Statistical Heterospectroscopy, An Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies

ARTICLE *in* ANALYTICAL CHEMISTRY · FEBRUARY 2006

Impact Factor: 5.64 · DOI: 10.1021/ac051444m · Source: PubMed

9 AUTHORS, INCLUDING:

Séverine Zirah
Muséum National d'Histoire Naturelle
44 PUBLICATIONS **1,235** CITATIONS
SEE PROFILE

Stephen J Bruce
University Hospital of Lausanne
30 PUBLICATIONS **1,130** CITATIONS
SEE PROFILE

Paul D Rainville
The Kings College
37 PUBLICATIONS **1,167** CITATIONS
SEE PROFILE

Jeremy K Nicholson
Imperial College London
740 PUBLICATIONS **43,846** CITATIONS
SEE PROFILE

Anal. Chem. **2006,** *78,* 363−371

Accelerated Articles

Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies

Derek J. Crockford,† Elaine Holmes,† John C. Lindon,† Robert S. Plumb,‡ Severine Zirah,† Stephen J. Bruce,† Paul Rainville,‡ Chris L. Stumpf,‡ and Jeremy K. Nicholson*,†

Biological Chemistry, Division of Biomedical Sciences, Sir Alexander Fleming Building, Imperial College London, SW7 2AZ, U.K., and Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757

***S*tatistical *h*eterospectroscop*y* (SHY) is a new statistical paradigm for the coanalysis of multispectroscopic data sets acquired on multiple samples. This method operates through the analysis of the intrinsic covariance between signal intensities in the same and related molecules measured by different techniques across cohorts of samples. The potential of SHY is illustrated using both 600-MHz ^{1}H NMR and UPLC-TOFMS data obtained from control rat urine samples ($n = 54$) and from a corresponding hydrazine-treated group ($n = 58$). We show that direct cross-correlation of spectral parameters, viz. chemical shifts from NMR and m/z data from MS, is readily achievable for a variety of metabolites, which leads to improved efficiency of molecular biomarker identification. In addition to structure, higher level biological information can be obtained on metabolic pathway activity and connectivities by examination of different levels of the NMR to MS correlation and anticorrelation matrixes. The SHY approach is of general applicability to complex mixture analysis, if two or more independent spectroscopic data sets are available for any sample cohort. Biological applications of SHY as demonstrated here show promise as a new systems biology tool for biomarker recovery.**

A variety of metabolic profiling tools are routinely used and also currently under development for metabonomic[1−3] and metabolomic[4,5] studies. Spectroscopic methods such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled to gas chromatography and liquid chromatography (LC) have proved to be of high value because of their capabilities in detecting and where appropriate quantifying a wide range of chemical substances in complex mixtures such as urine and plasma.[6,7] The high information densities generated by these techniques and their inherent multivariate profiling capabilities usually necessitate the use of combinations of chemometric and mathematical modeling methods to classify biochemical perturbations that reflect specific physiological or pathological states and to recover latent biomarker information. Applications of this technology include the construction of expert systems for the prediction of drug toxicity[8−10] and the characterization of a range of pathologies such as cardiovascular disease,[11] inborn errors of

* Corresponding author. Tel.: +44-20-7594-3195. E-mail: j.nicholson@imperial.ac.uk.

† Imperial College London.

‡ Waters Corp.

(1) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181−1189.
(2) Lindon, J. C.; Holmes, E.; Bollard, M. E.; Stanley, E. G.; Nicholson, J. K. *Biomarkers* **2004**, *9*, 1−31.
(3) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 153−161.
(4) Kell, D. B. *Curr. Opin. Microbiol.* **2004**, *7*, 296−307.
(5) Fiehn, O.; Weckwerth, W. *Curr. Opin. Biotechnol.* **2002**, *13*, 156−160.
(6) Plumb, R.; Granger, J.; Stumpf, C.; Wilson, I. D.; Evans, J. A.; Lenz E. M. *Analyst* **2003**, *128*, 819−823.
(7) Jonsson P.; Gullber J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738−1745.
(8) Ebbels, T.; Keun, H.; Beckonert, O.; Antti, H.; Bollard, M.; Holmes, E.; Lindon, J.; Nicholson, J. *Anal. Chim. Acta* **2003**, *490*, 109−122.
(9) Holmes, E.; Connor, S. C.; Nicholls, A.; Polly, S.; Nicholson, J. K.; Lindon, J. C.; Connelly, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 251−261.
(10) Gartland, K. P. R.; Sanins, S. M.; Nicholson, J. K.; Sweatman, B. C.; Beddell, C. R.; Lindon, J. C. *NMR Biomed.* **1990**, *3*, 166−172.

10.1021/ac051444m CCC: $33.50
Published on Web 12/08/2005

metabolism,[12] renal disease,[13] diabetes,[14] and neurodegenerative conditions,[15] and functional genomics studies.[16,17]

The complementarity of NMR and MS methods as structural tools has necessitated their parallel use in structure elucidation studies for natural product research, drug metabolite analysis, and other complex mixture analysis problems for many years.[18,19] Typically, NMR and MS spectra of various types are examined together, and structural parameters such as chemical shifts and coupling constants (NMR) and exact molecular mass and fragmentation patterns (MS) are compared, often for a single sample, to generate structural assignment information that is consistent with the outputs of both technologies. However, in many metabonomic studies, multiple samples with a wide range of biochemical variation are available for both NMR and MS analysis, and this creates the opportunity for statistical analysis of signal amplitude covariation between technologies and direct cross-correlation of data for assignment purposes. Recently we developed a new technique, *s*tatistical *to*tal *c*orrelation *s*pectroscop*y* (STOCSY), for use with NMR spectroscopic data.[20] STOCSY provides a means of extracting one- and two-dimensional spectral reconstructions of biomarker molecules from both patent and latent signals in one-dimensional NMR spectra.

