Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pharmaceutical and Biomedical Analysis 39 (2005) 156–163
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Progress toward automated metabolic profiling of human serum: Comparison of CPMG and gradient-filtered NMR analytical methods

Laura H. Lucas [a,1], Cynthia K. Larive [a,2], Patricia Stone Wilkinson [b], Stephen Huhn [b,*]

[a] *Department of Chemistry, University of Kansas, 1251 Wescoe Hall Drive, Malott Hall, Lawrence, KS 66045, USA*
[b] *Bruker BioSpin Corporation, NMR Applications Laboratory, 15 Fortune Drive, Billerica, MA 01821, USA*

Accepted 1 September 2004
Available online 10 May 2005

Abstract

The investigation of drug delivery and metabolism requires the analysis of molecules in complicated biological matrices such as human serum. In NMR-based metabonomic analysis, T_2 relaxation editing with a CPMG filter is commonly used to suppress background signals from proteins and other endogenous components. Radio frequency pulse imperfections and incomplete irradiation across the spectral bandwidth can cause phase and baseline distortions in CPMG spectra. These distortions are exacerbated by water suppression techniques. Baseline correction methods included in commercially available data processing software packages may be incapable of producing artifact-free spectra. To increase the analytical precision of metabolic profiling, one NMR spectroscopist may be responsible for manually phasing and baseline correcting hundreds of spectra individually to remove operator-dependent variations, significantly reducing throughput. For metabonomic analysis of human serum, it was observed that the application of a pulsed field gradient filter produced ^{1}H NMR spectra well suited to automatic phasing routines. Superior baseline characteristics, an increased tolerance to radio frequency pulse imperfections, and improved water suppression were achieved. A concomitant reduction in signal intensity compared with the CPMG method was easily recovered by increasing the number of scans. Principal component analysis (PCA) of spectra, acquired under a variety of experimental conditions, revealed the improved reproducibility and robustness of ^{1}H NMR pulsed field gradient-filtered metabonomic analyses of serum compared to the CPMG method.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Metabonomics; NMR spectroscopy of biofluids; T_2 relaxation; Diffusion; Principal component analysis (PCA); Spectral editing

1. Introduction

Metabonomic analysis of human biofluids is a rapidly developing area in clinical research [1]. Nuclear magnetic resonance spectroscopy (NMR) has played an important role in the growth of this field, aiding in the identification of drug metabolites [2] and disease biomarkers [3] in fluids such as whole blood, plasma, urine, bile, and cerebral spinal fluid [4–6]. NMR relaxation and diffusion-based gradient filtering methods have been widely employed to analyze these samples. An inherent challenge in many of these biological matrices is the presence of proteins and other biological macromolecules whose broad lines can mask the detection of low-level small molecule metabolites. The small molecules of interest can be selectively detected by relaxation editing since macromolecules generally have much faster T_2 relaxation times [7–11]. One way to achieve such selectivity is to use the Carr Purcell Meiboom Gill (CPMG) method, which utilizes a 180° pulse train to attenuate signals from fast relaxing species (Fig. 1A) [12,13]. The potential disadvantages of this method are the phase and baseline distortions resulting from pulse inaccuracies and magnetic field inhomogeneities.

* Corresponding author at Wyeth Pharmaceuticals, 200 Cambridgepark Drive, Cambridge, MA 02140. Tel.: +1 617 665 5665; fax: +1 617 665 5682.
E-mail address: shuhn@wyeth.com (S. Huhn).
[1] Present address: Department of Pharmaceutical Chemistry, University of Kansas, 2095 Constant Avenue, Lawrence, KS 66047, USA.
[2] Present address: Department of Chemistry, University of California, Riverside, CA 92521, USA.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.09.060

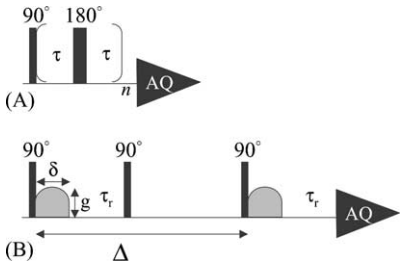

Fig. 1. Pulse sequences used for human serum analysis. In the CPMG experiment (A), the total pulse train length is $2\tau n$ and broad signals are selectively suppressed by T_2 relaxation. In the STE experiment (B), signal attenuation occurs as a function of molecular diffusion during the experimental diffusion delay time, Δ and by T_2 relaxation during the gradient pulses. The signal-to-noise ratio (S/N) is decreased for all signals due to gradient coherence selection. g = Gradient amplitude, δ = duration of gradient pulse, τ_r = gradient recovery delay time. AQ is the acquisition period. An optional homospoil gradient (not shown) can be applied after the second 90° pulse to destroy residual magnetization in the xy plane.

These distortions become increasingly significant as more pulses are applied [14,15] and can be exacerbated by water suppression techniques [16]. Phase and baseline errors can significantly complicate the identification and quantitation of small molecules, especially when the analytes of interest are present in the sample at low concentrations ($\leq\mu$M).

