Journal of Magnetic Resonance **144,** 35–44 (2000)
doi:10.1006/jmre.2000.2021, available online at http://www.idealibrary.com on IDEAL®

Automatic Correction for Phase Shifts, Frequency Shifts, and Lineshape Distortions across a Series of Single Resonance Lines in Large Spectral Data Sets

H. Witjes,* W. J. Melssen,* H. J. A. in 't Zandt,† M. van der Graaf,† A. Heerschap,† and L. M. C. Buydens*

**Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands; and*
†Department of Radiology, University Hospital Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

Received August 4, 1999; revised December 28, 1999

A new model-free method is presented that automatically corrects for phase shifts, frequency shifts, and additional lineshape distortions of one single resonance peak across a series of *in vivo* NMR spectra. All separate phase and frequency variations are quickly and directly derived from the common lineshape in the data set using principal component analysis and linear regression. First, the new approach is evaluated on simulated data in order to quantitatively assess the phase and frequency shifts which can be removed by the proposed correction procedure. Subsequently, the value of the method is demonstrated on *in vivo* ^{31}P NMR spectra from skeletal muscle of the hind leg of the mouse focusing on the phosphocreatine resonance which is distorted by the experimental procedure. Phase shifts, frequency shifts, and lineshape distortions with respect to the common lineshape in the spectral data set could successfully be removed. © 2000 Academic Press

Key Words: in vivo **NMR; automatic correction; phase and frequency shifts; lineshape distortions; principal component analysis.**

INTRODUCTION

In vivo NMR spectroscopy examinations may result in large data sets of spectra, e.g., spectra from time series or from spectroscopic imaging experiments. Separate analysis of each spectrum is cumbersome while automated analysis of complete data sets may pose some serious problems. Phase, line position, linewidth, and lineshape of resonance peaks are not identical for all spectra due to experimental variations. This may cause problems for peak integral estimations by various least-squares approaches proposed for this purpose (*1–4*). Lineshape variations throughout series of spectra hamper the correct peak area estimation since the least-squares approaches assume one particular model lineshape function (Lorentzian, Gaussian). Phase and frequency shifts throughout spectra demand user interaction to perform accurate quantitation on complete series of *in vivo* NMR spectra. Spurious spectral variations may also cause problems for classification purposes using pattern recognition methods like discriminant analysis and principal component analysis (PCA) (*5*). They may lead to misinterpretation and misclassification of the spectra because they will contribute to the nonnoisy part of the signal and hence will be described by the signal-related principal components (*6*). Thus, prior removal of these kind of spectral variations may be an important preprocessing step for both the accurate classification and the automatic quantitation of (large) spectral data sets.

First attempts have been made on the fast and automatic quantitation of complete *in vivo* NMR spectral data sets (*7–9*). These studies have shown that PCA is a promising technique for this purpose, that is, PCA estimates more quickly and more precisely the peak integral of a series of noisy peaks than least-squares approaches operating on each spectrum separately. However, the PCA quantitation method is sensitive to phase and frequency variations throughout the spectral data set; it is limited to spectral data sets containing one or more spectrally isolated peaks which may only vary in amplitude.

Therefore, Brown and Stoyanova developed a PCA-based correction method which automatically eliminates phase and frequency shift variations of one single resonance peak across a series of spectra (*10*). Their method is especially dedicated to the analysis of large spectral data sets because it simultaneously determines all phase and frequency shifts across the data set within one calculation step. Moreover, the shifts are derived in a model-independent way, i.e., without assuming a particular model lineshape function. They have tested their method both on simulated and on experimental data. Their results on simulated data have demonstrated that it can accurately handle phase shifts up to 45° and frequency shifts of less than half a linewidth. Their experimental results depict that other spectral variations than phase and frequency shifts, like lineshape variations, could not be compensated for by using their approach.

This paper deals with an improved method for quickly removing phase and frequency shifts in large spectral data sets. It is a simple, robust, and flexible method which combines PCA and linear regression to obtain all separate phase and frequency shifts within one calculation step. Phase and frequency shifts are determined unambiguously and precisely in

1090-7807/00 $35.00

(large) spectral data sets. Furthermore, larger phase shifts (up to 60°) and larger frequency shifts (up to two times the linewidth) can easily be removed. An additional advantage of the new method is that it can deal with lineshape variations across a series of spectral lines. The new method has been tested on simulated data as well as on time series of *in vivo* ^{31}P NMR spectra.

THEORY

General

Each complex spectrum S_k (without noise) in a large series of spectra which have the same lineshape function, but are possibly shifted in phase and/or frequency, can be described by

$$S_k(\omega_j) = A_k e^{i\varphi_k} f(\omega_j + \Delta\omega_k), \quad [1]$$

where $S_k(\omega_j)$ is the discrete signal of the kth spectrum at frequency ω_j, A_k is the amplitude of spectrum k, $i = \sqrt{(-1)}$, φ_k is the phase offset of spectrum k, $f(\omega_j)$ is the complex lineshape function, and $\Delta\omega_k$ is the frequency offset of spectrum k. The basic idea is to expand $f(\omega)$ in a Taylor series (*10*). Equation [1] then becomes

$$S_k(\omega_j) = A_k e^{i\varphi_k}\left\{ f(\omega_j) + \left|\frac{df}{d\omega}\right|_{\omega_j} \Delta\omega_k + \frac{1}{2}\left|\frac{d^2f}{d\omega^2}\right|_{\omega_j} \Delta\omega_k^2 + \cdots \right\}, \quad [2]$$

where the third and higher order terms are neglected for the sake of convenience. The complex terms can be rewritten as

$$e^{i\varphi_k} = \cos\varphi_k + i\sin\varphi_k \quad [3a]$$

$$f(\omega_j) = f_R(\omega_j) + i f_I(\omega_j), \quad [3b]$$

where $f_R(\omega_j)$ and $f_I(\omega_j)$ are the real and imaginary parts of $f(\omega_j)$. After substitution of Eqs. [3a] and [3b] in Eq. [2], the following expression for the real (physical) part of the signal is obtained:

$$S_k^R(\omega_j) = A_k\Big\{ \cos\varphi_k f_R(\omega_j) - \sin\varphi_k f_I(\omega_j) + \Delta\omega_k\cos\varphi_k f'_R(\omega_j) - \Delta\omega_k\sin\varphi_k f'_I(\omega_j) + \frac{1}{2}\Delta\omega_k^2\cos\varphi_k f''_R(\omega_j) - \frac{1}{2}\Delta\omega_k^2\sin\varphi_k f''_I(\omega_j) + \cdots \Big\}, \quad [4]$$