We have now applied an analogous approach (with some significant differences in the algorithms necessary for scaling and normalization) to the integration of high-resolution NMR and UPLC−MS[21] data collected on the same samples in order to achieve a more direct means of cross-assigning signals from these two methods. We have termed this approach *s*tatistical *h*eterospectroscop*y* (SHY), as it allows signals in one spectroscopic domain (e.g., NMR) to be dispersed in a second analytical spectroscopic domain (e.g., MS) in order to facilitate data coanalysis. This approach is of wide applicability and can be extended beyond NMR and LC−MS to include any spectroscopic, electrochemical, or other multivariate analytical measurements where multiple samples are measured by more than one technology. Furthermore, we show here that it is possible not only to generate meaningful NMR to m/z correlations to assist in structure assignment using SHY but also to derive connectivities between NMR signals and fragmentation patterns from the same molecules and, further, to cross-correlate NMR spectra and mass spectra from molecules that are not structurally identical but that have commonalities in terms of metabolic pathway regulation. Thus, SHY is not only a structural analytical tool but also a potentially powerful bioinformatics tool that can be used in metabonomic or systems biology investigations.

EXPERIMENTAL SECTION

Animal Studies and Sample Collection. Rat urine samples were obtained from a study of hydrazine toxicity conducted as part of a major toxicology project.[22] All animal experiments were conducted according to appropriate national guidelines. In this study, male Sprague−Dawley rats were randomly allocated to groups—control, low and high dose ($n = 10$ per group). Here we only consider the control and high-dose groups. The high-dose group received hydrazine hydrochloride in 0.9% saline administered orally at 90 mg/kg. Dose vehicles alone were administered to matched control animals. The time of dosing was defined as zero hours. Rats were housed in individual metabolic cages under controlled temperature, humidity, and light cycles. Urine samples were collected for time intervals from −16 to 168 h.[23] Data for the high-dose group used for this paper were obtained from 58 samples, of which 20 correspond to predose time points (showing no toxic response) and 38 to time points known to demonstrate a high toxic response. Data for the control group were obtained from $n = 54$ control samples.

NMR Samples and Spectra. All NMR spectra were acquired at 600.13 MHz on a Bruker[24] DRX-600 spectrometer using a pulse sequence derived from the two-dimensional NOESY experiment for detection and water suppression and 64K points in the time domain. Urine samples of 400 μL were buffered by addition of 200 μL of a sodium hydrogen phosphate buffer (pH 7.4) made up with D_2O to provide a lock signal. Spectral referencing was relative to 3-(trimethylsilyl)propionic acid-d_4 sodium salt (TSP), which was included in the buffer at 1 mM. All measurements used a constant volume of sample. All samples were analyzed in 96-well plates with a Bruker flow injection system.[23] A fast Fourier transform, with a prior exponential line-broadening factor of 0.3 Hz, was applied to transients using Bruker XWinNMR software. Referencing to TSP, and phase and baseline corrections, were applied using in-house MATLAB[25] software.

UPLC−MS Samples and Spectra. LC−MS analysis of urine was performed using a Waters[26] Acquity UPLC System Separations Module coupled to a Waters MicroMass Q-Tof Micro mass spectrometer equipped with an electrospray interface and leucine enkephalin lock-spray. Urine samples were diluted 4-fold with distilled water, and 2-μL injections were made from a well plate maintained at 4 °C into a 10 cm × 2.1 mm Waters Acquity BEH

(11) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethel, H. W. L.; Clarke, S.; Schofield, P. M.; McKilligin, E.; Mosedale, D. E.; Grainger, D. J. *Nat. Med.* **2002**, *8*, 12, 1439−1444.

(12) Moolenaar, S. H.; Engelke, U. F.; Wevers, R. A. *Ann. Clin. Biochem.* **2003**, *40*, 16−24.

(13) Neild, G. H.; Foxall, P. J. D.; Lindon, J. C.; Holmes E. C.; Nicholson J. K. *Nephrol. Dial. Transplant.* **1997**, *12*, 404−417.

(14) Nicholson, J. K.; O'Flynn, M.; Sadler, P. J.; Macleod, A.; Juul, S. M.; Sonksen. P. H. *Biochem. J.* **1984**, *217*, 365−375.

(15) Ghauri, F. Y. K.; Nicholson, J. K.; Sweatman, B. C.; Wood, J.; Beddell, C. R.; Lindon, J. C.; Cairns, N. *Biomedicine* **1993**, 163−167.

(16) Gavaghan, C. L.; Holmes, E.; Lenz, E., Wilson, I. D.; Nicholson, J. K. *FEBS Lett.* **2000**, *484*, 169−174.

(17) Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. *Nat. Biotechnol.* **2001**, *19*, 45−50.

(18) Shockcor, J.; Unger, S.; Wilson, I. D.; Foxall, P. J. D.; Nicholson, J. K.; Lindon, J. C. *Anal. Chem.* **1996**, *68*, 4431−4435.

