 doi:10.20944/preprints201810.0089.v1

Research Article

Low-Field, Benchtop NMR Spectroscopy as a Potential Tool for Point-of-Care Diagnostics of Metabolic Conditions: Validation, Protocols and Computational Models

Benita C. Percival[1], Yasan Osman[1], Martin Grootveld[1], Miles Gibson[1], Marco Molinari[2], Fereshteh Jafari[1], Mark Martin[3], Federico Casanova [4], Melissa L. Mather [5], Mark Edgar[6] and Philippe B. Wilson[1,*]

[1] Leicester School of Pharmacy, De Montfort University, The Gateway, Leicester, LE1 9BH, UK
[2] University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
[3] Greater Manchester NHS Trust, Stepping Hill Hospital, Poplar Grove, Hazel Grove, Stockport, SK2 7JE, UK
[4] Magritek GmbH, Philipsstraße 8, 52068, Aachen, Germany
[5] University of Nottingham, University Park, Nottingham, NG7 2RD
[6] University of Loughborough, Epinal Way, Loughborough, LE11 3TU, UK
* Correspondence: philippe.wilson@dmu.ac.uk; Tel.: +441162577631

Abstract: Novel sensing technologies for liquid biopsies offer a promising prospect for the early detection of metabolic conditions through -omics techniques. Indeed, high-field NMR facilities are routinely used for metabolomics investigations on a range of biofluids in order to rapidly recognize unusual metabolic patterns in patients suffering from a range of diseases. However, these techniques are restricted by the prohibitively large size and cost of such facilities, suggesting a possible role for smaller, low-field NMR instruments in biofluid analysis. Herein we describe selected biomolecule validation on a low-field benchtop NMR spectrometer (60 MHz), and present an associated protocol for the analysis of biofluids on compact NMR instruments. We successfully detect common markers of diabetic control at low-to-medium concentrations through optimized experiments, including glucose (≤ 2.6 mmol./L) and acetone (25 μmol./L), and additionally in readily-accessible biofluids. We present a combined protocol for the analysis of these biofluids with low-field NMR spectrometers for metabolomics, and offer a perspective on the future of this technique appealing to point-of-care applications.

Keywords: Metabolomics, Benchtop NMR, Biomarkers, Biomolecules, Validation, Protocol, Diabetes

1.0 Introduction

Although now recognized as a powerful tool in translational medicine, the principles of metabolomics were arguably first described by ancient Chinese scholars, who used ants to evaluate the glucose level in the urine of diabetic patients [1]. The ancient Egyptian and Greek societies (circa. 300 BC) developed this further to detect differences in the taste of urine as a means of disease diagnosis [2]. Etymologically-derived from the Greek language words for *change*, and *body* or *rule*, metabolomics and metabonomics respectively involve the measurement of metabolic responses to perturbation; metabolomics is centered on measurements of the entire metabolome, whereas metabonomics concentrates on longitudinal changes across the metabolome ascribable to interventional stimuli [3].

 doi:10.20944/preprints201810.0089.v1

The early metabolomics revolution effectively began with *De Statica Medicina,* published in 1614 by Santorio Santorio, who described his quantitative approach to modern medicine and the first systematic study of basal metabolism [4]. Both mass-spectrometric (MS) and nuclear magnetic resonance (NMR) strategies, two of the most common modern analytical techniques for metabolic measurements, were first described in the early part of the 20th century; MS by J. J. Thomson and F Aston in 1913 [5], and NMR by Bloch and Purcell in 1946 [6]. Although not described as such, the first analytical measurement of metabolites was published by Pauling in 1971 [7], before the pioneering work of Nicholson and Sadler in the early 1980s [8,9], and Willmitzer in 1987 [10].

Metabolic profiling now includes the measurement of a range of biofluids, including blood plasma and serum, saliva, urine, knee-joint synovial fluid, semen and cerebrospinal fluid [11]. Indeed, biological media will vary in suitability for each disease/condition investigated, with a wide range and high volume of metabolic data being extracted from each [12]. For example, common high-field (HF) NMR metabolomics studies performed at operating frequencies of 600 MHz and above can detect more than 150 metabolites in human urine, and quantify almost 70 metabolites in human blood serum [3]. Moreover, with these practices becoming commonplace, tools and repositories such as the Human Metabolome database (HMDB) [13] and Metaboanalyst [14] have been developed to serve the –omics community.

With the rapidly expanding functional improvements in both NMR and MS techniques, additional avenues for the incorporation of metabolite profiling in medicine have arisen [15], based on miniaturization of these technologies [16], biomarker discovery [17], and prognostic monitoring [18]. Although MS and NMR are both considered standard techniques for metabolomics and metabonomics studies [19], this investigation will focus on the latter, and more specifically, low-field (LF) compact, or mobile, benchtop NMR facilities. In general, these instruments operate at frequencies below 100 MHz, and are based on permanent magnets as opposed to the large, HF superconducting magnets commonly found in analytical characterization suites [20]. These facilities operating at LF suffer from the same issues of sensitivity which plagued early designs of NMR spectrometers [21]; however, with the employment of approaches such as solvent suppression [22], and magnet arrays [23], it is possible to simultaneously observe and monitor 20 or more metabolites in saliva and urine at LF, as demonstrated here for the first time. Although this limits the detection capabilities of NMR when compared to those of larger, high-field instruments, the advantages of more compact, mobile NMR instruments incorporating applications of chemometric/metabolomics approaches to the multivariate analysis of complex mixtures, have been demonstrated in fields such as materials science, [24] forensic chemistry, [25] chemical education [26] and biomedical sciences.

Herein, we present an updated protocol for the analysis of biofluids through compact, benchtop NMR measurements. Comprehensive protocols for metabolite measurement and profiling are currently available, and these rely on HF NMR or LC/MS techniques, such as that described in the pioneering works of Nicholson *et al* [27]. We describe a complete procedural development for the analysis of biofluids by LF NMR analysis, including validation and quantification experiments, experimental guidelines, and metabolomics data analysis, and these are experimentally demonstrated by an appropriate example based in the area of diabetes.

 doi:10.20944/preprints201810.0089.v1

2.0 Materials and Reagents

All materials were purchased from Sigma-Aldrich Ltd. (UK) unless otherwise stated. 5-mm Diameter NMR tubes were purchased from Norell. Sodium phosphate monobasic (99%) for analysis (anhydrous) and sodium phosphate, dibasic, heptahydrate 99%+ for analysis were purchased from Acros Organics, Fisher Scientific, (UK). Eppendorf micropipettes and tips were purchased from Eppendorf (UK), and sterile universal containers were purchased from Starlab Ltd. (UK).

3.0 Equipment

60 MHz Magritek Spinsolve Benchtop NMR spectrometer (similar instruments may be employed: Pulsar from Oxford Instruments, NMReady from Nanalysis, and PicoSpin from ThermoFisher Scientific). A Bruker 400 MHz Avance-I NMr spectrometer fitted with a Quadruple nucleus probe (A similar spectrometer can be utilised: JEOL) / Centrifuge / Rotamixer.

3.1 Sample acquisition

Samples were inserted into the instrument manually with an approximate acquisition period of 10 min. However, with recent advancements, automation is now a possibility within benchtop instrumentation through the use of a robotic arm autosampler. This allows for a larger number of samples to be prepared, and a reduction in commonplace onerous hands-on user periods between sampling.

Samples were acquired on a 60 MHz Magritek Spinsolve Ultra Benchtop system (Leicester School of Pharmacy, De Montfort University, Leicester UK). The instrument gives the option to shim to sample or shim to a standard 10%/90% D_2O/H_2O solution. Samples should ideally be acquired at a constant ambient temperature, i.e. between 18-30°C, and a magnet temperature of 18-30°C. Shimming to sample optimises resolution within a single minute, and therefore is recommended, but excellent results are achieved with the default shim settings. Both urine and saliva samples are dominated by a large water signal at ~4.8 ppm in ^{1}H NMR spectra; these require a preliminary 1D proton spectrum to be acquired (~10 seconds), in order to identify the exact position of this resonance. This intense water signal must be sufficiently suppressed in order to explore the dynamic range of metabolites, therefore identifying the signal (δ = ~4.80 ppm) and inputting the exact resonance into the water suppression sequence optimises signal-to-noise and significantly reduces signal overlap. Appropriate repetition times between scans are determined using a T_1 experiment on the sample in order for the optimum relaxation time to be determined ($5*T_1$). Signal-to-noise ratios can be improved by increasing the number of scans, scaling with a factor to $\sqrt{2}$, i.e. quadrupling the number of scans will result in a doubling of the signal-to-noise ratio. Spectra were acquired using a 1D PRESAT sequence to allow for efficient saturation of the water signal, without perturbing the remaining signals in the spectrum. The parameters used for these analyses are as follow: 64 scans, an acquisition time of 6.4 seconds, a repetition time of 10 seconds and a pulse angle of 90°. CPMG pulse sequences

 doi:10.20944/preprints201810.0089.v1

are recommended for plasma samples. The option is provided for these pulse sequences to be scripted and therefore full automation is possible.

High-field spectra of urine samples were also acquired using a Bruker Avance-I 400 MHz NMR (Leicester School of Pharmacy, De Montfort University, Leicester, UK) spectrometer operating at a frequency of 399.93MHz. The samples were analysed using the noesygppr1d pulse sequence in order to suppress the water signal (δ = ~4.80 ppm) in the urinary samples with irradiation at the water frequency during the recycle and mixing time delays. The free induction decay (FID) was acquired with 32K data points using 128 scans and 2 dummy scans, and 3 μs ^{1}H pulses, over a sweep width of 4,844 Hz (12.1 ppm) and a receiver gain automatically adjusted to each sample.

4.0 60 MHz Spectrometer Biomarker Validation

4.1 Calibration

Specific calibration samples of biomolecules relating to diabetes chemopathology were prepared at various concentrations including 1.00-400.00 mmol./L (and 1.00 mol./L), in HPLC-Grade water. For selected metabolites, concentration ranges of 15-800 μM were typically employed. Samples contained 50 μl of 0.05% (w/v) sodium azide, 50 μl of D_2O containing 0.05% (w/v) sodium 3-(trimethylsilyl)[2,2,3,3-d_4] propionate (TSP), 50 μl of 1.00 M phosphate buffer, and 500 μl of analyte. This mixture was then rotamixed and added to new 5-mm NMR tubes ready for analysis. $^{2}H_2O$ containing TSP was used in this case, since sample spectra were acquired on both the 400 and 60 MHz instruments. For the benchtop instrument employed here, $^{2}H_2O$ is not required as a field-frequency lock solvent. However, as we sought to compare the datasets acquired from both 60 and 400 MHz instruments, $^{2}H_2O$ was added to facilitate this. TSP is an appropriate internal standard for quantification, (∂ = 0.00 ppm); therefore it does not interfere with other signals. Sodium azide is added to suppress bacterial growth during periods of sample preparation and storage. Phosphate buffer is used to maintain a constant pH value in order to avoid pH-mediated signal shifts: pH 7.00 or 7.40 is recommended [28]. The purpose of the calibration curve is for analytical quantification, and also to determine the limit of detection (LOD); this represents a resonance intensity which is 3 times the background noise value, and the limit of quantification (LOQ), which is derived from a ratio of 10 times this noise level.