where $S_k^R(\omega)$ is the real part of $S_k(\omega)$, $f'_R(\omega)$ is the first derivative of $f_R(\omega)$, and so on. Equation [4] can be rewritten as

$$S_k^R(\omega_j) = b_{k,1} f_R(\omega_j) + b_{k,2} f_I(\omega_j) + b_{k,3} f'_R(\omega_j) + b_{k,4} f'_I(\omega_j) + b_{k,5} f''_R(\omega_j) + b_{k,6} f''_I(\omega_j) + \cdots, \quad [5]$$

where $b_{k,i}$ is the ith Taylor coefficient of spectrum k. The second and third Taylor coefficients ($b_{k,2}$ and $b_{k,3}$) contain the phase and frequency shift information of each spectrum k, respectively,

$$\frac{-b_{k,2}}{b_{k,1}} = \frac{A_k\sin\varphi_k}{A_k\cos\varphi_k} = \tan\varphi_k, \quad [6a]$$

$$\frac{b_{k,3}}{b_{k,1}} = \frac{\Delta\omega_k A_k\cos\varphi_k}{A_k\cos\varphi_k} = \Delta\omega_k. \quad [6b]$$

Brown and Stoyanova (B&S) Method for Removing Phase and Frequency Shifts

A short description of the existing method of Brown and Stoyanova is given below since its performance will be compared with that of our novel method. For further details the reader is referred to Ref. (*10*).

Brown and Stoyanova apply principal component analysis (*11*) to series of real-valued spectra containing phase and frequency shifts. They use the first principal component (PC_1) of the series of spectra to estimate the real part of the unknown resonance lineshape function in the data set. They observed that at least two more additional PC spectra of the data set contain the signal-related variance due to the phase and frequency shift differences among the spectra. Therefore, they use the second and third PC loading spectra of the series of spectra (PC_2 and PC_3) to estimate the phase and frequency shift of each individual spectrum. A transformation matrix is calculated which transforms the coordinate system defined by (PC_1, PC_2, PC_3) to a new coordinate system defined by (PC_1, PC'_1, $PC_{1,I}$). Subsequently, the scores of each spectrum on the PC_2 and PC_3 spectra are projected onto this new coordinate system exploiting the above transformation matrix. These projections (scores T) are used to approximate each spectrum k and its (possible) phase and frequency shifts according to

$$S_k^R(\omega_j) = T_{k_{PC_1}} PC_1(\omega_j) + T_{k_{PC_{1,I}}} PC_{1,I}(\omega_j) + T_{k_{PC'_1}} PC'_1(\omega_j), \quad [7]$$

with (see Eqs. [6a] and [6b])

$$\tan\varphi_k = \frac{-T_{k_{PC_{1,I}}}}{T_{k_{PC_1}}} \quad \text{and} \quad \Delta\omega_k = \frac{T_{k_{PC'_1}}}{T_{k_{PC_1}}}. \quad [8]$$

In this way, phase and frequency shifts across a series of resonances are simultaneously estimated via a target transformation of the PC_2 and PC_3 score vectors.

Novel Method for Removing Phase and Frequency Shifts

In this paper we propose the application of linear regression to derive the phase and frequency shifts directly from the PC_1 spectrum instead of deriving these from the PC_2 and PC_3 spectra. In this way, the data transformation is circumvented which obviously benefits the high precision and robustness of the novel method. In addition, the use of linear regression easily allows the incorporation of higher order terms of the Taylor series expansion (Eq. [5]) in order to extend the signal approximation to spectra containing larger phase and frequency shifts. If the shifts are relatively large, second and higher order terms in the Taylor series will be needed to approximate more accurately the phase- and frequency-shifted spectra. Then, the second and third Taylor coefficients better reflect the true phase and frequency shifts of each spectrum, respectively.

The proposed phase- and frequency-shift correction procedure is as follows.

Step 1: PCA is applied to the series of real-valued spectra for obtaining the PC_1 spectrum of the data set.

Step 2: Using linear regression, each spectrum k is approximated by a linear combination of the PC_1, the $PC_{1,I}$, and one or more of their signal-containing derivatives:

$$\begin{aligned} S_k^{R}(\omega_j) = {} & b_{k,1}PC_1(\omega_j) + b_{k,2}PC_{1,I}(\omega_j) \\ & + b_{k,3}PC_1'(\omega_j) + b_{k,4}PC_{1,I}'(\omega_j) \\ & + b_{k,5}PC_1''(\omega_j) + b_{k,6}PC_{1,I}''(\omega_j) + \cdots \end{aligned} \quad [9]$$

With the classical least-squares solution the coefficients in Eq. [9] for all spectra k are obtained within one calculation step.

Step 3: The numerical values for the phase and frequency shifts of each spectrum k in the data set are obtained from the regression coefficients on the $PC_{1,I}$ and the PC_1', spectra, respectively (see Eqs. [6a] and [6b]),

$$\tan \varphi_k = -\frac{b_{k,2}}{b_{k,1}},$$

$$\Delta\omega_k = \frac{b_{k,3}}{b_{k,1}}. \quad [10]$$

Step 4: Finally, the spectra are corrected for the calculated shifts. The phase corrections are applied in the frequency domain by multiplying the spectra with a factor $\exp(-i\varphi_k)$, while the frequency shift corrections are applied in the time domain. This is done by an inverse FFT (fast Fourier transform) of the spectra, multiplying each time signal with $\exp(-i\Delta\omega_k t_j)$ and transforming them back using FFT. In this way, frequency shift corrections exceeding the actual frequency resolution can be applied.

By repeating the whole procedure (steps 1–4) the spectra are updated until convergence occurs. After convergence, the PC_2 and PC_3 spectra only describe noise, indicating that the spectra have been corrected for the phase and frequency shifts (that is, PC_1 approximates f_R well).