Pulsed field gradient filtering is a promising alternative to relaxation-based methods for selectively detecting molecules of varying sizes. The stimulated echo (STE) experiment, shown schematically in Fig. 1B, can be used for gradient filtering using a single value of the gradient amplitude (g). Diffusion coefficients (D) are calculated by acquiring a pseudo-two-dimensional data set as a function of g and measuring the decay in signal intensity (I), according to Eq. (1) [17].

$$I = I_o \exp\left[-D(\gamma\delta g)^2\left(\Delta - \frac{\delta}{3}\right)\right] \quad (1)$$

In Eq. (1), I_o is the signal intensity in the absence of the gradient pulse and γ is the gyromagnetic ratio, an intrinsic property of the nucleus being analyzed. Experimental parameters like the gradient pulse duration, δ, and the diffusion time, Δ, also affect signal attenuation but are usually fixed for a given experiment.

Pulsed field gradients attenuate magnetization based on the diffusivity of the corresponding nuclei, leading to greater signal attenuation for quickly diffusing small molecules compared to slowly diffusing larger molecules [18]. Small molecules are edited from the spectrum when the diffusion experiment is performed at high gradient amplitude. This enables detection of large biological macromolecules (or small molecules bound to them), with fast T_2 relaxation times that are difficult to detect selectively by CPMG experiments. Small molecules themselves can be identified by subtracting a diffusion spectrum acquired at high gradient amplitude (large molecules only detected) from one acquired at low gradient amplitude (large and small molecules detected) [9,10,19–21]. Because fewer high power radio frequency (rf) pulses are applied in the STE experiment compared to the CPMG experiment, fewer phase anomalies in resulting spectra are observed. Resulting spectra are readily processed using programmed algorithms in the NMR software, thus reducing the data analysis time. Highly reliable metabolic profiles are also generated, an important consideration in the application of metabonomics for disease diagnostics [22]. The robustness of the CPMG and STE spectra to automatic phasing and baseline correction routines was investigated further to determine if the STE experiment is a viable alternative to CPMG for metabonomic analysis.

2. Experimental

2.1. Serum samples

Human serum was collected from a healthy female donor by venipuncture into a red top vacutainer (Becton, Dickinson and Company, Sparks, MD). The blood was allowed to clot (without additives) at room temperature for 30 min and then centrifuged to separate the serum. Sample aliquots were frozen at −70 °C until analysis. The serum (900 μL) was diluted by adding 100 μL D_2O (99.96% atom D, Cambridge Isotope Labs Andover, MA), used to obtain a deuterium lock signal for the NMR spectrometer. Inspection of a ^{1}H NMR spectrum revealed the sample to be stable at room temperature or lower for several days. Please note: Serum is a potential biohazard and should be handled and disposed of according to local and federal safety regulations.

2.2. NMR spectroscopy

All spectra were acquired on a Bruker BioSpin AVANCE 500 MHz spectrometer equipped with a 5 mm triple resonance $\{^1$H, ^{13}C, ^{15}N$\}$ inverse detection probe. On resonance presaturation [23–25] was used to suppress the residual HOD signal at 4.71 ppm. The CPMG and one-dimensional stimulated echo (STE) pulse sequences were used to measure CPMG and gradient-filtered spectra, respectively (Fig. 1). In the CPMG experiments analyzed by principal component analysis (PCA), the T_2 delay (τ) was set to 0.5 ms and the pulse train was repeated 20 times (n = 20), yielding a total pulse train length ($2\tau n$) of 20 ms. For the STE experiments subject to PCA, a constant gradient amplitude (g) of 11.3 G/cm was applied for 1.5 ms (δ), and a 200 μs gradient recovery time (τ_r) was used. The diffusion delay time (Δ) was held constant at 100 ms. A homospoil gradient pulse (not shown in Fig. 1B) with an amplitude of −7.74 G/cm and duration of 1.0 ms was applied during the diffusion delay time to destroy residual magnetization in the xy plane [26,27]. The temperature was regulated at either 298 or 277 K using a Bruker BioSpin variable temperature accessory. In these experiments, 8 or 64 scans were acquired into 16444 complex points and the residual HOD signal was referenced to the external HOD signal in a one-dimensional ^{1}H spectrum

acquired without water suppression (4.71 ppm). The NMR probe was tuned and matched to the proton transmitter resonance frequency (500.13 MHz) for each sample. Other experimental conditions are as indicated throughout the text.

2.3. Data processing

The free induction decays (FIDs) were baseline corrected for DC offset by applying a filter that subtracts a constant value calculated from the last 25% of the FID (default mode) [28]. The data sets were zero-filled to 32,768 points, apodized with an exponential multiplier equivalent to 1 Hz line broadening, and Fourier-transformed. Unless otherwise noted, data sets were either automatically or interactively phased. A polynomial function was subtracted from each spectrum for additional baseline correction (automatically or interactively) if necessary to achieve a flat baseline at zero intensity.