5.0 60 MHz Benchtop Spectrometer Protocol: Biofluid Analysis Recommendations

[1] Experimental Design and Research Ethics Approval

Prior to sample collection, ethical factors and experimental design should be carefully considered. All samples in our example study were collected with informed consent and approved by the appropriate Research Ethics Committee, specifically the Faculty of Health and Life Sciences Research Ethics Committee, De Montfort University, Leicester, UK (reference no. 1936). All participants were primarily provided with a participant information sheets (PISs), and were then required to sign a project consent form in the presence of a researcher witness. The PIS clearly informed those recruited

 doi:10.20944/preprints201810.0089.v1

that since their participation was voluntary, they had the freedom to withdraw from the investigation at any stage of its progress. Essentially, all ethics considerations were in accord with those of the Declaration of Helsinki of 1975 (revised in 1983).

A pre-established experimental design is crucial prior to starting sample collection, since factors such as age, sex, BMI, fasting, exercise, stress, drug intake and supplements all affect the metabolome. Samples can be collected over a period of time to monitor the pharmacokinetics and metabolism of drugs, and samples may be collected randomly, or under pre-fasted or non-fasted conditions. Essentially, the fewer lateral variables available, the more precise the study of the metabolome becomes for that particular disease manifestation. Validation of specific markers includes the performance of larger trials, and/or repetition of the trial by a different laboratory.

Contradictions in metabolomics experiments may commonly arise from experimental error, caused primarily from the high turnover rate of metabolites in terms of stability, solubility and volatility. Therefore, experimental design is critical for preparation of the experiment itself. Repeat freeze/thaw cycles should be avoided for all sample types [15], and factors such as maintaining samples at room temperature for a long period of time should be kept to a minimum. Buffering procedures should be kept uniform. It is also important to reduce the number of experimental steps, limit sample handling, process samples rapidly, maintain samples at a cold or frozen state prior to analysis, and analyse as soon as possible thereafter.

[2] Sample Collection and Sample Storage

Sample preparation can be automated or manually performed. Automation of samples involves a barcoding system which ensures participant anonymity in such studies. Robotic processing of the samples is also possible, allowing for samples to be prepared and analysed in bulk quantities. This increases productivity, but more importantly displays applicability and viability in a point-of-care setting. It is important to define and follow experimental design and Standard Operational Procedures (SOPs) in order to avoid erroneous results.

Urine

Participants should fast for a 12-hour period prior to providing a sample, since diet is known to affect the urinary metabolome, for example a higher intake of fruit leads to elevations in rhamnitol, 4-hydroxyhippurate, tartarate, hippurate and glycolate in urine [29]. Urine samples should be collected in sterile, plastic universal containers. Urine specimens should be transported to the laboratory on ice, and then centrifuged immediately (3,500 rpm at 4°C for 15 min.). The supernatants are then stored at -80°C prior to analysis, although temperatures below -25°C are usually adequate [30].

Saliva

Volunteers should fast for a 12-hour period prior to providing a sample, and it is preferable that participants provide samples immediately after awakening in the morning. Participants should

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

refrain from activities such as smoking, eating, drinking, brushing teeth *etc.* during the period between waking and sampling, in order to avoid analytical interference from external activities [31,32]. Saliva samples should be collected in sterile, plastic, universal containers. Specimens should be transported to the laboratory on ice and then centrifuged immediately (3,500 rpm for 15 min.) to remove cells and debris. The supernatants are then stored at -80°C prior to analysis.

Blood Plasma/Serum

Blood should be collected following a 12-hour fasting period. Blood samples are collected by a fully trained phlebotomist via venepuncture. For plasma, lithium heparin tubes should be used for collection to avoid analytical complications arising from the use of tubes containing EDTA or citrate anticoagulants. EDTA or citrate present in collection tubes will not only give rise to interfering resonances themselves, but will also chelate metal ions, i.e. Mg^{2+} and Ca^{2+}, which themselves generate interfering signals in the ^{1}H NMR spectra acquired; those of the Ca^{2+}-EDTA and Mg^{2+}-EDTA complexes, which are distinct from those of EDTA itself in view of a slow exchange of this chelator on the NMR timescale. However, often it is preferable to allow the blood to clot and isolate serum samples therefrom via a simple centrifugation step. Samples should be immediately centrifuged at 4°C, 4,300 rpm for 15 min. Serum and plasma samples are then stored at -80°C prior to analysis.

[3] Sample Preparation

Biofluid samples are thawed at ambient temperature and then immediately prepared. Preparation for the test set of samples described herein involved centrifuging 500 µl volumes of sample (plasma, serum, urine or saliva) and removing 450 µl supernatant for analysis. A 50 µl aliquot of phosphate buffer at pH 7.00 (1.00 mol./L) was added to the supernatant, which contained 0.05% (w/v) sodium azide (prepared in HPLC-grade water), and then 50 µl of $^{2}H_2O$, also containing 0.05% (w/v) TSP (Sigma-Aldrich, UK), was then added to the solution, so that the final v/v content of this deuterated solvent was 10%. The final added TSP concentration for these mixtures was therefore 264 µmol./L. Exceptionally for plasma and serum samples, and also other high protein content biofluids such as knee-joint synovial fluid, the TSP singlet resonance is substantially broadened in view of its binding to proteins therein, and hence is best avoided (although it may be included as a standard solution placed in a capillary insert within an NMR tube). Therefore, only $^{2}H_2O$ was added in this case. This mixture was then rotamixed and transferred to newly-purchased NMR tubes ready for analysis.

[4] Sample acquisition

Please refer to recommended acquisition parameters in section **4.2**.

[5] Preprocessing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

A variety of software modules can be used for the essential preprocessing of bioanalytical ^{1}H NMR data prior to multivariate metabolomics analysis: *Mestranova*, *JEOL Delta*, *Bruker Topspin 4.0* and *ACD Labs 12.0*. Free induction decays (FIDs) acquired by such instrumentation can be automatically preprocessed, including a correction of linewidth (apodisation), Fourier transformation, phase correction, baseline correction and data alignment. Fourier transformation is a mathematical process which converts the Free Induction Decay (FID) time functions into frequency ones, which yield the commonly observed format of NMR spectra where, depending on operating frequency, each signal is resolved in view of their differing chemical shift values. This allows for a comparison between a 60 and 400 MHz spectra. In this format, further preprocessing strategies can be applied to optimise the quality of acquired spectra.

Preferably, signals should be as symmetrical as possible, with a consistent baseline of ~ 0, and have clear, defined and narrow line shapes. Baseline and phase corrections can be manipulated manually for more precise corrections, or automatically using predefined coefficient parameters. Baseline correction allows for manipulation of the spectral baseline, adjusting the entire spectrum to start at a set value of 0. Without baseline corrections, data may show a non-uniform signal area when integrated, which is attributable to fluctuations in the baseline and does not arise from the concentration of particular metabolites present in samples investigated. Phase correction allows for the manipulation of the spectrum to reduce the effect of phase shifting. A poorly adjusted phase correction can result in the absorption signal dipping beneath the spectral baseline, which can lead to the erroneous integration of negative resonance areas. Using a phase correction, the signal region can be manipulated from its negative absorption into the desired pure absorption signal. Apodisation involves multiplying the free induction decay pointwise by a defined function in order to improve the line shape within the spectrum. A Gaussian function is commonly used in view of its ability to improve the resolution of signals by narrowing their linewidths; however this also increases the spectral noise intensity. On occasion, it is prudent to apply a Lorentzian function to improve the signal-to-noise ratio, at a cost of broadening the signals and hence reducing resolution and sine bell functions. The weakness of apodisation occurs when applying a correctional function, causing the resonance to broaden; as such, deformation of signals may occur, potentially leading to the misidentification of multiplets.

[6] Metabolite assignment

TSP acts as a reference signal (∂ = 0.00 ppm) for aqueous biofluid analyte samples in order to ensure that metabolites can be chemical shift-aligned accordingly, although alternative or study-specific NMR reference agents, which may also serve as quantitative internal standards, may also be employed. Databases for reference samples include the *Human Metabolome Database* (HMDB) [13], *Madison metabolomics consortium database* (MMCD) [33], and *COLMAR metabolomics web server* [34], which aid metabolite assignments. Further experiments for structural elucidation include the use of two-dimensional (2D) techniques such as COSY and TOCSY strategies, both of which are available on LF, benchtop instruments. Essentially, 2D techniques can show ^{1}H-^{1}H and where appropriate, ^{1}H-^{13}C correlations using a grid style map so that investigators may readily determine intramolecular

 doi:10.20944/preprints201810.0089.v1

connectivities between nuclei of interest therein. Limitations of databases for authentic biomolecule reference compounds include metabolites predominantly determined at 500-600 MHz operating frequencies, which could lead to problems interpreting spectral data at LF. However, the use of NMR-SIM and Guided Ideographic Spin System Model Optimisation (GISSMO) software strategies enables structural calculations, *i.e.* simulations of spins at different magnetic fields, and can also afford valuable information on both individual molecules and relatively complex mixtures [35,36]. Thus, a combination of databases and simulations are the most appropriate approach for structural elucidation strategies using LF instrumentation.

[7] Integration and data manipulation

Manual bucketing requires the user to observe signals and bucket accordingly; intelligent bucketing is automated and buckets integral datasets according to programming software and the parameters provided, using an algorithm to integrate selected bins. Manual bucketing was performed in the experiments used as examples herein. Since ^{1}H NMR shifts are critically dependent on temperature, pH and ionic strength, the bucketing step must be carefully performed. Hence, manual integration ensures that only one metabolite signal is incorporated into its corresponding bucket, although there are obviously limitations with may be achieved from this with LF benchtop NMR profiles. Uniform binning was not performed, since this can lead to issues such as signal splitting between and the inclusion of > 1 signal within single buckets.

The buckets were then imported into MS Excel in order to create a ^{1}H NMR data matrix, and data were normalised by constant sum accordingly. Any buckets attributed to baseline noise, water or other interfering agents were eliminated prior to analysis. Normalisation was also performed by the expression of resonance intensities relative to that of the internal reference (TSP), in order to calculate urinary metabolite concentrations. However, for diabetic urine samples, it was not possible to normalise biomolecule concentrations to those of urinary creatinine at a LF benchtop spectrometer operating frequency of 60 MHz, since the high glucose levels therein significantly overlapped with creatinine's characteristic >N-C$\underline{H}_3$ and -C$\underline{H}_2$- function resonances, unlike the ^{1}H NMR profiles obtained on 400 MHz facilities.

Data filtering was performed in order to primarily remove variables which are unlikely to be of value to the modelling of ^{1}H NMR datasets.

[8] Reproducibility

To ensure results are reproducible and reliable, within-assay run and between-assay run precisions should be assessed. This is usually performed via the computation of their corresponding components of variance in a pre-designed random effects or mixed model analysis of variance/covariance (ANOVA/ANCOVA) experimental design. Indeed, this ensures that the metabolites being monitored are consistently reported, and avoids any batch-related errors.

 doi:10.20944/preprints201810.0089.v1

[9] Univariate and Multivariate Statistical Analyses

Statistical analysis can be performed using *XLSTAT*, *Metaboanalyst 4.0*, *ROCCET*, *MetATT*, *Statistica*, *Python* and *R* software packages, amongst others. Data can be analysed in a univariate or multivariate manner, depending on whether one or more metabolites are targeted. Univariately, each metabolite can be compared using standard deviation, box-and-whisker plots, Student's t-tests and ANOVA or ANCOVA to assess any differences found. The 'step-down' Holm model of the Bonferroni correction, or corrections for false discovery rate (FDR) for the univariate analysis of multidimensional datasets should be adhered to. Multivariate data, which is acquired from ^{1}H NMR metabolomic studies, and considers many metabolites simultaneously, can be analysed using statistical techniques such as Principal Component Analysis (PCA), Partial Least Squares-Discriminant Analysis (PLS-DA), Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA), Self-Organising Maps (SOMs) and/or Correlated Component Regression (CCR), for example.