Novel Method for Additional Removal of Lineshape Variations

After phase and frequency shift correction each resonance line is phased and is shifted to the same spectral position in the data set. However, especially with *in vivo* NMR signals, the resonance lineshape function may be still distorted. Such a distorted line can be described by a sum of undistorted lines with varying amplitudes and slightly different frequencies. We propose here a simple and model-free method to quickly remove lineshape distortions in large spectral data sets. This is done by fitting each (distorted) single resonance line to the sum of a number of PC_1 lineshapes of the complete data set which may vary in amplitude A_i and in frequency offset $\Delta\omega_i$.

$$S_k(\omega_j) = \sum_{i=1}^{n} A_{k,i}PC_1(\omega_j + \Delta\omega_{k,i}), \quad [11]$$

where $PC_1(\omega_j)$ is used to approximate the underlying unaffected resonance lineshape. Here, n is the number of PC_1 lines, i denotes the ith PC_1 line with amplitude A_i and frequency offset $\Delta\omega_i$, and S_k denotes the real (measured) spectrum. The problem is to find the amplitudes and frequency offsets from a limited number of PC_1 lines for each spectrum k.

The proposed correction procedure for lineshape variations is as follows.

Step 1: Application of PCA to the series of phase- and frequency-shift corrected spectra for obtaining $PC_1(\omega_j)$.

Step 2: Based on the Taylor expansion discussed in the previous section, a first-order Taylor approximation of each spectrum k dedicated to the above problem is described:

$$S_k(\omega_j) = \sum_{i=1}^{n} c_{k,i}PC_1(\omega_j + \Delta\omega_{k,i}) + \sum_{i=1}^{n} d_{k,i}PC_1'(\omega_j + \Delta\omega_{k,i}). \quad [12]$$

A set of n starting frequency offsets is subjected to the n PC_1 lines for each spectrum k.

Step 3: Equation [12] is solved using linear regression. The coefficients $c_{k,i}$ contain the amplitude information of each PC_1 line, while the frequency offset adjustment δ of each PC_1 line is acquired in the same way as shown in Eq. [6b]:

$$\delta_{\Delta\omega_{k,i}} = \frac{d_{k,i}}{c_{k,i}}. \quad [13]$$

Step 4: The PC_1 lines are updated with the calculated δ-values and linear regression is started again. In an iterative way all spectra are simultaneously approximated by the sum of n PC_1 lines with varying amplitudes and frequency offsets for each spectrum k.

Step 5: After convergence, the removal of the peak distortions is accomplished by first subtracting the calculated spectrum from the measured spectrum. Then, the n PC_1 lines producing the calculated resonance line are shifted to the spectral position of the common lineshape of the data set (i.e., the PC_1 lineshape with $\Delta\omega = 0$). Finally, they are added to the remaining signal.

In this way, a major part of the lineshape variations across the data set is quickly eliminated by transforming simultaneously all distorted resonance lines to the common PC_1 lineshape of the data set.

EXPERIMENTAL

Data Processing

All data processing was performed using Matlab 4.2c software (The Mathworks, Inc.).

Simulated Data

Simulated data were generated to assess phase and frequency shift removal by the proposed method and by the existing B&S method (*10*). In the simulation study 100 spectra of 512 data points were generated. Each spectrum contained a single Lorentzian line centered at position 256 with linewidth $\tau = 30$ data points and amplitude A. Gaussian-distributed white noise ($\mu = 0$, $\sigma = 1$) was added to the spectra. This procedure was repeated several times in order to generate various simulated data sets. Each simulated data set was subjected to one amplitude variation ($A = 20$, 50, or 80, in arbitrary units) and to two uniform distributions of phase and frequency shifts. In this way, each spectrum in a data set of 100 spectra contained a single Lorentzian line of some signal-to-noise ratio (S/N, defined as $A/2{*}\sigma$) varying independently to a certain degree in phase and in spectral position. Figure 1a shows five spectra from one simulated example data set with $S/N = 25$ which vary in phase (between $-45°$ and 45°) and in line position (between $256 - \tau/2$ and $256 + \tau/2$).

Experimental Data

The experimental data set consists of a total of 108 *in vivo* ^{31}P NMR spectra of the skeletal muscle of the hind leg of four mice (M1 to M4), acquired at 7 T (*12*). During one measurement session 4 spectra were acquired during a rest period, 14 spectra under ischemic conditions, i.e., occlusion of the hind limb, and 9 spectra during recovery from ischemia. The time resolution was 108 s ($n = 12$, TR = 9 s). The occlusion by a diaphragm plate caused displacements of the leg in the coil and subsequently changed the shimming. This resulted in shifts and asymmetric broadening of the resonance lines. The spectral region around the single phosphocreatine (PCr) resonance was cut from the spectra and was first subjected to the novel phase- and frequency-shift correction procedure. The PCr lineshape is expected to be rather insensitive to biochemical changes and, hence, its variations throughout the time series are entirely due to instrumental or unwanted experimental variations (*13*). Therefore, the PCr peak is often used as a reference peak for eliminating undesired phase and frequency shifts in the individual ^{31}P NMR spectra. After removal of the undesired phase and frequency shifts additional lineshape distortions were removed from the PCr resonances.

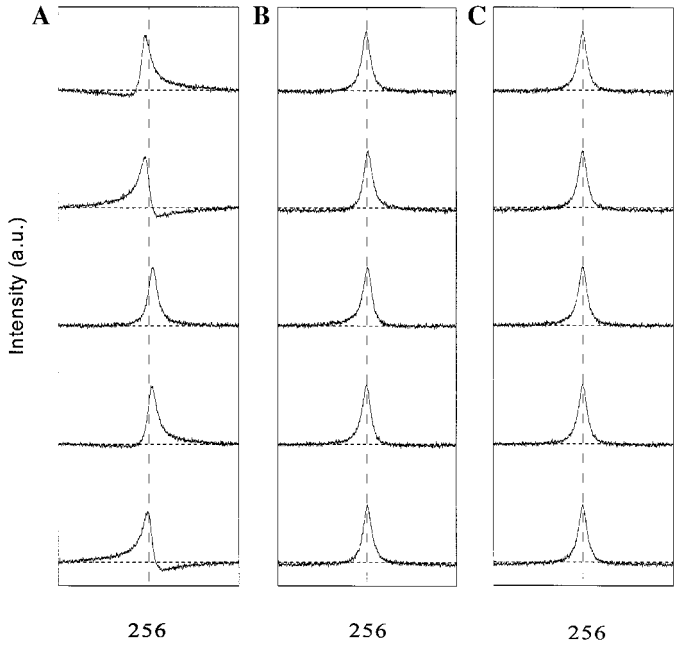

FIG. 1. Illustration of the phase- and frequency-shift correction procedure. (A) Five simulated spectra showing different phases and frequency offsets. (B) Same spectra after one cycle of the correction procedure. (C) Same spectra after three cycles of the correction procedure.

