Characterization of ^{1}H-NMR Plasma Glycoproteins as a new strategy to identify inflammatory patterns in Rheumatoid Arthritis

Journal:	*Journal of Proteome Research*
Manuscript ID	pr-2018-00411t.R1
Manuscript Type:	Article
Date Submitted by the Author:	26-Sep-2018
Complete List of Authors:	Fuertes-Martin, Rocío; Universitat Rovira i Virgili Escola Tecnica Superior Enginyeria, Departament of Electronic Engineering, Edifici E-4, 3 planta despatx 331; Biosfer Teslab S.L.; Institut d'Investigacio Sanitaria Pere Virgili Taverner, Delia; Universitat Rovira i Virgili Facultat de Medicina I Ciences de la Salut; Hospital Universitari Sant Joan de Reus, Unitat de Recerca en Lipids I Ateriosclerosi; CIBER de Diabetes y Enfermedades Metabolicas (CIBERDEM); Institut d'Investigacio Sanitaria Pere Virgili Vallvé, Joan-Carles; Universitat Rovira i Virgili Facultat de Medicina I Ciences de la Salut; Hospital Universitari Sant Joan de Reus, Unitat de Recerca en Lipids I Arteriosclerosi; CIBER de Diabetes y Enfermedades Metabolicas (CIBERDEM); Institut d'Investigacio Sanitaria Pere Virgili Paredes, Silvia; Universitat Rovira i Virgili Facultat de Medicina I Ciences de la Salut; Hospital Universitari Sant Joan de Reus, Unitat de Recerca en Lipids I Arteriosclerosi; CIBER de Diabetes y Enfermedades Metabolicas (CIBERDEM); Institut d'Investigacio Sanitaria Pere Virgili Masana, Lluís; Hospital Universitari Sant Joan de Reus, Unitat de Recerca en Lipids I Arteriosclerosi; Universitat Rovira i Virgili Facultat de Medicina I Ciences de la Salut; CIBER de Diabetes y Enfermedades Metabólicas (CIBERDEM); Institut d'Investigacio Sanitaria Pere Virgili Correig Blanchar, Xavier; Universitat Rovira i Virgili Escola Tecnica Superior Enginyeria, Department of Electronic Engineering, Edifici E-4, 3 planta despatx 331; Institut d'Investigacio Sanitaria Pere Virgili; CIBER de Diabetes y Enfermedades Metabolicas (CIBERDEM) Amigó Grau, Núria; Biosfer Teslab S.L.; CIBER de Diabetes y Enfermedades Metabólicas (CIBERDEM); Institut d'Investigacio Sanitaria Pere Virgili

Characterization of ^{1}H-NMR Plasma Glycoproteins as a new strategy to identify inflammatory patterns in Rheumatoid Arthritis

Rocío Fuertes-Martín,[†] *Dèlia Taverner,*[‡] *Joan-Carles Vallvé,*[*‡] *Silvia Paredes,*[‡] *Lluis Masana,*[‡] *Xavier Correig Blanchar,*[§] *and Núria Amigó Grau*[†]

[†]Biosfer Teslab SL; DEEEA, Metabolomics platform, Universitat Rovira i Virgili, IISPV; CIBERDEM

[‡] Unitat de Recerca en lipids i arteriosclerosi, Hospital Universitari Sant Joan de Reus; IISPV; CIBERDEM

[§]Metabolomics platform, Universitat Rovira i Virgili, IISPV; CIBERDEM

Corresponding Author

*E-mail: jc.vallve@urv.cat. Phone: 977759367. FAX: 977759322

ABSTRACT

Rheumatoid Arthritis (RA) is a chronic autoimmune inflammatory disease associated with a high index of morbidity and mortality from cardiovascular diseases (CVDs). In this study we used ^{1}H-NMR to characterize the plasma glycoprotein and lipoprotein profiles of a cohort of patients with RA (n=210) versus healthy individuals (n=203) to associate them with the RA disease and its severity. Using ^{1}H-NMR, we developed a lineshape method to characterize the two peaks associated with glycoproteins (GlycA and GlycB) and its derived variables: areas of GlycB (Area GlycB) and GlycA (Area GlycA), shape factors of these two peaks (H/W=Height/Width) and the distance between them (Distance GlycB-GlycA). It was also used the advanced lipoprotein test Liposcale® (CE) to characterize the lipoprotein subclasses. The standard lipid panel and traditional inflammatory markers such as C-reactive protein (CRP), the erythrocyte sedimentation rate (ESR), fibrinogen, the rheumatoid factor (RF), anti-citrullinated peptide antibodies (ACPA) and the DAS28 index has been also determined. RA patients presented a significant 10.65 % increase in the GlycA associated area compared to the control group ($p=2.21x10^{-10}$). They also presented significantly higher H/W GlycA and GlycB ratios than the control population (H/W GlycB $p=7.88x10^{-08}$; H/W GlycA $p=5.61x10^{-08}$). The prediction model that uses the traditional inflammatory variables and the ^{1}H-NMR-derived parameters presented an AUC that was almost 10% higher than the model that only uses the traditional inflammatory variables (from 0.7 to 0.79 AUC). In this study, we have demonstrated that GlycA and GlycB variables derived from ^{1}H-NMR, along with classic inflammatory parameters helps to improve the classification of individuals with high RA disease activity.

KEYWORDS: Glycoprotein, inflammation, Rheumatoid arthritis (RA), Proton Nuclear Magnetic Resonance (^{1}H-NMR).

INTRODUCTION

Rheumatoid arthritis (RA) is a systemic disease characterized by chronic inflammatory involvement of the synovial membrane of the joints and autoantibody production (rheumatoid factor [RF] and anti–citrullinated protein antibody [ACPA]).[1,2] This autoimmune process destroys cartilage and joint bones as well as ligaments and tendons. The exact cause of RA is not known but it is considered a multifactorial disease, resulting from the interaction of risk factors such as genetic susceptibility, sex, age, smoking, and infectious, hormonal, dietary, socioeconomic and ethnic agents. Patients with RA may have any of the traditional risk factors for cardiovascular disease (elevated levels of low-density lipoprotein [LDL] cholesterol or high-density lipoprotein [HDL] dysregulations). However, evidence suggests that the prevalence of these risk factors is no higher than in the general population.[3] Studies on the incidence and prevalence of RA in recent decades estimate a prevalence between 0.5 and 1% depending on the population studied, with an annual incidence of 0.02-0.05%.[4] Given that the course of the disease is chronic and progressive in most RA patients, there is a high socioeconomic cost as well as a significant impact on the physical function, productivity, quality and life expectancy of the patients.

Comorbidities may precede or accompany RA and there is evidence to suggest that the systemic inflammation and immune dysfunction that characterize RA play a major role in its development and acceleration.[5] Of these comorbidities, CVD is the leading cause of death in patients with RA. It is well established in the literature that the rate of CVD is higher in RA patients than in the

general population.[6,7] The chronic inflammatory state of RA contributes to the onset and development of accelerated atherosclerosis, since the inflammatory process in synovial and atherosclerotic plaques is very similar.[8] In addition, there is an increase in the synthesis of proinflammatory cytokines such as TNF-alpha and IL-6, which facilitates endothelial dysfunction as an initial step in atherosclerosis, leading to the formation and rupture of atherosclerotic plaques.[9]

The classical way of measuring the severity of inflammation through acute-phase reactants or serum or plasma proteins, mainly glycoproteins, was first described many years ago.[10] Serial measurements of the erythrocyte sedimentation rate (ESR), which is largely a measure of fibrinogen, have been used to monitor the progress of inflammatory disorders. Moreover, C reactive protein (CRP) reflects the clinical course of the disease over periods up to three years and is objective, simple and easy to record. However, the acute-phase protein response is not specific to rheumatoid arthritis. To measure the disease activity of patients with RA the DAS28 index is used. It is a composite score of four measures: the number of tender joints (out of 28), the number of swollen joints (out of 28), ESR, and the general health of the patients.[11]

Several studies have shown that glycoproteins play a key role in inflammatory and pathological processes.[12–15] Protein glycosylation is a post-translational process responsible for the attachment of glycan chains to the nitrogen of an asparagine residue (N-linkage) or to the oxygen of a serine or threonine residue (O-linkage) by a covalent bond.[16,17] During inflammation, there are changes in the number of antennary branches, increased sialylation and fucosylation, and decreased galactosylation.[18] The branches are rich in N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), N-acetylneuraminic acid (or sialic acid) and fucose residues.[16,18–20] The plasma levels of circulating glycoproteins rise (positive acute phase proteins) or fall (negative acute phase proteins) during the acute phase response to the presence of some

environmental stimuli in the organism.[21] The role of glycosylation in the structural changes of proteins means that they can be used as potential early biomarkers of disease.

Currently, the serum concentrations of glycoproteins such as antibodies, antiproteases, and binding or transport proteins are determined using a variety of methods: for example, enzyme-linked immunosorbent assays (ELISAs), electrochemiluminescence immunoassay (ECLIA), luminex-based assays, radioimmunoassays (RIA) and nephelometric assays. These methods quantify the amount of protein in biological samples.[22] Despite this, to date there is no high-performance, fast, or sufficiently sensitive technology that can measure the glycan portion of inflammatory proteins and quantify a measure of general glycosylation that can be useful for diagnosis, and the assessment of disease severity and treatment efficacy.

