Anal. Chem. **2007,** *79,* 7748−7752

Strategy for Automated Analysis of Dynamic Metabolic Mixtures by NMR. Application to an Insect Venom

Fengli Zhang,[†] Aaron T. Dossey,[‡] Cherian Zachariah,[‡] Arthur S. Edison,[†,‡] and Rafael Brüschweiler*[,†,§]

National High Magnet Field Laboratory and Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32310, and Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610

Elucidation of the composition of chemical−biological samples is a main focus of systems biology and metabolomics. Due to the inherent complexity of these mixtures, reliable, efficient, and potentially automatable methods are needed to identify the underlying metabolites and natural products. Because of its rich chemical information content, nuclear magnetic resonance (NMR) spectroscopy has a unique potential for this task. Here we present a generalization and application of a recently introduced NMR data collection, processing, and analysis strategy that circumvents the need for extensive purification and hyphenation prior to analysis. It uses covariance TOCSY NMR spectra measured on a 1-mm high-temperature cryogenic probe that are analyzed by a spectral trace clustering algorithm yielding 1D NMR spectra of the individual components for their unambiguous identification. The method is demonstrated on a metabolic model mixture and is then applied to the unpurified venom mixture of an individual walking stick insect that contains several slowly interconverting and closely related metabolites.

The chemical composition of biological systems changes in response to a multitude of factors such as genetics, age, pathology, development, environment, and stress. These chemical responses are the subject of systems biology and the emerging fields of metabolomics and metabonomics, which promise a better understanding of chemical processes in living systems.[1−3] Due to the inherent complexity of these mixtures, reliable and efficient methods are needed to identify the underlying natural products and their derivatives that are suitable for automation. Frequently, the chemical components occur at low concentrations and can undergo isomerization, which makes their isolation and identification difficult. Nuclear magnetic resonance (NMR) spectroscopy has a unique potential for this task as it does not a priori require potentially labor-intensive and costly hyphenation steps.[4] This characteristic is utilized in diffusion-ordered spectroscopy (DOSY),[5] differential analysis of 2QF-COSY spectra,[6] and other recently developed NMR-based techniques including STOCSY[7] and SHY[8] that identify components based on statistical covariations of resonance amplitudes across a large number of samples.

Biologically active molecules are often produced in small quantities, creating a challenge for NMR sample preparation and detection. This challenge has recently been alleviated by the introduction of a microsolenoid[9,10] and multicoil microsolenoid probe[11] and a 1-mm high-temperature superconducting (HTS) probe.[12] All of these technologies provide excellent mass sensitivity, and the HTS probe proved essential to a recent study of defensive secretions from a single insect.[13] However, because of the potentially large number of NMR resonances, the improved mass sensitivity can also lead to crowded spectra, which makes identification of individual chemical components difficult.

An approach to this problem has been introduced recently, termed DemixC,[14] which is generalized here to a NMR-based data collection, processing, and analysis platform (Figure 1). DemixC relies on covariance NMR total correlation spectroscopy (TOCSY)[15] to monitor spin−spin connectivity information across each molecule. This gives the means to deconvolute the spectrum

* To whom correspondence should be addressed. Tel.: 850-644-1768. Fax: 850-644-8281. E-mail: bruschweiler@magnet.fsu.edu.

[†] National High Magnet Field Laboratory, Florida State University.

[‡] University of Florida.

[§] Department of Chemistry and Biochemistry, Florida State University.

(1) Nicholson, J. K.; Wilson, I. D. *Nat. Rev. Drug Discovery* **2003**, *2*, 668−676.
(2) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245−252.
(3) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(4) Lenz, E. M.; Wilson, I. D. *J. Proteome Res.* **2007**, *6*, 443−458.
(5) Johnson, C. S. *Progress in Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203−256.
(6) Schroeder, F. C.; Gibson, D. M.; Churchill, A. C.; Sojikul, P.; Wursthorn, E. J.; Krasnoff, S. B.; Clardy, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 901−904.
(7) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282−1289.
(8) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Anal. Chem.* **2006**, *78*, 363−371.
(9) Olson, D. L.; Peck, T. L.; Webb, A. G.; Magin, R. L.; Sweedler, J. V. *Science* **1995**, *270*, 1967−1970.
(10) Schroeder, F. C.; Gronquist, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 7122−7131.
(11) Wang, H.; Ciobanu, L.; Edison, A. S.; Webb, A. G. *J. Magn. Reson.* **2004**, *170*, 206−212.
(12) Brey, W. W.; Edison, A. S.; Nast, R. E.; Rocca, J. R.; Saha, S.; Withers, R. S. *J. Magn. Reson.* **2006**, *179*, 290−293.
(13) Dossey, A. T.; Walse, S. S.; Rocca, J. R.; Edison, A. S. *Chem. Biol.* **2006**, *1*, 511−514.
(14) Zhang, F. L.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2007**, *46*, 2639−2642.
(15) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528.

