analytical chemistry

pubs.acs.org/ac
Article

Benchtop Proton NMR Spectroscopy for High-Throughput Lipoprotein Quantification in Human Serum and Plasma

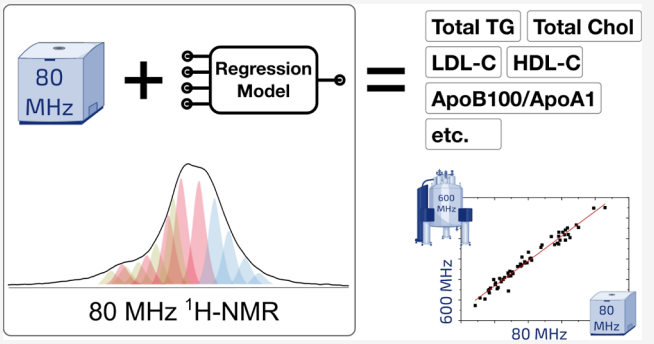
Julien Wist,*,○ Philipp Nitschke,○ Ricardo Conde, Angela de Diego, Maider Bizkarguenaga, Samantha Lodge, Drew Hall, Zhonglin Chai, Wenjun Wang, Sudhir Kowlessur, Marcio-Fernando Cobo, Niels Pompe, Birk Schütz, Hartmut Schäfer, Manfred Spraul, Claire Cannet, Tammo Diercks, Nieves Embade, Elaine Holmes, Oscar Millet, and Jeremy K. Nicholson*

Cite This: *Anal. Chem.* 2025, 97, 6399−6409

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We report the successful development and translation of high-field nuclear magnetic resonance (NMR) based comprehensive lipoprotein analysis to routine benchtop systems. This demonstrates the potential to reimagine population level cardiovascular disease risk analysis and individual level screening based on blood sampling. Using a quantitative calibration approach, we obtained stable and reproducible results from multiple sites, despite reduced spectral dispersion and sensitivity at lower field strengths. Our study shows that 25 out of 28 major lipoprotein parameters, including key cardiometabolic risk markers, were faithfully measured using benchtop NMR systems within 15 min. This development has significant implications for making a powerful diagnostic tool widely available, enhancing the potential for longitudinal personalized medicine through molecular phenotyping in the clinic.



INTRODUCTION

Lipoproteins (LP) are complex supramolecular structures made up of lipids and proteins that play a key role in lipid management and transport. High levels of low-density lipoprotein cholesterol (LDL-C) and low levels of high-density lipoprotein cholesterol (HDL-C) have long observed to associate with increased cardiovascular disease (CVD) risks,[1,2] although very high levels of HDL-C are also related to CVD and all-cause mortality.[3] Other lipoprotein related markers, such as apolipoproteins Apo-B100, Apo-A1[4] and Apo-A2 and their ratios,[5] have also gained attention as valuable CVD biomarkers. More recently, evidence linking the immune response with structural and functional modifications in lipoproteins and their composition,[6] suggests that chronic inflammatory conditions could be the root cause for the onset of secondary cardiometabolic disease, such as diabetes mellitus (T2D), chronic kidney disease, and rheumatoid arthritis.[2] Increased LDL serum concentrations[7] and a distribution shift toward smaller and denser LDL subfractions[8] have been observed in obesity, which is also associated with chronic inflammation.[9] HDL related parameters (e.g., HDL ApoA1) have been observed to drop in individuals infected by SARS-CoV-2.[10−12] Similarly, children with acute SARS-CoV-2 infection present a drop in HDL ApoA1 but no significant compositional changes in LP, whereas, in contrast, a strong LDL and HDL enrichment in triglycerides was observed in children with acute multi-inflammatory syndrome (MIS-C).[13] Molecular phenotyping studies on acute COVID-19 cohorts confirmed the association between acute inflammation and lipoproteins;[11] the latter have been shown to be accurate prognostic markers of severity[14,15] and recovery from infection.[15] Overall, these findings highlight the potential of lipoproteins for early detection of chronic inflammatory conditions and the need for further biological exploration.

Although the gold standard analytical techniques for the quantification and characterization of lipoproteins are ultracentrifugation and gradient gel electrophoresis,[16−18] nuclear magnetic resonance (NMR) spectroscopy quantifies over a hundred lipoprotein parameters in a single measurement (∼4 min) that can cover all aspects of lipoproteins e.g. lipid composition of main and subfractions, particle size and number, and protein content.[19−22] As a result, NMR lipoprotein analysis has established itself as a key technology for the biological exploration of lipoprotein related pathological conditions aiding disease risk stratification and long-term prognosis in large-scale

Received: August 30, 2024
Revised: February 14, 2025
Accepted: March 5, 2025
Published: March 17, 2025

epidemiological studies.[23] Current NMR-based lipoprotein analysis requires high field systems, from 400 to 600 MHz, which poses considerable challenges for its deployment in day-to-day applications e.g. in small laboratories or clinics. The requirement for large instrument footprints, high capital cost and the need for cryogenic liquids has limited the use of NMR lipoprotein analysis to central hubs for service measurements rather than promoting widespread instrument use.

As alternative to high field NMR spectrometers, benchtop NMR systems use permanent magnets to achieve medium field magnitudes ranging from 60 to 100 MHz. These instruments have a small footprint and operate at room temperature, thus there is no requirement for cryogenic liquids, keeping maintenance and operation simple and cost-effective. Over the past decade, benchtop NMR spectrometers have been used for applications as diverse as education, quality control and reaction monitoring.[24] Moreover, promising clinical applications have been reported using challenging biofluids like urine, serum or plasma. In urine, aminoadipic acid, citrate, creatine, creatinine, glucose, mannitol, phenylalanine, and hippurate were identified as a signature for tuberculosis at lower-field,[25] while biomarkers for diabetes (glucose and acetone) were also successfully detected in urine and plasma samples.[26]

Using pulsed field gradients to single out signals of interest, we have recently ported the measurement of inflammatory markers in serum from high field NMR spectrometers to benchtop systems.[27] In fact, part of this inflammatory signature, the supramolecular phospholipid composite (SPC) signal, provides a direct "model-free" reading of both HDL and LDL main fractions by peak integration. This observation raised the question as to how many lipoprotein parameters are accessible at medium field, bearing in mind that the first lipoprotein model was developed on a 250 MHz system.[19]

Therefore, we conducted a collaborative study across three research institutes aiming at establishing a lipoprotein serum model for benchtop NMR systems. Each institute recorded matched high field and benchtop spectral data for different serum cohorts (N_1 = 110, N_2 = 121, N_3 = 127; N_{total} = 358). Lipoprotein data were extracted using the Bruker IVDr Lipoprotein Subclass Analysis (B.I.-LISA) method based on 600 MHz spectra and used to build a benchtop NMR model by regression against the 80 MHz spectra for the same cohort set. As a result of the standardization of the acquisition protocol and external quantitative calibration all 3 cohort-models were similar, and a joint model was built that recovered 26 out of 28 (93%) main and 62 out of 112 (55%) parameters, overall. This successful translation of lipoprotein analysis to benchtop NMR systems is a major milestone toward deployment of NMR diagnostic and prognostic tools within the clinical and healthcare landscapes.

EXPERIMENTAL METHODS

Chemicals and Consumables. Phosphate buffer (75 mM Na_2HPO_4, 2 mM NaN_3, 0.08% sodium trimethylsilyl propionate-[2,2,3,3-$^2H_4$] (TSP) in H_2O/D_2O 4:1, pH 7.4 ± 0.1), all NMR tubes (5 mm outer diameter SampleJetTM NMR tubes and regular 7 in., 5 mm NMR tubes) with the corresponding sealing caps, the Bruker Fourier 80 5 mm shimming sample (Doped Water 5% H_2O 0.6 mM $CuSO_4$ in D_2O), the Bruker QuantRefC—5 mm sample, and 100 mg/mL Pamoic acid in DMSO-d_6 + 1% TMS 5 mm sample were purchased from Bruker Switzerland A.G. Fällanden. Beckmann OptiSeal tubes were purchased from Beckman Coulter. Amicon Ultra-4 centrifugal filters were purchased from Sigma-Aldrich.

Samples Analyzed and Analytical Laboratory Sites. *Germany—Bruker.* Samples from the US cohort (N = 97) were obtained under approval of the Advarra Institutional Review Board (approval no. CR00487335). Samples from the German cohort (N = 30) were obtained under the approval of the ethics committee Landesärztekammer Baden-Württemberg (approval no. F-2022-096). These samples were analyzed in Germany at Bruker Biospin (Ettlingen).