Given the imperfections of conventional biomarkers for the diagnosis, prognosis, risk prediction and disease prevention at the individual patient level, there is an ongoing effort using novel high-precision laboratory techniques to discover new biomarkers that could increase the sensitivity and specificity of the current clinical results. The increasing role of proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy in recent years has enabled compounds such as metabolites, lipoproteins and glycoproteins to be detected.[23] As many of the ^{1}H-NMR peaks are overlapped, the total-line-shape (CTLS) fitting strategy has been used to quantify the individual signals by modelling theoretical functions.[24]

It has been reported elsewhere that ^{1}H-NMR spectroscopy can quickly and accurately detect circulating levels of glycoproteins.[21,25] Thanks to the initial work by Bell et al.,[26] the quantification of glycosylated proteins is now technologically viable using the signals associated with the side-

chain protons of the N-acetyl-carbohydrate groups, which are covalently bound to plasma glycoproteins. [25]

Recently, a pro-inflammatory glycoprotein biomarker termed *GlycA* has been reported in the literature as a marker of inflammation that can be measured by proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy and is associated with cardiometabolic disease and mortality and other inflammatory diseases.[21,25,27–29] It can also be used to quantify plasmatic glycosylated proteins .[30,31] In particular, this signal is associated with the concentration of specific residues of N-acetylglucosamine and N-acetylgalactosamine in the branched side chains of glycosylated plasma proteins (mainly α1-antitrypsin, α1-antichymotrypsin, haptoglobin and transferrin). It has been shown that the quantification of these signals is a good indicator of the detection, prognosis and therapeutic monitoring of tissue injury marked by systemic inflammatory processes, as well as cardiovascular risk and type 2 diabetes.[25,32] Additionally, it is demonstrated that higher serum levels N-acetyl signal of glycoproteins has been revealed also as a biomarker of metastatic colorectal cancer.[33] Some associations with classical inflammation markers have been described previously as is the relationship of GlycA and CRP.[25]

Considering that there is a need for more reliable inflammatory risk assessment tools, in this study we aim first to use ^{1}H-NMR to characterize the plasma glycoprotein profile of a cohort of patients with RA versus healthy individuals. Our second aim is to model the activity of RA to identify patterns indicating the severity of the disease.

EXPERIMENTAL SECTION

Patients

The study population included 210 individuals (n=76 men and n=134 women) with a mean age of 58.00 (±12) and a body mass index of 27.79 (±6) who had been diagnosed with RA and were attending the rheumatology unit of the HUSJ (Hospital Universitari Sant Joan de Reus, Spain). We excluded patients older than 80 and younger than 20, and patients with acute intercurrent diseases. Another population of 203 apparently healthy individuals (CT) without AR and CVDs matched by sex, age, and body mass index were used as the control group. The study was approved by the ethical clinical research committee of the HUSJ and informed consent was obtained from each patient. Patients were visited between September 2011 and November 2014 and on the same day of the medical visit, blood was collected.

Plasma sample handling and analytical methods

Blood samples were withdrawn from the antecubital vein of each participant at the time of recruitment after a 12-hour overnight fast. EDTA plasma was prepared from venous blood collected into sterile, evacuated tubes (BD, Vacutainer). Plasma was immediately separated by low-speed centrifugation at 4 °C and frozen at -80 °C until biochemical and NMR analysis

Biochemical analysis

For the RA patients, we used traditional biochemical methods to determine the classic inflammatory markers: CRP, ESR, fibrinogen, rheumatoid factor (RF) and anti-citrullinated

peptide antibodies (ACPA). The index of clinical activity was assessed by the DAS28 index. Apolipoprotein A1 and B:100 concentrations and the standard lipid profile including total cholesterol, HDL cholesterol and triglycerides were also determined. LDL cholesterol was calculated using the Fridewald formula.[34]

Glycoprotein profiling

Plasma samples (250 µL) were previously diluted with 31 µl deuterated water and 244 µl of 50 mM phosphate buffer solution (PBS) at pH 7.4 consisting of 30.70 Na_2HPO_4 mM and 19.30 NaH_2PO_4 mM before NMR analysis.

^{1}H-NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz (14.1 T). One-dimensional ^{1}H-NMR pulse experiments were carried out using the nuclear Overhauser effect spectroscopy (NOESY)-presaturation sequence (RD-90° -s_1-90° -s_m-90° ACQ) to suppress the residual water peak. The time s_1 was set to 4 µs, and the time s_m (mixing time) was 100 ms. The 90° pulse length was calibrated for each sample[35] and varied from 12.3 to 20.2 µs. The spectral width was 30 ppm (18,000 Hz) and a total of 64,000 data points and four scans (NS) and a receiver gain value of 14.2 were acquired for each sample in NOESY pulse. The gradient pulse strength was 95% of the maximum strength of 53 G cm^{-1} (0.535 T m^{-1}). The glycoprotein profiling method was performed in the region between 2.15 and 1.90 ppm of chemical shift.[36]

Lipoprotein characterization

The diluted plasma samples used for the glycoprotein profiling were analyzed by using the NMR based Liposcale® test (CE) previously reported.[37] This protocol evaluates lipid concentrations (i.e., triglycerides and cholesterol), size and particle number of three different classes of lipoproteins (VLDL, LDL and HDL), as well as the particle number of nine subclasses (large, medium and small VLDL, LDL and HDL). Briefly, particle concentration and diffusion coefficients were obtained from the measured amplitudes and the attenuation of their spectroscopically distinct lipid methyl group NMR signals. To determine the lipoprotein size, the methyl signal was surface fitted using a previously reported procedure [37] with the numbers of functions so the nine lipoprotein subclasses could be determined. The NMR functions were associated with a given lipoprotein subclass (large, medium and small VLDL, LDL or HDL) according to their associated NMR size. The mean particle diameter for the subclasses (and the estimated ranges) were as follows: large VLDL particles (VLDL-P), 75.2 nm (>60 nm); medium VLDL-P, 52.1 nm (45-60 nm); small VLDL-P, 37.1 nm (35-45 nm); large LDL particles (LDL-P), 22.8 nm (22.5-27 nm); medium LDL-P, 20.5 nm (20-22.5 nm); small LDL-P 18.9 nm (18-20 nm); large HDL particles (HDL-P), 10.1 nm (9-13 nm); medium HDL-P, 8.7 nm (8.2-9 nm); small HDL-P, 7.8 nm (<8.2 nm) in agreement with previous literature.[38–40]

The particle numbers of each lipoprotein subclass were calculated by dividing the lipid volume by the particle volume of a given class. The lipid volumes were determined using common conversion factors to convert the concentration units of the esterified cholesterol (CE) and triglycerides (TG) contained in the lipoprotein core into volume units by using the averaged molecular volumes 1.058 g/ml and 1.021 g/ml for CE and TG and molar mass respectively.[41] Finally, weighted average VLDL, LDL and HDL particle sizes (in nm diameter units) can be calculated from the various subclass concentrations by summing the known diameter of each subclass multiplied by its relative

percentage of subclass particle number as estimated from the intensity of its methyl NMR signal.[42]

Statistical analysis

All statistical analyses were computed in MATLAB, Ver. 7.10.0 using PLS-Toolbox, Ver. 5.2.2 (Eigenvector Research). Firstly, univariate statistical analysis of the glycoprotein variables was conducted to identify differences between the RA patients and the healthy individuals. The Lilliefors test was used to study the normality of each variable. Not all the variables followed a normal distribution, therefore in this study the non-parametric Wilcoxon-Mann-Whitney test was applied. The false discovery rate (FDR) adjustment was applied in all the tests of the present study.[43] A Partial Least Squares-Discriminant Analysis (PLS-DA) with the glycoprotein variables measured by ^{1}H-NMR was conducted to identify the characteristic glycoprotein profiles associated with the AR disease and the healthy group. Additionally, to study the relationship between glycoprotein variables and lipids in the RA population we focused on the associations between the glycoprotein and lipoprotein variables determined by biochemical methods and the Liposcale® test. The associations between the glycoprotein variables and the inflammatory markers CRP, ESR, RF, ACPA and DAS28 were also studied. Spearman correlation coefficients were calculated for variable distributions.

Finally, various auto-scaled and cross-validated PLS-DA models by Venetian blinds cross validation were evaluated to build the best predictor model and identify individuals in the highest 20th percentile of disease activity according to the DAS28 index.[44] The input variables used were the traditional inflammatory markers and ^{1}H-NMR parameters (glycoprotein and lipoprotein).

Because of the high number of variables, the variable selection method Genetic Algorithms (GA)[45] was used to construct more reliable models. Receiver Operating Characteristic (ROC) curves were evaluated to assess the diagnostic ability of these models. Permutation test was applied on all PLS-DA models.

RESULTS

The RA group consisted of 210 individuals (n=76 men and n=134 women) with a mean age of 58 (±12.2) and a body mass index of 27.79 (±6) distributed in 4 groups according to their DAS28 index. The demographic, clinical and biochemical characteristics of patients are shown in table 1.

We deconvoluted mathematically the glycoprotein region as follows: the raw spectra was fitted with five analytical Voigt functions corresponding to the five signals: low molecular weight (LMW), GlycB, GlycA, G-lipid, and baseline based on their chemical shift. For each of these functions the total area (proportional to the concentration), the height, the position (characteristic of the magnetic environment) and the width (related to the flexibility and the aggregation state of the molecules generating the signal) were determined. The baseline function was used to fit the background within the region 1.80 – 1.92 ppm. The function LMW was used to fit the low molecular weight metabolites that resonate in the region 2.12 – 2.14 ppm. The functions GlycB and GlycA were used to fit the signal produced by the -$COCH_3$ acetyl groups of N-acetylglucosamine and N-acetylgalactosamine (GlycA) and N-acetylneuraminic acid (GlycB), which resonates in the region 2.02 – 2.07 [36]; and the G-lipid function was used to fit the signal produced by the CH_2=C protons of lipids in the sample, which resonates in the region 1.95 – 2.01 ppm. Figure 1 focuses on the glycoprotein profiling region of the spectrum.