(19) Albert, K., Ed. *On-line LC NMR and Related Techniques*; Wiley: New York, 2002.

(20) 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. K. *Anal. Chem.*, **2005**, *77*, 1282−1289.

(21) Plumb, R. S.; Granger, J. H.; Stumpf, C. L.; Johnson, K. A.; Smith, B. W.; Gaulitz, S.; Wilson, I. D.; Castro-Perez, J. *Analyst* **2005**, *130* (6), 844−849.

(22) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Antti, H.; Bollard, M. E.; Keun, H.; Beckonert, O.; Ebbels, T. M.; Reily, M. D.; Robertson, D.; Stevens G. J.; Luke, P.; Breau, A. P.; Cantor, G. H.; Bible, R. H.; Niederhauser, U.; Senn H.; Schlotterbeck, G.; Sidelmann, U. G.; Laursen, S. M.; Tymiak, A.; Car, B. D.; Lehman-McKeeman, L.; Colet J.-M.; Loukaci, A.; Thomas, C. *Toxicol. Appl. Pharmacol.* **2003**, *187*, 137−146.

(23) Bollard, M. E.; Keun, H. C.; Beckonert, O.; Ebbels, T. M. D.; Antti, H.; Nicholls, A. W.; Shockcor, J. P.; Cantor, G. H.; Stevens, G.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Toxicol. Appl. Pharmacol.* **2005**, *204* (2), 135−51.

(24) http://www.bruker-biospin.com/nmr/index.html.

(25) http://www.mathworks.com/products/matlab.

(26) http://www.waters.com.

Table 1. Nominal *m/z* Values for Selected Pure Compounds Using MS/MS

compound	CE/eV	expected parent	ions detected (with count)
2-aminoadipate	10	162	98(500), 116(660), 144(1090), 162(1060)
2-oxoglutarate	5	147	none
	10		none
creatine	10	132	90(400), 132(660)
creatinine	10	114	86(270), 114(1600)
	15		72(20), 86(160), 114(350)
hippuric acid	10	180	105(30)
	5		180(40)
N-methylnicotinamide	10	137	94(120), 137(5000)
	15		92(180), 94(660), 110(80), 137(2200)
N-methylnicotinic acid	10	138	94(50), 138(1700)
	15		92(150), 94(190), 138(1000)
spermine	10	203	112(30), 129(310), 203(700)

C18 1.7-μm UPLC column maintained in an oven at 40 °C. The column was eluted under gradient conditions at a flow rate of 500 μL/min; mobile phase component A consisted of 0.1% aqueous formic acid, and B of 0.1% formic acid in acetonitrile. The composition was linearly increased from 0 to 20% B over 4 min, then to 95% B over the next 5 min, returned to 100% A over 1 min, and then reequilibrated over a final 2.5 min prior to injection of the next sample. The mass spectrometer was operated in positive ion mode with a capillary voltage of 3 kV, cone voltage of 25 V, nebulizer gas flow of 600 L/h, desolvation temperature of 250 °C, and source temperature of 120 °C. The instrument was set to acquire over the mass range m/z 50−850 with acquisition time of 300 ms and an interscan delay of 100 ms.

MS/MS Experiments. LC−MS/MS analysis of solutions of pure compounds was performed on a Waters 2795 Separations Module linked to a MicroMass Q-Tof Micro mass spectrometer with an electrospray interface and leucine enkephalin lock-spray. Solutions were prepared by 100-fold dilution of 5 mg/mL stock solutions of reference compounds. The solvent was 0.1% aqueous formic acid, either pure or mixed with 0.1% formic acid in acetonitrile to a maximum proportion of 1:1, depending on the solubility of the compound. Lesser dilutions were prepared as necessary according to the strength of MS signal obtained. HPLC was conducted so as to broadly match the above UPLC retention times, which are the times quoted in this paper. The mass spectrometer parameters employed were substantially as described above. MS/MS was triggered by detection of an ion at the nominal m/z of the expected parent ion. Fragmentation was performed in a collision cell filled with argon. Data on the fragment ions obtained for each collision voltage used are provided in Table 1.

Statistical NMR−MS Correlations. NMR spectra from Bruker format data files were smoothed using a cubic spline and stored at a resolution of 0.001 ppm, which retains the native resolution of the NMR measurement.[27] Sections of the spectra containing representative aromatic and aliphatic compounds were considered for the correlation calculations. The region between δ 4.50 and 6.00 was excluded in order to remove residual baseline artifacts resulting from water signal suppression.

UPLC−MS data were converted from Waters format files into text files using the Waters DataBridge utility. These files were read into MATLAB where the data for each sample were integrated ("binned") to produce two-dimensional histograms with a resolution of 0.1 min in retention time and unit m/z. This resolution was found to be sufficient to produce meaningful results and to nullify the effects of any retention time variation that might occur over a long LC−MS run with many urine samples. The LC−MS histograms were summed over retention time windows prior to the correlation analysis, such that the NMR−MS correlation performed is representative of the total ion count for each nominal m/z in each sample. In this way, a pseudo direct injection spectrum is created for each sample but without the concomitant differential ionization or suppression problems associated with multiple analytes competing for ion current. It is necessary to do this as many of the highly polar metabolites that are readily observed in the NMR spectra are eluted close to the void volume and these correlations would otherwise be lost.