Spain—CIC bioGUNE. Samples (N = 127) were collected from Spain with Ethics Approval: Basque Committee of Ethics and Clinical Research CEIC PI-19-13. These samples were analyzed at CIC bioGUNE, Derio, Spain.

Mauritius—ANPC. Samples (N = 119) were collected from Mauritius with the Ethics Approval: Republic of Mauritius, Ministry of Health and Wellness, The National Ethics Committee MHC/CT/NETH/2021V3. These samples were analyzed at the Australian National Phenome Centre, Perth, Australia.

Sample Preparation and Data Acquisition. *Sample Preparation.* All 3 sites agreed to follow the same sample preparation procedure prior to starting the study. Serum samples were thawed on the day of measurement in a fridge at 278 K. Two aliquots, 300 μL each were taken out of the sample tube and each aliquot was prepared according to recommended procedures for in vitro analytical diagnostics procedures mixing 300 μL serum with 300 μL phosphate buffer for each aliquot. The first 600 μL of the aliquot intended for measurement at 600 MHz was transferred to a 5 mm outer-diameter SampleJet NMR tube. The second 600 μL aliquot intended for 80 MHz measurement was transferred to a standard 7-in. 5 mm outer diameter NMR tube.

For measuring samples at ANPC, in-house long-term plasma reference samples were prepared as regular samples and inserted after every 16 samples.

^{1}H NMR Spectroscopy Data Acquisition and Processing Parameters. *Acquisition at 600 MHz.* At each site, NMR spectroscopic analyses were performed on two 600 MHz Bruker AVANCE III HD (ANPC & CIC bioGUNE) and NEO (Bruker) spectrometers (one for 310 K—IVDr procedures, one for 298 K), each equipped with a 5 mm BBI probe and fitted with a Bruker SampleJet robot cooling system set to 278 K. A full quantitative calibration was completed for the 310 K IVDr measurements prior to analysis using a protocol described elsewhere. For the measurements at 298 K, the temperature was checked with a 99.8% MeOD sample. In addition, shimming quality and solvent suppression was checked with a 2 mM Sucrose sample in H_2O/D_2O 9:1.[21] The samples were first measured at 310 K—IVDr procedure. After completion, the samples were transferred to the second magnet set at 298 K. For each sample, a standard one-dimensional (1D) experiment with solvent suppression (pp: noesygppr1d) was acquired with 32 scans (+4 dummy scans), 98k data points, relaxation delay of 4.0 s, mixing time of 10 ms, presaturation of 25 Hz, and a spectral width of 30 ppm resulting in a total experimental time of 4 min 3 s (according to the Bruker In Vitro Diagnostics research IVDr methods).

Processing at 600 MHz. Time domain data for the 310 K—IVDr and ad-hoc 298 K measurements were Fourier transformed with an exponential line broadening function of 0.3 Hz and processed the same way in automation using Bruker TopSpin 3.6.3 (ANPC & CIC bioGUNE) and TopSpin 4.2.0

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

(Bruker) and ICON NMR to obtain phase- and baseline corrected spectra referenced to TSP at 0.0 ppm. Quantitative referencing of the 1D-NMR spectra was achieved using the (PUlse Length-based CONcentration) PULCON method in combination with an external quantification reference sample (Bruker's QuantRefC—5 mm sample) to add an artificial signal representing a predefined fixed proton concentration into the spectrum.[28] This procedure ensures that the observed intensities are quantitatively comparable across the different samples and platforms. Both, acquisition and processing procedures were done in full automation as described using Bruker's B.I.Methods 2.5 package.

A total of 112 lipoprotein parameters for each sample were generated using the Bruker IVDr Lipoprotein Subclass Analysis (B.I.-LISA) method, whereby the $-(CH_2)_n$ at 1.25 ppm and $-CH_3$ at 0.80 ppm peaks of the 1D spectrum—after normalization to the Bruker QuantRef reference sample within TopSpin—were quantified using a PLS-2 regression model. B.I.-LISA data consist of total plasma lipid analyte cholesterol, free cholesterol, phospholipids, triglycerides, apolipoproteins A1/A2/B100 and the B100/A1 ratio, and analyte distributions in different density classes of plasma-lipoproteins (see Table S2):[16−18,29,30] high-density lipoprotein (HDL, density 1.063−1.210 kg/L) further subdivided in four subclasses, low-density lipoprotein (LDL, density 1.09−1.63 kg/L) further subdivided into six subclasses, intermediate-density lipoprotein (IDL, density 1.006−1.019 kg/L), and very low-density lipoprotein (VLDL, 0.950−1.006 kg/L) further subdivided into 5 subclasses (further details in Supporting Information).

Acquisition at 80 MHz. At each site, NMR spectroscopic analyses were performed on an 80 MHz Bruker Fourier 80 spectrometer equipped with a Fourier 80 Two Channel (^1H + ^{13}C) probe including a nominal 25 G/cm gradient system and an automatic sample changer system. The system operates at a constant temperature of 298.15 K. The ANPC also used a second Fourier 80 with the same specifications, but additionally equipped with an adjustable temperature unit for measurements at 310 K. To ensure proper field homogeneity of the systems a shimming sample (Doped Water 5% H_2O 0.6 mM $CuSO_4$ in D_2O) was inserted every 1.5 h and the Bruker shimming algorithm operated until the minimum line shape criteria (0.8 Hz at 50% and 20 Hz at 0.55% of the resonance line) for the water signal were reached. As the Fourier 80 sample changer has no cooling system, a time series of a plasma sample inside the benchtop NMR was used to show that spectral deviation from potential degradation within 24 h was minimal (Figure S1).

The standard one-dimensional (1D) experiments with solvent suppression (pp: noesypr1d) were acquired with 96 scans (+4 dummy scans), 23,808 data points, relaxation delay of 4.0 s, mixing time of 50 ms, presaturation of 15 Hz and a spectral width of 30 ppm. To ensure proficient solvent suppression a "scout experiment" (pp: *zg*) with 4 scans (0 dummy scans) is acquired prior to the 1D experiment with solvent suppression to locate the water peak in automation. This is necessary, because the Fourier 80 uses an external rather than an internal lock system, resulting in slightly shifted spectra for every sample after insertion. The total experiment time of the scout experiment and subsequent 1D with solvent suppression was ∼16 min.

Quantitative referencing (QuantRef) was performed based on the PULCON principle similar to those established at 600 MHz.[28] Here, 600 μL of pure Millipore water in a 5 mm NMR tube was used as a reference standard equating to a molarity of 55.56 M. This QuantRef sample was measured at the beginning of each measurement day mimicking the 600 MHz procedure. After each measurement, a QuantRef file was automatically saved in each sample folder, which can be used for spectral correction. Note that no deuterated solvent needs to be used for the Fourier 80 as it uses an external lock system. In addition, the Fourier 80 operates at a fixed instrument receiver gain of 1, omitting the necessity to match this parameter for spectral acquisition between samples. Also, at lower field radiation damping effects are greatly diminished and the system can handle the measurement of pure water without spectral distortions.

For the lipoprotein main fractions (see Ultracentrifugation), a standard 1D experiment with solvent suppression was acquired with 512 scans and a relaxation delay of 2.0 s. All other parameters were equal to 1D experiments of the serum samples. First spectra were recorded at 310 K. After completion they were put into the fridge and the F80 system was adjusted to 298 K. Once stable, all fractions were measured again at 298 K with the same parameters.

Processing at 80 MHz. Time domain data were Fourier transformed and processed in automation using Bruker TopSpin 4.3.0 and ICON NMR to obtain phase and baseline corrected spectra. An exponential line broadening of 0.3 Hz was applied to the 1D water suppressed experiment and the spectra were referenced to TSP at 0.0 ppm. The QuantRef correction was applied to all spectra to ensure that the observed intensities were quantitative. The resulting spectra were interpolated to ensure they share a common x axis.

Ultracentrifugation. For the variable temperature experiments on lipoproteins, lipoprotein main fractions were isolated from one serum sample by sequential flotation ultracentrifugation based on established protocols[31,32] using a Beckman Coulter Optima L-100 XP Ultracentrifuge equipped with a Beckman Type NVT-100 near vertical angle rotor. Using 1.5 mL of serum in a Beckmann OptiSeal tube (V = 4.9 mL), the main fractions were collected according to their desired density (VLDL $d \leq 1.006$, IDL $d \leq 1.019$, LDL $d \leq 1.063$, and HDL $\leq$ 1.210 g/mL) using NaBr gradients. For each, the upper lipoprotein fraction was taken, whereas the lower residue was density adjusted using NaBr to isolate the next fraction by centrifugation. No further LDL, or HDL subfractions were isolated.