Correction for Phase and Frequency Shifts

The proposed phase- and frequency-shift correction procedure starts with a PCA of the series of spectra using the SVD (singular value decomposition) algorithm. The SVD algorithm is applied to the data matrix in which the rows represent the spectra and the columns represent the spectral frequencies. It calculates the first principal component (PC_1) spectrum normalized to unit length. The imaginary part of the PC_1 spectrum ($PC_{1,I}$) is obtained by the Hilbert transform, while the derivatives of the PC_1 and its imaginary part are obtained by numerical differentiation. Subsequently, the PC_1, the $PC_{1,I}$, and one or

more of their derivative spectra showing signal-related intensities are included in the regression analysis. After linear regression, each spectrum k is corrected for the calculated phase and frequency shifts (see Theory). The corrected spectra are once again subjected to the correction procedure until the shift corrections become negligibly small, i.e., phase corrections $<0.1°$ and frequency-shift corrections <0.1 data point. One full iteration cycle of the automatic correction procedure involving 100 spectra of 512 data points takes about 7 s on a SUN Ultra 10 workstation.

Quantitative Simulation Study

Phase- and frequency-shift estimations in a series of 100 simulated spectra were evaluated as a function of the size of the phase and frequency shifts, the S/N of the spectra investigated, and the PC_1 (signal-related) spectra included in the regression model. The existing B&S method was excluded from this study due to its poor convergence characteristics. However, one may assume that since the existing method indirectly uses the PC_1, $PC_{1,I}$, and PC_1' spectra for the signal approximation (see Theory), the size of the phase and frequency shifts which can be removed by the existing method is in potential similar to those which can be removed by the proposed regression method using only the PC_1, $PC_{1,I}$, and PC_1' spectra. In the quantitative simulation study the uniform phase shift distribution applied to a data set of 100 undistorted spectra varied from $[-5°, 5°]$ up to $[-90°, 90°]$ in which the interval size varied in steps of 5° in both directions. The applied frequency shift distribution differed from $[-\tau/10, \tau/10]$ up to $[-3\tau, 3\tau]$ in steps of $\tau/10$ in both directions. For each combination of one phase and one frequency shift interval five independent data sets were generated and, subsequently, were evaluated by the proposed correction procedure. It was decided that the novel correction procedure failed if 1 or more of the 100 spectra was not tuned accurately in phase and/or in frequency. This criterion could be handled since the application of larger phase and frequency shifts yielded more and more deteriorated spectra. If all data points in the performance plot depicting the true versus estimated shifts followed a straight line with slope equal to 1, e.g., in Fig. 2A, then the number of successful corrections was increased by 1. If not, the correction procedure had failed. The decision as to whether the correction procedure had succeeded or failed was based on the variance of the data points around the straight line. If the variance exceeded a predefined maximum threshold level then the correction procedure had failed. This decision criterion was sensitive enough to detect 1 false spectrum among 99 well-adjusted spectra. However, different threshold levels were defined for the different S/N data sets because it was found that the precision of estimation of the separate phase and frequency shifts decreased with decreasing S/N of the spectra.

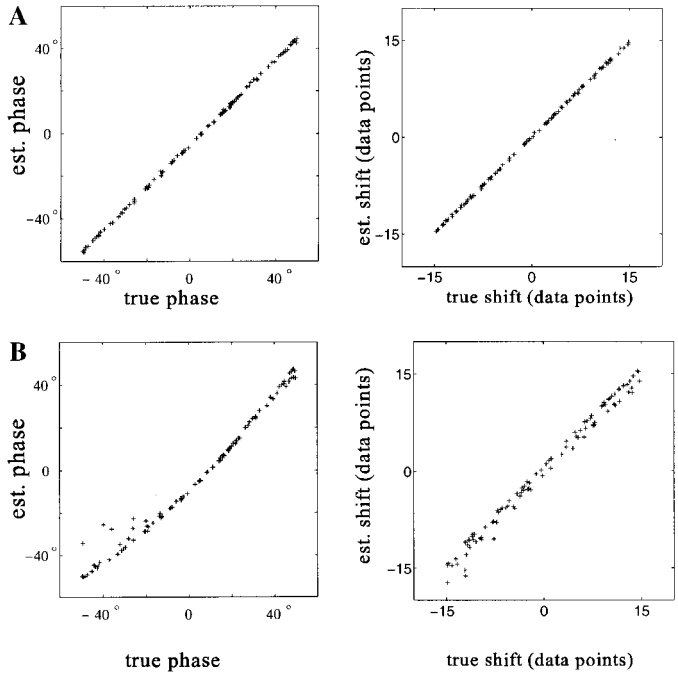

FIG. 2. Performance of the proposed correction procedure and the existing correction procedure on the entire example data set. (A) True versus estimated phase (left) and true versus estimated frequency offset (right) for each individual spectrum using the novel method. (B) True versus estimated phase (left) and true versus estimated frequency offset (right) for each individual spectrum using the B&S method.

Additional Correction for Lineshape Variations

The removal of significant lineshape distortions has been applied to the series of PCr resonances after phase- and frequency-shift correction. In our application study three PC_1 lines were sufficient for accurately fitting each spectrum to Eq. [11]: low-amplitude values were obtained for additional PC_1 lines with calculated frequency offsets significantly different from those already acquired. The initial frequency offsets subjected to the three PC_1 lines were systematically varied to be sure that each distorted resonance line is fitted well. A total of 11 experiments were performed, each starting with different initial frequency offsets (see Table 1). After convergence, the best of the 11 calculated spectra for each spectrum was selected using the minimum Euclidean distance between the PCr spectrum and each of these calculated spectra. Only a restricted spectral region was considered to prevent the possible fit of baseline noise. The best fit to each spectrum was used to remove the lineshape distortions in the way which is described under Theory.

Verification of Signal Corrections Using PCA Quantitation

The complete series of PCr resonances was quantitated before and after correction for phase shifts, frequency shifts, and lineshape variations using the alternative PCA quantitation method. The application of the PCA method is only justified if

TABLE 1

Fitting of the Complete Series of PCr Resonances Using Different Initial Combinations for the Three Frequency Shifts $\Delta\omega_1$, $\Delta\omega_2$, and $\Delta\omega_3$

Experiment number	$\Delta\omega_1$ (data points)	$\Delta\omega_2$ (data points)	$\Delta\omega_3$ (data points)
1	0	0	0
2	−2	0	2
3	−4	0	4
4	−6	0	6
5	−8	0	8
6	−10	0	10
7	−12	0	12
8	−14	0	14
9	−16	0	16
10	−18	0	18
11	−20	0	20

Note. The shifts are expressed in real data points; each PCr spectrum consists of a total of 256 data points.

the PCr resonance lines throughout the data set have been released from phase and frequency variations. Therefore, in order to verify the signal corrections made, the quantitation results before and after correction are compared with those obtained with a standard nonlinear curve-fitting approach using the complex Lorentzian lineshape function. To compare the quantitation results of the four mice, the peak areas in each of the four time series of spectra were multiplied by a factor which equated the peak areas of the first spectra in each of the four series. This was done in order to compensate for differences between the four mice due to the amount of muscular tissue which was measured.

