JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
COMMUNICATION
pubs.acs.org/JACS

# Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero $^{1}H-^{13}C$ HSQC ($HSQC_0$)

Kaifeng Hu, William M. Westler, and John L. Markley*

National Magnetic Resonance Facility at Madison, Biochemistry Department, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

S Supporting Information

**ABSTRACT:** Quantitative one-dimensional (1D) $^{1}$H NMR spectroscopy is a useful tool for determining metabolite concentrations because of the direct proportionality of signal intensity to the quantity of analyte. However, severe signal overlap in 1D $^{1}$H NMR spectra of complex metabolite mixtures hinders accurate quantification. Extension of 1D $^{1}$H to 2D $^{1}H-^{13}C$ HSQC leads to the dispersion of peaks along the $^{13}$C dimension and greatly alleviates peak overlapping. Although peaks are better resolved in 2D $^{1}H-^{13}C$ HSQC than in 1D $^{1}$H NMR spectra, the simple proportionality of cross peaks to the quantity of individual metabolites is lost by resonance-specific signal attenuation during the coherence transfer periods. As a result, peaks for individual metabolites usually are quantified by reference to calibration data collected from samples of known concentration. We show here that data from a series of HSQC spectra acquired with incremented repetition times (the time between the end of the first $^{1}$H excitation pulse to the beginning of data acquisition) can be extrapolated back to zero time to yield a time-zero 2D $^{1}H-^{13}C$ HSQC spectrum ($HSQC_0$) in which signal intensities are proportional to concentrations of individual metabolites. Relative concentrations determined from cross peak intensities can be converted to absolute concentrations by reference to an internal standard of known concentration. Clustering of the $HSQC_0$ cross peaks by their normalized intensities identifies those corresponding to metabolites present at a given concentration, and this information can assist in assigning these peaks to specific compounds. The concentration measurement for an individual metabolite can be improved by averaging the intensities of multiple, nonoverlapping cross peaks assigned to that metabolite.

The primary objective of metabolomics studies is to identify individual chemical components in mixtures and to relate their concentrations to the precise biological state of the system, such as stress, age, and disease.[1] Many methods have been developed to accurately and efficiently identify and profile changes in distinct sets of biomarkers.[2−4] Because NMR-based methods are unbiased, they have some advantages over more sensitive MS-based methods. The integrated intensities of resolved proton resonances (the area under the $^{1}$H NMR signal) in one-dimensional (1D) proton NMR spectra are directly proportional to the number of proton spins in the mixture,[5,6] and quantitative proton NMR is a routine analytical tool because of its universality, sensitivity, precision, and nondestructive nature.[6] However, this approach has shortcomings for signals that are overlapped.[7] Conventional 2D $^{1}H-^{13}C$ HSQC spectra of metabolite mixtures contain a much higher proportion of resolved peaks, but the signals are more difficult to quantify because of resonance-specific signal attenuation during the coherence transfer periods as the result of relaxation, imperfect pulses, and mismatch of the INEPT delay with specific *J*-couplings.[8,9] In theory, the various correction factors for different metabolites can be calculated, as suggested by Rai et al.,[9] provided that the correct relaxation parameters (such as $T_1$ and $T_2$) and *J*-couplings (specific to each metabolite and each functional group) are known. However, the overall signal attenuation resulting from imperfect pulses and the effect of the $^{13}$C offset are not taken into account, and all correction factors would require recalculation for each variation in the data collection scheme, such as changes in $d_1$, $\tau_1$, or $\tau_2$ delays.