Experimental animals are typically allowed to drink ad libitum, and this leads to variation in the dilution of the urine samples and causes an irrelevant correlation of every NMR peak to every MS peak. To compensate for this, the sections of NMR spectra used were normalized to unit total integral, and similarly, all mass histograms were normalized to unit sum. Pearson correlation coefficients were then calculated between the intensities of the NMR spectral points and MS bins using the MATLAB corrcoef function with the option of performing a *t*-test on the result enabled. All correlations failing to achieve a 99.9% confidence level were rejected. In addition, in each diagram, any correlation falling below the quoted correlation cutoff value has been assigned to zero. The insets along the axes of the correlation diagrams show the mean NMR and mass spectra for the sets of samples considered.

RESULTS AND DISCUSSION

Generally, the application of SHY can be expected to reveal latent relationships unapparent in the original data sets because the structure obtained from correlation is a statistical output involving the addition of genuine signals and the cancellation of noise interference from the biofluid spectra. By the same token, an intrinsically weak signal-to-noise ratio does not preclude SHY analysis because all that is required is covariation of one weak signal with another to generate cross-peaks between the analytical matrixes. In complex mixture analysis, a weakness of MS compared to NMR is the difficulty in obtaining molecular structures rapidly on complete unknowns and the variability in

(27) Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *Anal. Chem.* **2004**, *77*, 517−526.

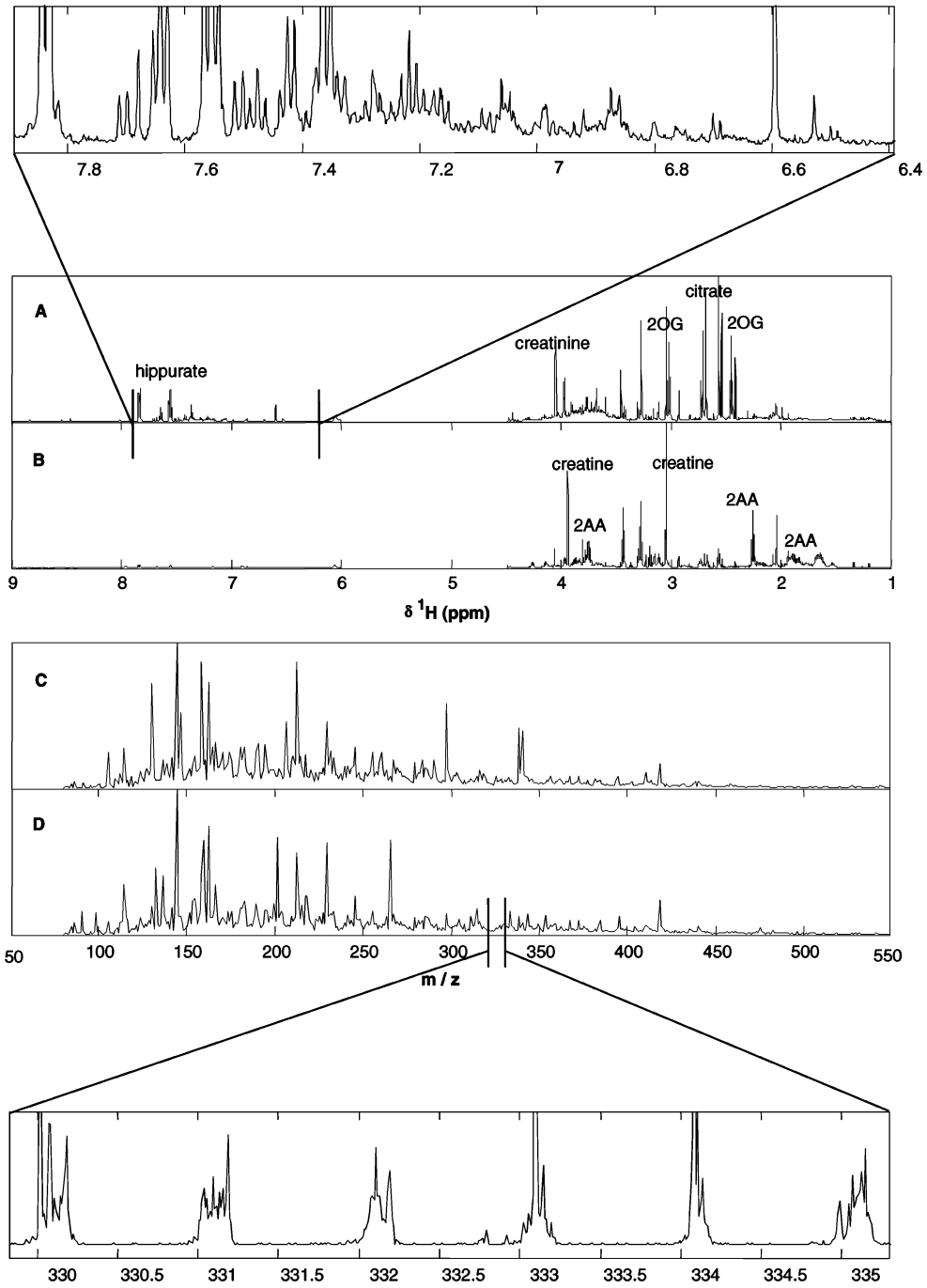

Figure 1. 600-MHz ^{1}H NMR spectra for single urine samples from control (A) and high-dose (B) animals showing depression of hippurate, creatinine, 2-oxoglutarate (2OG), and citrate in response to dose and elevation of creatine and 2-aminoadipate (2AA). The nonquantitative region δ 4.5−6.0 was set to zero. The expansion above shows the complexity of the NMR peak structure in the aromatic region. The corresponding mass spectra in (C) and (D) have many more peaks in comparison, as can be seen in the expansion below.

ionization efficiency between structurally diverse analytes. A weakness of conventional NMR compared to MS is its relative insensitivity and the possibility of extensive peak overlap. Using SHY, the complementary strengths of the two methods can be combined, and the covisualization of the NMR and MS data can yield greater structural and biological information than either used alone or applied pairwise between individual samples.