RESULTS

Simulated Data

Figure 1 illustrates the proposed phase- and frequency-shift correction procedure for five spectra (Fig. 1A) of a data set of 100 simulated spectra ($S/N = 25$). Also shown are the spectra after the first phase- and frequency-shift correction (Fig. 1B) and after three iterative corrections (Fig. 1C). In this case three iterations appear to be sufficient to align the spectra both in phase and in frequency. In this particular example, the regression model was defined by the signal-containing PC_1 spectra, i.e., the PC_1, the imaginary part of the PC_1 ($PC_{1,I}$), and their first-derivative spectra (PC_1' and $PC_{1,I}'$). The second and higher derivative spectra are omitted from the regression analysis because they only contain noise. Hence, they did not contribute to the signal approximation.

The performance of the automatic correction procedure for the entire data set is shown in Fig. 2A. It clearly demonstrates that all spectra are well-adjusted in phase and in frequency after three iterations. The PC_1 of the series of corrected spectra describes 98.4% of the total spectral variance. Note the high precision of the frequency-shift correction which exceeds the discrete spectral resolution. Figure 2B shows the results of the same data set based on the existing B&S method. These are obtained after applying a certain correction factor (R. Stoyanova, personal communication). It depicts that the precision of the phase- and frequency-shift estimations is less compared with those of the novel method. This is confirmed by the PC_1 of the series of corrected spectra which only describes 97.1% of the total data variance. In addition, Fig. 2B shows the best results of the existing method obtained after two iterations. However, after the next iteration the prediction of the shifts become worse, while a few iterations later again good approximations are obtained. In case linear regression is used stable (and more precise) phase- and frequency-shift solutions are obtained after three iterations. This is also the case when only the PC_1, $PC_{1,I}$, and PC_1' spectra are included in the regression model.

The results of the quantitative simulation study are summarized in the gray level images plotted in Fig. 3 to Fig. 5. Each so-called performance image is a 18×30 intensity matrix in which the pixel intensity varies between 0 and 5, representing

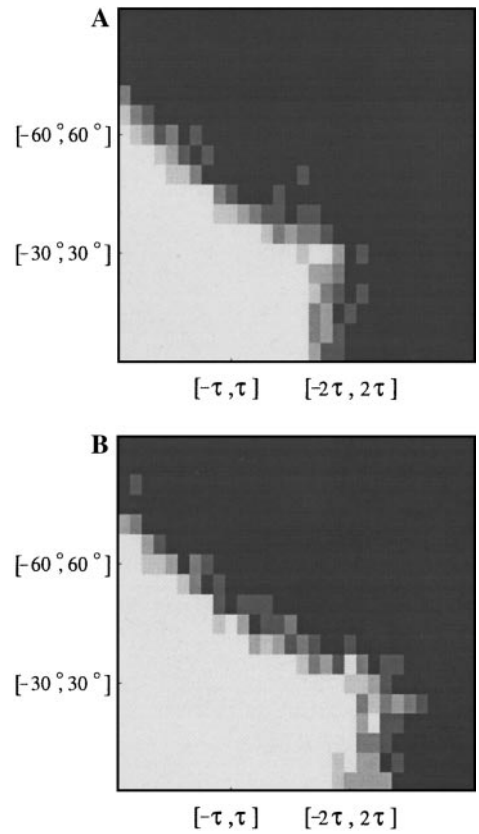

FIG. 3. Performance images showing the size of the phase and frequency shifts which can be removed by the proposed correction procedure; $S/N = 10$, τ = linewidth. (A) $\hat{S} = PC_1 + PC_{1,I} + PC_1'$. (B) $\hat{S} = PC_1 + PC_{1,I} + PC_1' + PC_{1,I}'$. White image regions denote a 100% performance.

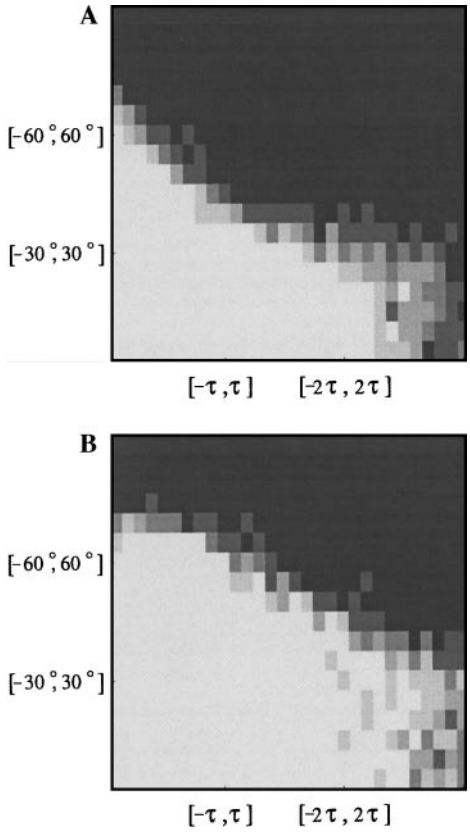

FIG. 4. Performance images for $S/N = 25$. (A) $\hat{S} = PC_1 + PC_{1,I} + PC'_1$. (B) $\hat{S} = PC_1 + PC_{1,I} + PC'_1 + PC'_{1,I}$. For more details see Fig. 3.