10.1021/ac0711586 CCC: $37.00
Published on Web 09/07/2007

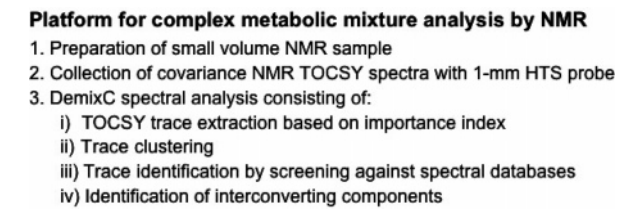

Figure 1.

into subspectra of the individual mixture components.[16] In the final step of Figure 1, the chemical components can be identified by screening the subspectra against a suitable NMR spectral database.

Covariance TOCSY[17−19] has been found to be particularly useful for mixture analysis.[14] The spectra exhibit high sensitivity and high spectral resolution along both frequency dimensions, which allows the selection of those cross sections through the spectrum that represent "fingerprint" traces of individual components. The combination of very high sensitivity NMR detection with covariance NMR allows identification of all components of a complex biological mixture from very small quantities of material. The platform is demonstrated for a metabolic test mixture and then applied to the analysis of the venom of the walking stick insect *Anisomorpha buprestoides*.

METHODS

Sample Preparation. A test mixture was prepared by mixing D-glucose, L- histidine, L-lysine, serotonin (5HT) hydrochloride, and D-sorbitol at final concentrations of 1.0, 0.4, 0.6, 0.8, and 0.2 mM in D_2O, respectively. A sample of a single defensive spray milking from a single adult female *A. buprestoides* was collected in the field from the University of Florida Natural Area (exact position 29° 38′ 1″ N, 82° 22′ 11″ W) and immediately frozen. Prior to NMR analysis, the sample was thawed and dissolved in D_2O, and a 10-μL sample was transferred to a 1-mm NMR tube.

NMR Data Collection and Analysis. 1D and 2D NMR data were collected at 600 MHz using a 1-mm triple-resonance HTS cryogenic probe.[12] Each sample volume was 10 μL, which was transferred to a 1-mm × 100-mm capillary NMR tube (Norell, Inc.) using a 10-μL syringe with a fixed 110-mm × 30-gauge blunt needle. The sample temperature was maintained at 300 K, and residual water signal was suppressed by presaturation. 1D and 2D NMR spectra were collected with a spectral width of 6600 Hz for the test mixtures and 7211.5 Hz for the *A. buprestoides* sample. 2D TOCSY experiments[15] were collected using the DIPSI-2 mixing sequence[20] with a 90-ms mixing time and 2048 t_2 and 512 t_1 (complex) data points. The radio frequency field strength during DIPSI-2 mixing was 8928 Hz. States-TPPI was used for quadrature detection along the indirect dimension. The NMR data were processed by NMRPipe.[21] The time-domain data were apodized using a squared cosine function, zero-filled by a factor 2, 2D Fourier transformed, and polynomially baseline corrected along both dimensions.

The covariance TOCSY spectrum (**C**) of the mixture was obtained from the standard 2D Fourier transform NMR spectrum (**F**) by the matrix square root operation $\mathbf{C} = (\mathbf{F}^T\mathbf{F})^{1/2}$ using the covNMR module for NMRPipe.[17,19,21] Because of its inherent symmetry, **C** displays the same high spectral resolution along both dimensions. For each trace (row) of matrix **C** an "importance index" is calculated as the sum of all elements of the corresponding row of $\mathbf{C}^2$, which is a measure of the cumulative overlap of this trace with all other traces of **C**.[14] Based on the importance index profile, a subset of traces of **C** is selected and clustered according to the agglomerative clustering algorithm.[22] For each cluster, a representative trace is identified as the one with a minimal importance index. In this way, the likelihood is optimized that the selected traces reflect an individual component free of spurious contributions from other spin systems.