Isolated fractions were concentrated using Amicon Ultra-4 centrifugal filters (molecular cutoff 100 kDa) and a mixture of phosphate buffer mixed with Millipore water 1:1 containing an additional 2 mM alanine for chemical shift referencing.

Data Modeling. The 80 MHz spectra were used to build a regression model for the 112 lipoprotein parameters obtained from the B.I.LISA method. A regularized generalized canonical correlation analysis[33,34] (RGCCA) and its Sparse implementation for variable selection (SGCCA)[35] was used to perform a multiblock linear regression. Therefore, the consolidated data matrix was divided in 3 data blocks, one for each spectral regions of interest:

(1) The aliphatic region from 0.4 to 1.8 ppm (CH block; 6109 complex points) (Figure 1c).

(2) The glyc[11,36] region from 1.8 to 2.5 ppm (GLYC block; 655 complex points).

(3) And the SPC region[11,37] from 3.1 to 3.4 ppm (SPC block; 611 complex points) (Figure 1b).

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

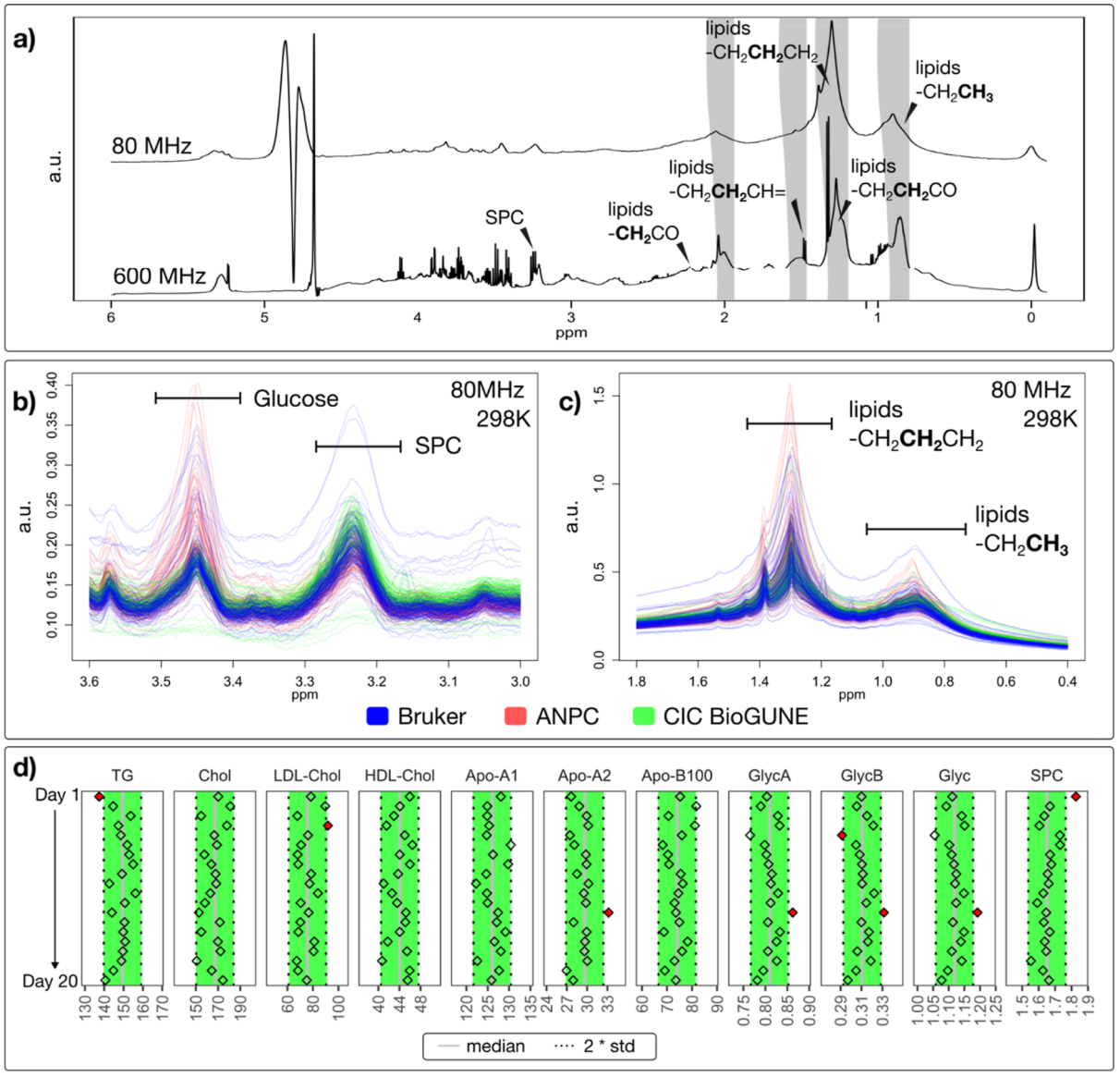


Figure 1. NMR spectra of serum at medium and high field across three institutes. (a) Comparison of an exemplar ^{1}H NMR spectrum of serum with water suppression at 80 MHz (upper trace) and 600 MHz (lower trace) highlighting the most pronounced lipid peaks and their field dependent dispersion behavior. (b) Selected spectral window with 358 ^{1}H spectra focusing on SPC at 80 MHz. (c) Excerpt of 358 ^{1}H spectra focusing on the methylene and methyl lipid peaks at 80 MHz. Spectra from the institutes are color coded: Bruker are in (blue), ANPC in (red) and CIC bioGUNE in (green) and show good agreement for spectral intensity after quantitative referencing (QuantRef). (d) Reproducibility and stability of acquisition over a 20 day period. The green area represents the 95% confidence interval (±2 standard deviations). Note that a significant shift in water peak position is observed after alignment to TSP in the top 2 spectra. This is due to the fact that the 80 MHz spectra was acquired at 298 K, while the lower trace (600 MHz) was acquired at 310 K.

A fourth *response* data block consisted of the 112 lipoprotein parameters extracted from the 600 MHz data using the B.I.LISA model from Bruker Biospin GMBH & Co. KG, Germany.

Considering J data blocks $X_1, ..., X_J$, the design matrix $C = \{c_{jk}\}$ that define their connections, the SGCCA method solves the following optimization problem

$$\text{argmax:} \sum_{j,k=1;j\neq k}^{J} c_{jk} g(\text{cov}(X_j a_j, X_k a_k))$$

$$\text{subject to: } \|a_j\|_2 = 1 \text{ and } \tau_j \|a_j\|_1 \leq s_j, \qquad j = 1, \ldots, J$$

All blocks were considered to be correlated ($\forall j, \forall k, c_{jk} = 1$), $g(x) = x^2$ (factorial scheme) was chosen, and the sparsity for each block was set to 0.4 to ensure that variables with very low contributions to the model are properly driven to 0. With these settings, the optimal number of predictive components was 11. The data set was divided into training (80%) and test sets (20%). To obtain average performance metrics, bootstrapping was performed 20 times by sampling the population with replacement. To detect and avoid overfitting of the data, the data set was split into training and test data sets. The model was built using only the training data and evaluated using the test data set only. This latter was achieved by computing the normalized root mean square error (nRMSE, also known as coefficient of variation) and R^2.

The analysis of the data was performed using in-house developed R scripts and the RGCCA[38] package.

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

RESULTS AND DISCUSSION

Experimental Design and NMR Spectral Analysis. To investigate the feasibility for translation of the lipoprotein model from a high field, 600 MHz NMR spectrometer to a lower field, 80 MHz benchtop system, three cohorts totalling 389 samples (including 16 Long-term References samples for quality control; LTR) were analyzed in three corresponding institutes. Each cohort contributed over 100 samples (after exclusion of 15 poor quality and 16 LTR spectra from the models) with a healthy/free-living population cohort (N_3 = 127) from Bruker Biospin GmbH & Co. KG, Germany, a free-living population cohort (N_2 = 121) from CIC bioGUNE and a diabetic/obese cohort (N_1 = 110) from the ANPC Australia which was collected in the Republic of Mauritius. The combination of normal population and pathological samples, was designed to generate an inherently larger variance across lipoprotein blood profiles to maximize the covariance range of benchtop the lipoprotein model.[22,39] A demographic description of the three cohorts can be found in the Supporting Information (Table S1). All samples were measured with two aliquots from the same sample at 600 MHz according to IVDr procedures and at 80 MHz. Examples of the standard solvent suppressed ^{1}H spectrum at 600 and 80 MHz are shown in Figure 1a.