the number of successful corrections applied to the five independently generated sets of 100 spectra. Figure 3 shows the performance images of the automatic phase- and frequency-shift correction procedure applied to the series of spectra with $S/N = 10$. Figures 4 and 5 show the performance images of the correction procedure applied to the series of spectra with $S/N = 25$ and 40, respectively. The images in Figs. 3A, 4A, and 5A have been obtained using the regression model $\hat{S} = b^*_1 PC_1 + b^*_2 PC_{1,I} + b^*_3 PC'_1$, where $\hat{S}$ represents the estimated signal. Exploiting this model the maximum number of iterations is set by experience to 10: after 10 iterations it became clear whether each spectrum is either aligned or totally misaligned. Figures 3B, 4B, and 5B show the performance images obtained with $\hat{S} = b^*_1 PC_1 + b^*_2 PC_{1,I} + b^*_3 PC'_1 + b^*_4 PC'_{1,I}$. In this case the maximum number of 15 iterations appeared to be sufficient. Addition of the $PC'_{1,I}$ spectrum to the regression model increases the number of iterations before convergence is reached. Figure 5C depicts the results of exploiting the regression model $\hat{S} = b^*_1 PC_1 + b^*_2 PC_{1,I} + b^*_3 PC'_1 + b^*_4 PC'_{1,I} + b^*_5 PC''_1 + b^*_6 PC''_{1,I}$. The second-derivative spectra of the PC_1 and $PC_{1,I}$ of the $S/N = 40$ spectra show some signal and hence and included in the regression model. Using this model, the iteration procedure is stopped after 20 iterations. Evidently, the more terms are added to the regression model, the more iterations are required until convergence is reached.

Figures 3–5 depict that the white image regions corresponding to a 100% performance (thus with a pixel intensity equal to 5) are enlarged if the signal-related $PC'_{1,I}$ spectrum is incorporated in the regression model. On the other hand, incorporation of the PC''_1 and $PC''_{1,I}$ spectra into the regression model does not improve significantly the performance of the correction procedure (Figs. 5b–5c). This was expected since these second-derivative spectra show a very low S/N. This demonstrates that

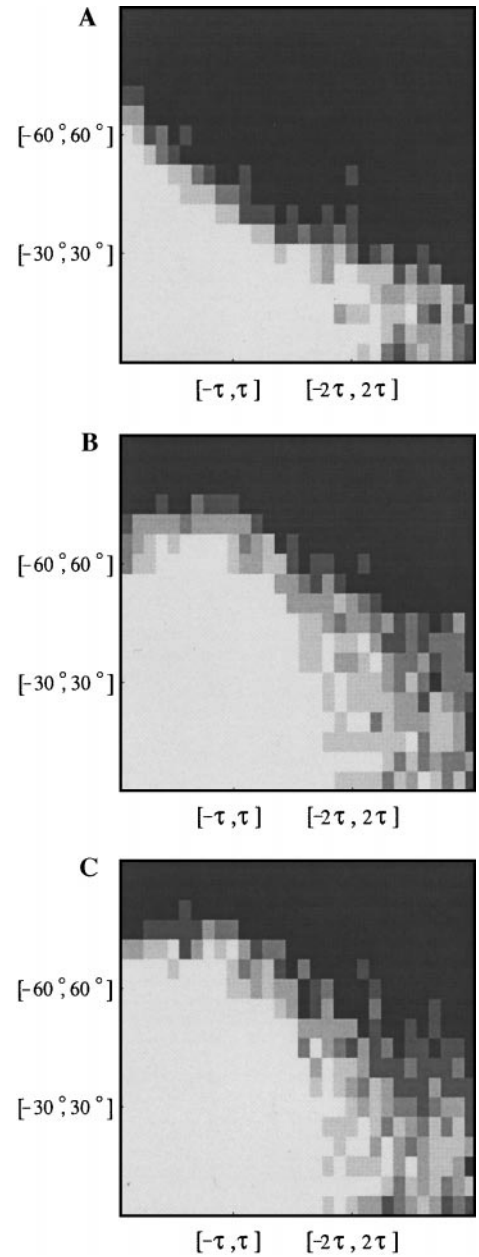

FIG. 5. Performance images for $S/N = 40$. (A) $\hat{S} = PC_1 + PC_{1,I} + PC'_1$. (B) $\hat{S} = PC_1 + PC_{1,I} + PC'_1 + PC'_{1,I}$. (C) $\hat{S} = PC_1 + PC_{1,I} + PC'_1 + PC'_{1,I} + PC''_1 + PC''_{1,I}$. For more details see Fig. 3.

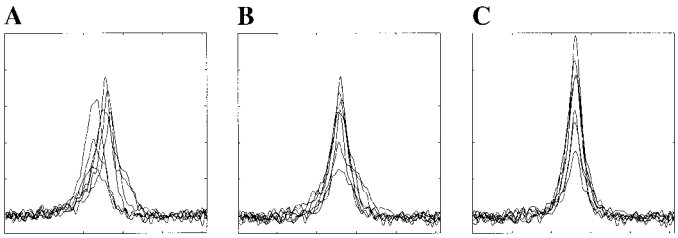

FIG. 6. (A) Seven representative ^{31}P NMR spectra of the mouse skeletal muscle containing the phosphocreatine (PCr) resonance. The same spectra after phase- and frequency-shift correction (B) and after transformation of the separate PCr resonances to the common lineshape of the data set (C). All spectra are plotted on the same arbitrary scale.

the experimental noise in the data set determines the number of useful PC_1 spectra defining the regression model. Appropriate filtering of the noisy PC_1 spectra would probably enable the regression method to deal with larger phase and frequency shifts. However, in this study filtering is omitted since our primary goal is to give only an impression of the sizes of the phase and frequency shifts which can be removed at least by the novel method. Figures 3 to 5 illustrate that addition of the signal-related $PC'_{1,I}$ spectrum to the regression model may improve the performance. On the other hand, more iterations are required until convergence is reached. This explains the failure of one of the five experiments for some combinations of phase- and frequency-shift intervals (e.g., Figs. 4A and 4B; $\Delta\omega \approx [-2\tau, 2\tau]$). In these cases the iterative procedure almost had adjusted all phase and frequency shifts toward negligible small values. However, the predefined maximum number of iterations did not allow convergence of the correction method for all shifted spectra.

In other words, the performance images in Figs. 3–5 give an impression of the phase and frequency shifts of a Lorentzian lineshape which can be treated with the proposed correction procedure. Obviously, addition of the $PC'_{1,I}$ spectrum enables the automatic correction procedure to correct for, in particular, larger phase shifts in the presence of small frequency shifts.

Experimental Data

PCr spectra obtained from the four different mice M1 to M4 show considerable line shift and line distortion effects, while the phases of the PCr resonances are rather similar (Fig. 6A). The PC_1 of the series of spectra describes 91.6% of the variance in the complete data set. The estimated noise variance is 1.6%.