PCA is an unsupervised technique which helps to detect any statistical 'outliers' in datasets, and following their removal, PC clusterings can be explored for each classification group, if indeed there are any distinguishing metabolic features present. PLS-DA, however, is a supervised technique which is able to seek and evaluate any significant distinctions between datasets using permutation and further testings. PLS-DA is able to distinguish between metabolites causing the most separation between the datasets, and these are statistically validated and cross-validated in order to confirm that the testing performed is sound. Moreover, the CCR technique has the ability to generate reliable metabolomics predictions from datasets in which the number of correlated explanatory metabolite variables (P) is greater or much greater than the sample size (n), irrespective of any multicollinearities between these predictor variables.
Validation of these statistical techniques can be performed using the leave-one-out cross-validation (LOOCV) approach, the Q^2 statistic and area under the receiver operating characteristic (AUROC) curve. AUROC describes the validity of a model based on sensitivity and specificity, i.e. the extent of objects correctly identified within a model, and that of samples correctly classified as 'foreign', respectively. A plot of sensitivity *vs.* (1-specificity), i.e. true positives *vs.* false positives, is described as a ROC curve, which extends to a multidimensional hypersurface should there be multiple classes considered. The area under the ROC curve (or AUROC value) acts as a measure of class separation, where a value of unity corresponds to optimal separation between two classes of classification, whilst a value of 0.50 indicates no separation whatsoever [37].

[10] Computational intelligence (Neural Networks)

Computational intelligence techniques (CITs), including support vector machines (SVMs), optimization algorithms (genetic algorithms, Ants) and machine-learning algorithms (supervised and unsupervised neural networks), can be employed for the analysis of multivariate (MV) ^{1}H NMR datasets acquired in order to recognize metabolic classification patterns for diseases, as well as

 doi:10.20944/preprints201810.0089.v1

disease severity indices for each of several biofluids investigated, *i.e.* as non-probabilistic binary classifiers. Optimization algorithms can be used to select the most relevant (discriminatory) biomarker variables from the datasets for use in constructing a machine-learning classification model. Random forest (RF) techniques can also be employed for classification and variable selection, in which data is divided into training and test sets. The performance is assessed using an out-of-the-bag (OOB) error value, and the accuracy, specificity and sensitivity can be obtained from the test set. Self-organising maps (SOMs) can explore self-similarities between spectra and hence clusterings that arise from each source of variation involved. The RF model can be repeated multiple times in order to prevent bias arising from the random sub-sampling of the training and test sets. Unsupervised learning methods can also be employed in order to perform data-mining and hence uncover any 'hidden' patterns.

[11] Pathway Identification

The *Kyoto Encyclopedia of Gene and Genomics* (KEGG), *MetaCyc* and *HMDB* each aim to aid in ascribing significant biomolecular modifications identified to imbalances of or perturbations to established metabolic pathways, and can also identify any linkages between cycles. Upregulation and downregulation of cycles can be analysed using these platforms which enable areas of the body/cell/organism to be assessed metabolically. *Metaboanalyst* also interfaces with these databases. Moreover, the connectivities of biofluid metabolites that potentially distinguish between diseases and corresponding healthy controls, and also disease severity classification groups, can, in principle, be explored through the Metabolomics software module of *Ingenuity Pathways Analysis* (IPA, Ingenuity Systems, www.ingenuity.com). Indeed, canonical pathways analysis can be employed to identify those that are the most significantly different between the above disease or severity group classifications.

Case Study: Urinary Profiles of Fasted Diabetic Patients with Elevated Glucose levels

As an example of the above protocol, a case study was performed monitoring glucose levels in type 2 diabetic urine samples, together with additional metabolites which parallel such a marked upregulation in patients with this conditions. Samples were collected and acquired using the above protocols and parameters.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

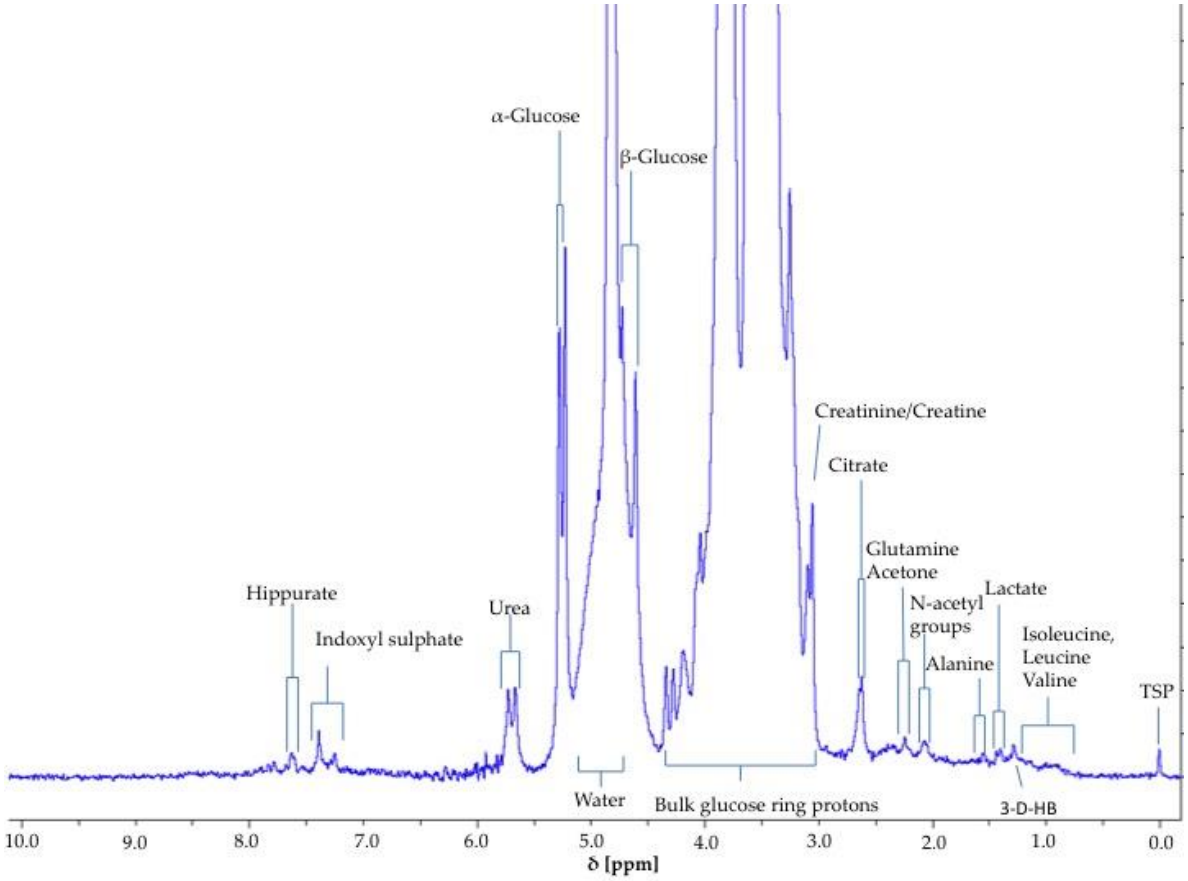


Figure 1: ^{1}H NMR Diabetic urinary profile acquired on a 60 MHz benchtop instrument ,highlighting a clearly distinguishable β-Glucose-CH resonance (*d*, ∂ = 4.63 ppm), in addition to the *α*-Glucose-CH one located at ∂ = 5.22 ppm (*d*), and all further bulk glucose ring structure protons within the 3.20-3.95 ppm chemical shift range for both anomers. Moreover, resonances arsing from a range of further metabolites such as hippurate-CH, indoxyl sulphate-CH, urea-NH_2, creatinine-CH_3/-CH_2, creatine-CH_3/-CH_2, citrate-$CH_{2A/B}$, glutamine-CH_2, acetoin-CH_3, acetate-CH_3, lactate-CH_3, 3-aminoisobutyrate-CH_3, alanine-CH_3,, isoleucine-CH_3 and leucine-CH_3 are also visible in this spectrum. Chemical shifts were referenced to internal TSP (∂ = 0.00 ppm). Abbreviations: 3-D-HB, 3-D-hydroxybutyrate.

The *α*-glucose anomer C1-H signal at ∂ = 5.23 ppm (*d*) was detected and integrated, and its intensity expressed relative to that of TSP (*s*, δ = 0.00 ppm). This particular resonance was employed for quantification purposes since urinary ^{1}H NMR profiles contain many overlapping signals within the crowded 3.00-4.00 ppm range where its C2-H to C6-H proton resonances are located, and in view of their poor resolution at 60 MHz, this was selected as the optimal resonance to monitor in diabetic and prospectively diabetic patients. Moreover, β-glucose's C1-H resonance (δ = 4.63 ppm) is too

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

closely overlapped with the residual water signal for quantification purposes (Figure 1). However, it should be noted that glucose's α-anomer only represents 36% of the total glucose concentration present (the remaining 64% being the β-anomer), and therefore the factor 100/36% (= 2.78) should be employed for converting α-anomer concentrations to total glucose ones. The limit of quantification for glucose was (3.60 mmol/L), and the limit of detection of glucose was (2.88 mmol/L). The calibration curve exhibited a clear linear relationship (r = 0.9911). The limitations of monitoring this α-glucose resonance using LF benchtop NMR analysis is ascribable to potential interference of the closely-located residual water signal. However, we successfully detected and created calibrations for other biomolecules at lower concentrations through optimized experiments, such as acetone, which we found had a LOQ value of ≤ 25 μmol./L using a 60 MHz benchtop facility.

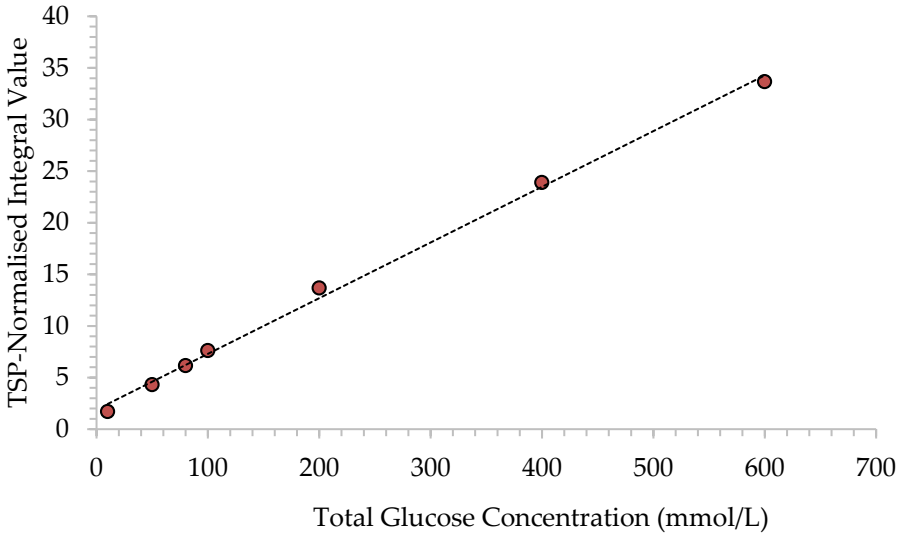


Figure 2: Calibration curve plot of the α-Glucose:TSP resonance integral ratio *vs.* total glucose concentration in phosphate-buffered aqueous solutions (pH 7.00) containing 10% (v/v) ^2H_2O. Glucose concentrations ranged from 10.0-600.0 mmol./L, and that of the TSP internal standard was maintained at 8% (v/v) ^2H_2O TSP concentration 0.05% w/v (223 μmol./L final concentration).