Hydrazine is a model hepatotoxin whose effects are well characterized and have been studied extensively in the rat.[28−32] Figure 1 shows typical ^{1}H NMR urine spectra from a control rat (A) and from a high-dose group rat (B). The mass histograms for the same samples from UPLC−MS are given in (C) and (D), respectively. The NMR spectra have been annotated with a

(28) Nicholls, A. W.; Holmes, E.; Lindon, J. C.; Shockcor, J. P.; Duncan Farrant, R.; Haselden, J. N.; Damment, J. P.; Waterfield, C. J.; Nicholson, J. K. *Chem. Res. Toxicol.* **2001**, *14*, 975−978.
(29) Holmes, E.; Nicholls, A. W.; Lindon, J. C.; Connor, S. C.; Connelly, J. C.; Haselden, J. N.; Damment, S. J.; Spraul, M.; Neidig, P.; Nicholson, J. K. *Chem. Res. Toxicol.* **2000**, *13* (6), 471−478.
(30) Sanins, S. M.; Timbrell, J. A.; Elcombe C.; Nicholson J. K. *Arch. Toxicol.* **1992**, *66*, 489−495.
(31) Waterfield, C. J.; Turton, J. A.; Scales, M. D. C.; Timbrell, J. A. *Arch. Toxicol.* **1993**, *67*, 244−254.
(32) Moloney, S. J.; Prough, R. A. *Rev. Biochem. Toxicol.* **1983**, *5*, 313−346.

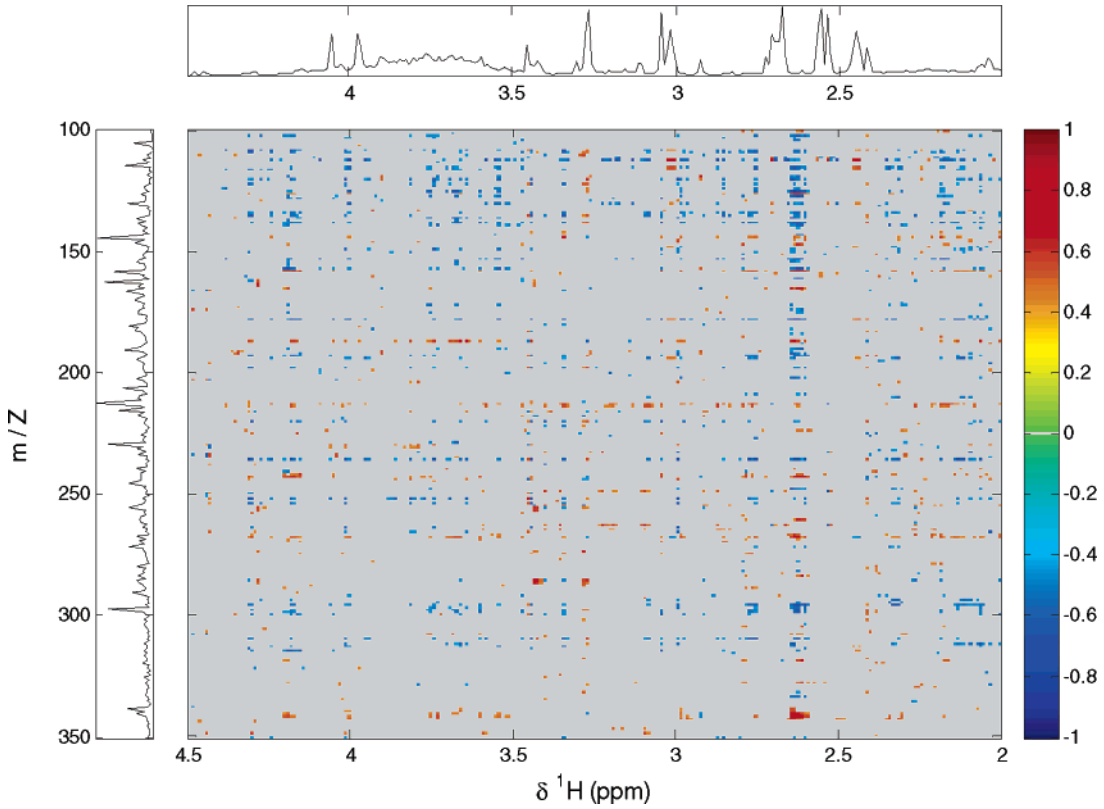

Figure 2. NMR−MS correlation for control samples at low resolution, showing part of the aliphatic region of the ^{1}H NMR spectra correlated with the total ion mass spectra in a low molecular weight range (no correlation cutoff applied).