Figure 6B shows the PCr spectra of Fig. 6A after phase- and frequency-shift correction using $\hat{S} = b_1^* PC_1 + b_2^* PC_{1,I} + b_3^* PC'_1 + b_4^* PC'_{1,I}$ (i.e., using the signal-containing PC_1 spectra). It depicts that all resonances are accurately phased and are shifted to the mean spectral position of the data set, irrespective of the lineshape distortions across the series of spectra. As a result, the PC_1 lineshape of the series of phase- and frequency-adjusted spectra explains 97.8% of the total variance in the data set. This indicates that for a greater part the unwanted spectral variations across the data set have been eliminated. The proposed method ended up with unique solutions for all phase and frequency shifts after four iterations. This is in contrast to the B&S method which produced phases and frequency shifts strongly varying iteration after iteration. Figure 6C shows the PCr spectra of Fig. 6B after transformation of the phase- and frequency-shift corrected PCr resonances to the common lineshape of the data set. Clearly, the lineshape distortions of the PCr resonances are substantially reduced. The PC_1 of the series of spectra finally describes 98.5% of the total variance present in the data set. This means that the remaining signal-related lineshape variations have successfully been removed. Figure 7 illustrates the lineshape correction procedure for one particular PCr resonance line.

The quantitation of the PCr resonances of the four mice before correction and after correction for phase shifts, frequency shifts, and lineshape distortions is shown in Fig. 8. It shows that the PCA quantitation method is more sensitive to phase and frequency variations than the curve-fitting approach. The peak areas of the spectra of M2 are underestimated (relative broad lines) and those of M4 are overestimated (relative narrow lines) before correction (Fig. 8A). However, after correction for phase shifts, frequency shifts, and lineshape distortions the peak areas of both series have become of the same order of magnitude as those of M1 and M3 (Fig. 8B). As a result, the quantitation results have become similar to those obtained with the nonlinear curve-fitting approach. Although the curve-fitting approach cannot accommodate lineshape variations, only minor changes in the peak area estimations are observed after correction for phase and frequency variations (Figs. 8C and 8D). These findings demonstrate that with the PCA quantitation method, which operates on the complete series of PCr resonances simultaneously, similar good results are obtained in much less time after appropriate signal correction.

DISCUSSION

The new method presented in this paper, i.e., the combination of PCA and linear regression, is a powerful method for removing phase and frequency shifts across a large series of single spectral lines in an automated way. All separate phase and frequency shifts are simultaneously identified and removed from the series of spectra in a model-independent way without requiring any user interaction. Compared to the original proposed method (*10*) the current method, as performed on simulated data, shows improved precision and robustness: accurate and unique phase solutions and frequency-shift solutions beyond the discrete spectral resolution are obtained. Furthermore, the proposed regression method enables the user to exploit more than three (signal-related) PC_1-derived spectra to correct for the individual phase and frequency shifts. This allows the

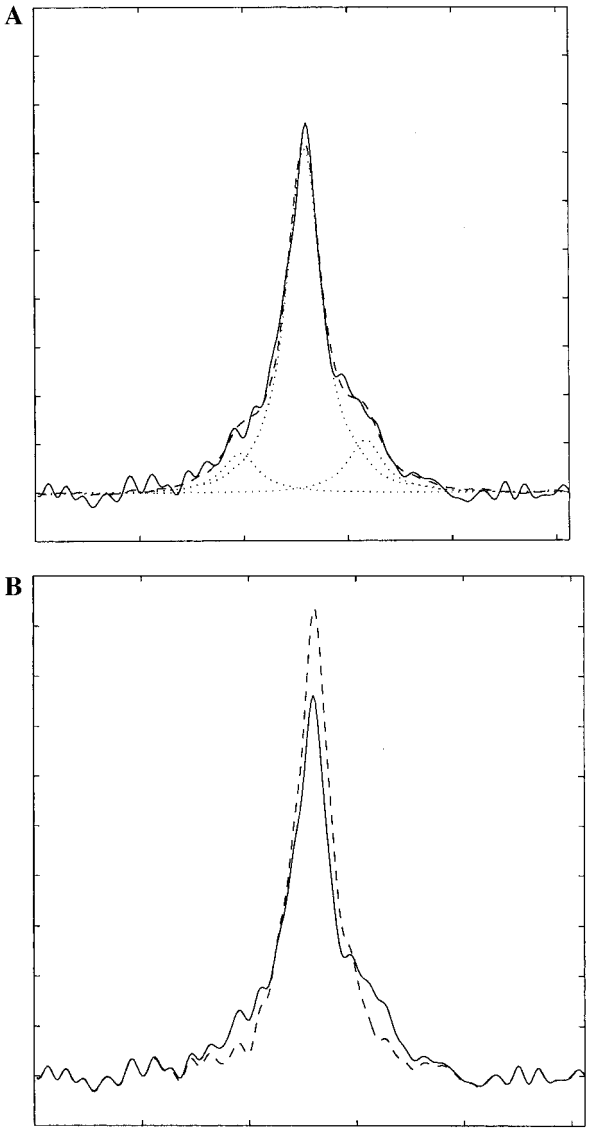

FIG. 7. Transformation of a disrupted PCr lineshape to the common lineshape of the data set. (A) Original spectrum (solid line) and calculated spectrum (dashed line) obtained with linear regression and PCA. The calculated spectrum is the sum of the three PC_1 lineshapes (dotted lines) which vary in line position and in amplitude. (B) Original spectrum (solid line) and transformed spectrum (dashed line). The transformed PCr spectrum is obtained in two steps. First, the three PC_1 lineshapes in (A) are subtracted from the original spectrum. Then, the resulting signal is added to the three PC_1 lineshapes positioned at the spectral position of the common lineshape of the data set.

novel method to correct for larger phase- and frequency-shift variations across series of spectra. From the time series of *in vivo* ^{31}P NMR spectra, phase and frequency shifts of the phosphocreatine resonance could successfully be removed in the presence of considerable lineshape distortions. The remaining lineshape distortions were subsequently removed using the novel regression method. The results of the experimental data demonstrate that postprocessing using the proposed correction methods compensates for unwanted, but inevitable, phase and frequency errors often occurring during *in vivo* NMR experiments. After these corrections all PCr resonances could be described by one arbitrary lineshape function. Variations in PCr amplitude can then be assessed by the PCA quantitation method for fast and model-free quantitation of a complete series of resonances.

