monocyte adhesion to the endothelium[121] and enhanced binding of apoB-containing lipoproteins to vascular proteoglycans[122]. High apo-CIII levels are associated with hypertriglyceridemia[123-125] and increased cardiovascular disease risk in the general population[123, 126, 127] and diabetes mellitus[128, 129]. Recently, the clinical interest for this protein has increased due to the promising results obtained from antisense oligonucleotide-based therapies for the reduction of apo-CIII and triglyceride levels[125, 130-132].

Apo-CIII exists in four major proteoforms: one non-glycosylated form (apo-$CIII_{0a}$) and three *O*-glycosylated variants with a core 1 (T-antigen) glycan structure, which is either non-sialylated (apo-$CIII_{0c}$), monosialylated (apo-$CIII_1$) or disialylated (apo-$CIII_2$)[133, 134]. Low-abundance fucosylated, non-sialylated apo-CIII forms have also been described[35]. It has been shown that not only the levels of apo-CIII but also the specific glycoforms and their relative expression control triglyceride metabolism[57, 58]. For example, an inverse association between apo-$CIII_2$/apo-$CIII_1$ ratio and triglyceride levels has been confirmed by two independent studies[57, 59]. It has also been shown that sialylation modulates the apo-CIII affinity for hepatic receptors that clear lipoprotein particles[59] and that different proteoforms of apo-CIII may affect the inhibition of LPL[135] and the interaction of LDL with the vascular wall[136]. Since the association of different apo-CIII proteoforms with specific cardiometabolic endpoints has not been fully elucidated, further research in large sample cohorts is warranted.

We have developed a high-throughput method based on magnetic-bead extraction and matrix-assisted laser desorption/ionization (MALDI) and ultrahigh-resolution Fourier

transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS) for the analysis of serum apo-CIII proteoforms[35, 137]. Apo-CIII contains methionine residues, which can be (partially) oxidised during biological processes *in vivo*[138], sample processing and freeze-thaw cycles[139]. The presence of different oxidoforms increases mass spectra complexity, which complicates MS data processing and affects the repeatability of measurements. Although the analyte oxidation may not pose a serious challenge in MALDI MS analysis of single samples, it can seriously impact the precision and accuracy of quantitative measurements in large sample cohorts.

In the current study, we have applied a modified workflow employing a previously established MALDI FT-ICR MS method preceded by a chemical oxidation step for complete oxidation of apo-CIII methionine residues. This results in highly reproducible high-throughput measurements for relative quantification of apo-CIII proteoforms in a large number of plasma samples varying in protein oxidation levels. Furthermore, we have adopted sinapinic acid (SPA) as a MALDI matrix to minimize the loss of sialic acid induced by MALDI. The high-throughput quantitation software, MassyTools[140], was here further developed to facilitate semi-automated MS data processing for intact proteins. The validity of the new workflow was tested on a clinical cohort comprised of 771 plasma samples, which allowed the evaluation of the relationship between apo-CIII glycoforms and metabolic biomarkers, such as BMI, cholesterol, and triglyceride levels.

MATERIALS AND METHODS

Clinical samples

Blood plasma samples from a group of individuals without diabetes of the DiaGene Study were used. The DiaGene Study is a case-control study comprising 1886 type-2 diabetes patients and 854 controls without diabetes, from the areas of Eindhoven and Veldhoven, in the Netherlands. The study is described in detail elsewhere[141]. For the current study, after quality control, apo-CIII glycosylation data were available for 771 samples, in 746 whereof, data on clinical characteristics were available. All participants gave their written informed consent. This study was approved by the Medical Ethics Committees of the Erasmus University Medical Center, Catharina Hospital and Maxima Medical Center.

Clinical information and blood samples were obtained at baseline, as described previously[141]. Triglycerides and cholesterol concentrations were measured using standard clinical chemistry essays and reported by the collecting clinic. Non-high-density lipoprotein (non-HDL)-cholesterol was calculated by subtracting the high-density lipoprotein (HDL)-cholesterol from the total cholesterol, body mass index (BMI) was calculated by dividing the body mass (in kg) by the square of the body length (in m). Triglyceride concentrations were logarithmically transformed before linear regression analysis, because of non-normal distribution.

Chemicals

Magnetic beads (Dynabeads RPC-18) were purchased from Invitrogen Dynal AS, Oslo, Norway. VisuCon-F plasma standard from Affinity Biologicals, Ancaster, Canada. Hydrogen peroxide 30%, ethanol and acetone were purchased from Merck, Darmstadt, Germany. Acetonitrile (ACN) was from Biosolve Chimie SARL, France. Trifluoroacetic acid (TFA) was purchased from Thermo Fisher Scientific, Tewksbury, MA. Sinapinic acid (SPA) and α-Cyano-4-hydroxycinnamic acid (HCCA) from Sigma-Aldrich. Ultrapure milliQ water (18 MΩ·cm at 25°C) was used throughout.

High-throughput RP-C18 solid-phase extraction (SPE) of plasma proteins

Plasma standards (VisuCon-F) were randomised over cohort sample plates. 10 µL of human blood plasma was transferred from the cohort sample plates into 96-well skirted PCR plates (4ti-0960/C, 4titude, Dorking, UK). 15 µL of an oxidizing solution (12% H_2O_2/0.5% TFA in water) was added to each sample. The plate was sealed with a pierce foil seal (4ti-0521, 4titude Ltd, Wotton, Surrey, UK) and incubated for 1 hour at 37°C. Subsequently, the plate was cooled at 4°C for 30 min and centrifuged briefly at 800 × *g*. The pierce foil was removed and the plate was transferred onto a liquid handling robot (Hamilton, Bonaduz, Switzerland) where solid-phase extraction (SPE) was carried out as follows: the RP-C18 beads were activated by three washes using acetonitrile (ACN) and trifluoroacetic acid (TFA) solution in water (first wash using 50% ACN/0.1% TFA followed by two washes with 0.1% TFA). Next, plasma samples were transferred to the activated beads and incubated for 10 min at room temperature. The incubation was followed by three washes: one wash using 15% ACN and two washes with 0.1% TFA. Proteins were eluted by adding 15 µL of 50% ACN/0.1% TFA in water and incubating for 5 min at room temperature. For MALDI spotting, 2 µL of sample eluates were mixed with either 16 µL of sinapinic acid solution (1.3 g/L in 2:1 v/v ethanol/acetone) or 15 µL alpha-cyano-4-hydroxycinnamic acid solution (1.4 g/L in 2:1 v/v ethanol/acetone). 1.5 µL of each sample mix was spotted in duplicate onto a MALDI AnchorChip target plate (800 µm anchor diameter; Bruker Daltonics, Bremen, Germany) and allowed to air-dry before MALDI MS analysis.

MALDI FT-ICR mass spectrometry and MS data analysis

All MALDI MS experiments were performed on a 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) equipped with a Smartbeam II™ laser system (355 nm wavelength) and a ParaCell detector. All spectra were acquired in the *m/z*-range 3495-30000, from the average of ten scans of 200 laser shots (at 500 Hz) each using 524288 data points. The analyser parameters were set as previously reported[142]. Briefly, measurements were performed with high trapping potentials (up to 8.5 V) and high ParaCell DC biases (up to 8.8 V) and with a Sweep excitation power of 57% for 13.5 µs.