The α-glucose signal is easily identified in the ^{1}H NMR spectral profiles of urine samples collected from non-rigorously controlled type 2 diabetic patients. However, there is a small level of overlap between the water signal and the α-glucose signal, a phenomenon complicating integration and hence quantification of this key biomarker at concentrations < 8.00 mmol./L. Optimisation was attempted by moving the driver signal to ∂ = 4.80 ppm to ensure the best clarity between these two signals; however, there was still a residual level of overlap. Despite these problems, we found that integration of glucose's α-anomeric proton resonance (δ = 5.26 ppm) was affected negligibly if the urinary concentration of this anomer was ≥ 2.8 mmol./L (corresponding to a total glucose level of ≥ *ca.* 8 mmol./L). To date, urinary profiles have not previously been acquired on LF, benchtop NMR systems for metabolomics analysis such as in this example. Indeed, additional ^{1}H NMR signals in addition to those of glucose and assignable in LF 60 MHz spectra may also be employed for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

metabolomics analysis, notably ketone bodies which arise from the use of lipid sources as a fuel in patients with poorly controlled diabetes. Such a multivariate metabolomics analysis of our LF 60 MHz ^{1}H NMR dataset was therefore performed, and results arising thereform are outlined below.

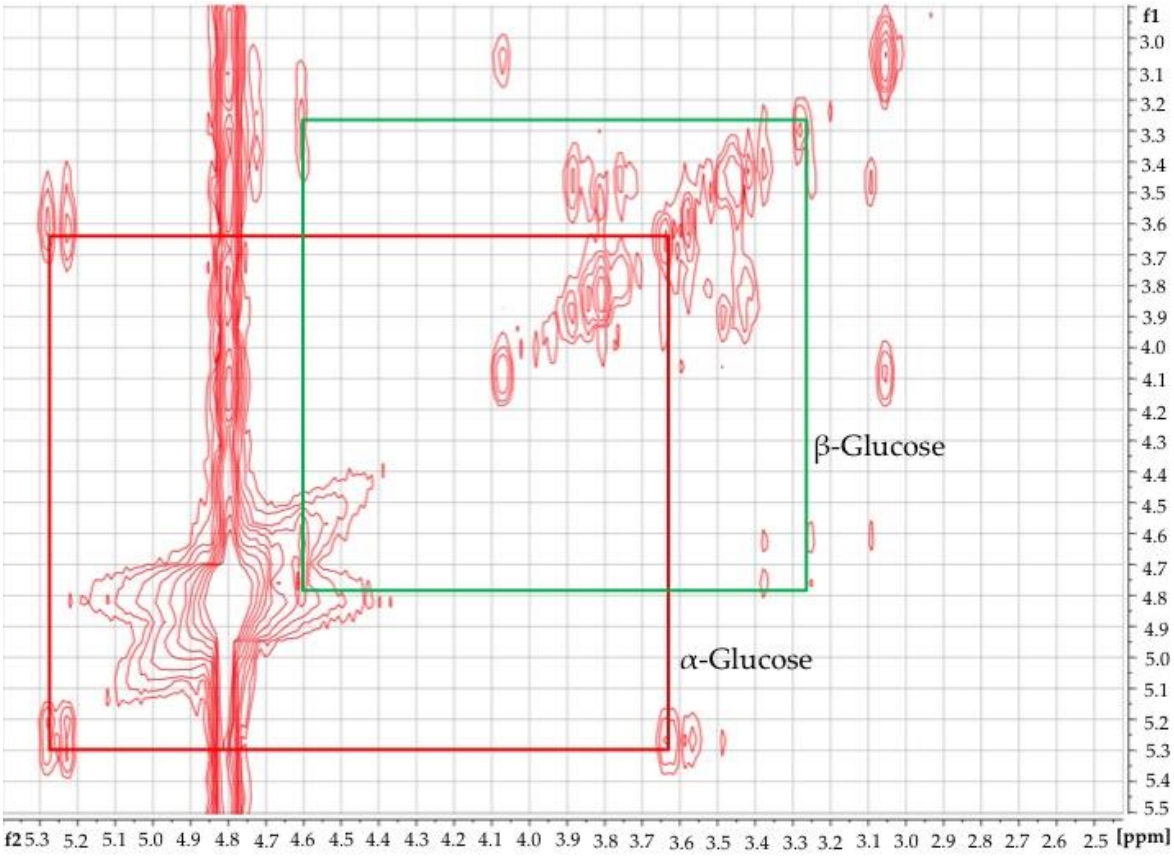


Figure 3: 2D ^{1}H-^{1}H COSY NMR diabetic urinary profile acquired on a 60 MHz instrument highlighting connectivities between α-Glucose C1-H and C2-H resonances at δ = 5.22 ppm (*d*) and 3.52 ppm (*dd*) respectively (labelled in red), and correspondingly those of the β-anomer at δ = 4.63 (*d*) and 3.23 ppm (*dd*) respectively (labelled in green. A further ^{1}H-^{1}H COSY connectivity between signals located at δ = 3.06 and 4.08 ppm is also clearly visible.

Using 2D ^{1}H-^{1}H COSY analysis, the identity of glucose in diabetic urine samples was readily confirmed, since this revealed connectivities between the C1-H and C2-H resonances of both its anomeric forms. Moreover, further glucose and other biomolecule connectivities were visible. This demonstrates the usefulness of 2D ^{1}H-^{1}H COSY analysis of human urine as a confirmatory tool for LF benchtop NMR-based metabolomics applications.

Both univariate and multivariate analysis of the LF benchtop 60 MHz NMR dataset revealed clear and highly statistically significant differences between of urine samples collected from a cohort of diabetic patients (n = 10) and those from healthy controls. The dataset acquired, comprising of 27 manually-selected and electronically integrated buckets ranging from 1.03-8.52 ppm, was normalized to the TSP internal standard (of final concentration 264 μmol./L), and then potential predictor

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

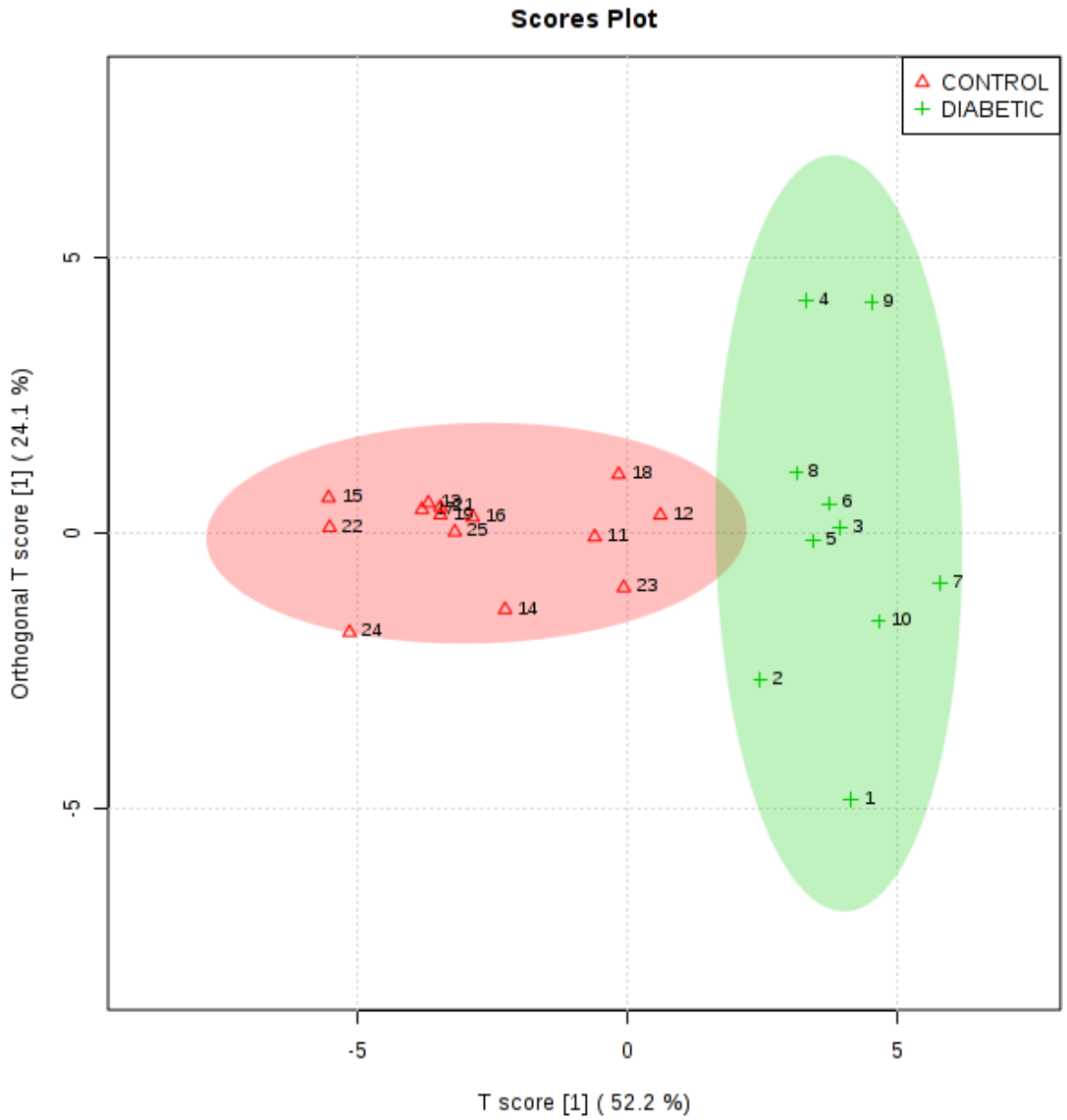
variables within chemical shift bucket columns were generalized-logarithmically (glog)-transformed, and Pareto-scaled (Pareto-scaling involves subtraction of the mean bucket value from all bucket observations followed by their division by the square root of that variable's standard deviation, so that each one has a mean value of 0 and a variance not equivalent but similar to unity.