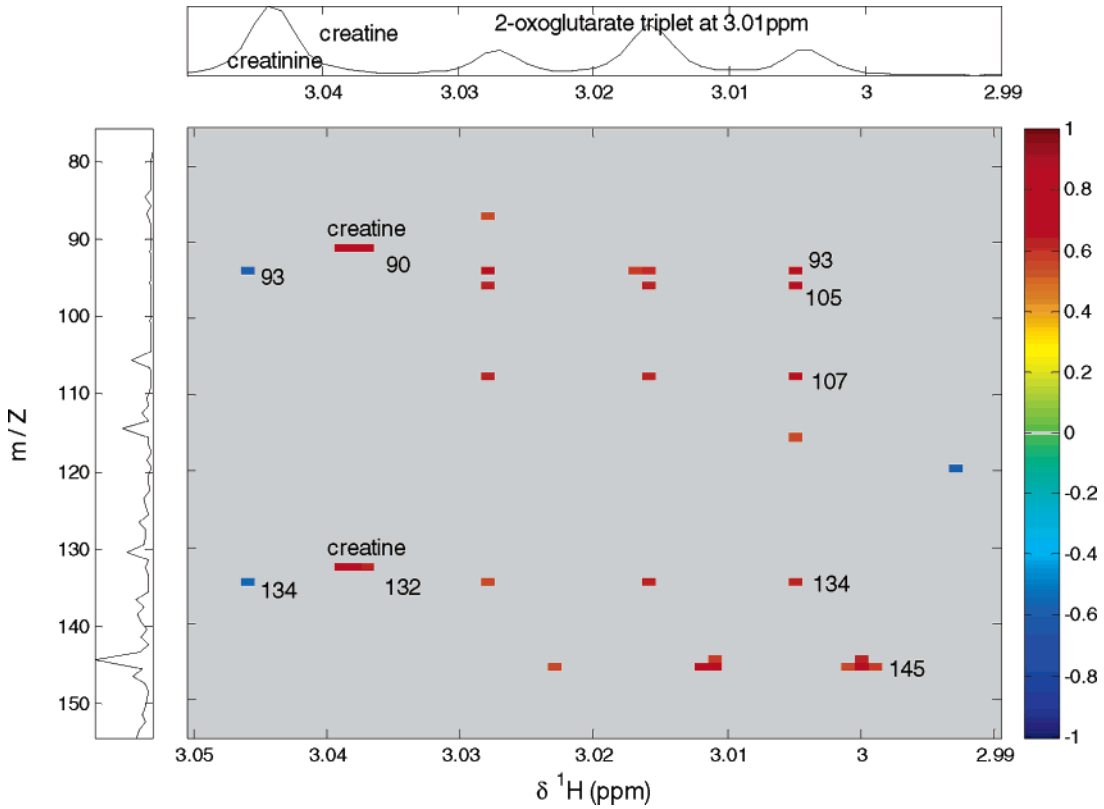

Figure 3. NMR−MS correlation for control samples expanded to show a 2-oxoglutarate region of the ^{1}H NMR spectra (cutoff, 0.55). The correlation of the creatine ions with the creatine NMR singlet is clear. There is a negative correlation of unidentified ions at *m*/*z* 93 and 134 to the overlapped creatinine peak, probably deriving from the same compound. The ions correlating directly with the 2-oxoglutarate NMR triplet have not been identified but may be related. The ion at *m*/*z* 145 correlates with a triplet overlapping 2-oxoglutarate, and we propose that it derives from *N*-acetyllysine.