RESULTS AND DISCUSSION

Analysis of Test Mixture. The methodology was first tested on a sample containing a mixture of five chemicals in aqueous solution that are commonly found in metabolic mixtures, namely, D-glucose, histidine, lysine, 5HT, and sorbitol. A blind test was set up as follows: the group at University of Florida prepared the mixture, collected the NMR TOCSY raw data using the 1-mm high-temperature cryoprobe at 600 MHz magnetic field strength, and sent the data, without revealing the identities of the components, to the group at Florida State University for processing and analysis. The groups then met and validated the results by comparing the selected traces extracted from the TOCSY spectra with the reference 1D spectra of the individual compounds.

From the covariance TOCSY spectrum (Figure 2A), the importance index for each trace is computed and plotted in Figure 2B. The covariance TOCSY traces belonging to the peaks picked in the importance index profile are then subjected to the clustering analysis, visualized by the data tree structure, a dendrogram, in Figure 2C. For each cluster, the trace with the lowest importance index is selected and plotted in Figure 3 (black) together with the corresponding 1D reference spectrum (red) taken from the BMRB database or measured for an individually prepared sample in the case of serotonin. Since serotonin consists of two disjoint spin systems, it is represented by the two separate TOCSY traces 2 and 3. The slowly interconverting α- and β-forms of glucose are represented by the separate traces 1 and 7, respectively, whereas the reference spectrum represents their superposition. The good agreement of the peak positions of the TOCSY traces with the reference spectra of the pure components clearly demonstrates the suitability of the platform of Figure 1. Differences in the peak amplitudes, which are caused by incomplete magnetization transfers among spins at a single TOCSY mixing time, do not affect spin system identification.

Insect Venom Analysis. The walking stick insect *A. buprestoides*, produces a defensive secretion,[23,24] which is a mixture of at least six compounds, some being isomers of one another.

(16) Zhang, F.; Brüschweiler, R. *Chemphyschem* **2004**, *5*, 794−796.
(17) Trbovic, N.; Smirnov, S.; Zhang, F.; Brüschweiler, R. *J. Magn. Reson.* **2004**, *171*, 277−283.
(18) Brüschweiler, R.; Zhang, F. *J. Chem. Phys.* **2004**, *120*, 5253−5260.
(19) Brüschweiler, R. *J. Chem. Phys.* **2004**, *121*, 409−414.
(20) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274−293.
(21) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−93.
(22) Jain, A. K.; Murty, M. N.; Flynn, P. J. *ACM Computing Surv.* **1999**, *31*, 264−323.
(23) Meinwald, J.; Chadha, M. S.; Hurst, J. J.; Eisner, T. *Tetrahedron Lett.* **1962**, 29−33.
(24) Eisner, T. *Science* **1965**, *148*, 966.