Here, we describe an approach for quantification of individual compounds in metabolite mixtures without the need to calibrate the 2D peak intensities against spectra of metabolites with known concentration under defined conditions.[7] This new approach, extrapolated time-zero $^{13}$C HSQC ($HSQC_0$), enables simultaneous quantification and identification of compounds in metabolite mixtures. We show here that resonance-specific signal attenuation scaling factors in 2D $^{13}$C HSQC can be determined simply through repetition of the pulse sequence from the point right after the first $^{1}$H excitation pulse to the point right before acquisition. Similar approaches have been used to study magnetization transfer efficiencies in heteronuclear NMR experiments[10] and to measure protein concentrations from $^{1}H-^{15}N$ HSQC data.[8,11] The $^{13}$C $HSQC_0$ spectrum, in which signal intensities are proportional to concentrations of individual metabolites, is obtained by extrapolating to zero time the series of HSQC spectra acquired with different repetition times. Relative concentrations determined from cross peak intensities can then be converted to absolute concentrations by reference to an internal standard of known concentration including the solvent,[12] or a synthesized electronic signal.[5,13] Absolute concentrations also can be determined from a single general external concentration reference through the application of PULCON techniques[8,11] or by analysis of the NMR receiving efficiency.[14] Furthermore, the $HSQC_0$ cross peaks can be clustered by their intensities into groups that

**Received:** October 22, 2010
**Published:** January 19, 2011

correspond to metabolites present at different concentrations. These clusters can assist in identifying peaks that correspond to different compounds. The concentration measurement for an individual metabolite can be improved by averaging the intensities of multiple, nonoverlapping cross peaks assigned to that metabolite.

In quantitative 1D proton NMR (qHNMR) (Figure 1A), the time-domain data are acquired immediately after a 90° excitation pulse. The integrated intensity of the acquired NMR signal is directly proportional to the number of proton spins in the mixture[6,12] provided that the relaxation delay is sufficiently long. The basic building block in the constant time 2D $^{13}C$ HSQC experiment is shown in parentheses in Figure 1B. As indicated, the density operator at point *f* for detection, $I_{-y}$, has the exact same form as the density operator generated immediately after the 90° excitation pulse in 1D qHNMR (Figure 1A) except for the amplitude attenuation factor $f_A$. The intensity can be expressed as

$$A_{1,n}(I_{-y}) = A_{0,n}(I_{-y})\bullet f_{A,n} \qquad (1)$$

in which $A_{1,n}$ is the integrated signal intensity (area under NMR signal) of peak *n* in HSQC, $A_{0,n}$ is the integrated signal intensity of corresponding isolated peak *n* in 1D qHNMR, and $f_{A,n}$ is the amplitude attenuation factor specific for peak *n*. The $f_{A,n}$ factor accounts for the signal losses during the coherence transfer periods from point *a* (immediately after the first $^1H$ excitation pulse) to point *f* (immediately before acquisition). It should be emphasized that this attenuation factor $f_{A,n}$ is specific to a particular cross peak (peak *n*) because of different chemical environments, dynamics, relaxation properties, and *J*-couplings. Therefore, even though peaks are better resolved in 2D $^{13}C$ HSQC, the peak intensities are not directly proportional to the number of spins giving rise to the signals. However, the scaling factor $f_{A,n}$ can be determined simply through repetition of the pulse sequence components included in the parentheses (Figure 1C) to acquire three 2D $HSQC_i$ ($i$ = 1, 2, 3) spectra in which the subscript *i* indicates the number of times the basic building block is repeated. Note that, in $HSQC_2$ and $HSQC_3$, additional phase cycling is applied on $\phi_4$ and $\phi_5$ and on the receiver phase $\phi_{rec}$.

The density operators detected in 2D $HSQC_i$ ($i$ = 1, 2, 3) are all $I_{-y}$, the exact same form as the density operator detected in 1D qHNMR. The NMR signal intensity is attenuated linearly as a function of the number of repetitions *i*, such that the peak intensity of peak *n* in $HSQC_i$ is

$$A_{i,n} = A_{0,n}\cdot f_{A,n}^{i} \qquad (2)$$

where $f_{A,n}$ is the amplitude attenuation factor specific for peak *n* and $A_{0,n}$ is the peak intensity for peak *n* in $HSQC_0$, the virtual 2D HSQC spectrum obtained through linear regression extrapolation:

$$\ln(A_{i,n}) = \ln(A_{0,n}) + i \times \ln(f_{A,n}) \qquad (3)$$

Figure 1D is a schematic representation of the extrapolation of 2D $HSQC_i$ ($i$ = 1, 2, 3) data to determine the 2D $HSQC_0$ peak intensities, $A_{0,n}$, for each peak *n*, which are free of attenuation during the coherence transfer period. These $A_{0,n}$ values are analogous to the peak intensities in 1D qHNMR, which are acquired immediately (at zero time) after the first $^1H$ excitation pulse. Provided that the 2D $HSQC_i$ ($i$ = 1, 2, 3) data are processed in