Primarily, 27 univariate two-sample t tests were performed, and when corrected for FDRs, these revealed that there were very highly significant increases in the urinary concentrations of a range of biomolecules in type 2 diabetic patients. Key biomarkers detected using the LF ^{1}H NMR technique were: citrate, i.e. -C$\underline{H}_2$CO$_2^-$ functions within the relatively spectroscopically-clear 2.53-2.70 ppm bucket (p = 1.87 x 10^{-6}); N-acetyl storage compounds, i.e. N-acetylsugar- and N-acetylamino acid-NHCOC$\underline{H}_3$ function protons in the relatively clear 1.99-2.13 ppm bucket (p = 1.87 x 10^{-6}); lactate as its -C$\underline{H}_3$ group protons in the 1.25-1.34 ppm region (p = 2.15 x 10^{-6}), which has potential interferences arising from threonine- and acetoin-C$\underline{H}_3$ functions; alanine as its -C$\underline{H}_3$ function doublet resonance at 1.48 ppm with minimal potential interferences; creatinine, i.e. as its >N-C$\underline{H}_3$ proton singlet within the 2.98-3.14 ppm bucket (p = 5.0 x 10^{-6}), with potential interferences arising from creatine-C$\underline{H}_3$, lysine-ε-C$\underline{H}_2$ and γ-aminobutyrate's γ-C$\underline{H}_2$ functions, and also β-glucose's C3-$\underline{H}$ 3.21 ppm signal; acetone as its -CH$_3$ groups' singlet resonance within the 2.14-2.29 ppm bucket (p = 5.53 x 10^{-6}), which has conceivable interferences from glutamine-C3-C$\underline{H}_2$ and acetoin-C$\underline{H}_3$ proton signals; acetate as its -C$\underline{H}_3$ function in the 1.87-1.99 ppm bucket (p = 4.53 x 10^{-5}), which has a potential interference from the thymine-C$\underline{H}_3$ resonance, although it should be noted that the latter metabolite has a substantially lower urinary concentration than the former; 3-D-hydroxybutyrate within the 1.14-1.25 ppm bucket (p = 4.55 x 10^{-5}), with potential interferences arising from 3-aminoisobutyrate-C$\underline{H}_3$ and L-fucose-C$\underline{H}_3$ function resonances; indoxyl sulphate in the 7.15-7.33 ppm bucket (p = 6.27 x 10^{-3}), with a potential interference from tyrosine's C2/C6 aromatic proton doublet signal; and hippurate as its signal localized within the 7.55-7.71 ppm bucket, the only potential interfering agent being 1-methylhistidine's C4 imidazole ring proton singlet (p = 0.037). Most importantly, glucose, which was determined firstly as a composite bulk carbohydrate ring proton (i.e. C2-$\underline{H}$ to C6-$\underline{H}$) bucket (δ = 3.14-3.99 ppm, p = 2.15 x 10^{-6}), and secondly as the more specific alpha-anomeric proton (i.e. α-C1-$\underline{H}$) resonance bucket (δ = 5.17-5.36 ppm, p = 0.038) was also found to be a key biomarker, as expected. However, as noted above, a complication of the α-C1-$\underline{H}$ glucose signal is its small fractional overlap with the residual water signal at 60 MHz operating signal. Although the bulk 3.14-3.99 ppm glucose sugar ring proton bucket intensity can be expected to be influenced by those of a range of further urinary metabolite signals also present within this spectral region, we found that when glucose concentrations were > 10 mmol./L, as indeed it was in all 6 of the type 2 diabetic urine samples explored which had detectable glucose levels (it was non-^{1}H NMR-detectable in 4/10 samples investigated), such interferences were limited in view of the much lower intensities of these further biomolecule signals within this broad spectral region (such as those arising from choline, betaine, trimethylamine N-oxide, taurine, glycine, creatine, glycolate, guanadinoacetate, etc.) than those of the relatively intense α- and β-glucose anomers (i.e. C2-H to C6-H resonances combined).

As an example of multivariate analysis, OPLS-DA was utilized to explore the ability of this strategy to distinguish between the type 2 diabetic and healthy control urinary ^{1}H NMR profiles acquired.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

Figure 4 shows an OPLS-DA scores plot with associated 95% confidence ellipses, and this demonstrates clearly distinctive clusterings for these two groups of participants. In order to evaluate the performance of this multivariate classification system, a 10-fold cross-validation procedure was applied. R^2X, R^2Y and Q^2 values obtained from this analytical model were 0.522, 0.674 and 0.634 respectively, and the Q^2 value obtained was highly significant (values of this index ≥ 0.40 are routinely employed as a cut-off for this model, as previously described by Worley and Powers [38]). Moreover, a permutation test conducted with 2,000 permutations gave p values of $< 5 \times 10^{-4}$ for both Q^2 and R^2Y. Variable importance in projection (VIP) scores and S-plots were utilized in order to identify the most important ^{1}H NMR bucket variables for discrimination between healthy and type 2 diabetic participants, and those assigned to methylsuccinate (upregulated) and formate (downregulated) were also found to serve as key biomarker features of this discrimination, in addition to the majority of those detected via the univariate t test analysis described above (including glucose itself, the ketone bodies acetoacetate, acetone and 3-D-hydroxybutyrate, acetate, N-acetyl storage compounds, citrate, creatinine and lactate).



 doi:10.20944/preprints201810.0089.v1

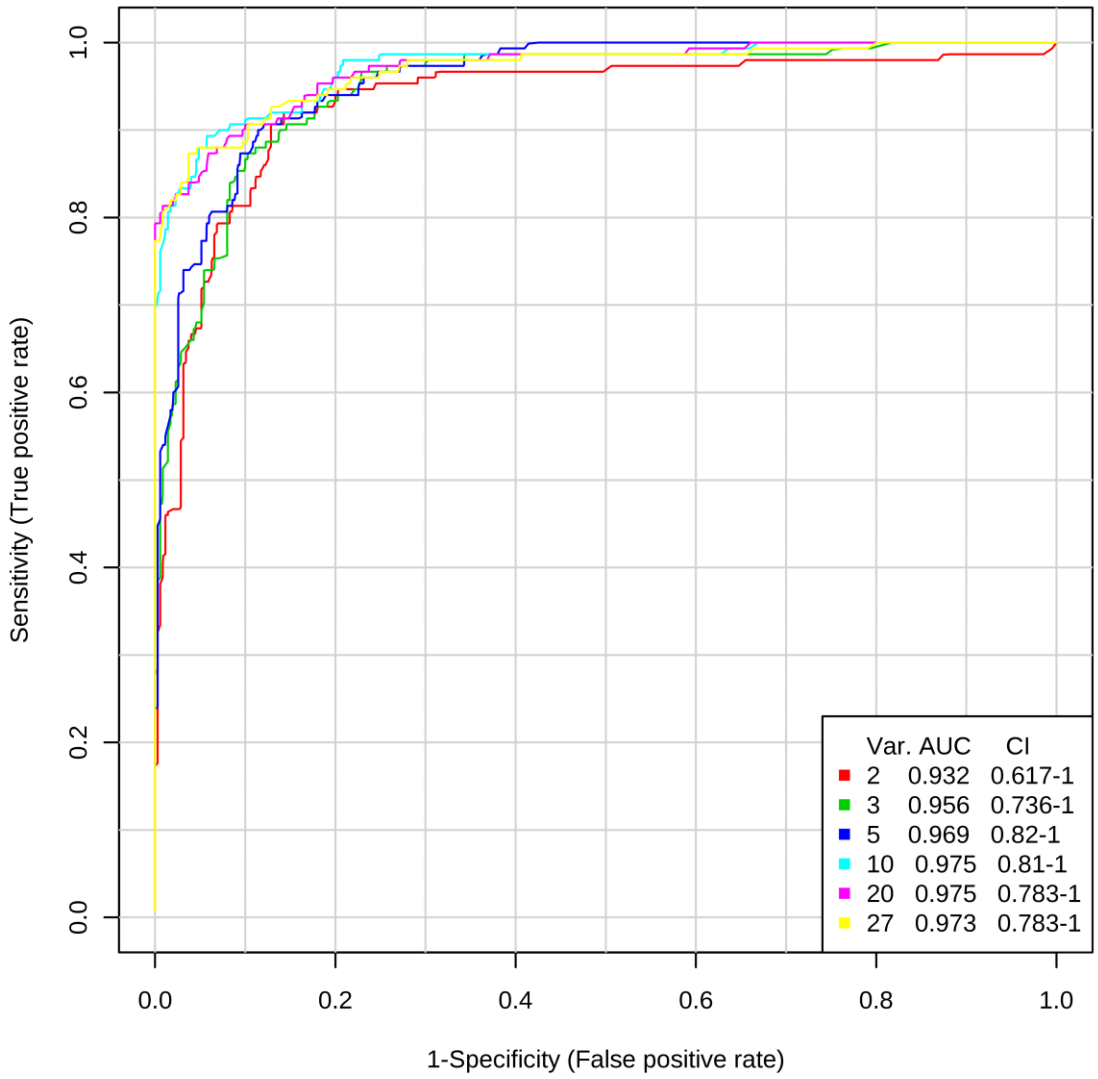
Figure 4. OPLS-DA scores plot of orthogonal T score *vs.* T score demonstrating a clear metabolomics-based distinction between type 2 diabetic patients and healthy controls. 95% confidence ellipsoids are also shown. The type 2 diabetic patient cluster sample T score values (+2.5 to +6) are all greater than those of the control cohort (-6 to +1).

Secondly, ROC curves produced via Monte Carlo Cross-Validation (MCCV) and based on the SVM strategy demonstrated that the overall mean classification success rate was 97.5% for this model.The most effective SVM models were those which incorporated the total number of 10 ISB intensity features, the AUROC value obtained being 0.975 (95% confidence intervals 0.81-1.00), as shown in Figure 5. Therefore, with the above overall classification reliability and AUROC values, this model applied served as one with a highly effective discriminatory ability (these values are considered effective, highly discriminatory and exceptional for models when they are > 0.70, 0.87-0.90 and > 0.90 respectively [39].

Key discriminatory biomarker variables identified from this form of multivariate analysis were citrate > 3-D-hydroxybutyrate > hippurate > N-acetyl storage compounds > alanine > total bulk glucose (C2-H to C6-H resonances only) > lactate > α-glucose (C1-H resonance only) > 3-(3-hydroxyphenyl)-3-hydroxypropanoate (C1/C6-CH resonances) > indoxyl sulphate > urea in that order of effectiveness.

(a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1



(b)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

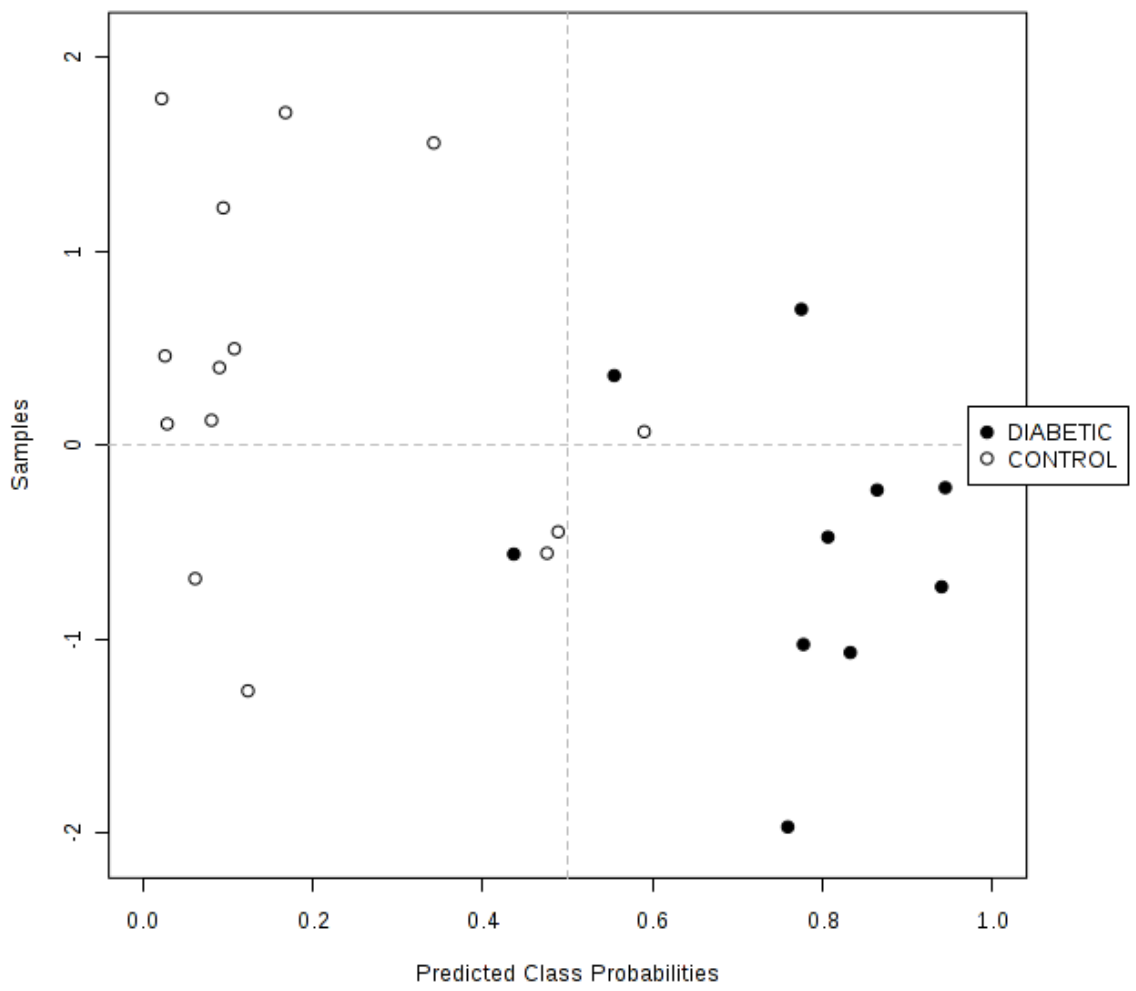


Figure 5. (a) ROC curve (plot of true positive *vs.* false positive rates) with an AUROC value of 0.975 obtained from the SVM model building system explored with 10 out of a possible 27 variables. ROC curves were developed via Monte Carlo Cross-Validation (MCCV) involving a balanced sub-sampling processes involving application of an SVM model builder. The inset shows mean AUC values estimated for increasing sampling sizes, together with 95% confidence intervals for these values. **(b)** Probability view arising from a balanced sub-sampling approach for SVM model training (predicted class probabilities for each sample employed the most effective AUC-based classification system).