2.4. Principal component analysis (PCA)[29]

The AMIX software program [30] (Bruker BioSpin Corporation) was used to perform PCA on the CPMG and STE data sets. The software utilizes bucketing, a method of data reduction to generate a manageable data set. For example, a ^{1}H NMR spectrum with 32,768 points over a spectral window of 10 ppm is reduced to 250 buckets 0.04 ppm wide. The peak intensity, normalized and scaled to the total intensity of the spectrum, excluding the solvent region from 4.62 to 6.00 ppm, was measured for each bucket, yielding a histogram representing the spectrum. Statistical analysis of bucketed spectra permitted comparison of the experimental reproducibility of the CPMG and STE results.

3. Results and discussion

Fig. 2 shows a direct comparison of the CPMG and gradient-filtered methods for analyzing serum samples. The spectrum in Fig. 2A was acquired with a CPMG filter and in Fig. 2B a pulsed field gradient filter was used. Overall, the spectra appear similar except for the reduced signal-to-noise ratio (*S*/*N*) in Fig. 2B due to the coherence selection of the pulsed field gradients [31]. While both methods suppressed broad components, superior water suppression was achieved with the gradient filter. When using the gradient filter there are two active modes of water suppression: on resonance saturation of the water signal (also active in the CPMG method) and diffusion editing, as water is generally the fastest diffusing molecule in the sample. Gradient filtering is advantageous for metabonomic analysis of biofluids since the chemical shift of the anomeric protons of sugars like β-glucose are near the residual HOD signal [5], and saturation methods affect the intensity of these signals.

Since metabonomic analysis of biofluids has medical diagnostic potential [22], rapid throughput is desirable. Accurate and precise calibration of the reference radio frequency (rf) pulse, necessary to achieve best results with the CPMG method, is a relatively lengthy procedure that often requires more time than needed to acquire a ^{1}H spectrum. In Fig. 3A, the reference 360° rf pulse was pre-calibrated on a buffered sucrose sample (5 mM in 0.2 M sodium phosphate, 90% H_2O/10% D_2O, v/v, pH* 7.4), chosen to closely mimic the ionic concentration of serum. The 180° pulse was calculated as half the length of the 360° pulse. The calibrated value serves as a reference 90° pulse length for samples in similar solvents, thus reducing the time required to optimize experimental conditions. When the calculated rf pulses are

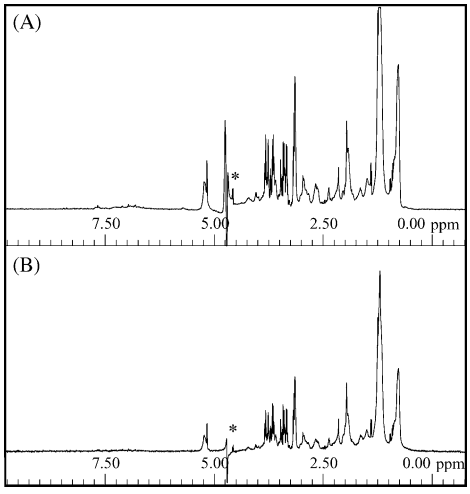

Fig. 2. Comparison of two NMR methods for human serum analysis. 500 MHz ^{1}H one-dimensional spectra of human serum acquired at 25 °C in eight scans using the CPMG (A) or STE (B) pulse sequence. Both methods employed presaturation to suppress the residual solvent signal at 4.71 ppm. The β-glucose anomeric proton resonance is indicated with an asterisk (*) for comparison with Fig. 6. For the CPMG data, $\tau = 0.5$ ms and $n = 20$. For the STE data, $g = 11.3$ G/cm, $\delta = 1.5$ ms, $\tau = 200$ μs, and $\Delta = 100$ ms.

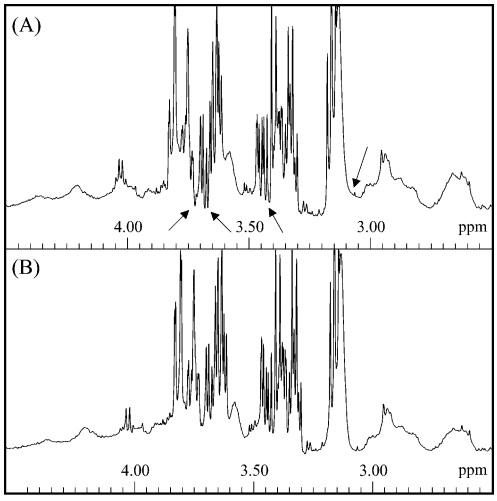

Fig. 3. ^{1}H CPMG spectra of human serum acquired at 25 °C. In (A), the 90° pulse was pre-calibrated on a buffered sucrose sample. The arrows indicate phase anomalies resulting from pulse inaccuracies. In (B), the 90° pulse was empirically optimized for the serum sample. Acquisition parameters are the same as those listed in Fig. 2.