A laser power of 20% and "medium" laser focus was used for MS measurements using HCCA, while a laser power of 30% and "ultra-large" focus was used with SPA.

MS data was converted as *.xy files using DataAnalysis (Bruker, ver. 5.0 SR1) before subsequent processing by MassyTools[140]. Originally developed for automated processing of MALDI MS data from released *N*-glycans, MassyTools was here adjusted to large molecules. Specifically, the modification entailed that the full isotopic envelope is calculated for each molecule, from which the largest contributing isotopes are selected for quantitation until a user-specified threshold (v1.0.2-alpha b180626a). The processing parameters listed below are explained in detail in work by Jansen et al.[140]. All spectra were internally calibrated using the most abundant isotopic peak of oxidoforms of the four major apo-CIII proteoforms (apo-CIII0a+2Ox, apo-CIII0c+2Ox, apo-CIII1+2Ox and apo-CIII2+2Ox; m/z values of 8797.228, 9162.361, 9453.456 and 9744.552, respectively). Data curation was performed in a semi-automated manner using Microsoft Excel and a custom-made script in RStudio (version 1.1.463). Spectra were excluded from further analysis if either calibration by MassyTools failed or if their "Fraction of Analyte Area - Background Area above S/N cut-off" parameter value was below 3*IQR of the mean value). Excluded spectra were spot-checked by visual inspection to confirm their low quality, especially in case of spectra which were close to the cutoff. Out of each sample spot duplicate, the spectrum with the higher absolute area sum of the four proteoforms was kept for further processing. All analytes of interest had to pass set cut-off values for minimum two quality parameters: signal-to-noise ratio (S/N) of the most-abundant isotopic peak, mass accuracy (MME) and isotopic pattern quality (IPQ). The cut-off values were as follows: S/N ≥ 9, MME ≤ 10 and IPQ ≤ 0.2. Spectra, in which an analyte met two out of three quality parameters, were subjected to additional manual inspection for the presence of interfering species. Peak intensities were measured as peak areas of the most abundant isotopic peak. MassyTools extraction parameters were chosen experimentally for optimal data extraction and were as follows: peak extraction width = 0.3, minimum fraction of total isotopic distribution used for extraction = 0.5, background detection window = 20.

Intra- and inter-plate repeatability were calculated based on 136 plasma standards (VisuCon) distributed over 31 sample plates. For intra-plate repeatability, the mean values of relative peak intensities of apo-CIII0a+2Ox, apo-CIII0c+2Ox, apo-CIII1+2Ox and apo-CIII2+2Ox, standard deviations (SD) and coefficients of variations (CV) were calculated for the standards per sample plate. To assess average intra-plate repeatability, the per-plate CVs were averaged. The per-plate means were averaged, SDs and CVs were calculated over all 31 plates to estimate the inter-plate repeatability. Statistical analyses were performed in SPSS version 25 and data visualization was performed in RStudio (version 1.1.463) and Microsoft Excel. To compare apo-CIII glycosylation in males and females, an independent-samples T-test was performed. The distribution of clinical variables was considered normal when Skewness and Kurtosis were within the range of -1 to +1. The association of apo-CIII glycoforms with age, BMI, HDL-cholesterol, non-HDL-cholesterol, low-density lipoprotein (LDL)-cholesterol, triglycerides and total cholesterol was analysed using univariate linear regression models, where apo-CIII glycoforms were entered as dependent variables. Triglycerides were logarithmically transformed prior to analysis, because of a non-normal distribution, to derive p-values and R-square. To investigate the influence of BMI on the association of apo-CIII glycosylation with triglyceride levels, a multiple linear regression model was used which included BMI and triglycerides as independent variables. $P < 0.05$ was considered statistically significant.

RESULTS AND DISCUSSION

The controlled oxidation of methionine residues reduces the complexity of mass spectra

A common event observed in proteomics studies is the oxidation of methionine residues due to biological and pathological processes occurring *in vivo*[138], sample storage and processing[139]. These reactions are so common that, in bottom-up studies, methionine oxidation is often included in database searches as a variable or even fixed modification. However, in general, peptides generated by enzymatic digestions (e.g. using trypsin) are small and often do not contain methionine residues and although a (partial) oxidation of methionine residues increases the number of peptides in a digest, these unwanted reactions do not significantly affect the analysis[143-145]. In a clinical setting, the use of fresh samples may be an ideal approach, especially for diagnostic purposes based on profiling of intact protein. Whereas in cohort studies, the collection of large numbers of clinical samples, their storage, transfer between institutions and multiple use may lead to oxidation processes that affect the analysis of intact proteins by increasing the heterogeneity of proteoforms detected in a spectrum. The higher complexity increases the chance of overlapping signals and reduces the sensitivity of measurements due to the spreading of the signal over a higher number of species. This results in MS spectra characterised by the presence of interfering species and very low abundant analyte peaks, which do not meet acceptable spectral quality criteria for consideration in the statistically significant quantitative analysis.

Apo-CIII contains two methionine residues which can be oxidised to form methionine sulfoxide (MetO) and methionine sulfone (MetO2) although this latter form requires harsher oxidizing conditions[146, 147]. Previously, MALDI-TOF MS methods have been used in analyses of apo-CIII proteoforms[148]. In such low-resolution methods, apo-CIII oxidoforms cannot be resolved, however, their presence results in the broadening and distortion of apo-CIII proteoforms' signals, which can eventually overlap or interfere with signals of other proteins affecting their quantification. The chance of signal interference increases when SPE is used for the enrichment of apo-CIII as it leads to the

co-enrichment of other small plasma proteins. The application of more specific enrichment methods, such as immunocapture, may help to reduce signals interfering with the various apo-CIII oxidoforms, but was not implemented in the present study for simplicity reasons. Of note, apo-CIII proteoforms have been analysed by methods employing LC systems[149, 150]. Despite certain advantages over MALDI-TOF MS, such as absolute quantification and enhanced resolution, the throughput of this approach remains relatively low.

Recently, we have developed a method for the analysis of apo-CIII proteoforms using ultrahigh-resolution MALDI FT-ICR MS[35]. Apo-CIII proteoforms were mainly detected as singly-charged ions. Thus, apo-CIII oxidoforms (1 and 2 times MetO) were detected at +15.995 Th and +31.990 Th from the non-oxidised forms (**Figure 1**). The degree of methionine oxidation of apo-CIII can vary greatly but the complete oxidation of apo-CIII (i.e. 100% conversion of the two methionine residues to MetO) is not commonly observed (**Supplementary Figure S1**). Therefore, for each of the four major proteoforms of apo-CIII, two additional oxidoforms were observed in MALDI FT-ICR MS spectra resulting in twelve proteoforms (**Figure 1**). In addition to that, we were able to detect C-terminal alanine cleaved and fucosylated proteoforms (**Supplementary Table S1**).