Table 1. Demographic, clinical and biochemical characteristics of patients and controls.

		RA group				Control group			
		Number	%	Mean	SD	Number	%	Mean	SD
Gender	Women	134	63.8%			131	64.5%		
	Men	76	36.2%			72	35.5%		
Age (years)				58	12			45	16
BMI (Kg/m2)				27.79	5.89			25.30	2.99
Waist circumference (cm)				92	15				
Systolic blood pressure				137	21				
Diastolic blood pressure				81	12				
HDL cholesterol (mg/dL)				65	19				
LDL cholesterol (mg/dL)				118	31				
Triglycerides (mg/dL)				105	56				
Glucose (mg/dL)				95	23				
Current smoker (n, %)	No	154	73.3%						
	Yes	56	26.7%						
Diabetes mellitus (n, %)	No	185	88.1%						
	Yes	25	11.9%						
Hypertension (n, %)	No	85	40.5%						
	Yes	125	59.5%						
Dyslipidemia (n, %)	No	124	59.0%						
	Yes	86	41.0%						
Disease duration (years)				9.3	9.2				
DAS28				3.45	1.27				
	Disease remission (<2.6)	57	27.1%						
	Low disease activity (2.6-3.2)	39	18.6%						
	Moderate disease activity (3.2-5.6)	95	45.2%						
	High disease activity (>5.1)	19	9.0%						
HAQ (0-2.5)				0.44	0.52				
Rheumatoid factor + (n, %)	Negative	58	27.6%						
	Positive	152	72.4%						
ACPA+	Negative	40	19.0%						
	Positive	170	81.0%						
ESR (mm/h)				37	26				
CPR (mg/dL)				0.72	0.84				
Fibrinogen (mg/dL)				442	98				
DMARs	No	51	24.3%						
	Yes	159	75.7%						
Biological agent	No	169	80.5%						
	Yes	41	19.5%						
Corticoids	No	105	50.0%						
	Yes	105	50.0%						
NSAIDs	No	89	42.4%						
	yes	121	57.6%						

HAQ=health assessment questionnaire index, ACPA=citrullinated anti-cyclic peptide antibodies, ESR=erythrocyte sedimentation rate, CRP=C-reactive protein, DAS28=disease activity score, DMARDs=disease-modifying antirheumatic drugs, NSAIDs=non-steroidal antiinflammatory drugs.

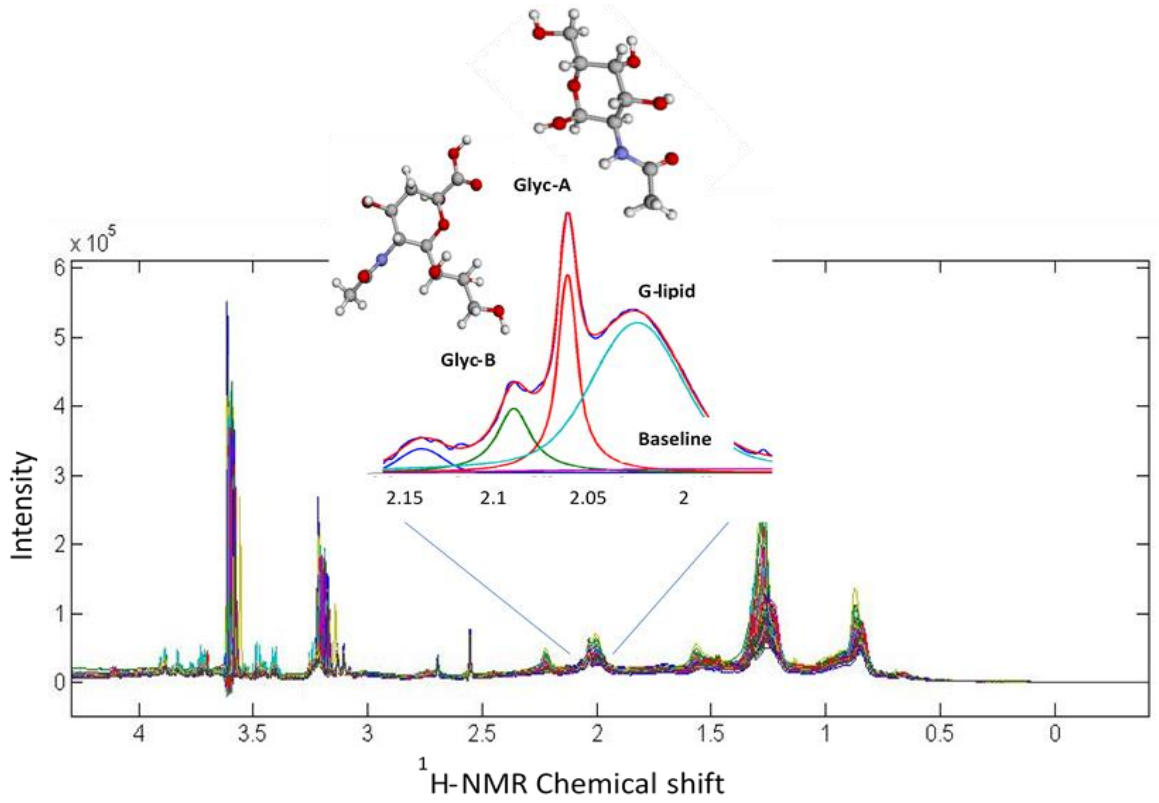


Figure 1. Mathematically treated proton nuclear magnetic resonance spectrum (phase corrected, referenced and scaled) in which the region where the signal produced by glycosilated proteins is indicated. The raw spectra were fitted with five analytic Voigt functions corresponding to five signals LMW (blue), GlycB (green), GlycA (red), G-lipid (cyan), and baseline (purple).

Additionally, for GlycB and GlycA functions we calculated the derived parameters H/W=Height/Width to capture the shape of the peaks and the parameter distance GlycB-GlycA to measure the distance between these two signals.

In this study, due to the high degree of correlation between the variables obtained by the peak deconvolution, we shall only consider the following five variables: areas of GlycB (Area GlycB)

and GlycA (Area GlycA), shape factors of these two peaks (H/W=Height/Width) and the distance between them (Distance GlycB-GlycA). The area of these functions gives us information about the concentration of acetyl groups of N-acetylglucosamine and N-acetylgalactosamine (GlycA) and N-acetylneuraminic acid (GlycB) in plasma. The shape factor ratio H/W provides information on what the function is like in each case depending on its height, which is related to the concentration, and its width, which is related to the flexibility and the aggregation of the molecules generating the signal. The variable Distance GlycB-GlycA indicates the relative distance between the chemical shift of both peaks.

The parameters mentioned above were extracted by a lineshape fitting process from the NOESY spectra using the normalized root mean squared error (NRMSE) in the interval 1.90 – 2.15 ppm. The error in all samples was smaller than 2.93 %.

The results of the univariate analysis showed that RA patients presented a significant 10.65 % increase in the GlycA associated area compared to the CT group ($p=2.21x10^{-10}$). They also presented significantly increase in the H/W GlycA ratio (12.70%) and GlycB ratio (16.76%) than the control population (H/W GlycB $p=7.88x10^{-08}$; H/W GlycA $p=5.61x10^{-08}$). The distance between the two functions in the RA group was significantly greater than in the control group (see Table 2). Area GlycB, Area GlycA and the distance parameters did not follow a normal distribution and, as has been explained in statistical analysis section, the non-parametric Wilcoxon-Mann-Whitney test was used to calculate the significance for all parameters.

Table 2. Results of univariate analysis between RA and healthy individuals (CT)

	AR		CT		
Variable	**Median**	**Iqr**	**Median**	**Iqr**	***p*-value**
Area GlycB (a.u.)	$2.14x10^{-01}$	$8.56x10^{-02}$	$2.22x1^{0-01}$	$7.39x10^{-02}$	$1.89x10^{-01}$
Area GlycA (a.u.)	$5.26x10^{-01}$	$1.12x10^{-01}$	$4.70x10^{-01}$	$1.09x10^{-01}$	$2.21x10^{-10}$*
H/W GlycB	$3.94x10^{-03}$	$1.81x10^{-03}$	$3.27x10^{-03}$	$1.03x10^{-03}$	$7.88x10^{-08}$*
H/W GlycA	$1.83x10^{-02}$	$4.27x10^{-03}$	$1.65x10^{-02}$	$5.46x10^{-03}$	$5.61x10^{-08}$*
Distance GlycB-GlycA (ppm)	$3.30x10^{-02}$	$9.16x10^{-04}$	$3.21x10^{-02}$	$9.17x10^{-04}$	$3.68x10^{-22}$*

Significant values ($p<0.05$) are marked (*).Median and interquartile range (Iqr) are reported. The Wilcoxon-Mann-Whitney test has been used to calculate significance. *P*-values adjusted by FDR.

Figure 2a shows the score plot of a PLS-DA model ($p=5.18x10^{-15}$, Q2=0.128) of three latent variables (LV) built to discriminate RA patients from CT individuals by using the 5 ^{1}H-NMR-derived glycoprotein parameters as the input dataset. Figure 2b shows the loading plot of the same PLS-DA model to illustrate the contribution of each variable to each LV.