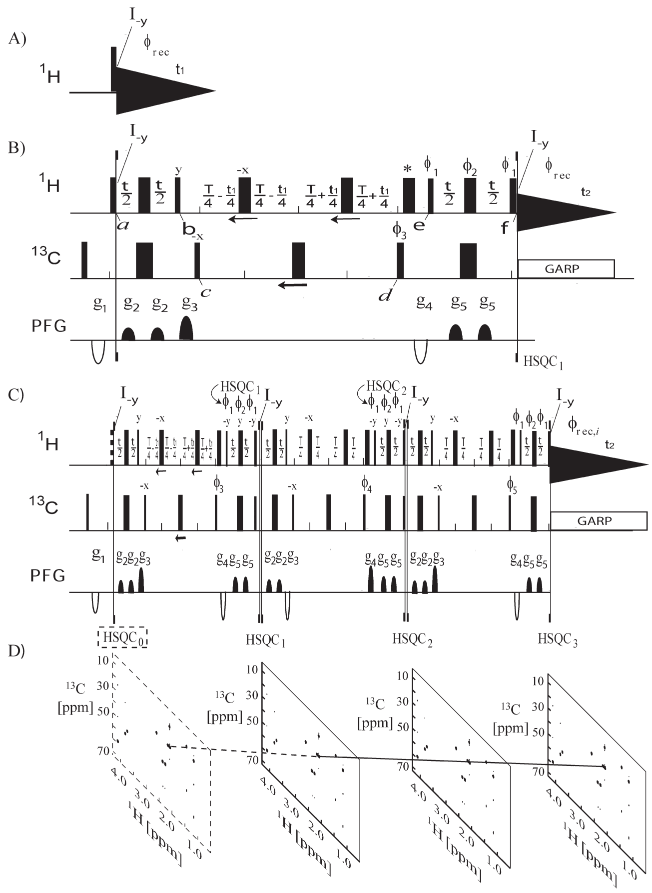


**Figure 1.** (A) Pulse sequence used for quantitative 1D $^1H$ NMR spectroscopy. (B) Pulse sequence used for 2D $^1H-^{13}C$ HSQC (heteronuclear single-quantum correlation) spectroscopy. Narrow and wide black bars indicate 90° and 180° pulses, respectively. The delays are $\tau$ = 3.3 (or 3.4) ms; $T$ = 5.3 (or 5.8) ms. The phase cycling is as follows: $\phi_1 = -y, y$; $\phi_2 = y, -y$; $\phi_3 = 2[x], 2[-x]$; $\phi_{rec} = x, -x, -x, x$. All other radio frequency pulses are applied with phase *x*, except as indicated. Quadrature detection in the $^{13}C$ ($t_1$) dimension is achieved using States-TPPI applied to the phase $\phi_3$. The duration and strength of the pulsed field gradients applied along the *z*-axis are as follows: ($g_1$) 3 ms, 15.9 G/cm; ($g_2$) 0.4 ms, 10.6 G/cm; ($g_3$) 1 ms, 18.6 G/cm; ($g_4$) 1 ms, −26.5 G/cm, and ($g_5$) 0.6 ms, 13.25 G/cm, followed by gradient recovery period of 200 μs. The 180° pulse marked with an asterisk serves to refocus proton chemical shift during the gradient delays for $g_3$ and $g_4$ to achieve better $t_1$ noise suppression. (C) Pulse sequence proposed here for time-zero extrapolated HSQC ($HSQC_0$) spectroscopy. Three data sets are collected: the first at the position indicated by $HSQC_1$, the second at the position indicated by $HSQC_2$, and the third at the position indicated by $HSQC_3$; each data collection proceeds as shown for $HSQC_3$. In all three pulse sequences, composite phase cycling of $\phi_1$ and $\phi_2$ is carried out in the HSQC unit that precedes data acquisition; otherwise, $\phi_1 = -y$ and $\phi_2 = y$. For $HSQC_1$, the phase cycling is the same as that in (B). $\phi_3 = 2[x], 2[-x]$; $\phi_{rec,1} = x, -x, -x, x$. The constant time chemical shift evolution period $T$ is replaced by the invariant constant time $T$ in the second and third HSQC units, respectively for $HSQC_2$ and $HSQC_3$. Additional phase cycling for $HSQC_2$ is $\phi_4 = 4[x], 4[-x]$; $\phi_{rec,2} = x, -x, -x, x, -x, x, x, -x$. Additional phase cycling for $HSQC_3$ is $\phi_5 = 8[x], 8[-x]$; $\phi_{rec,3} = x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x$. The density operators detected in 