Measuring at lower field strength leads to (1) a decreased spectral dispersion by a factor of 7.5 between 80 and 600 MHz and (2) a decreased signal-to-noise by a factor of ∼50 which is equivalent to a ∼2500 fold increase in experimental time (Figure S2). Nevertheless, all major lipid peaks are already visible at 80 MHz after 15 min acquisition time, providing one of the basic requirements for lipoprotein analysis at medium field. In addition, the spectrum is dominated by broad features throughout the whole spectrum stemming from plasma proteins (mostly albumin) at both field strengths. Under these conditions, the detection of most abundant small molecules such as glucose, lactate and alanine enables spectral calibration; alignment to alanine is commonly used as chemical shift reference for lipoprotein analysis. Here, the calibration of medium field spectra to the methyl peak of alanine was achieved manually, with some spectra showing barely distinguishable doublets from the methyl group (Figure S3b before and Figure S5b after calibration).

Field deployment of lipoprotein analysis requires comparable spectral quality and reproducibility across the 3 institutes ANPC, CIC bioGUNE and Bruker Biospin. This was achieved by introducing a quantitative referencing step based on the PULCON method,[28] as described in the Experimental Methods section. Superimpositions of all acquired spectra (N_{total} = 358) at 80 MHz highlighting key spectral regions for lipoprotein analysis are shown in Figure 1b,c. Samples from ANPC are in red, CIC BioGUNE in green and Bruker BioSpin in blue. In addition, 16 long-term references (LTR) samples were inserted during the run at the ANPC (Figure S5, in red). Integration of spectral regions across LTR samples showed a standard deviation of 2.3% at 80 MHz. This is higher than that which is usually found at 600 MHz (∼1%) and probably stems from the low signal-to-noise at 80 MHz taking into account that the variation is roughly ∼1/SN[40] and the SN of the lipoprotein regions is ∼100 in the investigated spectra.[40]

Figure 1c focuses on the lipoprotein methylene and methyl regions that are most commonly used for lipoprotein prediction using NMR spectroscopy. Due to the lower dispersion, both the methylene and methyl regions directly overlap at 80 MHz. Looking at the overall intensity of the low field methylene and methyl regions (Figure 1c) the spectral intensities generated at the three institutes showed a trend similar to that of their 600 MHz counterparts (data not shown) indicating no extreme deviations after QuantRef correction was applied across the 3 laboratories.

In addition, the SPC region is shown in Figure 1b as a potential, additional target for lipoprotein modeling as it presents the methyl groups of the mobile phospholipid, choline head groups from mainly LDL and HDL particles. The base of the broad lipoprotein peaks covers the same range in Hertz as at high field, but a much larger range in ppm at 80 MHz. Figure 1b reveals that at lower field SPC, the β-glucose C2−H and other, minor choline and choline derivatives merge into one broad composite peak. One can also detect multiple Glucose protons that are merged into one composite peak at 3.45 ppm. As expected, glucose is increased on average for the ANPC, diabetic cohort (red). Glyc is also present in the spectrum, as shown in our previous work[27] at 2.07 ppm. At 80 MHz, the Glyc signature merges into a composite peak with the allylic methylene moiety of the lipoproteins.

Successful water suppression was investigated as another quality criterion and the chemical shift region from 4.5 to 5.0 ppm is plotted in Figure S3a. Presaturation was used as a method of choice, but it should be mentioned that WET-suppression also showed effective performance (Figure S4). Some quality discrepancies among the 3 laboratories were observed, although successful suppression was achieved for all samples. However, several samples from CIC bioGUNE yielded a larger, residual distortion from the water peak. Yet, manual inspection of these spectra showed that the residual solvent peak did not lead to distortions of the spectral regions of interest. Although it goes beyond the scope of this manuscript, in a routine workflow quality check would have identified these spectra and, if necessary, the corresponding samples would have been reanalyzed to reacquire data.

Reproducibility and stability of the platform is illustrated in Figure 1d. The same sample was prepared and measured each day during 20 days. The green area between the 2 vertical dotted lines represent the 95% confidence interval (±2 standard deviations). Only 8 out of 220 (3.6%) measurements were found outside of these boundaries (red diamonds).

Prediction Performance of Regression Models. Prediction performance was investigated for the 112 lipoprotein parameters of the Bruker Biospin B.I.LISA platform using N_{total} = 358 spectra from the three centers ANPC (N_1 = 110), CIC bioGUNE (N_2 = 121), and BrukerBioSpin (N_3 = 127), after removing low quality and LTR spectra (15 samples and 16 LTR; see Figure S5).

Sparse regularized generalized canonical correlation analysis, SGCCA, was used to train a regression model. SGCCA is an extension of regularised generalized canonical correlation analysis (RGCCA) proposed by Tenenhaus as a framework that combines both the attractiveness of multiblock methods with the power of PLS[41] and other methods (CCA) for modeling linear relationships. Since this work aims at demonstrating the feasibility of the translation of the lipoprotein model from 600 to 80 MHz, we opted to use this recent framework that allows for the visual inspection of linear relationships between the blocks of data (611 (SPC), 655 (GLYC) and 6109 (CH) complex points) and benefits from the established loading interpretability of the PLS. However, other regression techniques such as LASSO,[42] ridge regression[43,44]