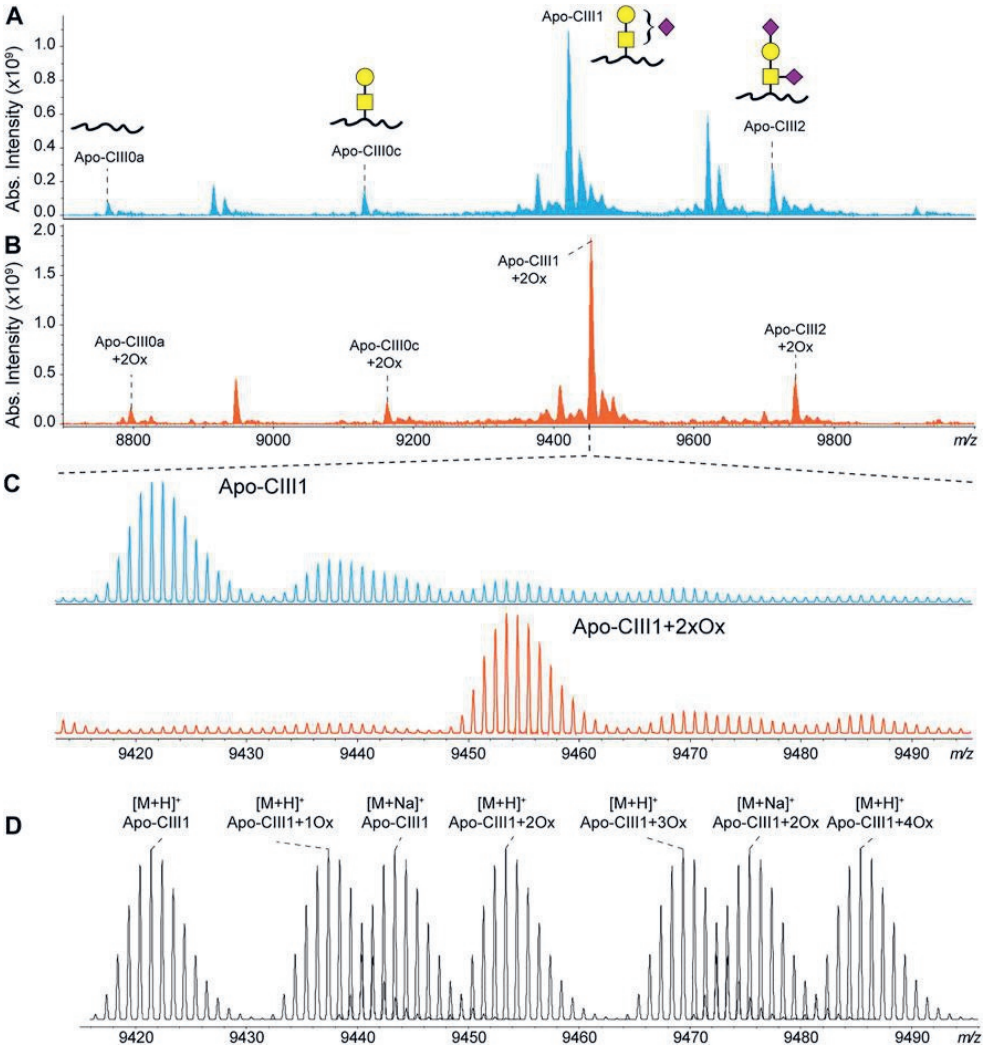

FIGURE 1. MALDI FT-ICR mass spectra of apo-CIII proteoforms from (A) non-treated (blue) and (B) H_2O_2 treated (orange) human blood plasma samples. The four most abundant proteoforms of apo-CIII are annotated. Mass spectra from non-treated samples are characterised by the presence of oxidoforms as exemplified for apo-CIII1 in panel (C). The H_2O_2 treatment allows the almost complete conversion of methionine residues to methionine sulfoxide with a minor conversion to methionine sulfone. Theoretical isotopic distribution of apo-CIII1 proteoforms are depicted in panel (D). For graphical representations of glycan structures: yellow square (*N*-acetylgalactosamine), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid).

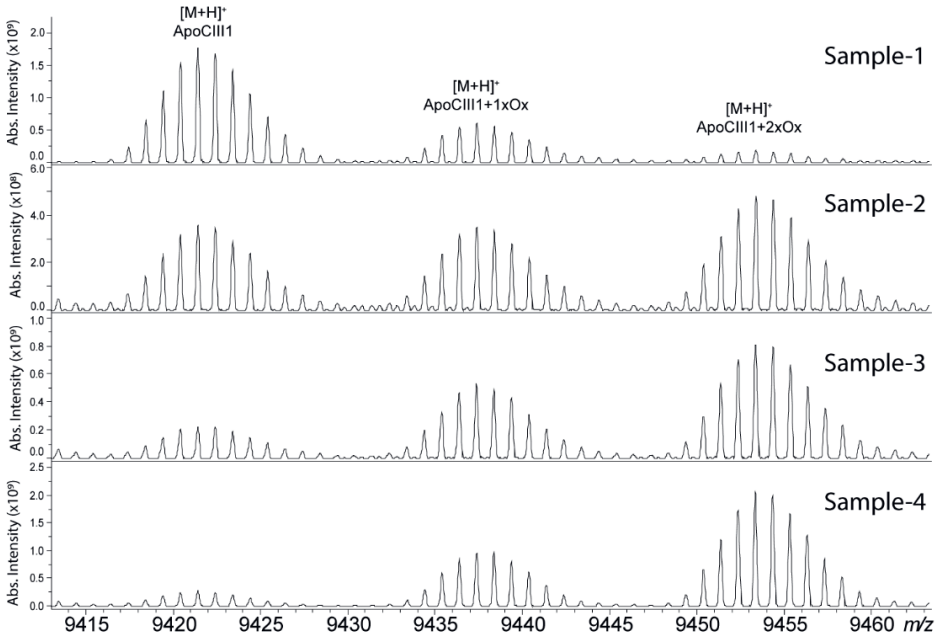

Figure S1. Degree of apo-CIII oxidation in four different plasma samples (without controlled oxidation). The top spectrum was obtained from a fresh aliquot of plasma sample stored at -80 °C after collection. The other spectra were obtained from plasma aliquots that underwent three freeze/thaw cycles.

To reduce sample complexity, we included an oxidation step with hydrogen peroxide to perform a controlled oxidation of both apo-CIII methionine residues to MetO (**Figure 1**). While the implementation of the oxidation step added two hours to the workflow for the analysis, it reduced the heterogeneity of the spectra and facilitated MS data processing using MassyTools software (see Implementation of MassyTools software for high-throughput MS data processing)[140]. The efficiency of the controlled oxidation was tested on 136 standard and 771 clinical plasma samples. The relative intensities between the non-, mono- and di-oxidised forms of apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$ are reported in **Supplementary Table S2 and S3**. Oxidation rates over 90% were found for apo-$CIII_1$ and apo-$CIII_2$. Oxidation efficiency seemed to be lower for apo-$CIII_{0a}$, apo-$CIII_{0c,}$ however, close inspection of the spectra revealed the presence of interfering species that contributed to the signal of the non- and mono-oxidised forms of apo-$CIII_{0a}$, apo-$CIII_{0c}$ thus increasing their apparent relative intensity (**Supplementary Figure S2**). Therefore, the controlled oxidation was considered efficient for all four

proteoforms by providing consistent oxidation rates across the standard and clinical plasma samples. These results supported our strategy of using only the signal of the di-oxidised apo-CIII proteoforms for further statistical analysis. The good efficiency and repeatability of the oxidation step allowed us to assess associations between apo-CIII glycosylation and different lipid markers using only the signal of the di-oxidised forms.