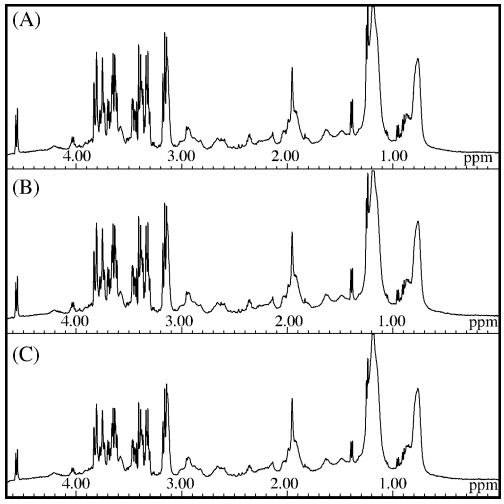

Fig. 4. ^{1}H STE spectra of human serum acquired at 25 °C with various experimental conditions. In (A), the gradient pulses were applied for 1.5 ms. The gradient amplitude (g) was 4.5 G/cm, the diffusion delay time (Δ) was 100 ms, and 32 scans were acquired. (B) Same as (A), g = 11.5 G/cm. (C) Same as (B), except g = 22.6 G/cm.

applied to human serum, phase artifacts (indicated by arrows in Fig. 3A) are evident near the baseline. These artifacts arise due to additive errors resulting from repeated application of imperfect 180° pulses, leading to incomplete refocusing of the magnetization. Incomplete irradiation across the bandwidth of the sample can also contribute to phase artifacts. Although the sample is homogeneous, not every physical region of the sample within the NMR coil receives the same pulse. The 360° pulse was optimized empirically on the serum sample to minimize the phase distortions, with the 180° pulse calculated as half the optimized value. As evident in Fig. 3B, the resulting spectrum is much easier to properly phase. The obvious phase artifacts in the CPMG spectra are easily remedied by calibrating the rf pulses for each sample, and incomplete irradiation across the bandwidth of the sample was not a significant problem in these studies.

Fig. 4 reveals that time-consuming 90° pulse optimization is unnecessary in the gradient-filtered experiments. All spectra shown in Fig. 4 were acquired with the reference 90° pulse pre-calibrated on buffered sucrose. It was evident that while the *S*/*N* varied concomitantly as the gradient parameters were changed, the phasing was relatively insensitive to the gradient amplitude (g) and diffusion time (Δ, data not shown). The gradient experiment has been demonstrated as more robust, providing spectra amenable to automatic phasing, and less operator time is required to optimize the data acquisition parameters compared to the CPMG method.

An important step toward automating biofluid analysis and increasing throughput is to define acquisition parameters and a data processing protocol that can be applied automatically to all experiments in the same data set. As illustrated in Figs. 2–4, CPMG and STE experimental parameters can be altered until the desired sample selectivity is achieved. Standard acquisition parameters can be saved in the NMR software and applied easily to a series of samples. For one-dimensional ^{1}H NMR experiments such as the CPMG and STE experiments, a standard processing protocol may include the following: (1) DC offset correction of the baseline in the time domain, (2) selection of an appropriate weighting function of the free induction decay (FID) to maximize *S*/*N* and/or resolution, (3) Fourier transformation of the time domain FID into a frequency domain spectrum, often with zero-filling or linear prediction to improve digital resolution, (4) adjustment of the phase of the resonances to yield a spectrum with positive intensity peaks, and (5) baseline correction in the frequency domain so that resonance integrals are measured with respect to a constant zero baseline across the whole spectrum. Phasing and baseline correction can be performed interactively until the spectroscopist observes the best result or automatically using programmed commands in the NMR processing software. The spectra shown in Figs. 2–4 were phased and baseline corrected interactively since the experimental parameters were being optimized. This becomes impractical when large numbers of spectra are acquired as a single spectroscopist may be responsible for all the data processing to reduce operator-dependent variations. To further assess the robustness of the CPMG and STE methods to automatic processing, groups of spectra were acquired by each method on a serum sample from one person.

The standard processing protocol outlined above was initially applied automatically to CPMG and STE data sets. DC offset, a vertical shift of the FID with respect to zero intensity, arises from a DC imbalance of the two channels of the quadrature detector [32]. This was corrected (unless otherwise noted) by calculating the average value of the last 25% of the FID and subtracting that constant from the FID [28]. An exponential weighting function equivalent to 1 Hz line broadening was applied to the FID to improve *S*/*N*. Resolution was enhanced by zero-filling the FIDs by a factor of 2, in which extra points of zero intensity are added to the end of the acquired data set. The automatic phasing routine used applies a zero-order (non-frequency dependent) phase correction to the tallest peak in the spectrum, in the case of serum, a lipid component at 1.2 ppm, followed by a first-order (frequency dependent) correction to the rest of the peaks in the spectrum [32]. The automatic baseline correction mode calculates up to a fifth-order polynomial function and subtracts it from the phase-corrected spectrum to yield a flat baseline across the entire spectral window [32].