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

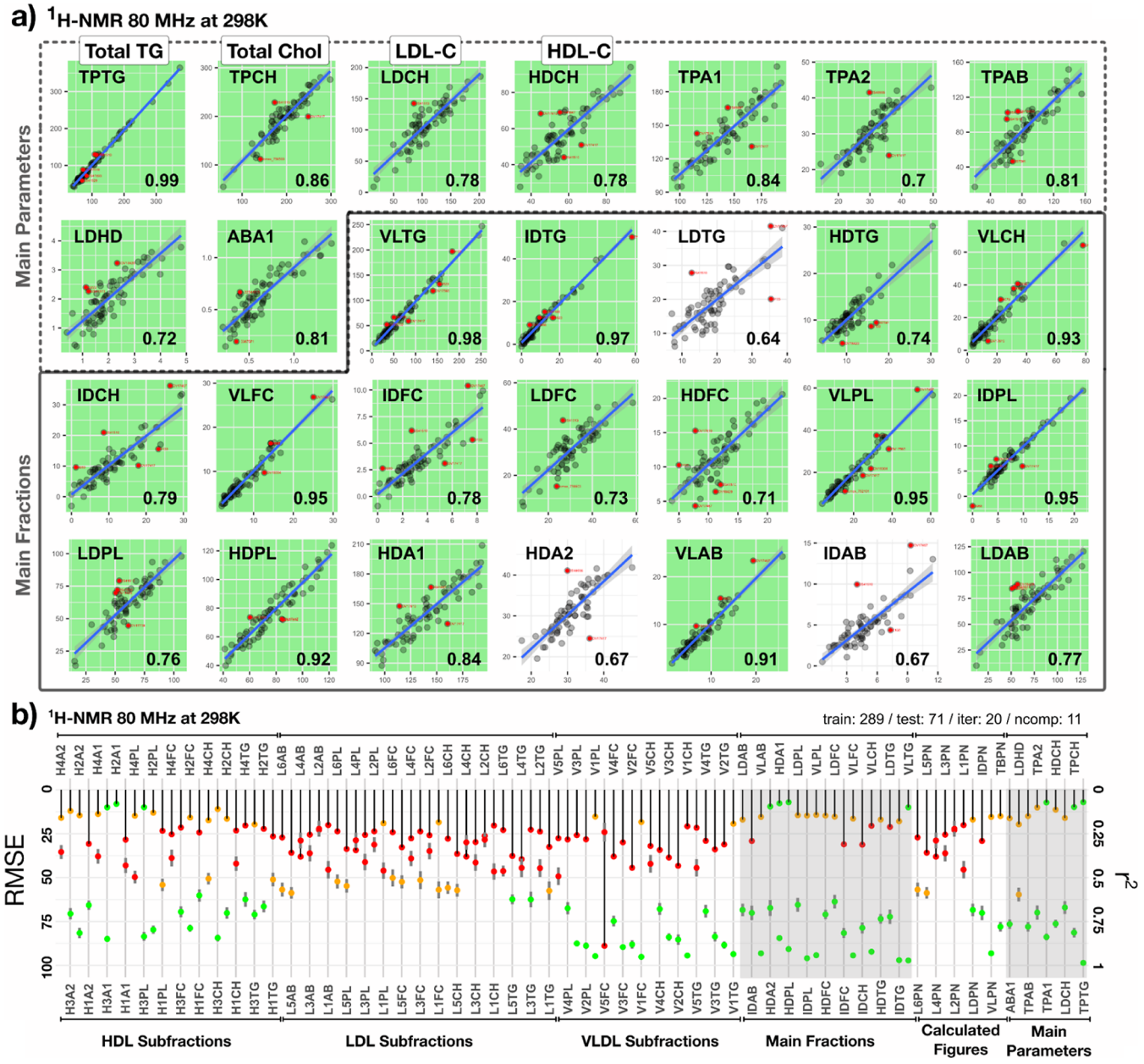


Figure 2. Prediction performance of the SGCCA lipoprotein model. (a) Performance of predicting the 28 lipoprotein main parameters and main fractions at 80 MHz. A green background indicates an r^2 of >0.70 (satisfactory), no highlight indicates r^2 values between 0.7−0.5 (moderate), a red background was given to parameters with $r^2 < 0.5$ (Table S3). (b) Downward sticks show the normalized RMSE (or coefficient of variance) while the green, orange and red dots show the averaged r^2 (goodness of fit) for 20 bootstrapped models using 80 MHz data. The most established clinical parameters, total triglycerides, total cholesterol, HDL-C and LDL-C are highlighted in gray. It can be observed that while a lot of information about subfractions is lost (red dots), the clinically most important parameters are recovered (green dots).

and gradient boosting methods[45,46] could have been chosen as an alternative.

After careful calibration to alanine (1.48 ppm), 3 spectra data blocks—referred to as CH, GLYC and SPC—were created as described in the Experimental Methods section. The inclusion of the GLYC block is justified by prior work showing the important role of the SPC/Glyc ratio to capture inflammatory conditions.[11,47] As we postulate that low intensity inflammatory conditions may cause the onset of secondary cardiometabolic disease, inclusion of the GLYC block into this analysis was deemed appropriate. A fourth "response" block containing the lipoprotein data described in Table S2 (B.I.LISA (600 MHz)). Each data block was split into a training (80%) and test set (20%). Using the four training blocks a model was built assuming that each block is correlated with every other. The parametrization of the SGCCA was chosen in a way that favors high variance within each block (new mode A33) which is equivalent to a multiblock PLS.

For each lipoprotein parameter, the correlation between observed and predicted values for the test set was evaluated using a linear model (Figure 2 and Table S3). Bootstrapping with replacement was used to train 20 models and obtain averaged statistics (normalized RMSE and r^2) for predictions. Best results were obtained using 11 predictive components. The cumulative explained variance for each block is shown in Figure S6.

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

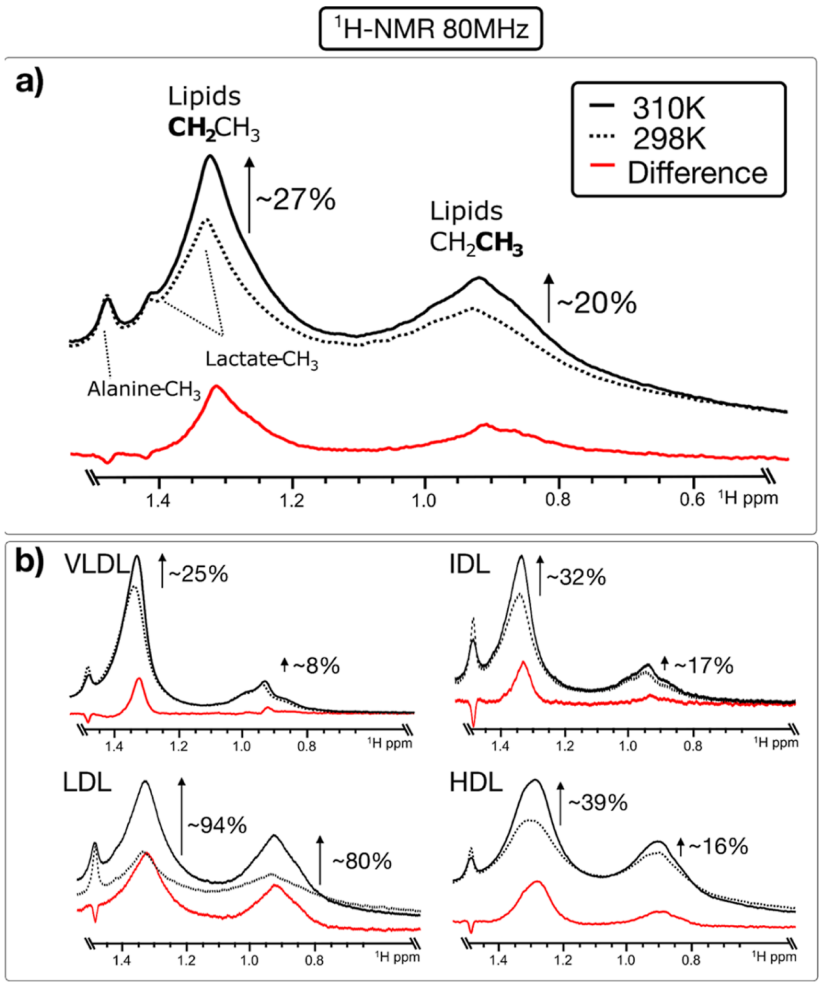


Figure 3. Evaluation of the impact of the temperature on the intensity of the lipoproteins signal at 80 MHz (a) for an arbitrarily chosen serum sample and (b) for all main fractions of that same sample. Black, dotted and red traces are for measurements performed at 310, 298 K and the difference between both. Increasing the temperature is expected to positively affect the signal-to-noise ratio, in particular for LDL fraction and subfractions,[29,48] which is confirmed by these first observations at benchtop NMR.

Figure 2a shows the linear regression obtained for the main parameters and fractions using benchtop test data. The correlation coefficients (r^2) ranged from 0.99 to 0.64 for all 7 main parameters, 19 main fraction parameters and the 2 Apo-B100/Apo-A1 and LDL/HDL ratios. All average normalized RMSE values were below 20% (Figure 2b, greyed areas) except for LDTG, VLCH, IDCH, IDFC and IDAB where found to be close to 30%. Notably, strong correlation scores were achieved for total triglycerides (TPTG, r^2 = 0.99), total cholesterol (TPCH, r^2 = 0.86), ApoA1 (TPA1, r^2 = 0.84), ApoB100 (TPAB, r^2 = 0.81), LDL-C (LDCH, r^2 = 0.78) and HDL-C (HDCH, r^2 = 0.78). All these parameters have been suggested as cardiovascular risk markers, which can be predicted with nRMSE errors below 20% using a benchtop NMR magnet within 15 min in this study.

Over half of the lipoprotein parameters (62 out of 112; 55%) reached a satisfactory $r^2 > 0.7$ (Figure 2b, greyed areas). This includes 25 out of 28 of the main parameters, main fractions and the 2 Apo-B100/Apo-A1 and LDL/HDL ratios. Further 30 parameters yielded a moderate prediction performance, $0.7 > r^2 > 0.5$ resulting in a total of 13 moderate values. Finally, 30 parameters failed to achieve a reasonable $r^2 < 0.5$, mainly LDL subfractions, but also VLDL subfraction 5 and HDL subfraction 1 and 4. This underlines the fact that certain parameters e.g. LDL and its subfractions are generally more difficult to predict due to their heavy spectral congestion and lower spectral visibility.

In addition, the two inflammatory signatures, SPC and Glyc were also successfully recovered with an r^2 of 0.95 and 0.78 respectively (Table S3).