Besides these improvements, it is also important to notice the simplicity and flexibility of the proposed regression method. For example, the new correction method is no longer necessarily committed to PCA. Instead of the PC_1 spectrum the mean spectrum of the data set or any other proper reference spectrum can also be used. In fact, each peak with a phase and/or frequency offset can be shifted to any appropriate reference peak. Once a reference peak is defined, all separate phase and frequency shifts are directly acquired in an elegant way using ordinary regression. Apart from this, each relevant spectrum, derived from the PC_1 or any other reference spectrum, can easily be included in the regression model for fitting adequately each shifted spectrum. This offers the user the possibility of extending the analysis to more complex spectra. In addition to phase- and frequency-shift variations linewidth

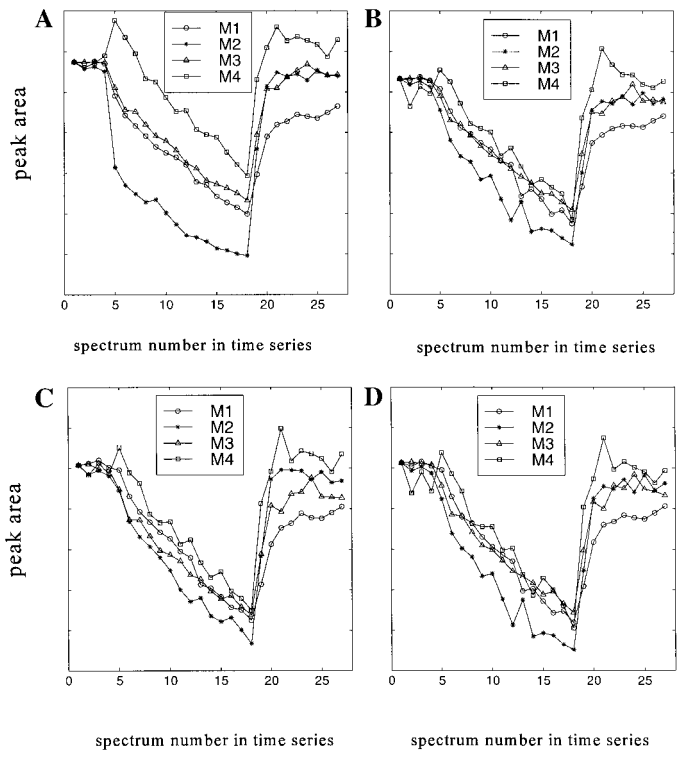

FIG. 8. Quantitation of time series of PCr resonances of the muscle of four different mice M1 to M4 using PCA (A, B) and nonlinear curve fitting (C, D). Quantitation before correction (A) and after automatic correction for phase shifts, frequency shifts, and lineshape variations (B) using PCA. Quantitation before correction (C) and after automatic correction for phase shifts, frequency shifts, and lineshape variations (D) using nonlinear curve fitting based on the complex Lorentzian model lineshape function. Ischemia was applied between spectra 4 and 18.

variations are also usually observed in (large) NMR spectral data sets.

The combination of PCA (or a proper reference spectrum) and linear regression will make it also possible to correct simultaneously for phase, frequency shift, and linewidth variations across a series of spectra containing a single lineshape function. This will be addressed in a forthcoming paper. Also spectral regions containing two or more (overlapping) resonance lines can be considered. Each spectrum will then be fitted to two or more reference lines which each may vary in amplitude and/or in phase and/or in frequency.

CONCLUSIONS

The proposed method is a promising preprocessing technique for the analysis of large spectral data sets, both for spectral quantitation and for classification purposes. It removes unwanted phase shifts, frequency shifts, and additional lineshape distortions throughout the spectral data set within one calculation step. Because it is a simple, robust, and flexible method it can easily be extended to the analysis of more complex spectra.

ACKNOWLEDGMENTS

The authors thank Radka Stoyanova and Truman Brown for giving information about their method and Arjan Simonnetti for acquiring the ^{31}P NMR spectra.

REFERENCES

1. R. de Beer and D. van Ormondt, Analysis of NMR data using time domain fitting procedures, *in* "NMR Basic Principles and Progress, Vol. 26" (P. Diehl, E. Fluck, H. Günther, R. Kosfeld, and J. Seelig, Eds.), pp. 201–248, Springer-Verlag, Berlin (1992).
2. H. Barkhuysen, R. de Beer, W. M. M. J. Bovée, and D. van Ormondt, Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a least-squares procedure, *J. Magn. Reson.* **61,** 465–481 (1985).
3. H. Barkhuysen, R. de Beer, and D. van Ormondt, Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals, *J. Magn. Reson.* **73,** 553–557 (1987).
4. A. Knijn, R. de Beer, and D. van Ormondt, Frequency-selective quantification in the time domain, *J. Magn. Reson.* **97,** 444–450 (1992).
5. G. Hagberg, From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods, *NMR Biomed.* **11,** 148–156 (1998).
6. R. Siuda, G. Balcerowska, and D. Aberdam, Spurious principal components in the set of spectra subjected to disturbances. I. Presentation of the problem, *Chemometrics Intell. Lab. Sys.* **40,** 193–201 (1998).
7. R. Stoyanova, A. C. Kuesel, and T. R. Brown, Application of principal-component analysis for NMR spectral quantitation, *J. Magn. Reson. A* **115,** 265–269 (1995).
8. A. C. Kuesel, R. Stoyanova, N. R. Aiken, C.-W. Li, B. S. Szwergold, C. Shaller, and T. R. Brown, Quantitation of resonances in biological ^{31}P NMR spectra via principal component analysis: Potential and limitations, *NMR Biomed.* **9,** 93–104 (1996).
9. M. A. Elliot, G. A. Walter, A. Swift, K. Vandenborne, J. C. Schotland, and J. S. Leigh, Spectral quantitation by principal component analysis using complex singular value decomposition, *Magn. Reson. Med.* **41,** 450–455 (1999).
10. T. R. Brown and R. Stoyanova, NMR spectral quantitation by principal-component analysis. II. Determination of frequency and phase shifts, *J. Magn. Reson. B* **112,** 32–43 (1996).
11. J. E. Jackson, "User's Guide to Principal Components," Wiley, New York (1991).
12. H. J. A. in 't Zandt, F. Oerlemans, B. Wierenga, and A. Heerschap, Effects of ischemia on skeletal muscle energy metabolism in mice lacking creatine kinase monitored by *in vivo* ^{31}P nuclear magnetic resonance spectroscopy, *NMR Biomed.* **12,** 327–334 (1999).
13. D. G. Gadian, "Nuclear Magnetic Resonance and Its Applications to Living Systems," Clarendon Press, Oxford (1982).