Table S1. **Apolipoprotein-CIII proteoforms detected using the MALDI-FT-ICR MS method**. The method allowed to detect four major glycoforms that occur in three different oxido-forms. The combination of glycosylation and oxidation variation results in a total of 12 proteoforms. Detectable C-terminal alanine-cleaved proteoforms are included in the table. TruncA, truncation of a single C-terminal alanine; TruncAA, truncation of two C-terminal alanine residues.

Apolipoprotein-CIII amino acid sequence: SEAEDASLLSFMQGYMKHATKTAKDALSSVQESQVAQ-QARGWVTDGFSSLKDYWSTVKDKFSEFWDLDPEVRPTSAVAA		
Observed apolipoprotein-CIII glycoform	**Calculated *m/z* value for the most abundant peak within an isotopic distribution $[M+H]^+$**	**Ref.**
Apo-CIII$_{0a}$	8765.238	[137]
Apo-CIII$_{0a}$+Ox	8781.233	-
Apo-CIII$_{0a}$+2Ox	8797.228	-
truncA Apo-CIII$_{0a}$+2Ox	8726.191	[137]
truncAA Apo-CIII$_{0a}$+2Ox	8655.154	-
Apo-CIII$_{0C}$	9130.371	[137]
Apo-CIII$_{0C}$+Ox	9146.366	-
Apo-CIII$_{0C}$+2Ox	9162.361	-
truncA Apo-CIII$_{0C}$+2Ox	9091.323	[137]
truncAA Apo-CIII$_{0C}$+2Ox	9019.284	-
Apo-CIII$_1$	9421.466	[137]
Apo-CIII$_1$+Ox	9437.461	-
Apo-CIII$_1$+2Ox	9453.456	-
truncA Apo-CIII$_1$+2Ox	9382.419	[137, 151]
truncAA Apo-CIII$_1$+2Ox	9311.382	-
Apo-CIII$_2$	9712.562	[35, 137]
Apo-CIII$_2$+Ox	9728.557	-
Apo-CIII$_2$+2Ox	9744.552	-
truncA Apo-CIII$_2$+2Ox	9673.514	[151]
truncAA Apo-CIII$_2$+2Ox	9602.477	-

Table S2. Relative intensity (and standard deviation) of the oxidoforms of the most abundant apo-CIII proteoforms in 136 standard plasma samples after controlled oxidation. The high value of the di-oxidised forms indicates a high oxidation efficiency. The relative intensities may be affected by the presence of interfering peaks as shown in Figure S2.

	Non-Ox	Mono-Ox	Di-Ox
apo-$CIII_{0a}$	4.3% (2.0%)	17.0% (5.0%)	78.7% (6.4%)
apo-$CIII_{0c}$	3.3% (0.8%)	11.6% (1.0%)	85.1% (1.7%)
apo-$CIII_{1}$	1.2% (0.4%)	6.6% (0.8%)	92.2% (1.0%)
apo-$CIII_{2}$	0.7% (0.4%)	6.5% (0.9%)	92.8% (1.0%)

Table S3. Relative intensity (and standard deviation) of the oxidoforms of the most abundant apo-CIII proteoforms in 771 control plasma samples after controlled oxidation. The high value of the di-oxidised forms indicates a high oxidation efficiency. The relative intensities of non- and mono-oxidised apo-$CIII_{0a}$ and apo-$CIII_{0c}$ were increased as a result of the presence of interfering species as exemplified in Figure S2.

	Non-Ox	Mono-Ox	Di-Ox
apo-$CIII_{0a}$	8.8% (6%)	22.7% (12.3%)	68.4% (16.3%)
apo-$CIII_{0c}$	4.3% (4.3%)	11.6% (1.8%)	84.1% (5.0%)
apo-$CIII_{1}$	2.0% (0.6%)	7.4% (0.9%)	90.6% (1.2%)
apo-$CIII_{2}$	1.7% (1.1%)	6.3% (1.4%)	91.9% (1.7%)

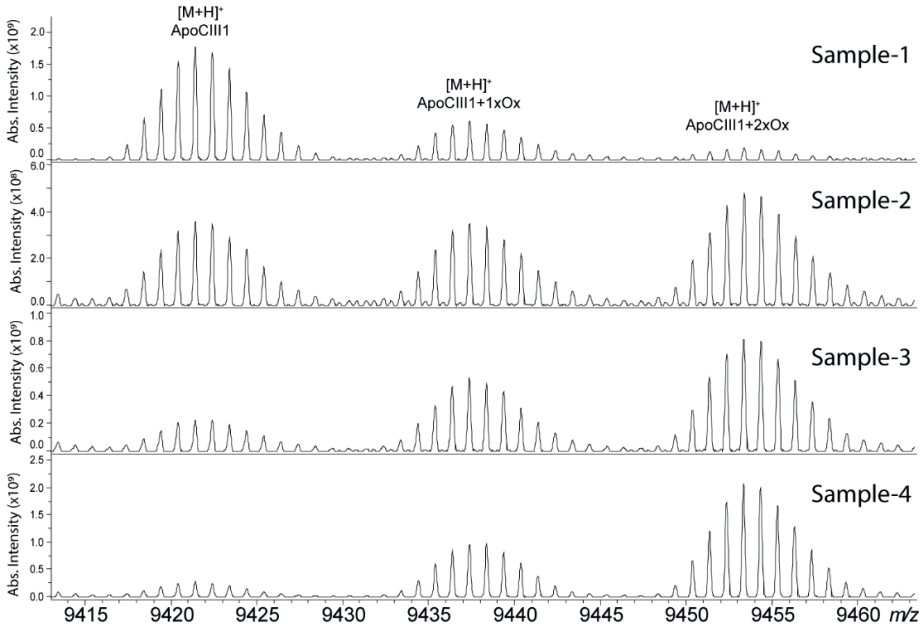

Figure S1. Degree of apo-CIII oxidation in four different plasma samples (without controlled oxidation). The top spectrum was obtained from a fresh aliquot of plasma sample stored at -80 °C after collection. The other spectra were obtained from plasma aliquots that underwent three freeze/thaw cycles.