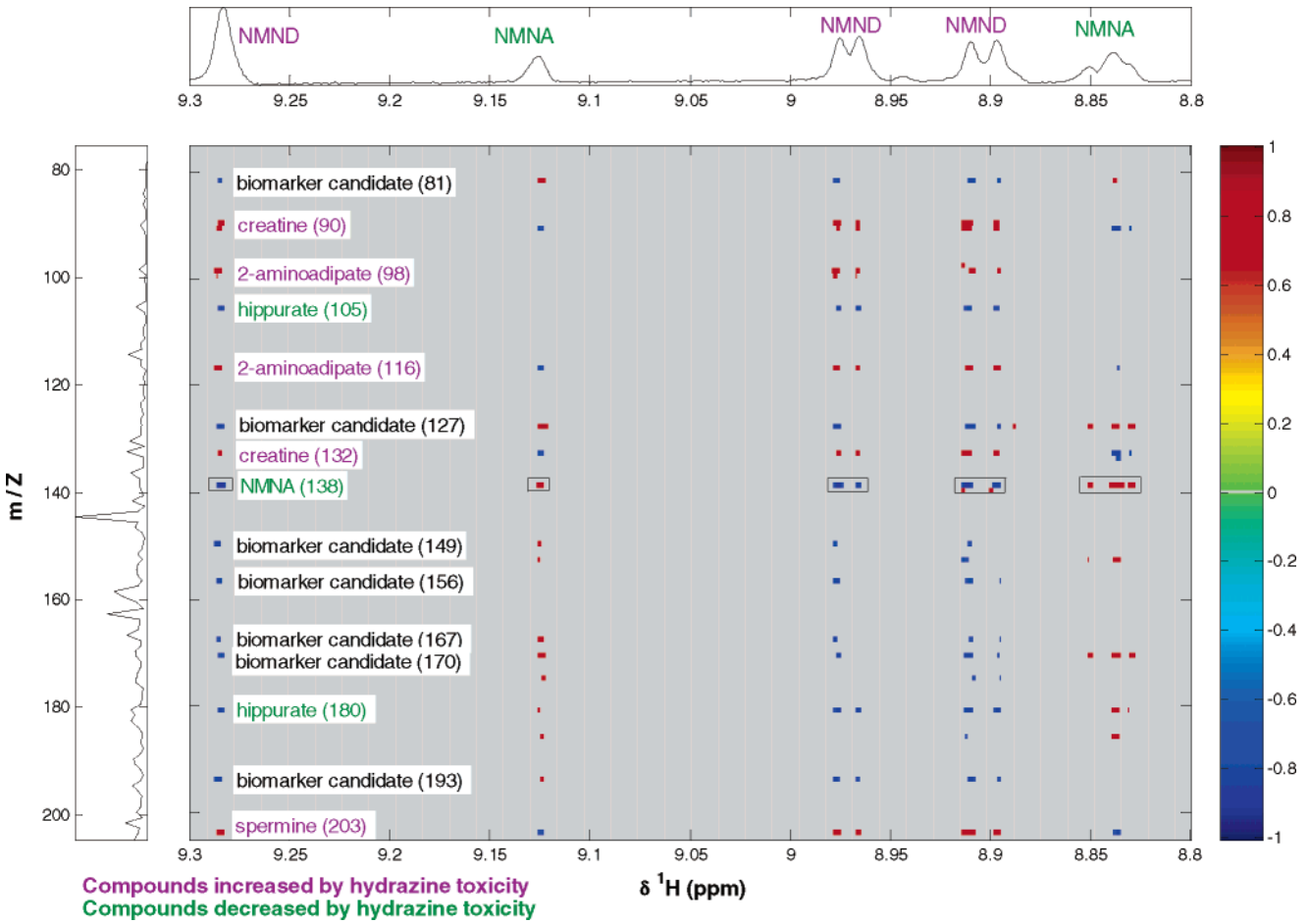

Figure 4. NMR−MS correlation for the high-dose hydrazine-treated sample group expanded to show an NMNA/NMND region of the ^{1}H NMR spectrum (cutoff, 0.75). The insets show mean NMR and mass spectra. All identified ions show directions of correlation consistent with known effects of hydrazine. For example, NMNA and NMND are negatively correlated, as they are transaminase-related (inhibited by hydrazine). Also, NMNA correlates positively with itself, as must be the case. These NMNA/NMND correlations are shown boxed. The newly identified spermine ion is shown to correlate positively with toxicity. All unidentified ions are *candidate biomarkers*.

number of resonances from metabolites that show a response to hydrazine dosing. The expansions show regions of NMR and MS spectra: it is apparent that the information density in both data sets is high but the MS spectrum is more dense.

In Figure 2, we show the NMR range δ 2.0−4.5 with intensity correlated against MS intensity for a region of low m/z. In this diagram, to improve visibility of the correlation points, the NMR spectra were first converted to histograms with a resolution of δ 0.005, and no minimum correlation cutoff was applied. However, for all other correlation diagrams, the full resolution of the NMR spectra has been retained, and the correlation cutoff for each is stated. The correlation scale is always from −1 to +1, where −1 indicates perfect anticorrelation between NMR and MS data, +1 indicates perfect correlation, and zero indicates no correlation. The mass histograms used in Figure 2 were obtained by summing over all retention time windows between 0 and 6 min, which cover virtually all of the signals of interest. Figure 2 shows a large number of points, and even with a confidence level of 99.9%, there will be some random correlations in a two-dimensional diagram containing thousands of points. However, it is extremely unlikely that multiple spurious points would arise adjacently or that they would occur in the form of NMR multiplet patterns. So, by expanding particular regions of interest, and distinguishing points that are clustered or arranged in patterns that coincide with the NMR spectra, it is possible to gain real information. Many of the ions detected in these expansions could be identified from prior knowledge and from MS/MS fragmentation patterns (Table 1).

Self-Molecular Correlation between NMR Signals and Parent Ions. A correlation diagram generated using data from undosed, control animals is given in Figure 3. The NMR region chosen illustrates a portion of the aliphatic spectrum, including peaks for creatine (singlet at δ 3.04), creatinine (singlet at δ 3.05), and 2-oxoglutarate (triplet at δ 3.01), but is a poor choice for studying creatine by NMR due to the overlap of the peak with the large creatinine signal. This overlap is made even less tractable by uncontrolled variation in creatine chemical shift resulting from differences in pH and metal ion concentration. Nevertheless, we can clearly distinguish the creatine NMR signal correlating positively with the creatine parent ion (m/z 132), and the width of the correlation is suggestive of the positional uncertainty of the creatine signal. The SHY approach has thus allowed us to deconvolve the two overlapping peaks (singlets) in a statistical fashion.

Because NMR spectroscopy and MS measure very different physicochemical properties and there is no setting for the UPLC−MS that allows detection optimized for every compound in a complex mixture, not all NMR signals will match those in the MS matrix. Moreover, NMR will pick up some substances that do not ionize well and therefore do not appear in the MS matrix.