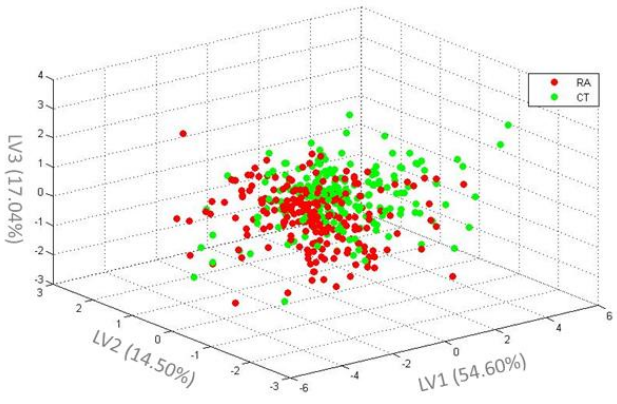


Figure 2a. 3D Score plot of the PLS-DA model (3 latent variables) for CT individuals and RA patients. RA patients are shown in red and control patients in green.

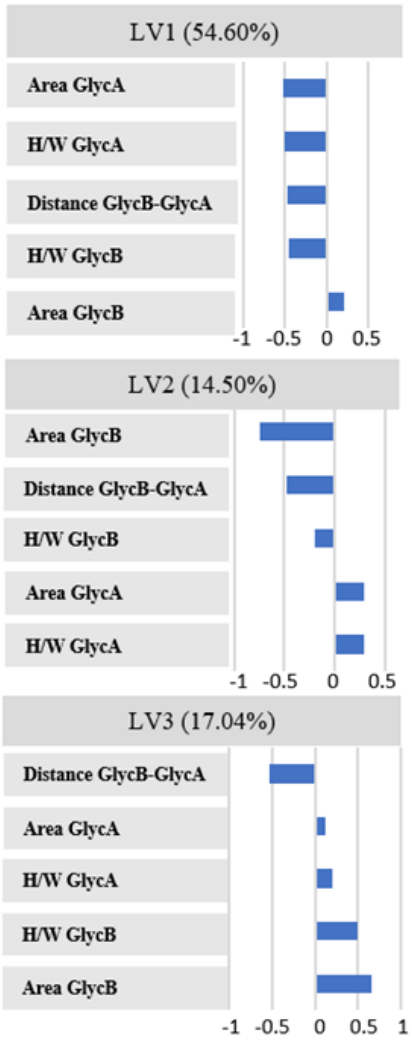


Figure 2b. PLS-DA loading plot showing the NMR glycoprotein-derived parameters contributing to the separation of the PLS-DA scores in RA patients and CT individuals.

The relationship between glycoprotein variables and clinical inflammatory markers was also studied. Table 3 shows the Spearman correlation coefficients for Fibrinogen, DAS28, ESR, CRP, RF, ACPA and Albumin. There was a significant, positive but modest correlation between the five glycoprotein variables and Fibrinogen, ESR, CRP and DAS28. There was also a significant and positive correlation between H/W GlycB and RF and ACPA. Correlations with Albumin are

negative but significant. Figure S-2 of the Supporting Information shows some dispersion plots of these results.

Table 3. Associations between ^{1}H-NMR glycoproteins variables and clinical inflammatory markers.

		Area GlycB	Area GlycA	H/W GlycB	H/W GlycA	Distance GlycB-GlycA
Fibrinogen (mg/dl)	r	0.247*	0.452*	0.477*	0.414*	0.397*
	p	1.08×10^{-03}	1.69×10^{-10}	2.14×10^{-11}	4.65×10^{-09}	2.29×10^{-08}
ESR (mm/h)	r	0.341*	0.424*	0.436*	0.381*	0.302*
	p	2.49×10^{-06}	1.89×10^{-09}	5.61×10^{-10}	9.07×10^{-08}	4.33×10^{-05}
CRP (mg/dl)	r	0.278*	0.456*	0.468*	0.457*	0.356*
	p	1.87×10^{-04}	1.34×10^{-10}	4.49×10^{-11}	1.34×10^{-10}	7.70×10^{-07}
RF	r	0.127	0.109	0.196*	0.068	0.047
	p	1.03×10^{-01}	1.65×10^{-01}	1.19×10^{-02}	4.02×10^{-01}	5.74×10^{-01}
ACPA	r	-0.019	0.090	0.143	0.042	0.019
	p	8.20×10^{-01}	2.56×10^{-01}	6.84×10^{-02}	6.09×10^{-01}	8.20×10^{-01}
DAS28	r	0.214*	0.285*	0.259*	0.229*	0.127
	p	5.35×10^{-03}	1.25×10^{-04}	5.59×10^{-04}	2.83×10^{-03}	1.01×10^{-01}
Albumin (g/dl)	r	-0.144	-0.168*	-0.132	-0.183*	-0.257*
	p	6.79×10^{-02}	3.18×10^{-02}	8.87×10^{-02}	1.86×10^{-02}	6.33×10^{-04}

Spearman correlation coefficients (r) and p-value (*p*) for each glycoprotein variable and the inflammatory markers obtained by traditional biochemistry. *P*-values adjusted by FDR. Significant values ($p<0.05$) are marked (*). ESR: Erythrocyte sedimentation rate; CRP: C-reactive protein; RF: Rheumatoid Factor; ACPA: anti-citrullinated peptide antibodies; DAS28: Disease Activity Score of 28 joints.

We also studied the relationship between the glycoprotein variables obtained by ^{1}H-NMR and the lipids obtained by traditional biochemistry. Table S-1 of the Supporting Information shows the Spearman correlation coefficients and *p*-values for total cholesterol, low-density lipoprotein

cholesterol (LDLc), high-density lipoprotein cholesterol (HDLc), very low-density lipoprotein cholesterol (VLDLc) and total triglycerides (TG). There was a significant but modest correlation between almost all glycoprotein variables and VLDLc, HDLc and TG. In addition, the glycoprotein variables were also compared with the variables of the Liposcale® test (see Table S-2). VLDLc was still the variable most positively and significantly associated with Area GlycA (0.36) and negatively associated with area GlycB (-0.43). Figure S-1 of the Supporting Information shows some scatter plots of the mentioned correlations.

To evaluate the ability of the models based on ^{1}H-NMR glycoprotein and lipoprotein and traditional inflammatory variables to predict the severity of the disease, two PLS-DA models were constructed. The objective of the two models was to evaluate whether the variables of the two models were able to separate individuals in the AR population with a high disease activity based on the DAS28 index. The two models had high RA activity patients (DAS28>80th percentile =4.53) as the y-block input binary variable (prediction block). The first PLS-DA model (Model A) had values of Fibrinogen, CRP, RF and ACPA as the x-block input variables. The second PLS-DA model (Model B) used a GA variable selection method to select the most important of the 5 ^{1}H-NMR-derived glycoprotein variables, 23 variables resulting from the ^{1}H-NMR lipoprotein profile by the Liposcale® test and the 4 variables in Model A as the x-block input variables. The variables selected by the GA were: Area GlycB, CRP, Fibrinogen, HDL-TG, VLDL-P, small VLDL-P, small HDL-P and LDL diameter (LDL-Z). Both models were auto-scaled and cross-validated using Venetian Blinds. The significance level of Model A was p=0.55x10$^{-0.2}$ (Q2=0.089), while model B was p= 1.1369 x 10^{-08} (Q2= 0.189). The discriminant ability of the models was evaluated using the Area Under the Curve (AUC) of the Receiving Operating Characteristic (ROC) analysis which considered high RA activity patients (above 4.53 on the DAS28 index) as positive cases,

and low-moderate RA activity patients (below 4.53) as negative cases. Figure S-3 in the Supporting Information shows the population distribution of DAS28 and a brief description of the population for the two AR groups. The AUC of Model B, which included the traditional variables (Model A) and ^{1}H-NMR-derived parameters, was almost 10% higher than that of Model A (from 0.7 to 0.79 AUC) indicating that Model B was much better at predicting the severity of the disease (see Figure 3).

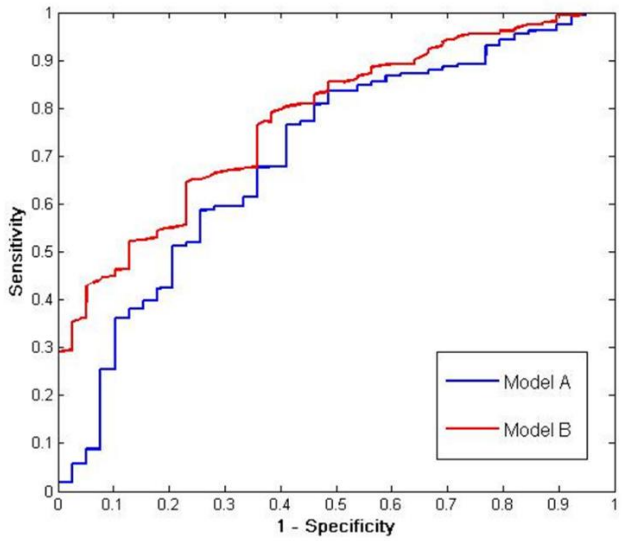


Figure 3. ROC curve of cross-validated Model A (blue) and Model B (red) for high/low AR activity. AUC Model A-value= 0.7017; AUC Model B-value= 0.7939.

DISCUSSION

High-throughput technologies such as ^{1}H-NMR enable circulating plasma or serum glycoproteins to be measured using the specific residues of N-acetylglucosamine, N-acetylgalactosamine and N-acetylneuraminic acid in the branched side chains of glycosylated plasma proteins. In this study we used ^{1}H-NMR to characterize the plasma glycoprotein profile in RA. The results of the

univariate study between RA and a control population confirmed expectations that RA patients have a significantly higher values of ^{1}H-NMR glycoprotein than the control group. These results are in line with the studies by J. Ormseth et al. and B. Barlett et al. in which GlycA concentrations were higher in patients with RA than in controls.[27,46] This may be because RA patients have a permanent inflammatory condition that increases the levels of glycoproteins. In this study, we found that the area of GlycA was significantly higher in RA patients than in controls, which suggests that there was a higher concentration of glycosylated proteins presenting a higher signal associated with acetyl N-acetylgalactosamine and N-acetylglucosamine methyl groups. In addition, the shape factor of GlycA and GlycB indicated that the ratio between the height and width (H/W) of the deconvoluted peaks was significantly higher in patients with RA, which could be explained by the lower glycoprotein aggregation in RA patients' plasma.