Finally, a Random Forest analysis performed with 1,000 trees and 7 distinguishing variables per node successfully classified 9 out of 10 type 2 diabetic samples, and 12 out of 14 healthy control ones, on the basis of their urinary ^{1}H NMR metabolic profiles, i.e. an overall classification accuracy of 0.875 (data not shown).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

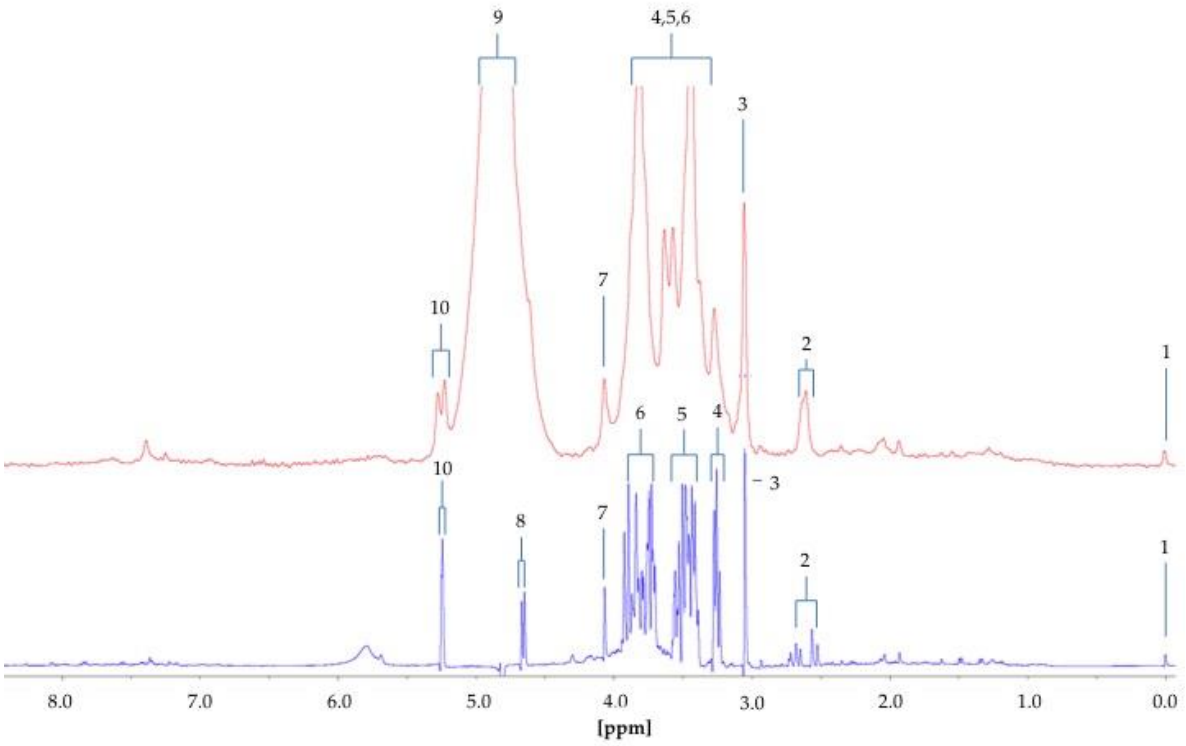


Figure 6. Diabetic urinary ^{1}H NMR profile acquired at 60 (red) and 400 MHz (blue**)** operating frequencies. Assignments: 1; TSP 2; Citrate-CH2A/B 3; creatinine/creatine >N-CH_3; 4; Glucose-C2-H 5; Glucose-C4/C2/C5-CH 6; Glucose-C3/C5/C6 –CH 7; creatinine-CH_2 8; β-glucose-C1-H; 9; Water-OH 10: α-glucose-C1-H.

The difference between the ^{1}H NMR metabolic profiles of urine acquired at 60 and 400 MHz operating frequencies for metabolomics analysis is substantial (Figure 6). Indeed, in 1D single-pulse spectra acquired at 60 MHz, the α-glucose signal is influenced by the residual water signal at concentrations < 2.8 mmol./L, and increasingly so with decreasing α-glucose level. Moreover, the difference in resolution from 400 MHz to 60 MHz causes significant expansion of multiplet resonances, especially complex second-order ones, at the lower operating frequency. This, in turn, gives rise to resonance overlap problems at 60 MHz, particularly for complex biofluid spectra. Indeed, this resonance δ value expansion would be 400/60 times greater for 60 MHz spectra than it is for 400 MHz ones. For example, the ethanol-CH_3 function triplet resonance located at δ = 1.19 ppm, which has a *J* value of 7.07 Hz, would wholly encompass 2 x 7.07 Hz/60 Hz = 0.235 ppm of the spectral profile at 60 MHz, but only 2 x 7.07 Hz/400 Hz = 0.035 ppm at 400 MHz. The more challenging signals to assign include those from higher-order multiplets and more highly split first-order ones, such as the lactate C-H quartet resonance (δ = 4.13 ppm).

Glucose was clearly ^{1}H NMR-detectable in 6 out of a total of n = 10 type 2 diabetic urine samples at an operating frequency of 60 MHz, and integration of the 5.26 ppm α-anomeric proton signal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

followed by its normalisation to internal TSP provided estimates of total urinary glucose levels in our cohort of type 2 diabetic patients. Mean±SEM glucose concentrations were 145±70 mmol./L (range 0-660 mmol./L). Using a mean urinary creatinine (Cn) concentration of 10.09 mmol./L [40], these mean±SEM values convert to 14.4±6.9 mmol./mmol. Cn. These urinary glucose level values concord with those previously reported in diabetic patients, as noted in Table 1, although it should be noted that these data correspond to type 1 diabetes or diabetic ketosis. This Table also lists mean glucose concentrations for healthy control subjects, which vary from 9-40 µmol./mol. Cn for adults, 32 µmol./mol. Cn for children, 7-144 µmol./mol. Cn for infants, and 15 µmol./mmol. Cn for new-borns.

A paired sample t test found no significant differences between total glucose concentration determinations performed on a LF benchtop 60 MHz spectrometer and our more conventional HF 400 MHz NMR facility (p = 0.079, n = 10). Moreover, there was an excellent correlation between determinations made on these two analytical systems (r = 9986).

Disease	**Age Group**	**Gender**	**Mean±Error* Creatinine-Normalised Concentration**	**Range**	**Reference**
Healthy Control	Adult (> 18 years old) healthy control	M/F	36.6 µmol./mmol. Cn	10.3-56.7 µmol./mmol. Cn	[41]
Healthy Control	Newborns (0-30 days old)	M/F	15.0 µmol./mmol. Cn	0.0-50.0 µmol./mmol. Cn	[42]
Healthy Control	Adult (> 18 years old) healthy control	M/F	9.0 µmol./mmol. Cn	0.0-19.0 µmol./mmol. Cn	[42]
Healthy Control	Adult (> 18 years old) healthy control	M/F	Unavailable	16.7-111.1 µmol./mmol. Cn	[43]
Healthy Control	Adult (> 18 years old)	M/F	37.5 µmol./mmol. Cn (error bars unavailable)	12.5-58.4 µmol./mmol. Cn	[41]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

	healthy control				
Healthy Control	Adult (> 18 years old) healthy control Male	M	31.1 µmol./mmol. Cn	Unavailable	[44]
Healthy Control	Infant (0-1 year old)		7.0 µmol./mmol. Cn	0.0-15.0 µmol./mmol. Cn	[42]
Healthy Control	Adult (> 18 years old) healthy control	M/F	25.8±13.8 µmol./mmol. Cn	Unavailable	[42]
Healthy Control	Infant (0-1 year old) Female	F	143.1±399.8 µmol./mmol. Cn	Unavailable	[45]
Healthy Control	Children (1-13 years old)	Unspecified	31.6±16.0 µmol./mmol. Cn	Unavailable	[13]
Diabetes	Adult (> 18 years old)	M/F	19.7 mmol./mmol. Cn (error bars unavailable)	Unavailable	[9]
Diabetic ketosis	Adult (> 18 years old)	M/F	79.6 mmol./mmol.Cn	68.3-19.9 mmol./mmol. Cn	[46]
Type 1 Diabetes	Adult (> 18 years old)	M/F	17.5 mmol./mmol. Cn	0.1-129.9 mmol./mol.Cn	[47]
Eosinophilic Esophagitis	Children (1 - 13 years old)	Unspecified	30.0±27.8 µmol/mmol. Cn		[13]
Fanconi Bickel Syndrome	Children (1 - 13 years old)	F	>5.55 mmol/mmol. Cn (error bars unavailable)		[48]

Table 1. Previously reported creatinine (Cn)-normalised urinary glucose concentrations in healthy controls and patients with diabetes, eosinophilic esophagitis and Fanconi Bickel syndrome. *Error bars unspecified. These data were obtained from the Human Metabolome Database (HMDB) [13].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

The significance of urinary glucose levels is not only evident in diabetes patients, but also in patients with conditions such as eosinophilic esophagitis and Fanconi Bickel Syndrome (Table 1). Therefore, potential applications for using LF-benchtop NMR reach far beyond screening for one disease, a metabolic 'fingerprint' is able to be formed to successfully diagnose multiple diseases. Moreover, monitoring urinary glucose levels is just a single example of what can be achieved using LF NMR analysis. Investigations of other biofluids and corresponding metabolites are possible, and therefore an abundance of metabolic disturbances may be explored with this novel technique.