Temperature Positively Impacts Prediction for LDL Subfractions. It is well established that the lipoprotein peak intensities in an NMR spectrum are significantly influenced by temperature due to increased/decreased motion of the different lipoprotein compartments within the lipoprotein particles at increased/decreased temperature.[48] Above phase transition (310 K), increased thickness of the outer surface shell leads to a structural change of the particle; probably due to conformational change in the Apo-B protein.[29] As a result, it is common to acquire spectra at temperatures ranging from 310 to 320 K to maximize the lipoprotein peak area. Although adjustable temperature exists as an optional feature for Fourier 80 NMR systems, only the most recent benchtop NMR spectrometer installed at the ANPC was equipped with adjustable temperature. Therefore, the experimental design favored the multisite validation experiment at 298 K and it was warranted to investigate how a temperature difference of 12 K (310 to 298 K) influences the spectra at lower-field as certain features could be lost due to restricted mobility of the underlying lipids in the

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

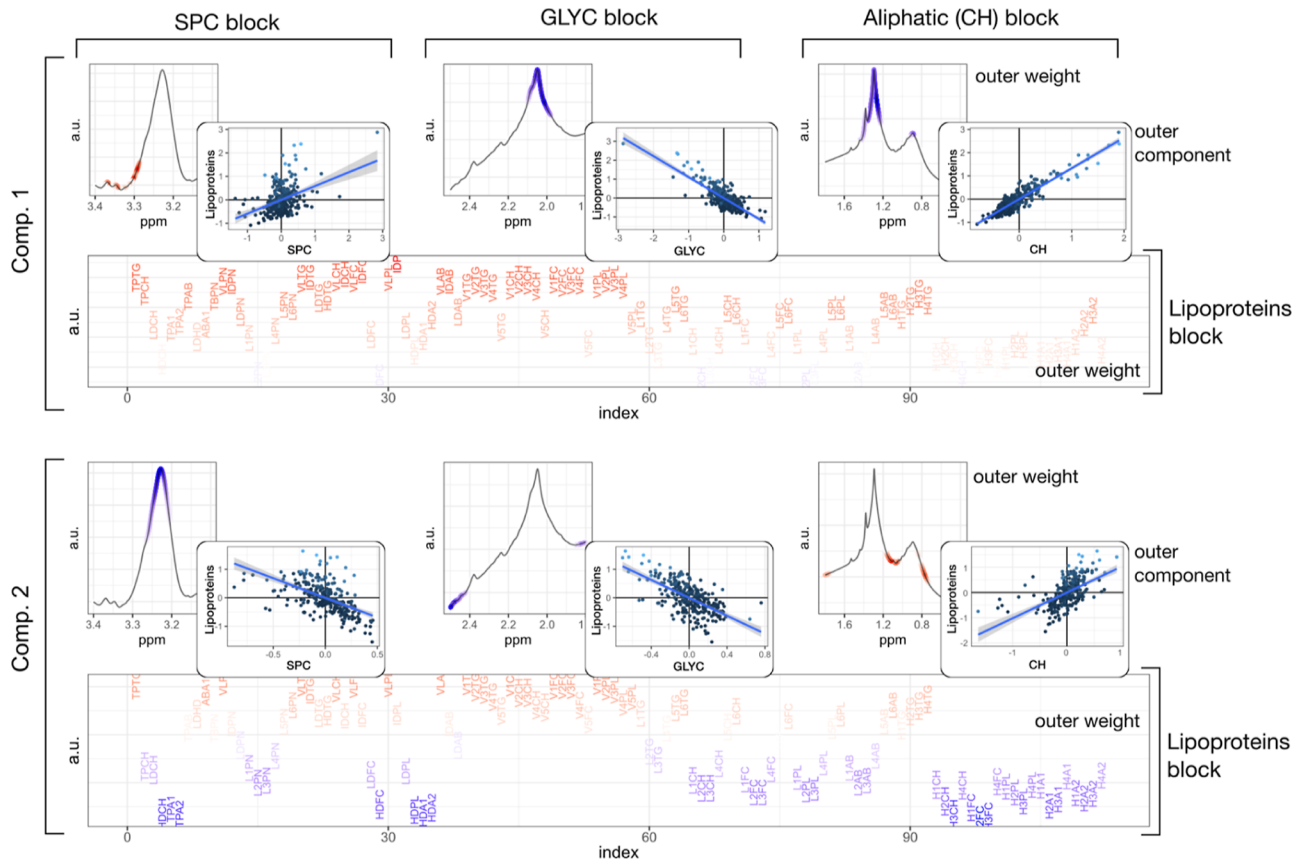


Figure 4. Outer weights and outer components for all 3 data blocks (SPC, GLYC and CH regions of 80 MHz spectra) and the "response block" (lipoprotein parameters) for the first 2 components. For the outer weights, the trace (black) is an arbitrarily chosen spectrum, while the color code indicates the positive (red) or negative contribution of the variable to the latent variable (comp. 1). The darker the color the larger the contribution, no contribution is indicated in white (no color). The regularisation term ensures that only the most relevant variables are selected, while all others are driven to 0. The same color code is applied for the lipoprotein's outer weight. As mentioned above, the GLYC block was included into this analysis based on prior work showing the important role of the SPC/Glyc ratio to capture inflammatory conditions.[11,47] Again, as we postulate that low intensity inflammatory conditions may cause the onset of secondary cardiometabolic disease, it was deemed appropriate to retain the GLYC block into this analysis. Comp. 1 displays the first "latent" variable that is found to covariate among all blocks. This latent variable mostly corresponds to the total particle numbers and VLDL parameters. The data block that contributes the most to this latent variable is the CH block as illustrated by the high correlation of its outer component. The spectral variables that drive this high correlation are highlighted in the corresponding outer weight (in blue). Not surprisingly, the SPC block does not contribute to this latent variable as VLDL particles do not contribute to this signal.[37] This approach offers the opportunity to visually explore the results, unlike other regression models that operate as "black boxes".

lipoproteins. Figure 3a shows the same sample prepared and measured at 298 K (dotted line) and 310 K (black line), while Figure 3b shows the results for the main fractions of that sample obtained by ultracentrifugation. Increasing the temperature yields a substantial gain in intensity for all fractions. As expected, it is the LDL fraction that benefits most from this 12 K temperature increase by almost doubling its intensity. Therefore, a strong limitation of this manuscript is that its conclusions are based on 298 K data, acquired at likely suboptimal temperature. On the other hand it means that there is still room for improvement, making this approach potentially more promising, in particular considering most of the parameters that were not successfully predicted for the benchtop NMR model come from LDL fractions and subfractions (Figures 2b, S7 and Table S3).

Confirming the Assignments of Spectral Regions to Lipoprotein Fractions. The attractiveness of projection methods over neural networks and other supervised machine learning algorithms is the fact that loadings or in this case outer weights are readily interpretable. Figure 4 shows the contribution of specific regions of the spectra to a selected group of lipoprotein parameters (outer weights). For the sake of visualization, an arbitrarily chosen spectrum is depicted (solid black line). A color code is superimposed that ranges from negative (anticorrelated; blue; $r = -1$) to positive (correlated; red; $r = 1$) contribution, with no or low contribution depicted in white. The same color code is used for the response block. The first component in Figure 4 displays the first "latent" variable that is found to be correlated across all blocks. It is found to correspond to the total particles and VLDL parameters (lipoprotein outer weight). The strongest correlation between blocks is found in the CH outer component indicating that this block contributes the most to the prediction of this latent variable. The color code in the corresponding outer weight (CH block) indicates which spectral variables are driving this contribution. Visual inspection of the outer weights allows one to check whether selected variables make spectroscopic sense. For example, the middle part of the SPC peak and the HDL lipoprotein parameters are highlighted in blue (anticorrelated) in their respective outer weights (component 2 in Figure 4). Since the outer component between those two blocks displays a negative correlation, it means that high values in that spectral

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

region are associated with high values for the corresponding HDL parameters. Indeed, this region of the SPC peak has been associated with HDL subfractions 1 to 3.[37] As another example, the outer weight for the second component and third block (CH) highlights the contribution of cholesterol although no peaks are resolved in those regions.

CONCLUSIONS

While high field NMR spectroscopy has been, and will remain, the method of choice for the biological exploration of lipoproteins, this work demonstrates that benchtop NMR systems are ideal platforms for clinical translation. This collaborative study across 2 research institutes and a company laboratory evidenced the feasibility of translating NMR lipoprotein analysis from high field to benchtop systems, where high field associated limitations in terms of cost, maintenance and accessibility can be vastly reduced. Using a quantitative calibration approach similar to the one deployed for the 600 MHz IVDr harmonized procedure, stable and reproducible results were obtained from multiple sites. The ability to deploy harmonized measurements across multiple spectroscopic platforms and multiple sites is an obvious prerequisite for operating within the clinical realm. Despite expected challenges such as a decreased spectral dispersion and a lower sensitivity at lower field strengths, 25 out of 28 major parameters were successfully predicted, including the main parameters that are commonly used (LDL-C, HDL-C, total cholesterol, Apo-A1, Apo-B100, Apo-B100/Apo-A1) as cardiometabolic risk markers. Combined with recent advances suggesting that lipoprotein parameters can be accurately be measured from self-administered capillary blood, this work represents a significant milestone toward making a powerful diagnostic widely available. Accessibility is crucial for transitioning from population-derived risk scores to molecular phenotyping in the clinic, enhancing the potential for longitudinal personalized medicine.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c04660.