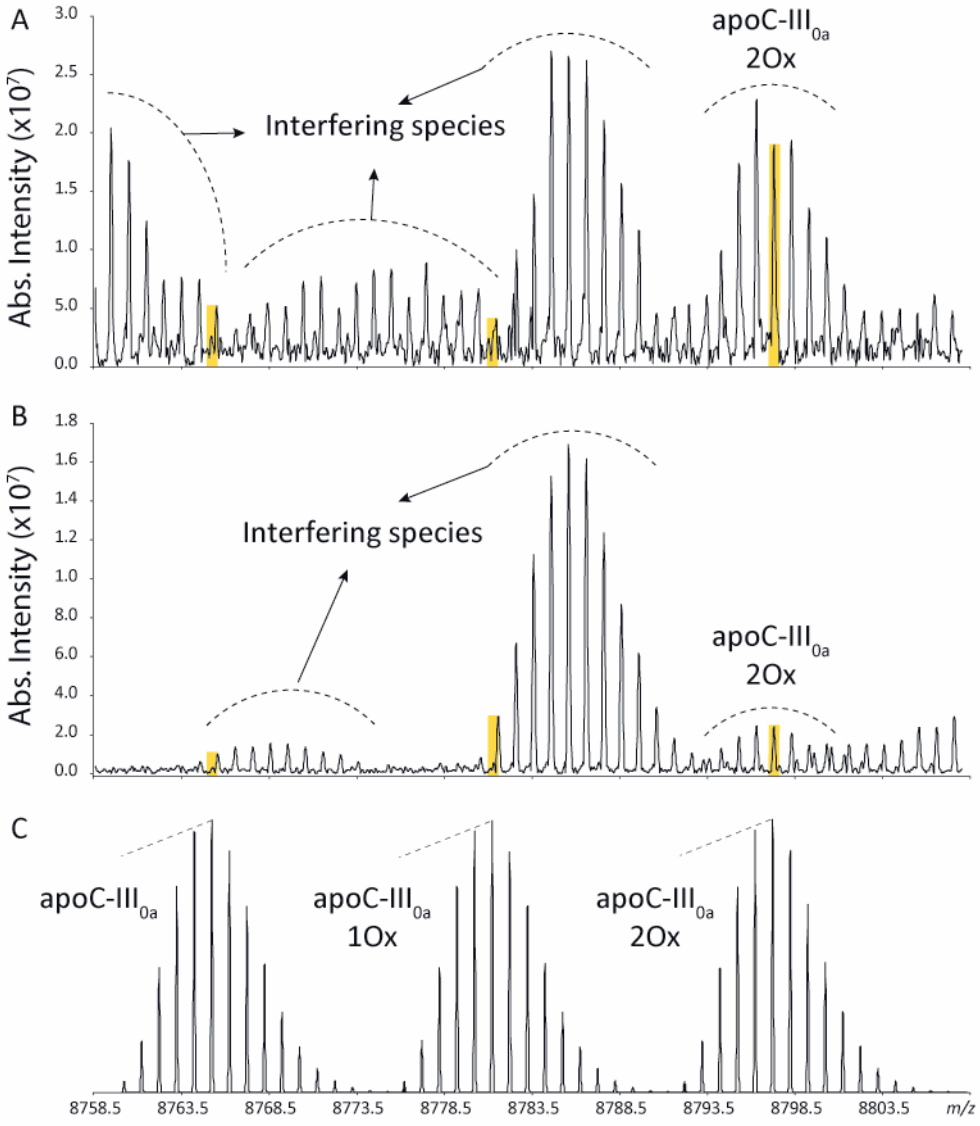

Figure S2. A and B) Enlargements of MALDI FT-ICR MS spectra in the *m/z*-region of apo-CIII$_{0a}$ oxidoforms. The presence of interfering species can affect the determination of the intensity of the most intense isotopic peaks which is calculated in the regions of the spectra highlighted in yellow. In fact, the high relative abundance of Non-Ox and Mono-Ox apo-CIII0a forms (namely 8.8% and 22.7%) in Table S3 was attributed to the presence of interfering species rather than an incomplete oxidation. C) Theoretical isotopic distributions.

Minimizing sialic acid loss using sinapinic acid (SPA) as MALDI matrix

In our previously reported ultrahigh-resolution MALDI FT-ICR MS method for the analysis of apo-CIII proteoforms HCCA was used as a MALDI matrix[35, 137]. This compound was chosen to increase the sensitivity for other serum peptides and small proteins present in C18-SPE eluates obtained from the high-throughput enrichment step using magnetic beads. However, it is known that sialic acid loss can result from in-source decay fragmentation events of glycan structures even when linked to peptides and proteins. In fact, previous reports on the analysis of apo-CIII by MALDI-TOF MS were based on the use of a MALDI matrix colder than HCCA, namely SPA[57-59]. The use of SPA allowed to minimize the loss of sialic acid, as evidenced by an increased relative intensity of both the mono- and the disialylated apo-CIII proteoforms and leading to reproducible apo-CIII glycosylation profiles (**Figure 2 and Supplementary Table S4**). Importantly, compared to other matrices previously used for profiling of apo-CIII glycoforms such as DHB[152, 153], SPA provides more desirable matrix/analyte co-crystallization in the context of high-throughput, automated MALDI measurements.

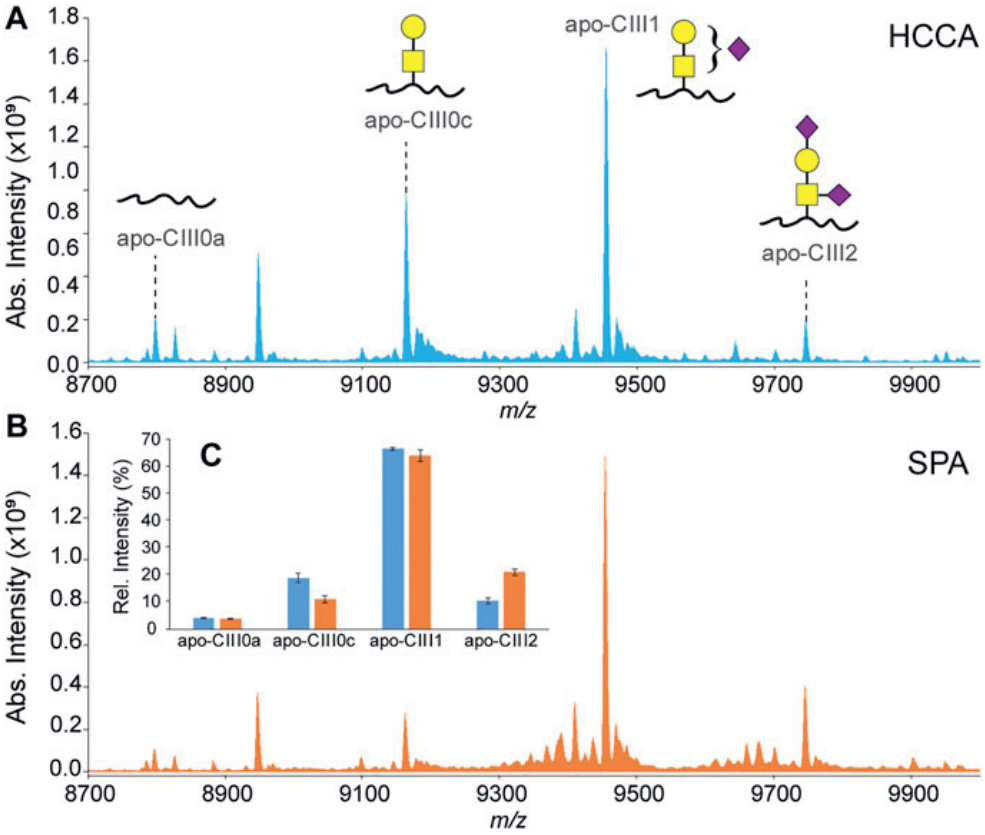

FIGURE 2. MALDI-FT-ICR mass spectra of apo-CIII proteoforms from H_2O_2-treated human blood plasma samples obtained using either HCCA (A) or SPA (B) as MALDI matrices. Relative peak intensities of the four most abundant proteoforms of apo-CIII are shown for both sets (n = 3 in duplicate) of samples in the inset (C).