2D $HSQC_i$ ($i$ = 1, 2, 3) are all $I_{-y}$, as is the density operator detected in quantitative 1D proton NMR (A). (D) Shown are examples of $HSQC_1$, $HSQC_2$, and $HSQC_3$ spectra following Fourier transformation to yield frequency domain spectra. Peak intensities in the virtual $HSQC_0$ spectrum (bordered by the dashed line) are derived from linear ln extrapolation of the peak intensities of the corresponding peaks in the $HSQC_1$, $HSQC_2$, and $HSQC_3$ spectra.

exactly the same way, the analysis can be carried out with peak heights, instead of peak volumes, to determine $H_{0,n}$:

$$\ln(H_{i,n}) = \ln(H_{0,n}) + i \times \ln(f_{A,n}) \qquad (4)$$

in which $H_{i,n}$ is the measured peak height of peak $n$ in HSQC$_i$ and $f_{A,n}$ is the attenuation factor for peak $n$. The integrated peak intensity for peak $n$ in the virtual 2D HSQC$_0$, $A_{0,n}'$, can then be calculated from

$$A_{0,n}' = A_{1,n} \cdot H_{0,n}/H_{1,n} \qquad (5)$$

in which $A_{1,n}$ and $H_{1,n}$ are the integrated peak intensity and peak height of peak $n$ in HSQC$_1$, respectively.

We demonstrated this approach using a 63.75 mM/88.47 mM/124.36 mM mixture of alanine/methionine/sodium 3-hydroxybutyrate in $D_2O$. A relaxation enhancing agent, $[Fe(EDTA)]^-$, was added to a final concentration of about 1.8 mM to shorten the interscan delay. We collected the NMR data at 25 °C on a 500 MHz Bruker DMX spectrometer equipped with a $z$-gradient triple resonance cryogenic TCI probe and on a 600 MHz Bruker DMX spectrometer equipped with a $z$-gradient triple resonance room-temperature TXI probe.

The 500 MHz data were collected with $^1$H and $^{13}$C radio frequency pulses applied at 4.7 and 48 ppm, respectively, and delays $\tau$ and $T$ were set to 3.4 and 5.8 ms, respectively. GARP $^{13}$C-decoupling was at a field strength of $\gamma B_2$ = 2.08 kHz. 2048 × 64 complex data points with spectral widths of 14 and 80 ppm, respectively, were collected along the $^1$H and $^{13}$C dimensions, with 16 scans per FID and an interscan delay of 2.5 s, resulting in a total experimental time of about 1.5 h for each HSQC. The interscan relaxation delay was set to 5 times the longest measured $T_1$ in the sample (0.5 s).