Additional information on cohort demographics (Table S1); full overview lipoprotein parameters (Table S2), regression results for benchtop NMR model (Table S3); additional experimental data on sample stability (Figure S1), and general benchtop system stability/performance (Figures S2−S5); variance of each block for the 80 MHz model (Figure S6) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Julien Wist − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia; Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, U.K.; Chemistry Department, Universidad del Valle, Cali 76001, Colombia;* orcid.org/0000-0002-3416-2572; Email: Julien.Wist@murdoch.edu.au

Jeremy K. Nicholson − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia; Institute of Global Health Innovation, Faculty of Medicine, Imperial College London, London SW7 2NA, U.K.;* orcid.org/0000-0002-8123-8349; Email: Jeremy.Nicholson@murdoch.edu.au

Authors

Philipp Nitschke − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia;* orcid.org/0000-0002-5814-7529

Ricardo Conde − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain*

Angela de Diego − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain*

Maider Bizkarguenaga − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain*

Samantha Lodge − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia;* orcid.org/0000-0001-9193-0462

Drew Hall − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia*

Zhonglin Chai − *Department of Diabetes, School of Translational Medicine, Monash University, Melbourne VIC3800, Australia*

Wenjun Wang − *Department of Diabetes, School of Translational Medicine, Monash University, Melbourne VIC3800, Australia*

Sudhir Kowlessur − *Health Promotion and Research Unit, Ministry of Health and Wellness, Port Louis 11321, Mauritius*

Marcio-Fernando Cobo − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Niels Pompe − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Birk Schütz − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Hartmut Schäfer − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Manfred Spraul − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Claire Cannet − *Bruker Biospin GmbH & Co. KG, Ettlingen 76275, Germany*

Tammo Diercks − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain;* orcid.org/0000-0002-5200-0905

Nieves Embade − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain;* orcid.org/0000-0001-9878-3290

Elaine Holmes − *Australian National Phenome Centre and Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth WA6150, Australia; Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, U.K.;* orcid.org/0000-0002-0556-8389

Oscar Millet − *Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Parque Tecnológico de Bizkaia, Derio 48160, Spain;* orcid.org/0000-0001-8748-4105

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c04660

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

Author Contributions

○J.W. and P.N. contributed equally to this paper.

Notes

The authors declare the following competing financial interest(s): Author M-FC, NP, BS, HS, MS, CC, are employees of Bruker Biospin, which manufactures the Fourier 80 benchtop NMR used in this study. While this affiliation could be perceived as a potential conflict of interest, the authors confirm that all experimental design, data collection, analysis, and interpretation were conducted objectively and independently to ensure the integrity and transparency of the research.

■ ACKNOWLEDGMENTS

We thank The Spinnaker Health Research Foundation, WA, The McCusker Charitable Foundation, WA, The Western Australian Department of Health, National Health and Medical Research Council and the Medical Research Future Fund (EPCD000037 and MRF2014349) for financial support. We thank the Department of Jobs, Tourism, Science and Innovation, Government of Western Australian Premier's Fellowship for funding E.H. We thank the Australian Research Council for Laureate Fellowship funding for E.H. We thank the Department of Industry of the Basque Country and the Agencia Estatal de Investigación (Spain) for grants Elkartek: KK-2023/00009 and PID2021-124171OB-I00. The authors from Bruker thank Prof. Dr. med. Matthias Nauck, University Medicine Greifswald, Germany, and PD Dr. Manfred Baumstark, Department of Medicine, Albert Ludwigs-University, Freiburg, Germany, for model samples, ultracentrifugation and related metadata and information used by Bruker in the later development process of the B.I.LISA method released in 2016. The authors are grateful to Prof. Paul Zimmet to facilitate this new collaboration and for insightful discussions and recommendations.