Table S4. Average relative peak intensities and standard diviation (SD) values obtained from the analysis of three plasma samples in duplicate using HCCA and SPA as a MALDI matrix.

		apo-CIII$_{0a}$	**apo-CIII$_{0c}$**	**apo-CIII$_1$**	**apo-CIII$_2$**
HCCA	Relative Intensity	4%	19%	67%	10%
	SD	0%	2%	1%	1%
SPA	Relative Intensity	4%	11%	64%	21%
	SD	0%	1%	2%	1%

Implementation of MassyTools software for high-throughput MS data processing

One of the advantages of using ultrahigh-resolution MS is that measurements at isotopic resolution provide more spectra information compared to broad-peak detection in linear mode MALDI-TOF MS. Previously, we showed that the goodness of the observed isotopic distributions can be used as a quality control parameter for the selection of high-quality spectra generated from the analysis of a large cohort of samples[154]. This concept was then implemented in a more powerful software – namely, MassyTools – developed for the high-throughput processing of MALDI mass spectra[140]. MassyTools allows the determination of a series of quality control parameters that can be used to perform a curation of MS data at different levels. Mass spectra with unacceptable internal calibration quality and low intensity were discarded at first. Then, the quality of the signal of each apo-CIII proteoform was assessed using the S/N and MME values determined for the most intense peak within an isotopic distribution. Additionally, the quality of such distribution (i.e. IPQ value) was taken into account. The distributions of values of these parameters over 136 standard and 771 clinical plasma samples are reported in **Supplementary Figures S3 and S4**, and **Table S5**. The analytes passing the curation process were then used for statistical analysis.

Table S5. Quality parameters of four major apo-CIII proteoforms: signal to noise (S/N), isotopic pattern quality (IPQ) and mass measurement error (MME; in part-per-million (ppm)). The average values and standard deviations (SD) were calculated based on 136 VisuCon plasma standards and 771 controls used in the cohort study. H_2O_2 treatment provided good quality spectra with analyte peaks meeting minimum two of the set quality criteria: PPM ≤ 10, IPQ ≤ 0.2 and S/N ≥ 9.

			Apo-$CIII_{0a}$+2Ox *m/z* = 8797.228	**Apo-$CIII_{0c}$+2Ox *m/z* = 9162.361**	**Apo-$CIII_1$+2Ox *m/z* = 9453.456**	**Apo-$CIII_2$+2Ox *m/z* = 9744.552**
Plasma standards (n = 136)	**S/N**	**Mean**	**38.37**	**70.29**	**127.48**	**62.42**
		SD	17.80	21.45	27.69	17.11
	IPQ	**Mean**	**0.03**	**0.02**	**0.01**	**0.02**
		SD	0.02	0.01	0.01	0.01
	MME (ppm)	**Mean**	**0.65**	**1.31**	**1.41**	**0.48**
		SD	1.66	0.37	0.39	0.13
Controls (n = 771)	**S/N**	**Mean**	**22.03**	**53.59**	**84.45**	**44.86**
		SD	15.18	17.96	17.13	10.84
	IPQ	**Mean**	**0.09**	**0.04**	**0.02**	**0.03**
		SD	0.09	0.03	0.01	0.03
	MME (ppm)	**Mean**	**1.64**	**0.98**	**1.04**	**0.37**
		SD	4.36	0.6	0.5	0.46

As assessed on 136 standard plasma samples, which were distributed over 17 MALDI target plates measured over 28 days, the method provided good repeatability for relative quantitation of all four proteoforms with CVs in a range of 1% to 18% for average intra-plate and 6% to 16% for inter-plate variability (**Table 1**). While we reduced in-source decay by selecting SPA as a MALDI matrix, we expect that partial sialic acid loss from apo-$CIII_1$ during MS analysis may lead to a slight, artificial increase in the apo-$CIII_{0c}$ glycoform abundance. Hence, fluctuations in the extent of sialic acid loss may contribute to the larger CVs for apo-$CIII_{0c}$.

Table 1. Inter- and intra-plate variability. Relative peak intensities, standard deviation (SD) and coefficient of variation (CV) are given for the intra- and inter-plate variability based on 136 plasma standards.

		Apo-$CIII_{0a}$ +2Ox	**Apo-$CIII_{0c}$ +2Ox**	**Apo-$CIII_1$ +2Ox**	**Apo-$CIII_2$ +2Ox**
Inter-plate	Relative peak intensity	0.039	0.114	0.636	0.203
	SD	0.003	0.018	0.04	0.02
	CV	8%	16%	6%	10%
Average intra-plate	CV	7%	18%	1%	10%

Associations between apo-CIII sialylation and lipid markers

We used this approach to determine non-glycosylated and the glycosylated non-sialylated, mono-sialylated and disialylated apo-CIII glycoforms within a cohort of 746 individuals without diabetes (cohort characteristics in **Table 2**) and test their association with a range of metabolic biomarkers. We found the association of disialylated apo-$CIII_2$ with overall improved lipid profiles and decreased BMI (**Table 3 and Supplementary Table S6**), which is in accordance with some of the previous reports[57, 59]. A subgroup analysis in participants not using statins or fibrates, did not change these associations (**Supplementary Table S7A and 7B**).

Table 2. DiaGene cohort characteristics.

	Individuals without diabetes
Participants, n	746
Male sex, n (%)	290 (38.9)
Age, year	65.7 (6.7)
Age within males, year	66.0 (6.7)
Age within females, year	65.5 (6.7)
BMI, kg/m2	25.4 (4.5)
HDL-cholesterol, mmol/l	1.48 (0.36)
non-HDL-cholesterol, mmol/l	4.09 (0.97)
LDL-cholesterol, mmol/l	3.56 (0.90)
Triglycerides, mmol/l	1.20 (0.68)
Total cholesterol, mmol/l	5.57 (0.99)
Use of lipid lowering therapy, n (%)	93 (12.5)

Mean (and standard deviation) for normal distribution and median (and interquartile range) for non-normal distributions (BMI and triglycerides).