6.0 Discussion

Whilst metabolite analysis in intact biofluids by LF NMR spectroscopy is a newly-advanced technique, it has the potential to impact on 'point-of-care' clinical chemistry diagnosis and monitoring, especially with the increasingly rapid development of spectrometers with operating frequencies > 60 MHz. In comparison, high-field NMR analysis at operating frequencies of 400-800 MHz has become a standard probing tool for the multicomponent analysis of complex biofluids collected from humans and other organisms. Indeed, a variety of important biological information regarding the molecular nature and concentrations of a wide range of endogenous biomolecules, together with exogenous agents present in such fluids, can be obtained from such investigations. Moreover, biomedical NMR analysis serves as a virtually non-invasive technique since it often has little or no requirement for the pre-treatment of samples, and generally requires only a limited knowledge of sample composition prior to analysis. These approaches offer significant potential regarding the investigation of metabolic processes, and can be coupled with multidimensional data analysis techniques within metabolomics workflows, serving as an extremely powerful means of probing, for example, the biochemical basis of human disease aetiology. These investigations can also provide substantial diagnostic and/or prognostic disease monitoring information, including the identification and validation of reliable biomarker molecules. This combined NMR-based metabolomics approach is also readily applicable to the simultaneous analysis of a wide range of metabolites in tissue biopsies and cultured cells (either intact via ^{1}H high-resolution magic angle spinning (MAS)-NMR-based metabolic profiling strategies, or correspondingly as appropriate solution-state extracts), and/or cell culture media.

Here, we have report for the first time the rapid, essentially or completely non-invasive analysis of a human biofluid sample (human urine) by a LF 60 MHz benchtop NMR facility. Whilst compact, LF NMR analyses of human plasma, serum, and whole blood can reveal trends in disease development and prognosis through inspection of longitudinal (T_1) and transverse (T_2) relaxation times,[49] the major purpose of these pilot studies reported herein was to (a) establish whether 60 MHz benchtop ^{1}H NMR spectra may be reliably employed to detect urinary biomolecules, and to determine their concentrations in this biofluid for future diagnostic and prognostic monitoring purposes, despite

 doi:10.20944/preprints201810.0089.v1

some inherent resonance overlap phenomena arising from the lower operating frequency involved, and (b) provide a protocol 'blueprint' for future NMR-based metabolomics strategies available for the analysis of biofluids. Since the resonance frequencies of nuclei are dependent on the static magnetic field strength, resonances of the same linewidth (in Hz) will appear increasingly broader at decreasing field strengths when expressed relative to the overall spectral window range in ppm; as noted here, this gives rise to an enhancement of resonance overlap, particularly in spectra obtained from complex multianalyte samples such as biofluids and tissue biopsy extracts.

From the protocols and case study described here, a schematic representation has been proposed (Figure 7) in order to allow researchers to follow a protocol for the LF benchtop NMR spectrometroscopc analysis of biolfuids, and which is combined with multivariate metabolomics analysis techniques. This aims to provide full considerations for metabolomics investigations to improve uniformity across the field when using LF benchtop NMR analysis, in order to ensure that some level of protocol uniformity is maintained.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2018 doi:10.20944/preprints201810.0089.v1

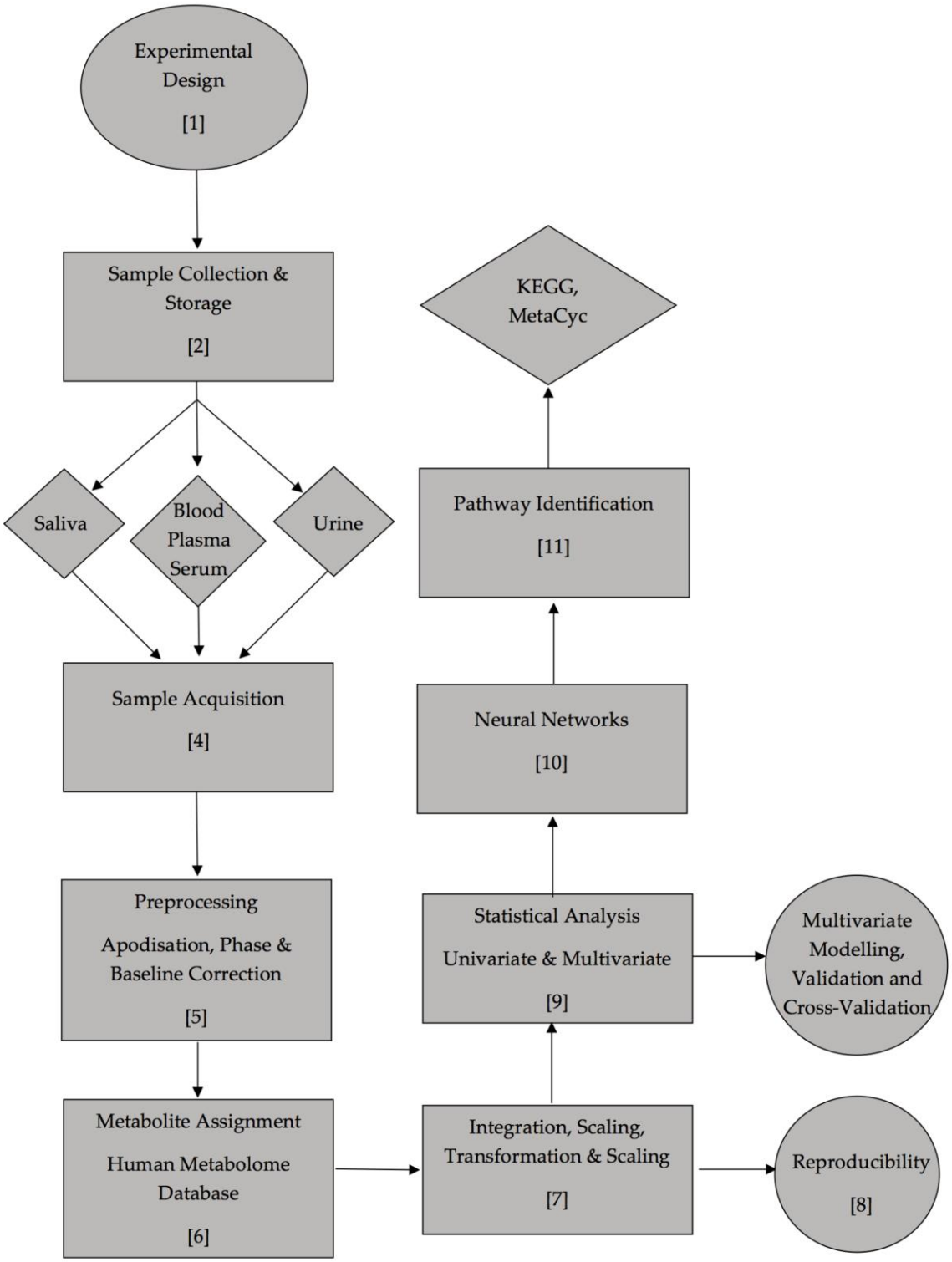


Figure 7. Proposed scheme for biofluid analysis by LF benchtop NMR spectroscopy. Numbers refer to sections of the protocol relevant to each respective section of the scheme.

 doi:10.20944/preprints201810.0089.v1

7.0 Conclusions

Biomarkers can serve as powerful predictive indicators of the presence of metabolic diseases, together with a myriad of other conditions, and here we have presented, for the first time, the potential applications of LF benchtop ^{1}H NMR analysis to the rapid diagnosis of type 2 diabetes in human urine samples. Therefore, this technique acts as a sensitive means of monitoring the urinary metabolic status of diabetes at 'point-of-care' sites, either type 1, type 2 or pre-diabetic conditions, together with gestational diabetes. In view of the very large, inconvenient sizes of HF NMR instruments, hospitals and laboratories do not tend to use these for the rapid, routine and multicomponent analysis for the detection of diseases, but instead may employ such small size LF instruments, rendering them suitable for the clinical chemistry analysis of patient samples directly on-site at healthcare settings. Herein, we have demonstrated the detection and quantification capabilities of LF 60 MHz, H_2O-suppressed spectra acquired on such an NMR instrument, which is based on a rare earth magnet arranged within a Halbach array. In addition to its small size and portability, advantages of the technique regarding the diagnostic and prognostic screening of biofluids for selected human diseases include (1) no requirement for expensive deuterated NMR solvents for sample preparation purposes for some spectrometers available, which (2) most importantly, benefits from ease-of-use, with no requirement for specialist operational staff, and which can, at least in principle, detect very low concentrations of biomolecules present in intact biofluids within short timescales (less than 15 min). The applications of this technique continue to expand, and will likely develop further in the field of disease diagnosis and severity monitoring screening.

8.0 Future Perspectives

LF benchtop NMR has potential to be used as a point-of-care diagnostic and prognostic screening facility, if suitable protocols and the validation of biomarker analytes as outlined here are adhered to in clinical practices. The technique is advantageous in view of the low running costs and lack of cryogens required when compared to those of higher resolution, HF facilities commonly found in bioanalytical laboratories. Scripting and robotic sampling can ensure full automation of sample preparation and acquisition, which could later be linked to artificial intelligence techniques.

LF benchtop NMR facilities are becoming more sophisticated, enabling multiple solvent suppression regimens on samples which have been extracted in different solvent systems, for example. This would be particularly applicable to extractions using methanol or chloroform from water-based samples subsequent to a lyophilisation procedure. Furthermore, gradients, which are now available on benchtop NMR systems, extend the availability of such techniques for further methodologies to be developed, such as pure shift, which can assist in the confirmation of signal assignments.

Further NMR-active nuclei such as ^{13}C can also be readily monitored in both simple and complex sample matrices using LF benchtop NMR facilities. However, at present sensitivity for such nuclei is poor, although LOD values for ^{19}F are reasonably high. Using techniques such as hyperpolarisation could increase the sensitivity of such nuclei, and may be employed adjunctly with this technology.

Internal standard in instrument ERETIC (Electronic Reference To access *In Vivo* Concentrations) has emerged to remove the addition of internal standard reference agents completely, and reliance on external references such as TSP may become routine in the near future. Issues with internal reference standards may arise from their volatility (as in tetramethylsilane, TMS) and potential to bind to macromolecules in high protein content biofluids such as blood plasma/serum, factors which represent major analytical/bioanalytical disadvantages. The field of applications for these LF NMR techniques is substantial, ranging from food technology through to 'at crime scene site' forensics, and now biomedical imaging in addition to spectroscopy. Whilst miniaturisation of technology has led to improvements in the analytical performance of these facilities, improvements in sensitivity, and the field strength of the permanent rare earth magnets featured in LF benchtop NMR spectrometers, will lead to further advances in the range of applications of these instruments, and their effectiveness therein.

Abbreviations

AI; Artificial Intelligence, ANCOVA; Analysis of Covariance, ANOVA; Analysis of Variance AUROC; Area under the receiver operating characteristic, CCR; Correlated Component Regression, CIT; Computational intelligence techniques, COSY; Correlation Spectroscopy, EDTA; Ethylenediaminetetraacetic acid, ERETIC; Electronic Reference To access In Vivo Concentrations, FID; Free Induction Decay, GISSMO; Guided Ideographic Spin System Model Optimisation, HF; High-field, HMDB; Human Metabolome Database, HPLC; High Performance Liquid Chromatography, IPA; Ingenuity Pathways Analysis, KEGG; Kyoto Encyclopaedia of Gene and Genomics, LF; Low-field, LOD; Limit of Detection, LOOCV; Leave one out Cross validation, LOQ; Limit of Quantification, MAS; Magic Angle Spinning MCCV; Monte Carlo Cross Validation, MMCD; Madison metabolomics consortium database, MV; Multivariate, NMR; Nuclear Magnetic Resonance, OOB; Out-of-the-bag, PCA; Principle Component Analysis, PLS-DA; Partial Least Squares Discriminant Analysis, SOM; Self Organising Map, SOP; Standard Operating Procedures, SVM; Support Vector Machines, TCOSY; Total Correlation Spectroscopy, TMS; Tetramethylsilane, TSP; sodium 3-(trimethylsilyl)[2,2,3,3-d_4] propionate

Acknowledgements

The authors thank Magritek GmbH for the use of a Spinsolve 60 Spectrometer and our 6 week internship students, Caleb Anderson and Funke Sangowawa for their contributions.