In this study we used ^{1}H-NMR to characterize the glycoprotein profile as well as the advanced lipoprotein profile (Liposcale®). This has shown whether the information obtained by ^{1}H-NMR can help from a clinical point of view to evaluate patients with inflammatory processes. Comparable findings were described in the metabolic phenotyping study of Lauridsen et al. [47] where a cohort of patients with RA where measured by ^{1}H-NMR. They report that ^{1}H-NMR plasma metabolic profiles of RA patients (including cholesterol, lactate, acetylated glycoprotein, and lipid signatures) were significantly different from healthy subjects, indicating that the state of inflammation in RA patients is reflected in the ^{1}H-NMR spectra.

Table 3 shows that there were significant positive associations between Fibrinogen, ESR and CRP and the 5 ^{1}H-NMR glycoprotein variables. This is in line with other studies carried out in this field. In 2015, Otvos et al. showed that the concentration of GlycA correlated with high-sensitivity CRP and fibrinogen.[25,30] In this context, the question arises as to whether these glycoproteins are found

in high concentrations when there is an acute inflammatory process, such as rheumatoid arthritis, just as CRP, ESR and Fibrinogen are.[48,49] It should be noted that the H/W GlycA and H/W GlycB ratios were undoubtedly key variables throughout this study. They were a solid indicator of the shape of the glycoprotein peak in inflamed patients, capable of distinguishing RA from healthy individuals and also significantly correlated with lipid and inflammatory variables. Correlations were also found with the DAS28, which were very similar to the correlations with the inflammatory variables mentioned above since the DAS28 includes ESR measurements in its formula. Additionally, the glycoprotein variables are significantly associated in a negative way with albumin. This may be explained by the fact that hypoalbuminemia is characteristic of inflammatory states.[15] Furthermore, hypoalbuminemia in also related with juxta-articular erosions or with the incidence of peptic ulcer in RA.[50]

The representation of the ROC curve comparing the two PLS-DA models (Model A and Model B) indicated that by adding the glycoprotein and lipoprotein ^{1}H-NMR variables to classical inflammation parameters such as Fibrinogen, CRP, RF and ACPA we were able to make a better classification model (Model B) of the severity of the disease according to the 80th percentile of the DAS28 from the RA population.

The results of the correlations between glycoproteins and lipid variables in Table S-1 of the Supporting Information show that GlycA and the H/W ratio of GlycA and GlycB parameters correlated negatively with HDL-cholesterol, which makes sense since HDL-C is related to a cardio-protective function during inflammatory processes that occurs in cardiovascular events.[51–53] Furthermore, with the exception of Area GlycB, all the ^{1}H-NMR glycoprotein variables were significantly and positively correlated with VLDL-cholesterol, which, along with LDL-c, is associated with cardiovascular events. [54] These significant correlations were also noted in table S-

2 when we studied the correlations with the variables of the Liposcale® test. In addition, some studies highlight the importance of considering strategies to reduce VLDL-cholesterol as a therapeutic intervention to reduce the residual risk of atherosclerotic cardiovascular disease,[55,56] beyond the well stablished strategy of reducing LDL-c.

In recent years, the advances in knowledge at genetic level (through the study of epigenetic modifications) and at proteomic level (at the level of the study of post-translational modifications) are contributing to a better characterization of the functional diversity of genes and proteins in the pathological states.[57] As we mentioned earlier in this study, glycosylation represents one of the most important and complex of these post-translational modifications. Studying the glycosylation of proteins could allow a more precise knowledge of the mechanisms of the disease and how is the implication of the proteins in the pathogenesis. The study of glycans has been called *glycomics* and it has received increasing interest as a novel tool for identifying markers and potential mediators of disease pathogenesis.[31] As we mentioned earlier, several authors have demonstrated that a serum glycan signal is associated with type 2 diabetes,[58] atherosclerotic CVD,[25,27,28,59] as well as longitudinal CVD and cancer mortality risk[33,60] highlighting the importance of inflammation as a shared risk pathway. It should be noted that some authors have focused their study in a specific glycoprotein; for example, Menni et al.[61] studied the glycomic profiling of IgG to identify CVD risk; and Väänänen et al.[62] studied YKL-40, a chitinase-like glycoprotein associated with inflammation and tissue remodeling as a potential biomarker of disease activity in patients with early RA. The cited study by Lauridsen et al.[47] highlights that the metabolic profile may provide additional information to the clinical evaluation. In the present study we have evaluated the information provided by glycoprotein variables analyzed by ^{1}H-NMR as non-

specific markers of inflammation as part of the current trend to seek new inflammatory markers that improve the limitations of classical inflammation biomarkers.[63–65]

This study confirms that the composite ^{1}H-NMR signal is associated with some clinical parameters of inflammation and the results reported are in line with the studies mentioned above in which the GlycA NMR-base assay has been used. Nevertheless, various points need to be borne in mind about this technique. On the one hand, ^{1}H-NMR is not a selective technique and does not enable individual proteins to be identified and quantified. The areas of GlycB and GlycA that we have obtained correspond to the overall concentration of glycosylated proteins in the plasma. On the other hand, one of its important advantages is that the ^{1}H-NMR spectrum enables a lipoprotein profile (Liposcale®) and a glycoprotein profile to be identified in a single experiment. This is a great advantage because the information is obtained with minimal manipulation of the samples and involves a considerable saving in experimental time. Finally, although AR patients are being treated, this does not prevent the glycoprotein signals detected by ^{1}H-NMR being sensitive to the level of inflammation, as we have seen in the results. Bearing this in mind, in the future this first exploratory study should carefully study if medication can modify this ^{1}H-NMR spectrum region.

CONCLUSIONS

In this study, we have demonstrated that new markers of inflammation can be characterized using the ^{1}H-NMR signal of glycosylated proteins. In addition, the glycoprotein and lipoprotein ^{1}H-NMR variables, along with classic inflammatory parameters, provide information about the high activity of the disease that is more accurate than if we only consider the classic inflammatory parameters, as is conventional in clinical practice.

ASSOCIATED CONTENT

Supporting Information (SI)

The following supporting information is available free of charge at ACS Publications website http://pubs.acs.org.

Table S-1. Spearman and Pearson correlation coefficients for each glycoprotein variable detected by ^{1}H-NMR according to its relation to the lipids obtained by traditional biochemistry (total cholesterol, high density lipoprotein cholesterol, low density lipoprotein cholesterol, very low - density lipoprotein and total triglycerides).

Figure S-1. Scatter plots of Lipids variables obtained by traditional biochemistry and glycoprotein variables.

Table S-2. Spearman correlation coefficients for each glycoprotein variable detected by ^{1}H-NMR and the variables resulting from the Liposcale® test.

Figure S-2. Scatter plots of 1H-NMR glycoproteins variables and clinical inflammatory markers.

Figure S-3. Population distribution of DAS28 and a brief description of the population for the AR groups.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jc.vallve@urv.cat. Phone: 977759367. FAX: 977759322

Conflict of interest disclosure: X. Correig, L. Masana and N. Amigó are stockowners of Biosfer Teslab, and have the patent for the Liposcale® NMR method for Lipoprotein characterization.

ORCID

Rocío Fuertes-Martín: 0000-0002-1558-8535

AKNOWLEDGMENTS

We acknowledge the Generalitat de Catalunya Industrial Ph.D plan (2014DI0029) for support this project.

ABBREVIATIONS

^{1}H-NMR, Proton Nuclear Magnetic Resonance; RA, Rheumatoid Arthritis; CVDs, Cardiovascular Diseases; CRP, C-Reactive Protein; ESR, Erythrocyte sedimentation rate; RF, Rheumatoid Factor; ACPA, Anti-Citrullinated Peptide Antibodies; PLS-DA, Partial Least Squares-Discriminant Analysis; AUC, Area under the curve; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; ELISAs, enzyme-linked immunosorbent assays; ECLIA, electrochemiluminescence immunoassay; RIA, radioimmunoassays; CTLS, Total-line shape fitting; VLDL, Very Low Density Lipoprotein; LDL, Low Density Lipoprotein; HDL, High

Density Lipoprotein; TG, Tryglycerides; GA, Genetic Algorithms; ROC-curve, Receiver Operating Characteristic-curve; SD, Standard Deviation; Iqr, Interquartile range; LV, Latent Variable.

REFERENCES

(1) McInnes, I. B.; Schett, G. The Pathogenesis of Rheumatoid Arthritis. *N. Engl. J. Med.* **2011**, *365* (23), 2205–2219.

(2) van de Stadt, L. A.; van der Horst, A. R.; de Koning, M. H. M. T.; Bos, W. H.; Wolbink, G. J.; van de Stadt, R. J.; Pruijn, G. J. M.; Dijkmans, B. A. C.; van Schaardenburg, D.; Hamann, D. The Extent of the Anti-Citrullinated Protein Antibody Repertoire Is Associated with Arthritis Development in Patients with Seropositive Arthralgia. *Ann. Rheum. Dis.* **2011**, *70* (1), 128–133.