■ REFERENCES

(1) Kannel, W. B.; McGee, D. L. *JAMA* **1979**, *241* (19), 2035−2038.
(2) Chiesa, S. T.; Charakida, M. *Cardiovasc. Drugs Ther.* **2019**, *33* (2), 207−219.
(3) Liu, C.; Dhindsa, D.; Almuwaqqat, Z.; Sun, Y. V.; Quyyumi, A. A. *Am. J. Cardiol.* **2022**, *167*, 43−53.
(4) Walldius, G.; Jungner, I. *Curr. Opin. Cardiol.* **2007**, *22* (4), 359−367.
(5) Walldius, G.; Jungner, I. *J. Int. Med.* **2006**, *259* (5), 493−519.
(6) Bonacina, F.; Pirillo, A.; Catapano, A. L.; Norata, G. D. *Cells* **2021**, *10* (5), 1061.
(7) Urbina, E. M.; McCoy, C. E.; Gao, Z.; Khoury, P. R.; Shah, A. S.; Dolan, L. M.; Kimball, T. R. *J. Clin. Lipidol.* **2017**, *11* (4), 1023−1031.
(8) James, R. W.; Marini, M.; Ruiz, J.; Pometta, D. *Diabet. Med.* **1995**, *12* (1), 61−65.
(9) Dandona, P.; Aljada, A.; Chaudhuri, A.; Mohanty, P.; Garg, R. *Circulation* **2005**, *111* (11), 1448−1454.
(10) Kimhofer, T.; Lodge, S.; Whiley, L.; Gray, N.; Loo, R. L.; Lawler, N. G.; Nitschke, P.; Bong, S.-H.; Morrison, D. L.; Begum, S.; Richards, T.; Yeap, B. B.; Smith, C.; Smith, K. G. C.; Holmes, E.; Nicholson, J. K. *J. Proteome Res.* **2020**, *19* (11), 4442−4454.
(11) Lodge, S.; Nitschke, P.; Kimhofer, T.; Wist, J.; Bong, S.-H.; Loo, R. L.; Masuda, R.; Begum, S.; Richards, T.; Lindon, J. C.; Bermel, W.; Reinsperger, T.; Schaefer, H.; Spraul, M.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* **2021**, *93* (8), 3976−3986.
(12) Bruzzone, C.; Bizkarguenaga, M.; Gil-Redondo, R.; Diercks, T.; Arana, E.; García de Vicuña, A.; Seco, M.; Bosch, A.; Palazón, A.; San Juan, I.; Laín, A.; Gil-Martínez, J.; Bernardo-Seisdedos, G.; Fernández-Ramos, D.; Lopitz-Otsoa, F.; Embade, N.; Lu, S.; Mato, J. M.; Millet, O. *iScience* **2020**, *23* (10), 101645.
(13) Lawler, N. G.; Yonker, L. M.; Lodge, S.; Nitschke, P.; Leonard, M. M.; Gray, N.; Whiley, L.; Masuda, R.; Holmes, E.; Wist, J.; Fasano, A.; Nicholson, J. K. Children with Post COVID-19 Multisystem Inflammatory Syndrome Display Unique Pathophysiological Metabolic Phenotypes. 2024. https://papers.ssrn.com/abstract=4904694 (accessed Aug 22, 2024).
(14) Lodge, S.; Lawler, N. G.; Gray, N.; Masuda, R.; Nitschke, P.; Whiley, L.; Bong, S.-H.; Yeap, B. B.; Dwivedi, G.; Spraul, M.; Schaefer, H.; Gil-Redondo, R.; Embade, N.; Millet, O.; Holmes, E.; Wist, J.; Nicholson, J. K. *Int. J. Mol. Sci.* **2023**, *24* (14), 11614.
(15) Ruffieux, H.; Hanson, A. L.; Lodge, S.; Lawler, N. G.; Whiley, L.; Gray, N.; Nolan, T. H.; Bergamaschi, L.; Mescia, F.; Turner, L.; de Sa, A.; Pelly, V. S.; Kotagiri, P.; Kingston, N.; Bradley, J. R.; Holmes, E.; Wist, J.; Nicholson, J. K.; Lyons, P. A.; Smith, K. G. C.; Richardson, S.; Bantug, G. R.; Hess, C.; Cambridge Institute of Therapeutic Immunology and Infectious Disease-National Institute of Health Research (CITIID-NIHR) BioResource COVID-19 Collaboration. *Nat. Immunol.* **2023**, *24* (2), 349−358.
(16) Anderson, D. W.; Nichols, A. V.; Forte, T. M.; Lindgren, F. T. *Biochim. Biophys. Acta* **1977**, *493* (1), 55−68.
(17) Lindgren, F. T. J., Lin, C. H., Frederick, T. The Isolation and Quantitative Analysis of Serum Lipoproteins. https://escholarship.org/content/qt6q48593k/qt6q48593k.pdf (accessed June 28, 2024).
(18) Lindgren, F. T. Preparative Ultracentrifugal Laboratory Procedures and Suggestions for Lipoprotein Analysis. In *Analysis of Lipids and Lipoproteins*; American Oil Chemists' Society, 1975.
(19) Otvos, J. D.; Jeyarajah, E. J.; Bennett, D. W. *Clin. Chem.* **1991**, *37* (3), 377−386.
(20) Ala-Korpela, M.; Korhonen, A.; Keisala, J.; Hörkkö, S.; Korpi, P.; Ingman, L. P.; Jokisaari, J.; Savolainen, M. J.; Kesäniemi, Y. A. *J. Lipid Res.* **1994**, *35* (12), 2292−2304.
(21) Dona, A. C.; Jiménez, B.; Schäfer, H.; Humpfer, E.; Spraul, M.; Lewis, M. R.; Pearce, J. T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2014**, *86* (19), 9887−9894.
(22) Khakimov, B.; Hoefsloot, H. C. J.; Mobaraki, N.; Aru, V.; Kristensen, M.; Lind, M. V.; Holm, L.; Castro-Mejía, J. L.; Nielsen, D. S.; Jacobs, D. M.; Smilde, A. K.; Engelsen, S. B. *Anal. Chem.* **2022**, *94* (2), 628−636.
(23) Nicholson, J. K.; Holmes, E.; Kinross, J. M.; Darzi, A. W.; Takats, Z.; Lindon, J. C. *Nature* **2012**, *491* (7424), 384−392.
(24) Castaing-Cordier, T.; Bouillaud, D.; Farjon, J.; Giraudeau, P. Chapter Four—Recent Advances in Benchtop NMR Spectroscopy and Its Applications. In *Annual Reports on NMR Spectroscopy*; Webb, G. A., Ed.; Academic Press, 2021; Vol. *103*, pp 191−258.
(25) Izquierdo-Garcia, J. L.; Comella-Del-Barrio, P.; Campos-Olivas, R.; Villar-Hernández, R.; Prat-Aymerich, C.; De Souza-Galvão, M. L.; Jiménez-Fuentes, M. A.; Ruiz-Manzano, J.; Stojanovic, Z.; González, A.; Serra-Vidal, M.; García-García, E.; Muriel-Moreno, B.; Millet, J. P.; Molina-Pinargote, I.; Casas, X.; Santiago, J.; Sabriá, F.; Martos, C.; Herzmann, C.; Ruiz-Cabello, J.; Domínguez, J. *Sci. Rep.* **2020**, *10* (1), 22317.
(26) Percival, B. C.; Grootveld, M.; Gibson, M.; Osman, Y.; Molinari, M.; Jafari, F.; Sahota, T.; Martin, M.; Casanova, F.; Mather, M. L.; Edgar, M.; Masania, J.; Wilson, P. B. *High Throughput* **2019**, *8* (1), 2.
(27) Nitschke, P.; Lodge, S.; Hall, D.; Schaefer, H.; Spraul, M.; Embade, N.; Millet, O.; Holmes, E.; Wist, J.; Nicholson, J. K. *Analyst* **2022**, *147* (19), 4213−4221.
(28) Wider, G.; Dreier, L. *J. Am. Chem. Soc.* **2006**, *128* (8), 2571−2576.
(29) Baumstark, M. W.; Kreutz, W.; Berg, A.; Frey, I.; Keul, J. *Biochim. Biophys. Acta* **1990**, *1037* (1), 48−57.
(30) Wiemer, J.; Winkler, K.; Baumstark, M.; März, W.; Scherberich, J. E. *Nephrol. Dial. Transplant.* **2002**, *17* (12), 2231−2238.
(31) Monsonis-Centelles, S.; Hoefsloot, H. C. J.; Engelsen, S. B.; Smilde, A. K.; Lind, M. V. *Clin. Chem. Lab. Med.* **2019**, *58* (1), 103−115.
(32) Ordovas, J. M. *Methods Mol. Biol.* **1998**, *110*, 93−103.
(33) Tenenhaus, A.; Tenenhaus, M. *Psychometrika* **2011**, *76* (2), 257−284.

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409

(34) Tenenhaus, M.; Tenenhaus, A.; Groenen, P. J. F. *Psychometrika* **2017**, *82*, 737.
(35) Tenenhaus, A.; Philippe, C.; Guillemot, V.; Le Cao, K.-A.; Grill, J.; Frouin, V. *Biostatistics* **2014**, *15* (3), 569−583.
(36) Connelly, M. A.; Otvos, J. D.; Shalaurova, I.; Playford, M. P.; Mehta, N. N. *J. Transl. Med.* **2017**, *15* (1), 219.
(37) Masuda, R.; Lodge, S.; Whiley, L.; Gray, N.; Lawler, N.; Nitschke, P.; Bong, S.-H.; Kimhofer, T.; Loo, R. L.; Boughton, B.; Zeng, A. X.; Hall, D.; Schaefer, H.; Spraul, M.; Dwivedi, G.; Yeap, B. B.; Diercks, T.; Bernardo-Seisdedos, G.; Mato, J. M.; Lindon, J. C.; Holmes, E.; Millet, O.; Wist, J.; Nicholson, J. K. *Anal. Chem.* **2022**, *94* (10), 4426−4436.
(38) Girka, F.; Camenen, E.; Peltier, C.; Gloaguen, A.; Guillemot, V.; Le Brusquet, L.; Tenenhaus, A. *Regularized and Sparse Generalized Canonical Correlation Analysis for Multiblock Data*. [R package RGCCA version 3.0.3]. https://CRAN.R-project.org/package=RGCCA (accessed June 04, 2024).
(39) Eskildsen, C. E.; Næs, T.; Skou, P. B.; Solberg, L. E.; Dankel, K. R.; Basmoen, S. A.; Wold, J. P.; Horn, S. S.; Hillestad, B.; Poulsen, N. A.; Christensen, M.; Pieper, T.; Afseth, N. K.; Engelsen, S. B. *Chemometr. Intell. Lab. Syst.* **2021**, *213*, 104311.
(40) Horwitz, W.; Kamps, L. R.; Boyer, K. W. *J. Assoc. Off. Anal. Chem.* **1980**, *63* (6), 1344−1354.
(41) Wold, S.; Martens, H.; Wold, H. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In *Proceedings of the Conference on Matrix Pencils*, 1983, pp 286−293..
(42) Tibshirani, R. *J. Roy. Stat. Soc. B Stat. Methodol.* **1996**, *58* (1), 267−288.
(43) Hoerl, A. E.; Kennard, R. W. *Technometrics* **1970**, *12* (1), 55.
(44) Hoerl, A. E.; Kennard, R. W. *Technometrics* **1970**, *12* (1), 69.
(45) Schapire, R. E. *Mach. Learn.* **1990**, *5* (2), 197−227.
(46) Friedman, J. H. *Ann. Stat.* **2001**, *29* (5), 1189−1232.
(47) Nitschke, P.; Lodge, S.; Kimhofer, T.; Masuda, R.; Bong, S.-H.; Hall, D.; Schäfer, H.; Spraul, M.; Pompe, N.; Diercks, T.; Bernardo-Seisdedos, G.; Mato, J. M.; Millet, O.; Susic, D.; Henry, A.; El-Omar, E. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K.; Wist, J. J. *Anal. Chem.* **2022**, *94* (2), 1333−1341.
(48) Baumstark, D.; Kremer, W.; Boettcher, A.; Schreier, C.; Sander, P.; Schmitz, G.; Kirchhoefer, R.; Huber, F.; Kalbitzer, H. R. *J. Lipid Res.* **2019**, *60* (9), 1516−1534.

https://doi.org/10.1021/acs.analchem.4c04660
Anal. Chem. 2025, 97, 6399−6409