Table S6. Associations of apo-CIII sialylation with clinical characteristics. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

Characteristics	Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value
Sex (mean ± SD)	0.114 ± 0.026	0.116 ± 0.025	0.531	0.667 ± 0.028	0.674 ± 0.031	**0.001**	0.218 ± 0.036	0.209 ± 0.039	**0.003**
	beta		p-value	beta		p-value	beta		p-value
Age	-9.22E-05		5.05E-01	-2.96E-04		7.06E-02	4.06E-04		5.11E-02
BMI	8.34E-04		**8.15E-04**	1.55E-03		**1.16E-07**	-2.42E-03		**6.13E-11**
HDL cholesterol	-1.09E-03		6.80E-01	-6.35E-03		**4.21E-02**	7.66E-03		5.30E-02
non-HDL cholesterol	1.39E-03		1.64E-01	7.60E-03		**6.54E-11**	-9.00E-03		**1.11E-09**
LDL cholesterol	1.66E-03		1.20E-01	7.55E-03		**1.59E-09**	-9.23E-03		**6.45E-09**
Triglycerides	1.92E-03		1.01E-01	1.50E-02		**1.99E-20**	-1.69E-02		**8.31E-17**
Total cholesterol	1.18E-03		2.26E-01	6.42E-03		**2.01E-08**	-7.59E-03		**1.75E-07**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

Table S7a. Subgroup analysis in non-diabetic controls, not using statins or fibrates. The sum of the glycoforms Apo-CIII0a, Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0a}$		Apo-$CIII_{0c}$		Apo-$CIII_1$		Apo-$CIII_2$	
	beta	p-value	beta	p-value	beta	p-value	beta	p-value
Age	2.37E-05	8.44E-01	-1.18E-04	3.84E-01	-2.63E-04	1.02E-01	3.60E-04	1.04E-01
BMI	-5.76E-04	**8.24E-03**	6.48E-04	**8.19E-03**	1.82E-03	**1.86E-10**	-1.91E-03	**1.49E-06**
HDL cholesterol	1.46E-03	5.26E-01	5.45E-05	9.83E-01	-9.24E-03	**2.20E-03**	7.67E-03	6.56E-02
non-HDL cholesterol	9.09E-04	3.19E-01	1.86E-03	6.75E-02	6.23E-03	**1.55E-07**	-8.85E-03	**6.12E-08**
LDL cholesterol	1.08E-03	2.72E-01	2.51E-03	**2.25E-02**	6.14E-03	**1.80E-06**	-9.56E-03	**6.63E-08**
Triglycerides	2.12E-03	**4.65E-02**	7.37E-04	4.89E-01	1.19E-02	**7.20E-14**	-1.46E-02	**4.68E-12**
Total cholesterol	1.13E-03	2.14E-01	1.86E-03	6.72E-02	4.74E-03	**6.65E-05**	-7.59E-03	**3.40E-06**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

Table S7b. Subgroup analysis in non-diabetic controls, not using statins or fibrates. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0c}$		Apo-$CIII_1$		Apo-$CIII_2$	
	beta	p-value	beta	p-value	Beta	p-value
Age	-1.20E-04	4.07E-01	-2.49E-04	1.51E-01	3.81E-04	8.52E-02
BMI	5.97E-04	**2.25E-02**	1.47E-03	**1.96E-06**	-2.13E-03	**7.33E-08**
HDL cholesterol	1.96E-04	9.43E-01	-8.52E-03	**8.87E-03**	8.50E-03	**4.10E-02**
non-HDL cholesterol	1.99E-03	6.74E-02	7.04E-03	**3.80E-08**	-9.05E-03	**2.96E-08**
LDL cholesterol	2.69E-03	**2.19E-02**	6.99E-03	**4.41E-07**	-9.73E-03	**3.67E-08**
Triglycerides	9.42E-04	4.00E-01	1.41E-02	**4.21E-17**	-1.49E-02	**2.14E-12**
Total cholesterol	2.01E-03	6.43E-02	5.66E-03	**1.01E-05**	-7.66E-03	**2.73E-06**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

So far, the inhibitory effect of apo-CIII on LPL has been linked to total apo-CIII concentration, but not to the relative proportion of apo-CIII glycoforms[150]. Recent studies proposed that the presence of apo-CIII on triglyceride-rich lipoproteins (TRLs) alters the affinity between TRLs and their receptors in the liver. Kegulian *et al.* demonstrated that the degree of apo-CIII sialylation directs TRLs to different hepatic clearance pathways, as shown in mice[59]. In detail, apo-$CIII_1$-enriched very low-density lipoproteins (VLDLs) are preferentially cleared by faster-acting low-density lipoprotein (LDL) receptor (LDLR) and LDL receptor-related protein 1 (LRP1), whereas apo-$CIII_2$

directs VLDLs to syndecan 1 (SDC1) receptors that are characterised by a slower but larger capacity metabolism of TRLs. The same study also showed that a 13-week antisense oligonucleotide treatment for apo-CIII, which, as expected, reduced plasma TG levels, also altered relative abundances of these two glycoforms leading to an increase of apo-$CIII_2$ and a decrease of apo-$CIII_1$. The increase of the apo-$CIII_2$/apo-$CIII_1$ ratio in a response to the antisense oligonucleotide therapy was explained by a differing capacity and clearance speed of the hepatic TRL receptors. In support of this, we observed in our cohort study of individuals without diabetes a positive association of the relative abundance of apo-$CIII_1$ glycoform with TG levels, and a negative association for apo-$CIII_2$ (**Table 3 and Figure 3**).

Table 3. Associations of apo-CIII glycosylation with clinical characteristics.

	Apo-CIII$_{0a}$			Apo-CIII$_{0c}$			Apo-CIII$_1$			Apo-CIII$_2$		
Seks	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value
(mean ± SD)	0.05 ± 0.02	0.05 ± 0.02	0.605	0.11 ± 0.02	0.11 ± 0.02	0.558	0.63 ± 0.03	0.64 ± 0.03	**0.002**	0.21 ± 0.04	0.20 ± 0.04	**0.004**
	beta	p-value		beta	p-value		beta	p-value		beta	p-value	
Age	1.28E-05	9.10E-01		-8.84E-05	4.96E-01		-2.97E-04	5.12E-02		3.84E-04	6.47E-02	
BMI	-4.61E-04	**2.23E-02**		8.60E-04	**2.31E-04**		1.81E-03	**1.74E-11**		-2.21E-03	**2.41E-09**	
HDL cholesterol	1.52E-03	4.80E-01		-1.20E-03	6.29E-01		-7.23E-03	**1.24E-02**		6.87E-03	8.27E-02	
non-HDL cholesterol	1.36E-03	9.56E-02		1.21E-03	1.95E-01		6.53E-03	**1.46E-09**		-8.88E-03	**1.81E-09**	
LDL cholesterol	1.55E-03	7.55E-02		1.45E-03	1.49E-01		6.41E-03	**3.32E-08**		-9.15E-03	**8.51E-09**	
Triglycerides	2.61E-03	**7.45E-03**		1.60E-03	1.53E-01		1.25E-02	**9.02E-16**		-1.66E-02	**1.44E-16**	
Total cholesterol	1.51E-03	5.82E-02		9.98E-04	2.75E-01		5.28E-03	**6.77E-07**		-7.58E-03	**1.76E-07**	