(3) Steiner, G.; Urowitz, M. B. Lipid Profiles in Patients with Rheumatoid Arthritis: Mechanisms and the Impact of Treatment. *Semin. Arthritis Rheum.* **2009**, *38* (5), 372–381.

(4) Alamanos, Y.; Drosos, A. A. Epidemiology of Adult Rheumatoid Arthritis. *Autoimmun. Rev.* **2005**, *4* (3), 130–136.

(5) Gabriel, S. E. Why Do People with Rheumatoid Arthritis Still Die Prematurely? *Ann. Rheum. Dis.* **2008**, *67* (Suppl 3), iii30-iii34.

(6) Gabriel, S. E.; Crowson, C. S.; Kremers, H. M.; Doran, M. F.; Turesson, C.; O'Fallon, W. M.; Matteson, E. L. Survival in Rheumatoid Arthritis: A Population-Based Analysis of

Trends over 40 Years. *Arthritis Rheum.* **2003**, *48* (1), 54–58.

(7) Castañeda, S.; Martín-Martínez, M. A.; González-Juanatey, C.; Llorca, J.; García-Yébenes, M. J.; Pérez-Vicente, S.; Sánchez-Costa, J. T.; Díaz-Gonzalez, F.; González-Gay, M. A. Cardiovascular Morbidity and Associated Risk Factors in Spanish Patients with Chronic Inflammatory Rheumatic Diseases Attending Rheumatology Clinics: Baseline Data of the CARMA Project. *Semin. Arthritis Rheum.* **2015**, *44* (6), 618–626.

(8) Stevens, R. J.; Douglas, K. M. J.; Saratzis, A. N.; Kitas, G. D. Inflammation and Atherosclerosis in Rheumatoid Arthritis. *Expert Rev. Mol. Med.* **2005**, *7* (07), 1–24.

(9) Choy, E. Understanding the Dynamics: Pathways Involved in the Pathogenesis of Rheumatoid Arthritis. *Rheumatology* **2012**, *51* (suppl 5), v3–v11.

(10) Amos, R. S.; Constable, T. J.; Crockson, R. A.; Crockson, A. P.; McConkey, B. Rheumatoid Arthritis: Relation of Serum C-Reactive Protein and Erythrocyte Sedimentation Rates to Radiographic Changes. *BMJ* **1977**, *1* (6055), 195–197.

(11) Prevoo, M. L. L.; Van'T Hof, M. A.; Kuper, H. H.; Van Leeuwen, M. A.; Van De Putte, L. B. A.; Van Riel, P. L. C. M. Modified Disease Activity Scores That Include Twenty-Eight-Joint Counts Development and Validation in a Prospective Longitudinal Study of Patients with Rheumatoid Arthritis. *Arthritis Rheum.* **1995**, *38* (1), 44–48.

(12) Gruys, E.; Toussaint, M. J. M.; Niewold, T. A.; Koopmans, S. J. Acute Phase Reaction and Acute Phase Proteins. *J. Zhejiang Univ. Sci. B* **2005**, *6* (11), 1045–1056.

(13) Lyons, J. J.; Milner, J. D.; Rosenzweig, S. D. Glycans Instructing Immunity: The Emerging

Role of Altered Glycosylation in Clinical Immunology. *Front. Pediatr.* **2015**, *3*, 54.

(14) Marth, J. D.; Grewal, P. K. Mammalian Glycosylation in Immunity. *Nat. Rev. Immunol.* **2008**, *8* (11), 874–887.

(15) Gabay, C.; Kushner, I. Acute-Phase Proteins and Other Systemic Responses to Inflammation. *N. Engl. J. Med.* **1999**, *340* (6), 448–454.

(16) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126* (5), 855–867.

(17) van Kooyk, Y.; Rabinovich, G. a. Protein-Glycan Interactions in the Control of Innate and Adaptive Immune Responses. *Nat. Immunol.* **2008**, *9* (6), 593–601.

(18) Gornik, O.; Lauc, G. Glycosylation of Serum Proteins in Inflammatory Diseases. *Dis. Markers* **2008**, *25* (4–5), 267–278.

(19) McCarthy, C.; Saldova, R.; Wormald, M. R.; Rudd, P. M.; McElvaney, N. G.; Reeves, E. P. The Role and Importance of Glycosylation of Acute Phase Proteins with Focus on Alpha-1 Antitrypsin in Acute and Chronic Inflammatory Conditions. *J. Proteome Res.* **2014**, *13* (7), 3131–3143.

(20) Hart, G. W.; Copeland, R. J. Glycomics Hits the Big Time. *Cell* **2010**, *143* (5), 672–676.

(21) Connelly, M. A.; Gruppen, E. G.; Wolak-Dinsmore, J.; Matyus, S. P.; Riphagen, I. J.; Shalaurova, I.; Bakker, S. J. L.; Otvos, J. D.; Dullaart, R. P. F. GlycA, a Marker of Acute Phase Glycoproteins, and the Risk of Incident Type 2 Diabetes Mellitus: PREVEND Study. *Clin. Chim. Acta* **2016**, *452*, 10–17.

(22) Connelly, M. A.; Gruppen, E. G.; Otvos, J. D.; Dullaart, R. P. F. Inflammatory Glycoproteins in Cardiometabolic Disorders, Autoimmune Diseases and Cancer. *Clin. Chim. Acta* **2016**, *459*, 177–186.

(23) Soininen, P.; Kangas, A. J.; Würtz, P.; Tukiainen, T.; Tynkkynen, T.; Laatikainen, R.; Järvelin, M.-R.; Kähönen, M.; Lehtimäki, T.; Viikari, J.; et al. High-Throughput Serum NMR Metabonomics for Cost-Effective Holistic Studies on Systemic Metabolism. *Analyst* **2009**, *134* (9), 1781.

(24) Soininen, P.; Haarala, J.; Vepsäläinen, J.; Niemitz, M.; Laatikainen, R. Strategies for Organic Impurity Quantification by 1H NMR Spectroscopy: Constrained Total-Line-Shape Fitting. *Anal. Chim. Acta* **2005**, *542* (2), 178–185.

(25) Akinkuolie, A. O.; Buring, J. E.; Ridker, P. M.; Mora, S. A Novel Protein Glycan Biomarker and Future Cardiovascular Disease Events. *J. Am. Heart Assoc.* **2014**, *3* (5), e001221.

(26) Bell, J. D.; Brown, J. C. C.; Nicholson, J. K.; Sadler, P. J. Assignment of Resonances for 'Acute-Phase' Glycoproteins in High Resolution Proton NMR Spectra of Human Blood Plasma. *FEBS Lett.* **1987**, *215* (2), 311–315.

(27) Gruppen, E. G.; Riphagen, I. J.; Connelly, M. A.; Otvos, J. D.; Bakker, S. J. L.; Dullaart, R. P. F. GlycA, a Pro-Inflammatory Glycoprotein Biomarker, and Incident Cardiovascular Disease: Relationship with C-Reactive Protein and Renal Function. *PLoS One* **2015**, *10* (9), e0139057.

(28) Otvos, J. D.; Guyton, J. R.; Connelly, M. A.; Akapame, S.; Bittner, V.; Kopecky, S. L.; Lacy, M.; Marcovina, S. M.; Muhlestein, J. B.; Boden, W. E. Relations of GlycA and

Lipoprotein Particle Subspecies with Cardiovascular Events and Mortality: A Post Hoc Analysis of the AIM-HIGH Trial. *J. Clin. Lipidol.* **2018**, *12* (2), 348–355.e2.

(29) Lawler, P. R.; Akinkuolie, A.; Buring, J.; Ridker, P.; Glynn, R.; Mora, S. A Novel Biomarker of Circulating Glycoproteins and Cardiovascular and All-Cause Mortality among 39,521 Initially Healthy Adults. *J. Am. Coll. Cardiol.* **2015**, *65* (10), A1358.

(30) Otvos, J. D.; Shalaurova, I.; Wolak-Dinsmore, J.; Connelly, M. A.; Mackey, R. H.; Stein, J. H.; Tracy, R. P. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin. Chem.* **2015**, *61:5*, 714–723.

(31) Lawler, P. R.; Mora, S. Glycosylation Signatures of Inflammation Identify Cardiovascular Risk. *Circ. Res.* **2016**, *119* (11), 1154–1156.

(32) Akinkuolie, A. O.; Pradhan, A. D.; Buring, J. E.; Ridker, P. M.; Mora, S. Novel Protein Glycan Side-Chain Biomarker and Risk of Incident Type 2 Diabetes Mellitus. *Arterioscler. Thromb. Vasc. Biol.* **2015**, *35* (6).

(33) Bertini, I.; Cacciatore, S.; Jensen, B. V.; Schou, J. V.; Johansen, J. S.; Kruhoffer, M.; Luchinat, C.; Nielsen, D. L.; Turano, P. Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer. *Cancer Res.* **2012**, *72* (1), 356–364.

(34) Bairaktari, E.; Hatzidimou, K.; Tzallas, C.; Vini, M.; Katsaraki, A.; Tselepis, A.; Elisaf, M.; Tsolas, O. Estimation of LDL Cholesterol Based on the Friedewald Formula and on Apo B Levels. *Clin. Biochem.* **2000**, *33* (7), 549–555.

(35) Wu, P. S. C.; Otting, G. Rapid Pulse Length Determination in High-Resolution NMR. *J. Magn. Reson.* **2005**, *176* (1), 115–119.

(36) Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C. 750 MHz 1H and 1H-13C NMR Spectroscopy of Human Blood Plasma. *Anal. Chem.* **1995**, *67* (5), 793–811.