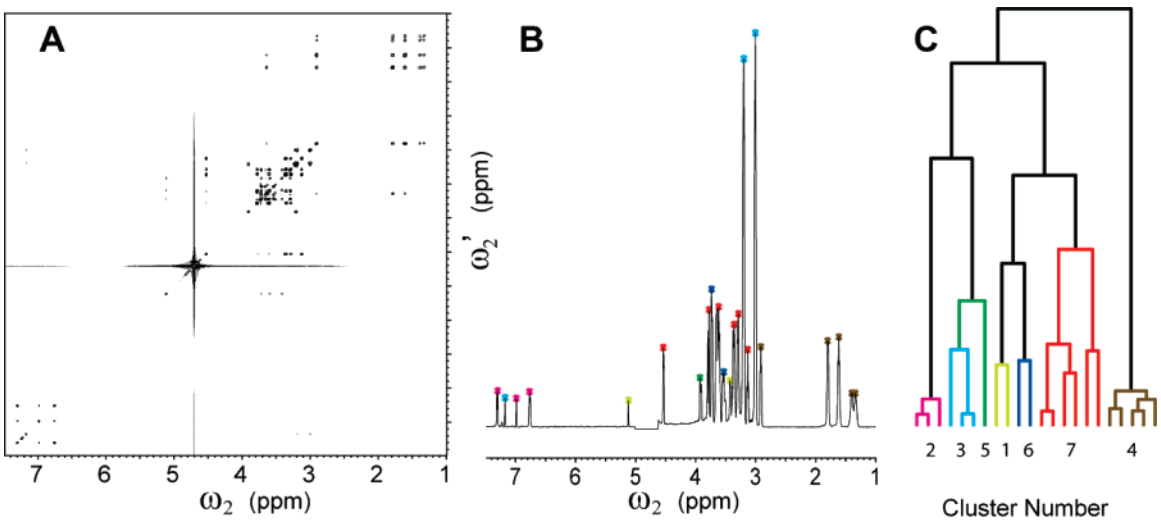

Figure 2. (A) Covariance proton TOCSY NMR spectrum of mixture containing D-glucose, histidine, lysine, 5HT, and sorbitol in aqueous solution. (B) Importance index profile. Peaks that are picked to represent traces of (A) are indicated by colored × symbols. (C) Dendrogram representation of cluster analysis according to similarity of the traces picked in (B). Each edge at the bottom corresponds to a peak picked in (B) of the same color.

Previous work showed that the defensive secretions vary from insect to insect, making analysis of samples from individual insects essential.[13] Because sample quantity is limited and the chemical components are highly overlapped in the experimental NMR spectra, analysis of these mixtures is challenging by traditional means. For a 1-mm NMR sample containing a single defensive spray milking, a 2D TOCSY experiment was recorded with a 90-ms mixing time and covariance processed with the result shown in Figure 4A. Based on the importance index profile, depicted in Figure 4B, traces are selected and subjected to the clustering (see dendrogram in Figure 4C). The traces representing each of the six clusters of the dendrogram are automatically selected by the algorithm and are shown in black in Figure 5. They are sorted according to their importance index with the traces at the bottom having the lowest index. Because spectral overlap inflates the importance index, the bottom traces are more likely to represent pure components than the top traces. Traces 1 and 2 are readily assigned to α-glucose and β-glucose, which are present in slow equilibrium. Traces 4 and 5 correspond to anisomorphal and peruphasmal, respectively, which are two monoterpene dialdehyde stereoisomers that are components of the defensive secretions of *A. buprestoides*. In aqueous solution, each monoterpene dialdehyde is in chemical equilibrium with a geminal diol form, and the diols from anisomorphal and peruphasmal correspond to traces 3 and 6, respectively (their chemical structures are depicted in Figure 5). These components have been independently identified and assigned by gas chromatography, mass spectrometry, and traditional NMR.[13] *A. buprestoides* from the previous study contained only a small amount of peruphasmal, but the insect used for this work was collected from a different subpopulation and developmental life stage that produces more peruphasmal (unpublished). These compounds are automatically and unambiguously identified by DemixC as can be seen by comparison with the 1D reference spectra of Figure 5.

Based on previous studies,[13] the total glucose concentration is estimated to be ~15 mM, which amounts to a total of 27 μg in the present 10-μL sample. A further reduction in mass by a factor