The 600 MHz data were collected with radio frequency pulses applied on $^1$H and $^{13}$C at 4.7 and 43 ppm, respectively, and the delays $\tau$ and $T$ were set to 3.3 and 5.3 ms, respectively. GARP $^{13}$C-decoupling used a field strength of $\gamma B_2$ = 2.5 kHz. 2048 × 64 complex data points with spectral width of 16 ppm and 80 ppm, respectively, were collected along the $^1$H and $^{13}$C dimensions, with 16 scans per FID and an interscan delay of 3.0 s, resulting in a total acquisition time of 2 h for each HSQC. The time interscan delay was determined by multiplying the longest measured $T_1$ at (0.58 s) by 5.

Different $\tau$ and $T$ values were used with the two spectrometers (500 and 600 MHz) to verify whether the concentration could be determined accurately regardless of the relaxation property and the mismatch of the INEPT transfer delay with any specific $J$-coupling. By collecting data on spectrometers operating at two different fields, we tested the effects of different conditions, such as imperfect power level or pulses during the coherence transfer, on the accuracy of the concentration measurements.

NMRPipe[15] software was used to process all data sets. Prior to Fourier transformation, a squared sine-bell window function was applied to both dimensions of the time-domain data, and the data were zero-filled to give at least five data points above the half width for each resonance to allow for precise and reliable integration.[16] The Fourier transformed spectra were phased manually, and automated polynomial baseline correction was applied to improve the accuracy of the integral. The integrated peak intensity (peak volume) was calculated by direct summation over a rectangular box using SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) NMR analysis software. Peak heights were obtained from the

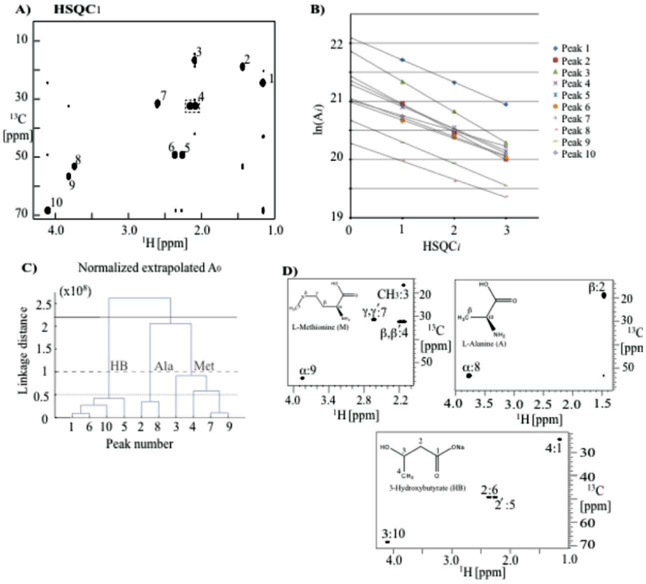


**Figure 2.** (A) 500 MHz 2D $^1$H−$^{13}$C HSQC$_1$ spectrum of the metabolite mixture: alanine (Ala), methionine (Met), and 3-hydroxybutyrate (HB); the 10 cross peaks are labeled arbitrarily. (B) Extrapolation of the 500 MHz 2D HSQCi ($i$ = 1, 2, 3) peak intensities from integrated peak volumes ($A_i$) to yield $A_0$ values. (C) Clustering of the $A_0$ values normalized by the number of contributing protons. (D) Structures of the three metabolites and identification of cross peaks in the $^1$H−$^{13}$C HSQC spectrum assigned to specific groups in the molecules (atom designator: cross peak number).

local maximum with the SPARKY peak 'center' command. The region of integration for each peak was determined by including the HSQC$_1$ peak contour line obtained by setting the threshold at 1/162.5 of the peak height. Assuming a Lorentzian line shape, this procedure captures 95% of the total area under the peak.

Peaks in the 500 MHz 2D HSQC$_1$ spectrum (Figure 2A) of the metabolite mixture were picked manually and numbered arbitrarily from 1 to 10. The peak volumes ($A_i$) and peak heights ($H_i$) measured in each 500 MHz 2D HSQC$_i$ ($i$ = 1, 2, 3) expressed in ln format are listed in Table S1. These values were used to extrapolate the HSQC$_0$ peak volumes $A_0$ (Figure 2B) and peak heights $H_0$ (Figure S1A). Corresponding 600 MHz results are listed in Table S2 and Figure S2. The volume and height regression lines for each peak $n$ were approximately parallel as indicative of similar attenuation factors.