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

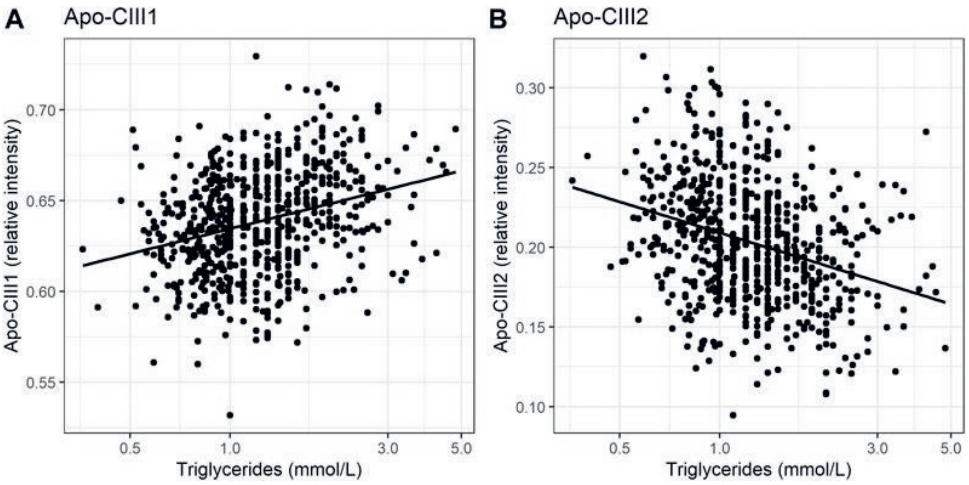

Figure 3. Relationship of apo-CIII1 (A) and apo-CIII2 (B) with triglycerides.

Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced, by apo-CIII glycosylation, specifically sialylation. However, other aspects have to be considered. For instance, defective LDLR/LRP1-driven metabolic pathways might lead to decreased clearance of TGs. Expression and stability of LDLR and LRP1 in the liver might be affected by naturally occurring genetic

variants[132, 155-157]. Moreover, it has been shown in mice that a high-fat diet can lead to the down-regulated expression of hepatic LRP1 by causing hyperglycemia with a high level of plasma triglycerides[158]. In humans, obesity is associated with increases in plasma triglycerides[159]. We hypothesised that the association of apo-CIII glycoforms with triglycerides could be confounded by BMI. Surprisingly, after adjustment for BMI, the direction of effect and goodness-of-fit did not evidently change (**Supplementary Table S8a and 8b**). This indicates that the association of apo-CIII sialylation with triglycerides is largely independent of BMI, and that it is not obesity-associated physiological changes that determine apo-CIII sialylation and its association with triglycerides.

Expression levels of apo-CIII were not investigated in this study. The differences in apo-CIII glycosylation profiles observed between individuals may be caused by varying expression levels of apo-CIII[150] or apo-CIII glycoforms[135], or the accumulation of certain glycoforms due to dysfunctional clearance pathways, based on recent findings by Kegulian et al.[59]. It may also be a combination of the listed factors, which should be explored in further research. Nevertheless, from the results of this study, we cannot determine whether apo-CIII sialylation influences triglyceride levels, or vice versa. Further studies are needed to elucidate the genetic and environmental factors that determine apo-CIII sialylation in health and disease.

Table S8a. Associations of apo-CIII glycoforms with triglycerides, with and without adjustment for BMI. The sum of the glycoforms Apo-CIII0a, Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0a}$			Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2
Triglycerides	2.61E-03	**7.45E-03**	0.01	1.60E-03	1.53E-01	0.00	1.25E-02	**9.02E-16**	0.09	-1.66E-02	**1.44E-16**	0.09
Triglycerides adjusted for BMI	3.69E-03	**3.64E-04**	0.02	1.66E-05	7.96E-01	0.02	1.05E-02	**3.70E-11**	0.11	-1.43E-02	**2.72E-12**	0.09

Blue: negative associations, red: positive associations, bold: significant p-value. P-value and R-square (R2) are calculated for the logarithmically transformed values of triglycerides. 'Triglycerides adjusted for BMI' represent a multivariate linear model, BMI and triglycerides as independent variables.

Table S8b. Associations of apo-CIII glycoforms with triglycerides, with and without adjustment for BMI. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2
Triglycerides	1.92E-03	1.01E-01	0.00	1.50E-02	**1.99E-20**	0.12	-1.69E-02	**8.31E-17**	0.10
Triglycerides adjusted for BMI	4.12E-04	5.54E-01	0.01	1.38E-02	**1.10E-16**	0.12	-1.42E-02	**4.78E-12**	0.11

Blue: negative associations, red: positive associations, bold: significant p-value. P-value and R-square (R2) are calculated for the logarithmically transformed values of triglycerides. 'Triglycerides adjusted for BMI' represent a multivariate linear model, BMI and triglycerides as independent variables.

CONCLUSIONS

Apo-CIII is a novel potential drug target in the management of cardiovascular disease driven by multiple studies demonstrating that plasma levels of apo-CIII are predictive of coronary heart disease and the risk of disease-related events[160]. Previously, it has been shown that sialylated apo-CIII glycoforms are differentially cleared by hepatic receptors and that a higher apo-$CIII_2$/apo-$CIII_1$ ratio is associated with improved triglyceride levels[59]. In humans, the production rates of these two glycoforms are comparable[161], therefore varying apo-CIII2/apo-CIII1 ratios between individuals in healthy and disease groups might suggest various dysfunctional mechanisms involved in their production and clearance. This is the first large-scale study of apo-CIII glycosylation. Clinical cohort studies employing large numbers of individuals will provide more insight into this topic, and the development of highly robust and accurate analytical methods enabling such large-scale studies is warranted.

Here, we present a workflow for high-throughput MALDI-FT-ICR MS analysis of apo-CIII glycosylation in human plasma samples varying in protein oxidation levels. The controlled oxidation of apo-CIII methionine residues, the use of sinapinic acid as a MALDI matrix, and the use of MassyTools software for semi-automated, standardised spectra processing have been implemented to achieve highly repeatable measurements of intact apo-CIII proteoforms. The new analytical workflow allowed us to overcome the problem of the high spectral heterogeneity produced by methionine oxidation thus allowing the robust screening of a large cohort of plasma samples for the relative quantitation of apo-CIII proteoforms. Importantly, the evaluation of MS spectra-derived quality parameters was implemented to minimize biases and ensure accuracy of collected data.

The cohort analysis confirmed that the level of apo-CIII sialylation is strongly associated with lipid markers, especially with triglyceride levels. The relation between relative abundances of apo-CIII glycoforms and cardiovascular disease development should be further explored. More insight into the role of apo-CIII glycosylation in disease

pathophysiology could provide new drug targets. Also, understanding of the mechanisms of existing drugs might increase by considering apo-CIII glycosylation. The methods presented, will enable such large-scale studies.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095.