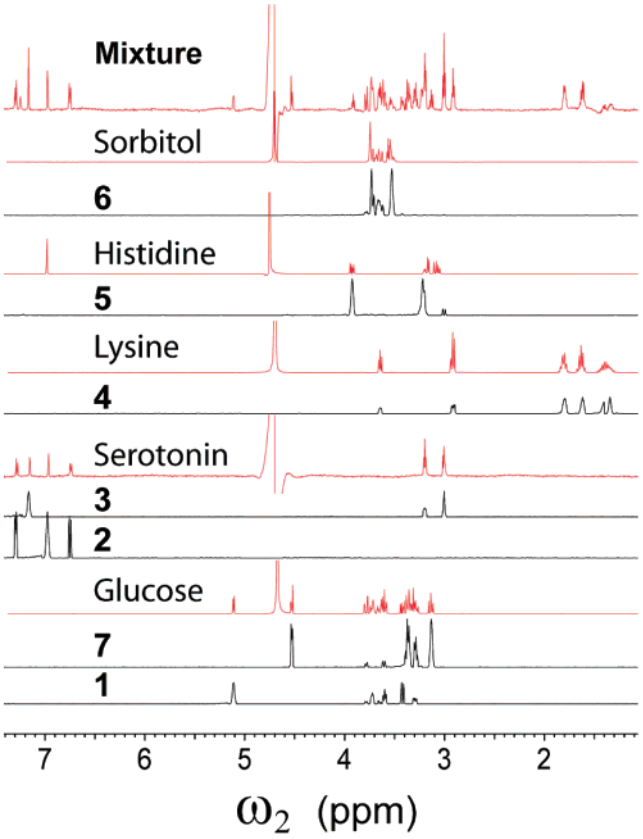

Figure 3. (A) Top spectrum: 1D ^{1}H NMR spectrum of the mixture of Figure 2. Seven black spectra: covariance TOCSY traces representing the clusters of Figure 2C with the numbers corresponding to the cluster numbers at the bottom of Figure 2C. The bottom five red spectra are reference 1D spectra of the pure components. Except for serotonin, the reference spectra were taken from the BMRB databank (the large peaks in the center correspond to the water signal). Traces 1 and 7 correspond to α-D-glucose and β-D-glucose, respectively. Traces 2 and 3 correspond to the 2 disjoint spin systems of serotonin.

10−100 should be possible. In fact, a COSY spectrum and ^{13}C-HMQC spectrum at natural abundance with 1.7 μg of ibuprofen

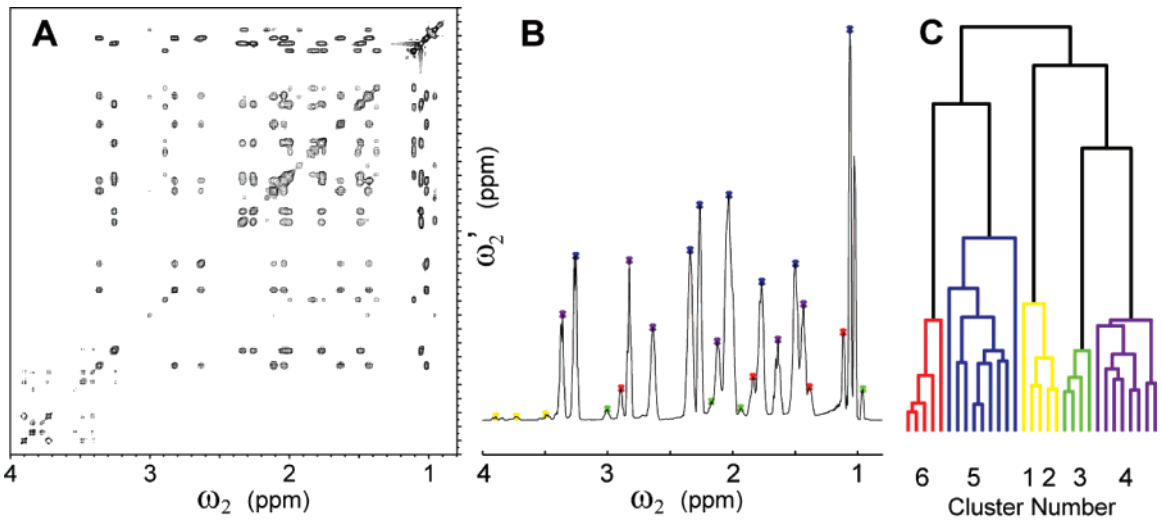

Figure 4. (A) Aliphatic section of covariance proton TOCSY spectrum of defensive secretion of a single walking stick insect. (B) Importance index profile. Peaks that are picked to represent traces of panel A are indicated by the colored × symbols. (C) Dendrogram representation of cluster analysis according to similarity of the traces picked in (B). Each edge at the bottom corresponds to a peak picked in (B) of the same color.

were previously reported using the 1-mm HTS probe.[12] The ^{13}C-HMQC data corresponded to 17 ng or ~82 pM NMR-active substance.

This study demonstrates the effectiveness of DemixC to analyze the composition of metabolic mixtures under realistic conditions, in this case, a field milking of a single insect. Some of the advantages of measuring mixtures of metabolites before purification have been demonstrated previously by characterizing 13 steroids in fireflies,[25] and DemixC provides a general and robust method for such problems. A prerequisite for its successful application is that the chemical components possess proton spin systems that permit sufficient magnetization transfer under TOCSY mixing, which is fulfilled for the majority of natural products, and that the components to be detected have at least one resonance that does not strongly overlap with other spin systems. Due to the narrow NMR line width of small molecules, the latter condition is fulfilled with high probability even in mixtures that contain a much larger number of components. Because covariance NMR can operate with small numbers of evolution times t_1,[26] the methodology described here has excellent potential for high-throughput applications, including metabolomics and metabonomics. The identified traces can be readily screened against spectral databases that are rapidly growing, such as the metabolomics/metabonomics database of the Biological Magnetic Resonance Data Bank (BMRB) (http://www.bmrb.wisc.edu), the Human Metabolome database (http://www.hmdb.ca),[27] and the Swedish NMR metabolomics database of Lingkoping (http://www.liu.se/hu/mdl/main), which are worldwide in the public domain.