(37) Mallol, R.; Amigó, N.; Rodríguez, M. A.; Heras, M.; Vinaixa, M.; Plana, N.; Rock, E.; Ribalta, J.; Yanes, O.; Masana, L.; et al. Liposcale: A Novel Advanced Lipoprotein Test Based on 2D Diffusion-Ordered 1H NMR Spectroscopy. *J. Lipid Res.* **2015**, *56* (3), 737–746.

(38) Jeyarajah, E. J.; Cromwell, W. C.; Otvos, J. D. Lipoprotein Particle Analysis by Nuclear Magnetic Resonance Spectroscopy. *Clin. Lab. Med.* **2006**, *26* (4), 847–870.

(39) Rosenson, R. S.; Brewer, H. B.; Chapman, M. J.; Fazio, S.; Hussain, M. M.; Kontush, A.; Krauss, R. M.; Otvos, J. D.; Remaley, A. T.; Schaefer, E. J. HDL Measures, Particle Heterogeneity, Proposed Nomenclature, and Relation to Atherosclerotic Cardiovascular Events. *Clin. Chem.* **2011**, *57* (3), 392–410.

(40) Mallol, R.; Rodríguez, M. A.; Heras, M.; Vinaixa, M.; Plana, N.; Masana, L.; Morris, G. A.; Correig, X. Particle Size Measurement of Lipoprotein Fractions Using Diffusion-Ordered NMR Spectroscopy. *Anal. Bioanal. Chem.* **2012**, *402* (7), 2407–2415.

(41) Schumaker, V. N.; Phillips, M. L.; Chatterton, J. E. Apolipoprotein B and Low-Density Lipoprotein Structure: Implications for Biosynthesis of Triglyceride-Rich Lipoproteins. *Adv. Protein Chem.* **1994**, *45*, 205–248.

(42) Mallol, R.; Rodríguez, M. A.; Heras, M.; Vinaixa, M.; Cañellas, N.; Brezmes, J.; Plana, N.; Masana, L.; Correig, X. Surface Fitting of 2D Diffusion-Edited 1H NMR Spectroscopy Data for the Characterisation of Human Plasma Lipoproteins. *Metabolomics* **2011**, *7* (4), 572–582.

(43) Yekutieli, D.; Benjamini, Y. Resampling-Based False Discovery Rate Controlling Multiple Test Procedures for Correlated Test Statistics. *J. Stat. Plan. Inference* **1999**, *82* (1–2), 171–196.

(44) Fransen, J.; Stucki, G.; van Riel, P. L. C. M. Rheumatoid Arthritis Measures: Disease Activity Score (DAS), Disease Activity Score-28 (DAS28), Rapid Assessment of Disease Activity in Rheumatology (RADAR), and Rheumatoid Arthritis Disease Activity Index (RADAI). *Arthritis Rheum.* **2003**, *49* (S5), S214–S224.

(45) Chipperfield, A. J. The MATLAB Genetic Algorithm Toolbox. In *IEE Colloquium on Applied Control Techniques Using MATLAB*; IEE, 1995; Vol. 1995, pp 10–10.

(46) Bartlett, D. B.; Connelly, M. A.; AbouAssi, H.; Bateman, L. A.; Tune, K. N.; Huebner, J. L.; Kraus, V. B.; Winegar, D. A.; Otvos, J. D.; Kraus, W. E.; et al. A Novel Inflammatory Biomarker, GlycA, Associates with Disease Activity in Rheumatoid Arthritis and Cardio-Metabolic Risk in BMI-Matched Controls. *Arthritis Res. Ther.* **2016**, *18* (1), 86.

(47) Lauridsen, M. B.; Bliddal, H.; Christensen, R.; Danneskiold-Samsøe, B.; Bennett, R.; Keun, H.; Lindon, J. C.; Nicholson, J. K.; Dorff, M. H.; Jaroszewski, J. W.; et al. ^{1} H NMR Spectroscopy-Based Interventional Metabolic Phenotyping: A Cohort Study of Rheumatoid Arthritis Patients. *J. Proteome Res.* **2010**, *9* (9), 4545–4553.

(48) Choy, E. H. S.; Panayi, G. S. Cytokine Pathways and Joint Inflammation in Rheumatoid Arthritis. *N. Engl. J. Med.* **2001**, *344* (12), 907–916.

(49) Dawes, P. T.; Fowler, P. D.; Clarke, S.; Fisher, J.; Lawton, A.; Shadforth, M. F. Rheumatoid Arthritis: Treatment Which Controls the C-Reactive Protein and Erythrocyte Sedimentation Rate Reduces Radiological Progression. *Rheumatology* **1986**, *25* (1), 44–49.

(50) Niwa, Y.; Iio, A.; Niwa, G.; Sakane, T.; Tsunematsu, T.; Kanoh, T. Serum Albumin Metabolism in Rheumatic Diseases: Relationship to Corticosteroids and Peptic Ulcer. *J. Clin. Lab. Immunol.* **1990**, *31* (1), 11–16.

(51) Barter, P. J.; Puranik, R.; Rye, K.-A. New Insights into the Role of HDL as an Anti-Inflammatory Agent in the Prevention of Cardiovascular Disease. *Curr. Cardiol. Rep.* **2007**, *9* (6), 493–498.

(52) Barter, P.; Gotto, A. M.; LaRosa, J. C.; Maroni, J.; Szarek, M.; Grundy, S. M.; Kastelein, J. J. P.; Bittner, V.; Fruchart, J.-C. HDL Cholesterol, Very Low Levels of LDL Cholesterol, and Cardiovascular Events. *N. Engl. J. Med.* **2007**, *357* (13), 1301–1310.

(53) Navab, M.; Reddy, S. T.; Van Lenten, B. J.; Fogelman, A. M. HDL and Cardiovascular Disease: Atherogenic and Atheroprotective Mechanisms. *Nat. Rev. Cardiol.* **2011**, *8* (4), 222–232.

(54) MacDougall, E. D.; Kramer, F.; Polinsky, P.; Barnhart, S.; Askari, B.; Johansson, F.; Varon, R.; Rosenfeld, M. E.; Oka, K.; Chan, L.; et al. Aggressive Very Low-Density Lipoprotein (VLDL) and LDL Lowering by Gene Transfer of the VLDL Receptor Combined with a Low-Fat Diet Regimen Induces Regression and Reduces Macrophage Content in Advanced

Atherosclerotic Lesions in LDL Receptor-Deficient Mice. *Am. J. Pathol.* **2006**, *168* (6), 2064–2073.

(55) Lawler, P. R.; Akinkuolie, A. O.; Chu, A. Y.; Shah, S. H.; Kraus, W. E.; Craig, D.; Padmanabhan, L.; Glynn, R. J.; Ridker, P. M.; Chasman, D. I.; et al. Atherogenic Lipoprotein Determinants of Cardiovascular Disease and Residual Risk among Individuals with Low Low-Density Lipoprotein Cholesterol. *J. Am. Heart Assoc.* **2017**, *6* (7), e005549.

(56) Lawler, P. R.; Akinkuolie, A. O.; Harada, P.; Glynn, R. J.; Chasman, D. I.; Ridker, P. M.; Mora, S. Residual Risk of Atherosclerotic Cardiovascular Events in Relation to Reductions in Very-Low-Density Lipoproteins. *J. Am. Heart Assoc.* **2017**, *6* (12), e007402.

(57) Lawler, P. R. Glycomics and Cardiovascular Disease. *Circ. Res.* **2018**, *122* (11), 1488–1490.

(58) Akinkuolie, A. O.; Pradhan, A. D.; Buring, J. E.; Ridker, P. M.; Mora, S. Novel Protein Glycan Side-Chain Biomarker and Risk of Incident Type 2 Diabetes MellitusSignificance. *Arterioscler. Thromb. Vasc. Biol.* **2015**, *35* (6), 1544–1550.

(59) McGarrah, R. W.; Kelly, J. P.; Craig, D. M.; Haynes, C.; Jessee, R. C.; Huffman, K. M.; Kraus, W. E.; Shah, S. H. A Novel Protein Glycan-Derived Inflammation Biomarker Independently Predicts Cardiovascular Disease and Modifies the Association of HDL Subclasses with Mortality. *Clin. Chem.* **2017**, *63* (1), 288–296.

(60) Lawler, P. R.; Akinkuolie, A. O.; Chandler, P. D.; Moorthy, M. V.; Vandenburgh, M. J.; Schaumberg, D. A.; Lee, I.-M.; Glynn, R. J.; Ridker, P. M.; Buring, J. E.; et al. Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality RiskNovelty and Significance.

Circ. Res. **2016**, *118* (7), 1106–1115.

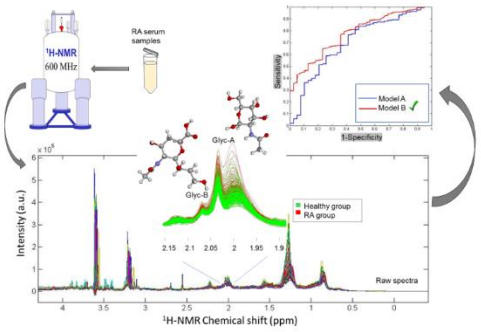
(61) Menni, C.; Gudelj, I.; Macdonald-Dunlop, E.; Mangino, M.; Zierer, J.; Bešić, E.; Joshi, P. K.; Trbojević-Akmačić, I.; Chowienczyk, P. J.; Spector, T. D.; et al. Glycosylation Profile of Immunoglobulin G Is Cross-Sectionally Associated With Cardiovascular Disease Risk Score and Subclinical Atherosclerosis in Two Independent Cohorts. *Circ. Res.* **2018**, *122* (11), 1555–1564.