PCA calculation in the AMIX software was used to characterize the overall variation in the individual sets of processed CPMG and gradient-edited STE spectra (Fig. 5). The results are displayed as two-dimensional scores plots, with the PC1 variable, accounting for the most variation in the data, in the *x* dimension and the PC2 variable in the y dimension. The scores plots shown in Fig. 5 reveal clustering of data sets according to a specific distinguishing variable; the dispersion of the data points within the cluster serves as a measure of the reproducibility of the spectra within a particular data set. Fig. 5A shows the data clusters for a set of 10 CPMG and

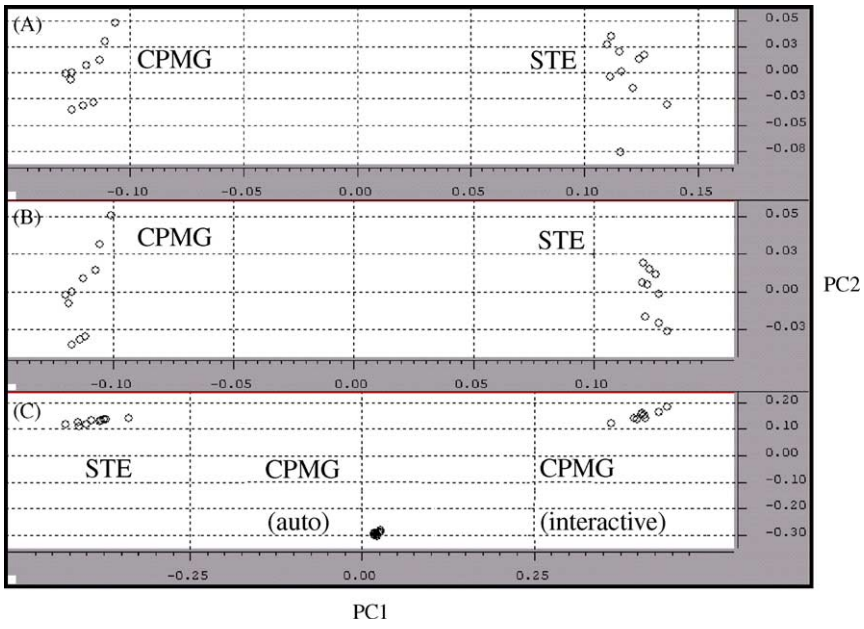

Fig. 5. Score plots resulting from PCA calculations to show the clustering, or reproducibility, of a series of spectra acquired under the same experimental conditions. The plots are shown as the first principal component (PC1) versus the second principal component (PC2). (A) Ten CPMG spectra and 10 STE spectra (ns = 8) were acquired at 25 °C and processed with automatic phasing and baseline correction routines. (B) Same as (A), except nine STE experiments were *S*/*N* normalized by acquiring 64 scans. One STE spectrum was eliminated from the data set using the *t*-test on the data in PC1 at a 99% confidence interval. (C) Nine CPMG experiments and 10 STE experiments (ns = 8) were acquired at 4 °C and processed automatically with the default DC offset correction mode (STE), automatically with a DC offset correction mode that incorporates subtraction of the solvent resonance (CPMG auto), or phased interactively with the default DC offset correction mode (CPMG interactive). One CPMG spectrum was eliminated from the data set using the *t*-test on the data in PC2 at a 99% confidence interval. Acquisition parameters are the same as those listed in Fig. 2.

10 STE spectra each acquired with eight scans (representative spectra shown in Fig. 2). The dispersion of the points in the two data clusters is similar, even though the *S*/*N* is less in the STE data due to application of the pulsed field gradients. Fig. 5B shows a significant improvement in the reproducibility of the STE data when the *S*/*N* ratio is normalized to that obtained with CPMG by acquiring more scans. The STE experiment is superior to the CPMG in terms of reproducibility of spectral replicates for the same sample, even when the experiments are interleaved. The loadings plot (not shown) revealed that the spectral region contributing most to the observed differences between the CPMG and STE data was that between 0.00 and 2.00 ppm. This region contains resonances of macromolecular components of serum such as lipids and albumin, whose integrals are expected to be significantly different in the two experiments due to the different experimental techniques employed.

An inherent problem in biofluid analysis is masked detection of resonances near the residual water signal. Presaturation-based water suppression methods may not solve this problem, since selective irradiation of a specific spectral region can attenuate nearby analyte resonances as well [16]. When analyzing serum at low temperature (4 °C), the chemical shift of the residual solvent signal shifts downfield, affording improved resolution for the anomeric proton doublet of β-glucose (4.34 ppm, compare Figs. 2 and 6). Lower temperatures should favor CPMG suppression of broad components because of reduced molecular mobility,

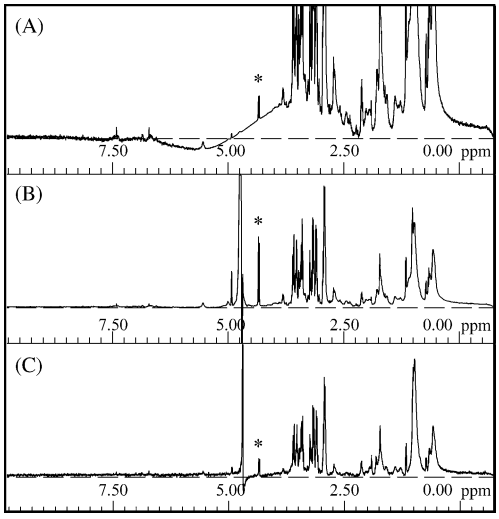

Fig. 6. Representative spectra for the PCA data shown in Fig. 5C of human serum analyzed at 4 °C. The β-glucose anomeric proton resonance is indicated with an asterisk (*). (A) Eight scans acquired by CPMG and processed automatically with the DC offset correction mode incorporating removal of the solvent resonance. (B) Same as (A), except the spectrum was phased interactively and the default mode was used for DC offset correction. (C) Eight scans acquired by STE using automatic phasing and baseline correction with the default DC offset correction mode. The broken gray lines indicate a zero offset (i.e., flat) baseline. Acquisition parameters are the same as those listed in Fig. 2.