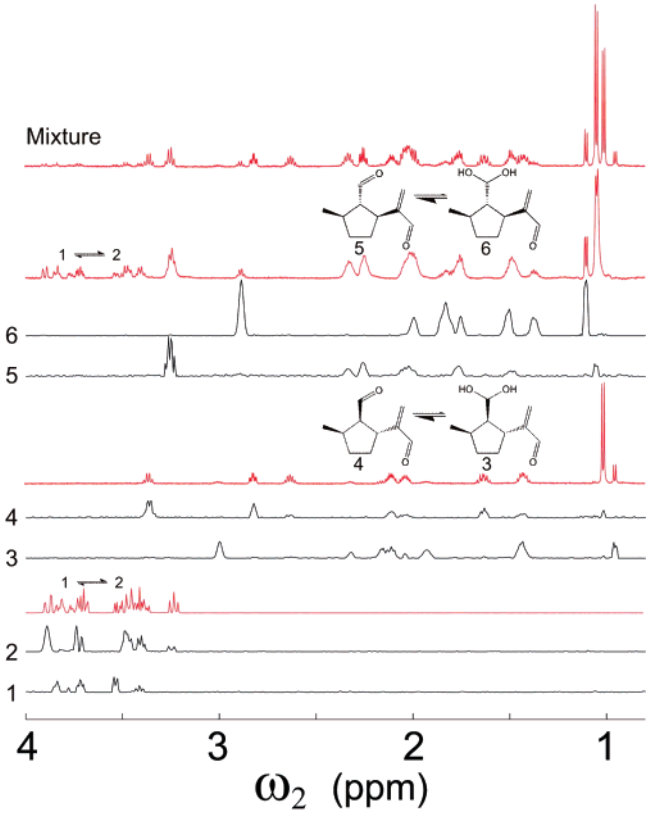

Figure 5. (A) Top spectrum: 1D ^{1}H NMR spectrum of the mixture. Six black spectra: covariance TOCSY traces representing the clusters of Figure 4C with the numbers corresponding to the cluster numbers at the bottom of Figure 4C. The bottom three red spectra are reference 1D spectra of purified components. Each reference spectrum contains two species, α-glucose (trace 1) and β-glucose (trace 2); dialdehyde and diol forms of the anisomorphal (traces 4 and 3, respectively), and the peruphasmal (traces 5 and 6, respectively) monoterpenes. The top reference trace also contains some glucose signals (on the left side). Chemical structures of the anisomorphal and peruphasmal and their corresponding geminal diols are shown as insets.

(25) Gronquist, M.; Meinwald, J.; Eisner, T.; Schroeder, F. C. *J. Am. Chem. Soc.* **2005**, *127*, 10810−10811.
(26) Chen, Y.; Zhang, F.; Bermel, W.; Brüschweiler, R. *J. Am. Chem. Soc.* **2006**, *128*, 15564−15565.
(27) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35*, D521−D526.

We have demonstrated here the feasibility of the DemixC approach for the identification of chemical components, some of which are in dynamic equilibrium and are present in small quantities in a single metabolic sample. The technique does not require different rates of translational diffusion of components[5] or measurements on multiple samples.[7,8] Because DemixC can be applied to microgram (or less) samples in just a few microliter volumes, it has a unique potential for automation in studying a wide range of biological mixtures.

ACKNOWLEDGMENT

We thank Dr. Bill Brey for his expertise in HTS probe technology, Dr. Fatma Kaplan for her help to develop the test mixtures, and Mr. Walter Thorner for technical support. This work was supported by the National Institutes of Health (grant R01 GM 066041 to R.B.), the Human Frontiers Science Program (to A.S.E.), and the National High Magnetic Field Laboratory (NHMFL).

Received for review June 1, 2007. Accepted July 26, 2007.

AC0711586