The normalized signal intensity is calculated by dividing the integrated intensity, from either direct extrapolation of peak volumes ($A_{0,n}$) or peak heights $H_0$ by eq 5 ($A_{0,n}'$), by the number of magnetically equivalent protons contributing to the signal (for example, 1 for a −CH− group and 3 for a $-CH_3$ group (see Figure 2 and Table S3)). If the geminal protons of a $-CH_2-$ group have the same chemical shift (peak 7 in Figure 2) or have very close chemical shifts and are grouped together (peak 4 in Figure 2), the normalization factor is 2; if the geminal protons of a $-CH_2-$ group have different chemical shifts and are treated as two different peaks, the normalizing factor is 1 (peaks 5 and 6 in Figure 2). The type of C−H group, CH, $CH_2$, or $CH_3$, can be determined by DEPT experiments or nonrefocused 2D $^{13}$C HSQC.

A Matlab (http://www.mathworks.com/products/matlab/) routine was used to cluster HSQC$_0$ cross peaks into groups on

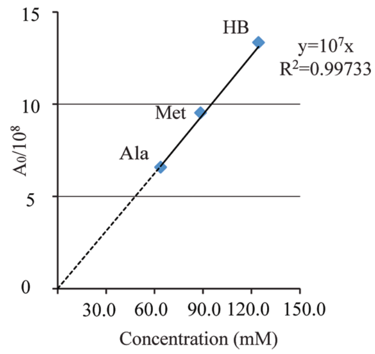


**Figure 3.** Regression of the averaged normalized, extrapolated 500 MHz $HSQC_0$ peak volumes determined from integrated peak volumes ($V_0$) and assigned to the three metabolites vs the concentrations of the three metabolites: alanine (Ala), methionine (Met), and 3-hydroxybutyrate. The correlation coefficient was better than 0.99.

the basis of their normalized extrapolated peak volumes $A_{0,n}$ (Figure 2C, Figure S3A) or $A_{0,n}'$ (Figure S1B, Figure S3B) by setting the threshold (shown as dashed lines) of the linkage distance to $1.0 \times 10^8$ and $0.7 \times 10^8$, respectively. In both cases, because the compounds in the mixture were at different concentrations, the peaks clustered into three separate groups, which corresponded to the chemical shifts of standard compounds in the BMRB database (http://www.bmrb.wisc.edu/): peaks of 1, 5, 6, and 10 (3-hydroxybutyrate); peaks 2 and 8 (alanine); and peaks of 3, 4, 7, and 9 (methionine).

The normalized extrapolated $HSQC_0$ peak intensities assigned to the same metabolite showed a standard deviation of about 7% (Table S2 and Table S4). However, the concentration measurement accuracy was improved greatly by averaging the normalized intensities of the cross peaks assigned to a given metabolite (Figure 3 and Figures S1C and S4). Previous studies have shown that an $S/N$ of at least 150 is required to achieve a target uncertainty of 1%.[16] Linear regression of the averaged normalized peak volumes $A_{0,n}$ (or $A_{0,n}'$) for the three metabolites vs their concentrations gave very high correlation coefficients (>0.99).

The approach is applicable to $^{13}$C-labeled compounds simply by adjusting the length of the constant time period to $1/J_{CC}$. As with 1D qHNMR,[6] the experimental conditions used in quantitative 2D $^{13}$C HSQC must be optimized for high measurement accuracy. In our experience, the interscan delay should be at least 5 times the longest $T_1$ in the sample to allow all polarization to reach equilibrium, and manual phasing and optimized baseline correction should be used. Slightly incorrect phase and baseline corrections can result in low measurement precision and inaccuracy. In order to achieve higher measurement efficiency, a relaxation enhancing agent can be added as was done here to shorten the relaxation delay. Usually NMR peak heights alone cannot be used for quantification because of different line widths. Instead, integration of the signals must be performed.[7] Because HSQC spectra of metabolites are not sufficiently dispersed to enable integration that would capture 99% of the peak intensity (integration over 64 times the line width at half height in each dimension),[16] we chose a compromise value of 95% of the total area under the peak.