(62) Väänänen, T.; Vuolteenaho, K.; Kautiainen, H.; Nieminen, R.; Möttönen, T.; Hannonen, P.; Korpela, M.; Kauppi, M. J.; Laiho, K.; Kaipiainen-Seppänen, O.; et al. Glycoprotein YKL-40: A Potential Biomarker of Disease Activity in Rheumatoid Arthritis during Intensive Treatment with CsDMARDs and Infliximab. Evidence from the Randomised Controlled NEO-RACo Trial. *PLoS One* **2017**, *12* (8), e0183294.

(63) Rietzschel, E.; De Buyzere, M. High-Sensitive C-Reactive Protein: Universal Prognostic and Causative Biomarker in Heart Disease? *Biomark. Med.* **2012**, *6* (1), 19–34.

(64) Gerszten, R. E.; Wang, T. J. The Search for New Cardiovascular Biomarkers. *Nature* **2008**, *451* (7181), 949–952.

(65) Hoefer, I. E.; Steffens, S.; Ala-Korpela, M.; Bäck, M.; Badimon, L.; Bochaton-Piallat, M.-L.; Boulanger, C. M.; Caligiuri, G.; Dimmeler, S.; Egido, J.; et al. Novel Methodologies for Biomarker Discovery in Atherosclerosis. *Eur. Heart J.* **2015**, *36* (39), 2635–2642.



Abstract graphic: For Table of Contents Only

Table 1. Demographic, clinical and biochemical characteristics of patients and controls.

		RA group				Control group			
		Number	%	Mean	SD	Number	%	Mean	SD
Gender	Women	134	63.8%			131	64.5%		
	Men	76	36.2%			72	35.5%		
Age (years)				58	12			45	16
BMI (Kg/m2)				27.79	5.89			25.30	2.99
Waist circumference (cm)				92	15				
Systolic blood pressure				137	21				
Diastolic blood pressure				81	12				
HDL cholesterol (mg/dL)				65	19				
LDL cholesterol (mg/dL)				118	31				
Triglycerides (mg/dL)				105	56				
Glucose (mg/dL)				95	23				
Current smoker (n, %)	No	154	73.3%						
	Yes	56	26.7%						
Diabetes mellitus (n, %)	No	185	88.1%						
	Yes	25	11.9%						
Hypertension (n, %)	No	85	40.5%						
	Yes	125	59.5%						
Dyslipidemia (n, %)	No	124	59.0%						
	Yes	86	41.0%						
Disease duration (years)				9.3	9.2				
DAS28				3.45	1.27				
	Disease remission (<2.6)	57	27.1%						
	Low disease activity (2.6-3.2)	39	18.6%						
	Moderate disease activity (3.2-5.6)	95	45.2%						
	High disease activity (>5.1)	19	9.0%						
HAQ (0-2.5)				0.44	0.52				
Rheumatoid factor + (n, %)	Negative	58	27.6%						
	Positive	152	72.4%						
ACPA+	Negative	40	19.0%						
	Positive	170	81.0%						
ESR (mm/h)				37	26				
CPR (mg/dL)				0.72	0.84				
Fibrinogen (mg/dL)				442	98				
DMARs	No	51	24.3%						
	Yes	159	75.7%						
Biological agent	No	169	80.5%						
	Yes	41	19.5%						
Corticoids	No	105	50.0%						
	Yes	105	50.0%						
NSAIDs	No	89	42.4%						
	yes	121	57.6%						

HAQ=health assessment questionnaire index, ACPA=citrullinated anti-cyclic peptide antibodies, ESR=erythrocyte sedimentation rate, CRP=C-reactive protein, DAS28=disease activity score, DMARDs=disease-modifying antirheumatic drugs, NSAIDs=non-steroidal antiinflammatory drugs.

Table 2. Results of univariate analysis between RA and healthy individuals (CT)

	AR		CT		
Variable	**Median**	**Iqr**	**Median**	**Iqr**	***p*-value**
Area GlycB (a.u.)	$2.14x10^{-01}$	$8.56x10^{-02}$	$2.22x1^{0-01}$	$7.39x10^{-02}$	$1.89x10^{-01}$
Area GlycA (a.u.)	$5.26x10^{-01}$	$1.12x10^{-01}$	$4.70x10^{-01}$	$1.09x10^{-01}$	$2.21x10^{-10}$*
H/W GlycB	$3.94x10^{-03}$	$1.81x10^{-03}$	$3.27x10^{-03}$	$1.03x10^{-03}$	$7.88x10^{-08}$*
H/W GlycA	$1.83x10^{-02}$	$4.27x10^{-03}$	$1.65x10^{-02}$	$5.46x10^{-03}$	$5.61x10^{-08}$*
Distance GlycB-GlycA (ppm)	$3.30x10^{-02}$	$9.16x10^{-04}$	$3.21x10^{-02}$	$9.17x10^{-04}$	$3.68x10^{-22}$*

Significant values ($p<0.05$) are marked (*).Median and interquartile range (Iqr) are reported. The Wilcoxon-Mann-Whitney test has been used to calculate significance. *P*-values adjusted by FDR.

Table 3. Associations between ^{1}H-NMR glycoproteins variables and clinical inflammatory markers.

		Area GlycB	Area GlycA	H/W GlycB	H/W GlycA	Distance GlycB-GlycA
Fibrinogen (mg/dl)	r	0.247*	0.452*	0.477*	0.414*	0.397*
	p	$1.08x10^{-03}$	$1.69x10^{-10}$	$2.14x10^{-11}$	$4.65x10^{-09}$	$2.29x10^{-08}$
ESR (mm/h)	r	0.341*	0.424*	0.436*	0.381*	0.302*
	p	$2.49x10^{-06}$	$1.89x10^{-09}$	$5.61x10^{-10}$	$9.07x10^{-08}$	$4.33x10^{-05}$
CRP (mg/dl)	r	0.278*	0.456*	0.468*	0.457*	0.356*
	p	$1.87x10^{-04}$	$1.34x10^{-10}$	$4.49x10^{-11}$	$1.34x10^{-10}$	$7.70x10^{-07}$
RF	r	0.127	0.109	0.196*	0.068	0.047
	p	$1.03x10^{-01}$	$1.65x10^{-01}$	$1.19x10^{-02}$	$4.02x10^{-01}$	$5.74x10^{-01}$
ACPA	r	-0.019	0.090	0.143	0.042	0.019
	p	$8.20x10^{-01}$	$2.56x10^{-01}$	$6.84x10^{-02}$	$6.09x10^{-01}$	$8.20x10^{-01}$
DAS28	r	0.214*	0.285*	0.259*	0.229*	0.127
	p	$5.35x10^{-03}$	$1.25x10^{-04}$	$5.59x10^{-04}$	$2.83x10^{-03}$	$1.01x10^{-01}$
Albumin (g/dl)	r	-0.144	-0.168*	-0.132	-0.183*	-0.257*
	p	$6.79x10^{-02}$	$3.18x10^{-02}$	$8.87x10^{-02}$	$1.86x10^{-02}$	$6.33x10^{-04}$

Spearman correlation coefficients (r) and p-value (*p*) for each glycoprotein variable and the inflammatory markers obtained by traditional biochemistry. *P*-values adjusted by FDR. Significant values ($p<0.05$) are marked (*). ESR: Erythrocyte sedimentation rate; CRP: C-reactive protein; RF: Rheumatoid Factor; ACPA: anti-citrullinated peptide antibodies; DAS28: Disease Activity Score of 28 joints.

1

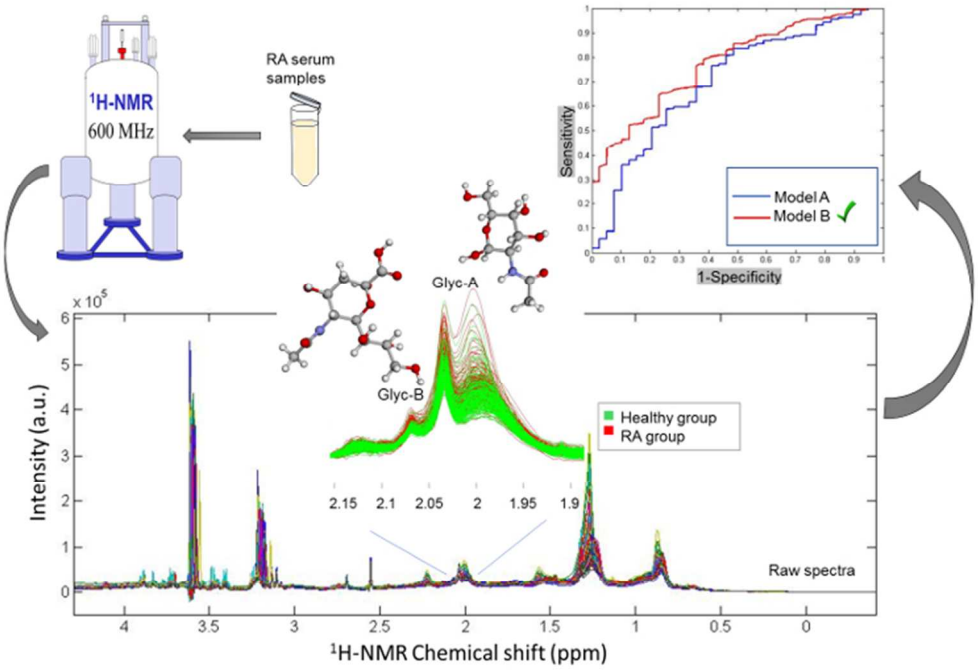
2

3

4

5

6



Abstract graphic: For Table Of Contents Only

64x44mm (300 x 300 DPI)