leading to decreased T_2 relaxation times and diffusion coefficients. Fig. 6 shows, however, that while baseline roll and negative baselines are observed in the automatically-processed CPMG data (Fig. 6A), relatively flat baselines are observed for the STE data (Fig. 6C). Slightly different water suppression parameters were used for the CPMG and STE experiments acquired at low temperature. The differences observed in the spectra are due in part to the efficiency of the water suppression, which in itself is dependent on the quality of the shimming (adjustment of magnetic field homogeneity). For both experiments at low temperature, the residual water resonance was the tallest peak in the spectrum. After Fourier transformation of the STE data, the water peak dipped slightly below the baseline due to presaturation, but most of the solvent resonance was at positive intensity (data not shown). In the CPMG data, the water peak was also the most intense peak in the transformed spectra; however, the resonance shape was dispersive, giving rise to a large negative component that caused many of the serum peaks to be phased negative when the automatic phasing routine used the solvent resonance as the reference peak for phasing. The CPMG data could be automatically processed when an alternate FID baseline correction routine incorporating removal of the solvent resonance was applied to the FID before Fourier transformation [33]. As shown in Fig. 6A, this method subtracts the solvent resonance from the spectrum but does not minimize baseline roll. Baseline roll is only minimized when the data is phased interactively, as shown in Fig. 6B. Fig. 6C shows that the STE data is amenable to the standard automatic processing routine, yielding a spectrum containing less overall distortion. These results illustrate the sensitivity of the resulting spectra to the solvent suppression schemes and emphasize the importance of optimizing solvent suppression for each set of experimental conditions. For quantitative metabonomic analysis, consistent solvent suppression parameters should be used to avoid spectral anomalies such as those shown in Fig. 6, although optimizing these parameters for drastically different samples or samples analyzed on different days may be necessary.

Fig. 5C shows the PCA results for the data acquired at 4 °C. Interestingly, the automatically-processed CPMG data was the most reproducible. Fig. 6A shows, however, that despite this precision, the data acquired and processed in this manner are unacceptable for metabonomic analysis due to the rolling baseline that skewed many of the resulting resonance integrals to higher values. These observations reveal the importance of optimizing the processing parameters to yield reliable spectra for PCA and that different parameters may be necessary for data acquired at different temperatures. Less precise but more accurate spectra are obtained when the CPMG data are processed interactively (Fig. 5C), showing that even a single operator may bias the spectral reproducibility as each spectrum is processed until a desired result is observed visually. Figs. 5C and 6C reveal that the STE data can be automatically processed and yield comparable accuracy and reproducibility to the CPMG spectra processed interactively, thus saving time and increasing throughput while reducing user interaction. These results show that the STE method is a viable alternative to CPMG for metabonomic analysis of biofluids, such as serum, by high resolution NMR.

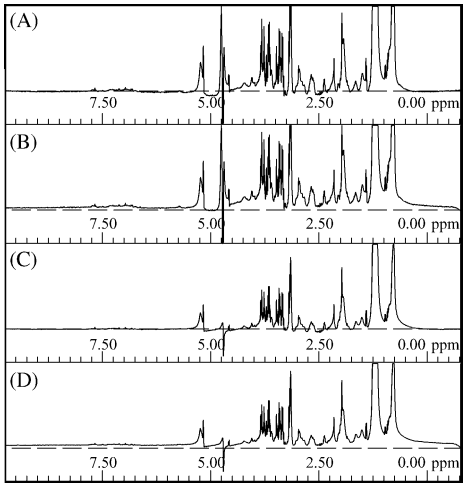

Fig. 7. Human serum analyzed at 25 °C with (A, C) and without (B, D) automatic baseline correction of the phased spectra. CPMG data are presented in spectra A and B while STE data are displayed in spectra C and D. Acquisition parameters are the same as those listed in Fig. 2.

Further investigation of Fig. 6 shows that the spectral noise is centered along the baseline (broken gray line) on the left-hand side of the spectrum while there is a slight positive offset of the right-hand edge of the spectrum relative to the baseline. This is corrected by an automatic determination of the fifth-order polynomial coefficients for the baseline correction routine of the processed and phased data. Baseline correction of serum data is somewhat complicated in this regard since the farthest upfield resonances arise from the aliphatic components of lipids and proteins and are inherently broad, making it difficult to determine the exact point of baseline resolution.