We have demonstrated here a simple method for quantifying 2D HSQC spectra. By extrapolating peak intensities from a series of measured $HSQC_i$ spectra, the time-zero $^{13}$C HSQC ($HSQC_0$) spectrum is constructed, whose peaks are linearly proportional to the concentrations of the compounds. In addition to enabling the quantification of components in mixtures, the approach can assist in compound identification by allowing peaks to be sorted by their normalized intensities so that those arising from the same compound are grouped. Proper clustering depends on the precision of the measured intensities and the concentration differences of the compounds to be distinguished in the mixture. The quantification accuracy can be improved by averaging values from all resolved peaks corresponding to a single compound, and relative concentrations can be converted to absolute concentrations by reference to data from an internal reference compound of known concentration. We expect that the $HSQC_0$ approach to metabolite identification and quantification can be easily adapted for automatic or semiautomatic data analysis.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Peak volumes and peak heights measured from 500 and 600 MHz 2D $HSQC_i$ ($i = 1, 2, 3$) spectra; analysis of 500 MHz and 600 MHz $HSQC_i$ ($i = 1, 2, 3$) peak volume and peak height data to yield clustering of normalized intensities; and averaged normalized intensities regressed vs the known concentrations of the compounds in the mixture. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

**Corresponding Author**
markley@nmrfam.wisc.edu

## ■ ACKNOWLEDGMENT

The authors thank James Ellinger for preparing the metabolite mixture. This work was supported by NIH Grant P41 RR02301 from the Division of Biomedical Technology, National Center for Research Resources and by the DOE-funded Great Lakes Biofuels Research Center.

## ■ REFERENCES

(1) Wishart, D. S. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 228.
(2) Piotto, M.; Moussallieh, F. M.; Dillmann, B.; Imperiale, A.; Neuville, A.; Brigand, C.; Bellocq, J. P.; Elbayed, K.; Namer, I. J. *Metabolomics* **2009**, *5*, 292.
(3) Zhang, F.; Bruschweiler, R. *Angew. Chem., Int. Ed.* **2007**, *46*, 2639.
(4) Zhang, F.; Dossey, A. T.; Zachariah, C.; Edison, A. S.; Bruschweiler, R. *Anal. Chem.* **2007**, *79*, 7748.
(5) Akoka, S.; Barantin, L.; Trierweiler, M. *Anal. Chem.* **1999**, *71*, 2554.
(6) Pauli, G. F.; Jaki, B. U.; Lankin, D. C. *J. Nat. Prod.* **2005**, *68*, 133.
(7) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385.
(8) Dreier, L.; Wider, G. *Magn. Reson. Chem.* **2006**, *44*, S206.
(9) Rai, R. K.; Tripathi, P.; Sinha, N. *Anal. Chem.* **2009**, *81*, 10232.
(10) Braun, D.; Wuthrich, K.; Wider, G. *J. Magn. Reson.* **2003**, *165*, 89.
(11) Wider, G.; Dreier, L. *J. Am. Chem. Soc.* **2006**, *128*, 2571.
(12) Mo, H. P.; Raftery, D. *Anal. Chem.* **2008**, *80*, 9835.
(13) Michel, N.; Akoka, S. *J. Magn. Reson.* **2004**, *168*, 118.
(14) Mo, H. P.; Harwood, J.; Zhang, S. C.; Xue, Y.; Santini, R.; Raftery, D. *J. Magn. Reson.* **2009**, *200*, 239.
(15) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(16) Malz, F.; Jancke, H. *J. Pharm. Biomed. Anal.* **2005**, *38*, 813.

## ■ NOTE ADDED AFTER PRINT PUBLICATION

The original publication, published on January 19, 2011, did not include the correct ACS copyright credit line. This has been corrected to show that this work is Copyright 2011 American Chemical Society.