Author Contribution

B Percival, Y Osman and M Gibson carried out the experiments. F Jafari assisted in the experimental preparation. M Martin provided clinical input. F Casanova assisted in resourcing. M Mather, M Molinari and M Edgar provided theoretical considerations and assisted in experimental design. P Wilson and M Grootveld designed experiments and analytical procedures. All authors contributed to editing the manuscript.

Conflicts of Interest

F. Casanova is employed by Magritek GmbH.

References

1. Teng, Q. *Structural Biology: Practical NMR Applications*; 2013 ISBN 978-1-4614-3964-6.
2. Shen, B.; Tang, H.; Jiang, X. *Translational Biomedical Informatics*; Springer, 2016; ISBN 9811015023.
3. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC - Trends Anal. Chem.* **2008**, *27*, 228–237, doi:10.1016/j.trac.2007.12.001.
4. Santorio, S. *De statica medicina*; 1713
5. Thomson, J. J. Bakerian Lecture:—rays of positive electricity. *Proc. R. Soc. Lond. A* **1913**, *89*, 1–20.
6. Purcell, E. M.; Pound, R. V; Bloembergen, N. Nuclear magnetic resonance absorption in hydrogen gas. *Phys. Rev.* **1946**, *70*, 986.
7. Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci.* **1971**, *68*, 2374–2376.
8. Nicholson, J. K.; Buckingham, M. J.; Sadler, P. J. High resolution 1H n.m.r. studies of vertebrate blood and plasma. *Biochem. J.* **1983**, *211*, 605–15.
9. Nicholson, J. K.; O'Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sönksen, P. H. Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochem. J.* **1984**, *217*, 365–75.
10. Logemann, J.; Schell, J.; Willmitzer, L. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* **1987**, *163*, 16–20.
11. Percival, B.; Wann, A.; Masania, J.; Sinclair, J.; Sullo, N.; Grootveld, M. Detection and Determination of Methanol and Further Potential Toxins in Human Saliva Collected from Cigarette Smokers: A 1H NMR Investigation. *JSM Biotechnol. Biomed. Eng.* **2018**.
12. Visentin, S.; Crotti, S.; Donazzolo, E.; D'Aronco, S.; Nitti, D.; Cosmi, E.; Agostini, M. Medium chain fatty acids in intrauterine growth restricted and small for gestational age pregnancies. *Metabolomics* **2017**, *13*, 1-9. doi:10.1007/s11306-017-1197-8.
13. Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *4*, 608-617. doi:10.1093/nar/gkx1089.
14. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *2*, 486-494. doi:10.1093/nar/gky310.
15. Trivedi, D. K.; Hollywood, K. A.; Goodacre, R. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons Transl. Med.* **2017**, *3*, 294-305. doi: 10.1016/j.nhtm.2017.06.001
16. Blümler, P.; Casanova, F. Chapter 5. Hardware Developments: Halbach Magnet Arrays. In Mobile NMR and MRI, **2015**; pp. 133–157.
17. Qiu, Y.; Rajagopalan, D.; Connor, S. C.; Damian, D.; Zhu, L.; Handzel, A.; Hu, G.; Amanullah,

A.; Bao, S.; Woody, N.; MacLean, D.; Lee, K.; Vanderwall, D.; Ryan, T. Multivariate classification analysis of metabolomic data for candidate biomarker discovery in type 2 diabetes mellitus. *Metabolomics* **2008**, *4*, 337–346, doi:10.1007/s11306-008-0123-5.

18. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. *Anal. Chem.* **2006**, *78*, 4430–4442.
19. Blekherman, G.; Laubenbacher, R.; Cortes, D. F.; Mendes, P.; Torti, F. M.; Akman, S.; Torti, S. V.; Shulaev, V. Bioinformatics tools for cancer metabolomics. *Metabolomics* **2011**, *7*, 329–343, doi:10.1007/s11306-010-0270-3.
20. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692.
21. Blümich, B.; Casanova, F.; Dabrowski, M.; Danieli, E.; Evertz, L.; Haber, A.; Van Landeghem, M.; Haber-Pohlmeier, S.; Olaru, A.; Perlo, J.; Sucre, O. Small-scale instrumentation for nuclear magnetic resonance of porous media. *New J. Phys.* **2011**, *13*, 015003, doi:10.1088/1367-2630/13/1/015003.
22. Gouilleux, B.; Charrier, B.; Akoka, S.; Giraudeau, P. Gradient-based solvent suppression methods on a benchtop spectrometer. *Magn. Reson. Chem.* **2017**, *55*, 91–98, doi:10.1002/mrc.4493.
23. Danieli, E.; Perlo, J.; Blümich, B.; Casanova, F. Small magnets for portable NMR spectrometers. *Angew. Chemie - Int. Ed.* **2010**, *49*, 4133–4135, doi:10.1002/anie.201000221.
24. Schaeler, K.; Roos, M.; Micke, P.; Golitsyn, Y.; Seidlitz, A.; Thurn-Albrecht, T.; Schneider, H.; Hempel, G.; Saalwaechter, K. Basic principles of static proton low-resolution spin diffusion NMR in nanophase-separated materials with mobility contrast. *Solid State Nucl. Magn. Reson.* **2015**, *72*, 50–63, doi:10.1016/j.ssnmr.2015.09.001.
25. Singh, K.; Blümich, B. Desktop NMR for structure elucidation and identification of strychnine adulteration. *Analyst* **2017**, *142*, 1459–1470, doi:10.1039/C7AN00020K.
26. Isaac-Lam, M. F. Analysis of Bromination of Ethylbenzene Using a 45 MHz NMR Spectrometer: An Undergraduate Organic Chemistry Laboratory Experiment. *J. Chem. Educ.* **2014**, *91*, 1264–1266, doi:10.1021/ed400365p.
27. Chang, W. H.; Chen, J. H.; Hwang, L. P. Single-sided mobile NMR with a Halbach magnet. *Magn. Reson. Imaging* **2006**, *24*, 1095–1102, doi:10.1016/j.mri.2006.04.005.
28. Mickiewicz, B.; Vogel, H. J.; Wong, H. R.; Winston, B. W. Metabolomics as a novel approach for early diagnosis of pediatric septic shock and its mortality. *Am. J. Respir. Crit. Care Med.* **2013**, *187*, 967-976. doi:10.1164/rccm.201209-1726OC.
29. Garcia-Perez, I.; Posma, J. M.; Gibson, R.; Chambers, E. S.; Hansen, T. H.; Vestergaard, H.; Hansen, T.; Beckmann, M.; Pedersen, O.; Elliott, P.; Stamler, J.; Nicholson, J. K.; Draper, J.; Mathers, J. C.; Holmes, E.; Frost, G. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol.* **2017**, *5*, 184-195. doi:10.1016/S2213-8587(16)30419-3.
30. Lauridsen, M.; Hansen, S. H.; Jaroszewski, J. W.; Cornett, C. Human urine as test material in 1H NMR-based metabonomics: Recommendations for sample preparation and storage. *Anal. Chem.* **2007**, *79*, 1181-1186. doi:10.1021/ac061354x.
31. Grootveld, M.; Silwood, C. J. L. 1H NMR analysis as a diagnostic probe for human saliva.

Biochem. Biophys. Res. Commun. **2005**, *329*, 1-5. doi:10.1016/j.bbrc.2005.01.112.

32. Yin, P.; Lehmann, R.; Xu, G. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal. Bioanal. Chem.* **2015**, *407*, 4879-4892. doi:10.1007/s00216-015-8565-x.
33. Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.* **2008**, *26*, 162-164, doi:10.1038/nbt0208-162
34. Robinette, S. L.; Zhang, F.; Brüschweiler-Li, L.; Brüschweiler, R. Web server based complex mixture analysis by NMR. *Anal. Chem.* **2008**, *80*, 3606-3611, doi:10.1021/ac702530t.
35. Dashti, H.; Westler, W. M.; Tonelli, M.; Wedell, J. R.; Markley, J. L.; Eghbalnia, H. R. Spin System Modeling of Nuclear Magnetic Resonance Spectra for Applications in Metabolomics and Small Molecule Screening. *Anal. Chem.* **2017**, *89*, 12201–12208, doi:10.1021/acs.analchem.7b02884.
36. Dashti, H.; Wedell, J. R.; Westler, W. M.; Tonelli, M.; Aceti, D.; Amarasinghe, G. K.; Markley, J. L.; Eghbalnia, H. R. Applications of Parametrized NMR Spin Systems of Small Molecules. *Anal. Chem.* **2018**, *90*, 10646–10649, doi:10.1021/acs.analchem.8b02660.
37. Lamanna, R. Proton NMR Profiling of Food Samples. *Annu. Reports NMR Spectrosc.* **2013**, *80*, 239–291, doi:10.1016/B978-0-12-408097-3.00004-4.
38. Worley, B.; Halouska, S.; Powers, R. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Anal. Biochem.* **2013**, *433*, 102-104, doi:10.1016/j.ab.2012.10.011.
39. Hanley, J. A.; McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29-36, doi:10.1148/radiology.143.1.7063747.
40. Hou, H.; Xiong, W.; Zhang, X.; Song, D.; Tang, G.; Hu, Q. LC-MS-MS measurements of urinary creatinine and the application of creatinine normalization technique on cotinine in smokers' 24 hour urine. *J. Anal. Methods Chem.* **2012**, 245415, doi:10.1155/2012/245415.
41. Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. The Human Urine Metabolome. *PLoS One* **2013**, *8*, e73076, doi:10.1371/journal.pone.0073076.
42. Lentner, C. (Cornelius); CIBA-GEIGY Limited. *Geigy scientific tables*; Ciba-Geigy, 1981; ISBN 9780914168508.
43. Putman, D. F. Composition and Concentrative Properties of Human Urine. *NASA Contract. Report* **1971**, doi:NASA CR-1802.
44. Shaykhutdinov, R. A.; MacInnis, G. D.; Dowlatabadi, R.; Weljie, A. M.; Vogel, H. J. *Metabolomics* **2009**, *5*, 307-317, doi: 10.1007/s11306-009-0155-5.
45. Shoemaker, J. D.; Elliott, W. H. Automated screening of urine samples for carbohydrates, organic and amino acids after treatment with urease. *J. Chromatogr.* **1991**, *562*, 125–138.
46. Hoppel, C. L.; Genuth, S. M. Urinary excretion of acetylcarnitine during human diabetic and fasting ketosis. *Am. J. Physiol. Metab.* **1982**, *243*, E168–E172, doi:10.1152/ajpendo.1982.243.2.E168.
47. Ştefan, L. I.; Nicolescu, A.; Popa, S.; Mota, M.; Kovacs, E.; Deleanu, C. 1H-NMR URINE Metabolic Profiling In Type 1 Diabetes Mellitus. *Rev. Roum. Chim* **2010**, *55*, 1033–1037.

48. Gupta, N.; Nambam, B.; Weinstein, D. A.; Shoemaker, L. R. Late Diagnosis of Fanconi-Bickel Syndrome. *J. Inborn Errors Metab. Screen.* **2016**, *4*, 232640981667943, doi:10.1177/2326409816679430.

49. Cistola, D. P.; Robinson, M. D. Compact NMR relaxometry of human blood and blood components. *TrAC - Trends Anal. Chem.* **2016**, *83*, 53–64, doi:10.1016/j.trac.2016.04.020.