The CPMG and STE data sets acquired at 25 °C were reprocessed without automatic baseline correction in the frequency domain to determine what effect, if any, this had on the final spectra. Fig. 7 compares CPMG (A, B) and STE (C, D) spectra with (A, C) and without (B, D) automatic baseline correction. The results are quite striking: while automatic baseline correction flattens the baseline at the edges of the spectrum, many resonances of serum in the middle of the spectra dip below the baseline and would thus be erroneously integrated in metabonomic analysis (Fig. 7A and C). Bowing of the baseline suggests that the baseline adjustment algorithm overcorrected the baseline in some regions. The order of the polynomial applied for baseline correction can be set by the user (fifth order is the default). When less than a fifth order correction was used, the baseline did not improve significantly. Alternatively, as shown in Fig. 7B and D, there remains a slight offset of the spectra above the baseline when no baseline correction is applied. The offset appears greater farther upfield, where the spectral S/N is highest. In this spectral region (<2.00 ppm) there is also a greater con-

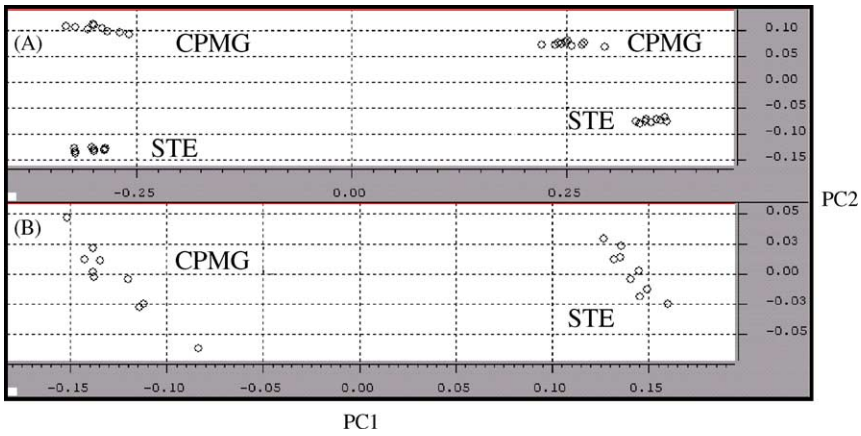

Fig. 8. Score plots resulting from PCA on data sets acquired with or without baseline correction of the phased spectra. The plots are shown as the first principal component (PC1) vs. the second principal component (PC2). (A) Ten CPMG spectra (ns = 8) and nine STE spectra (ns = 64) were acquired at 25 °C and processed with baseline correction (left set of data clusters) or without baseline correction (right set of data clusters). (B) Data without baseline correction (shown on the left in (A) for comparison with Fig. 5B (baseline correction applied). Acquisition parameters are the same as those listed in Fig. 2.

tribution to the spectra from the protein background (such as the aliphatic resonances of albumin) and lipids. Similar results (not shown) were observed for the data acquired at 4 °C. These observations indicate that standard baseline correction routines may be undesirable if DC offset correction of the time domain data is sufficient to yield a flat baseline in the transformed data.

Fig. 8 shows the PCA scores plots for the data acquired at 25 °C and processed without automatic baseline correction in the frequency domain. In Fig. 8A, the data processed with baseline correction is displayed on the left and the data processed without baseline correction is displayed on the right. Clustering in both cases is seen with respect to experiment type, with the CPMG data towards the top and the STE data towards the bottom of the scores plot. A narrowing between the clusters in PC2 is observed when baseline correction is not used, showing that the CPMG and STE data sets are more similar when the polynomial function is not subtracted from the processed data. Fig. 8B is analogous to Fig. 5B, except the results for spectra without baseline correction are displayed. The reproducibility of the gradient-filtered STE data is superior to the CPMG data.

4. Conclusions

The STE method was demonstrated as a robust alternative to CPMG for the analysis and automatic processing of ^{1}H NMR data of human serum. At first consideration it would appear inconsistent that a gradient filter would be useful for characterizing small organic molecules as their signals are attenuated more greatly by diffusion than the macromolecular serum components. The implementation of the gradient filter has a concomitant reduction in the *S*/*N* of approximately a factor of two, but signals of small molecules were detected when the gradient parameters were carefully selected (e.g., $g < 50\%$ maximum amplitude). The *S*/*N* was readily regained with more scans. Although this increases the data acquisition time for each sample, the total acquisition time using STE is similar to the total experimental time using CPMG, given that the rf pulse length must be optimized for best results with the latter method. In addition, the spectra obtained from the gradient-filtered experiment are well suited to automatic processing routines. This advantage makes the STE experiment attractive for analysis of biofluids, especially when broad signals do not need to be completely suppressed and quantitation of metabolites, the accuracy of which depends upon properly phased spectra, is important in characterizing the metabolic profile.

Implementing automatic processing schemes removes the need for a single individual to process the data. This represents an important opportunity in reducing labor costs as it is not unusual for these studies to encompass hundreds of samples, with all of the processing performed by one person to minimize variations in the spectra. To maximize this opportunity, however, acquisition and processing parameters should be selected carefully to ensure both the precision and accuracy of NMR data subject to metabonomic analysis by PCA.

Note added in proof

The authors recently became aware of another report on diffusion-based NMR analysis of serum [34].

Acknowledgements

The authors thank Marielle Delnomdedieu, Pfizer Global R & D, Drug Safety Evaluation for donating serum samples and for useful discussion. LHL acknowledges the support of the NIH Training Grant in the Dynamic Aspects of Chemical Biology (2 T32 GM08545) and the ACS Division of Analytical Chemistry for a graduate fellowship (sponsored by Eli Lilly).

References

[1] J.C. Lindon, E. Holmes, J.K. Nicholson, Anal. Chem. 75 (2003) 385A–391A.
[2] J.R. Everett, Bioanalysis of Drugs, Including Anti-allergics and Anti-asthmatics, in: Proceedings of the Ninth International Bioanalytical Forum, Guildford, UK, September 3–6, 1991; The Royal Society of Chemistry: Cambridge, 1992, A1, pp. 3–10.
[3] J. Vion-Dury, F. Nicoli, G. Torri, J. Torri, M. Kriat, M. Sciaky, A. Davin, P. Viout, S. Confort-Gouny, P.J. Cozzone, Biochimie 74 (1992) 801–807.
[4] J.K. Nicholson, P.J.D. Foxall, M. Spraul, R.D. Farrant, J.C. Lindon, Anal. Chem. 67 (1995) 793–811.
[5] J.C. Lindon, J.K. Nicholson, J.R. Everett, Annu. Rep. NMR Spectrosc. 38 (1999) 1–88.
[6] J.C. Lindon, J.K. Nicholson, E. Holmes, J.R. Everett, Concepts Magn. Reson. 12 (2000) 289–320.
[7] D.L. Rabenstein, J. Biochem. Biophys. Meth. 9 (1984) 277–306.
[8] D.L. Rabenstein, K.K. Millis, E.J. Strauss, Anal. Chem. 60 (1988) 1380A–1391A.
[9] M. Liu, J.K. Nicholson, J.C. Lindon, Anal. Chem. 68 (1996) 3370–3376.
[10] P.J. Hajduk, E.T. Olejniczak, S.W. Fesik, J. Am. Chem. Soc. 119 (1997) 12257–12261.
[11] Q.N. Van, G.N. Chmurny, T.D. Veenstra, Biochem. Biophys. Res. Commun. 301 (2003) 952–959.
[12] H.Y. Carr, E.M. Purcell, Phys Rev. 94 (1954) 630–638.
[13] S. Meiboom, D. Gill, Rev. Sci. Instr. 29 (1958) 688–691.
[14] J. Simbrunner, R. Stollberger, J. Magn. Reson., Ser. B 109 (1995) 301–309.
[15] Y.-Q. Song, J. Magn. Reson. 157 (2002) 82–91.
[16] B.C.M. Potts, A.J. Deese, G.J. Stevens, M.D. Reily, D.G. Robertson, J. Theiss, J. Pharm. Biomed. Anal. 26 (2001) 463–476.
[17] J.E. Tanner, J. Chem. Phys. 52 (1970) 2523–2526.
[18] C.S. Johnson Jr., Prog. NMR Spectrosc. 34 (1999) 203–256.
[19] M. Liu, J.K. Nicholson, J.A. Parkinson, J.C. Lindon, Anal. Chem. 69 (1997) 1504–1509.
[20] X. Zhang, C.-G. Li, C.-H. Ye, M.-L. Liu, Anal. Chem. 73 (2001) 3528–3534.
[21] M. Liu, H. Tang, J.K. Nicholson, J.C. Lindon, Magn. Reson. Chem. 40 (2002) 83–S88.
[22] J.T. Brindle, J.K. Nicholson, P.M. Schofield, D.J. Grainger, E. Holmes, Analyst 128 (2003) 32–36.
[23] D.I. Hoult, J. Magn. Reson. 21 (1976) 337–347.
[24] E.R.P. Zuiderweg, K. Hallenga, E.T. Olejniczak, J. Magn. Reson. 70 (1986) 336–343.
[25] P.J. Hore, Methods Enzymol. 176 (1989) 64–77.
[26] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, J. Magn. Reson., Ser. A 102 (1993) 241–245.
[27] B.A. Messerle, G. Wider, G. Otting, C. Weber, K. Wüthrich, J. Magn. Reson. 85 (1989) 608–613.
[28] Xwin-NMR Software Manual, Bruker Analytik GmbH, Ettlingen, Germany, 1997. pp. P-18-P-23, P-144-P-145.
[29] J.E. Jackson, A User's Guide to Principal Components, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc., New York, 1991.
[30] M. Spraul, P. Neidig, U. Klauck, P. Kessler, E. Holmes, J.K. Nicholson, B.C. Sweatman, S.R. Salman, R.D. Farrant, E. Rahr, C.R. Beddell, J.C. Lindon, J. Pharm. Biomed. Anal. 12 (1994) 1215–1225.
[31] B. Antalek, Concepts Magn. Reson. 14 (2002) 225–258.
[32] P. Bigler, NMR Spectroscopy: Processing Strategies, Second updated ed., Wiley-VCH, New York, 2000, pp. 183–184.
[33] D. Marion, M. Ikura, A. Bax, J. Magn. Reson. 84 (1989) 425–430.
[34] R.A. de Graaf, K.L. Behar, Anal. Chem. 75 (2003) 2100–2104.