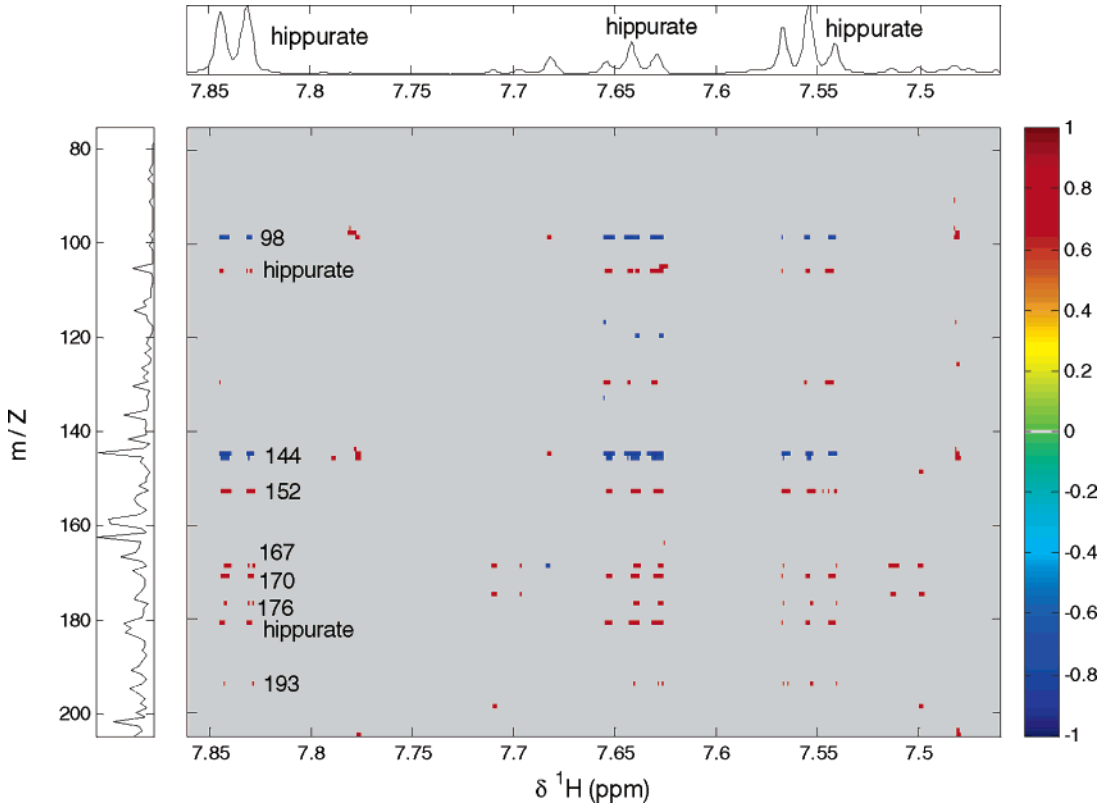

Figure 5. NMR−MS correlation for the high-dose hydrazine-treated sample group expanded to show a hippurate region of the ^{1}H NMR spectra against partial ion mass spectra for RT ≥ 1 (cutoff, 0.75). The filtering effect of the limited retention time range is clear, with hippurate being the only urinary component directly observable in a single one-dimensional NMR experiment.

For example, no correlation was observed between NMR and MS signals for 2-oxoglutarate in Figure 3 (and no signal was observed in MS/MS on the pure compound in positive ion mode—see Table 1). However, Figure 3 does show a correlation between an NMR triplet slightly displaced (by 3.00 Hz) from the 2-oxoglutarate triplet and an ion at *m/z* 145. This indicates that the 2-oxoglutarate NMR signal is overlapped with that of another compound, consistent with the NMR properties of *N*-acetyl lysine, which is not at all evident from the NMR spectrum itself. The assignment is reinforced by the correlated ion at *m/z* 145, the putative decarboxylated species, a very reasonable neutral loss from the parent ion at *m/z* 189. *N*-Acetyl lysine is readily produced from lysine by acetylation (a common biochemical transformation). Again, the SHY approach has deconvolved overlapping NMR peaks (triplets), in this case revealing a compound not previously observed in NMR spectra of control rat urine. So despite deficiencies in detection of both spectroscopic modalities, useful cross-correlation data at the assignment and pathway level can be obtained.

Part of the high-frequency region (δ 8.8−9.3) of the NMR spectrum and its correlations to a region of the mass spectrum (*m/z* 75−205) that includes many small, polar compounds is shown in Figure 4. Included are several compounds frequently detected in urine by ^{1}H NMR. The diagram was produced using data from urine samples in the high-dose group, but this includes samples taken at predose time points as well as peak response times. It clearly shows a high positive correlation between all of the NMR peaks for *N*-methylnicotinic acid (NMNA) (singlet at δ 9.13 and two overlapping doublets at δ 8.84, forming a pseudo-triplet) and the NMNA parent ion at *m/z* 138. Similarly, Figure 5 shows a positive correlation between all of the NMR peaks for hippurate (doublet at δ 7.84, and triplets at δ 7.64 and 7.56) and the hippurate parent ion at *m/z* 180. The other correlations are to other species (fragment ions and ions from other compounds) whose concentrations change as a result of toxin administration (see below).

In two cases, we can increase our confidence of a parent ion assignment due to the presence of an additional correlation with a known fragment ion (Table 1). So in Figure 3, we observe a positive correlation with the creatine fragment ion at *m/z* 90, and in Figure 5, we see a positive correlation with the hippurate fragment at *m/z* 105. The other creatine peak (at δ 3.94) also gives a clear correlation with both creatine MS signals (not shown).

Correlation as a Result of Toxic Perturbation. The driving variation in dosed animals is much stronger than in the controls, and this is reflected in the correlation cutoffs employed: 0.75 for high dose in Figures 4−6 and 0.55 for controls in Figure 3. The variation of many compounds in urine as a response to dosing with hydrazine has been studied previously,[28,29] and this knowledge can be used to check the consistency of the observed correlations. It is known that *N*-methylnicotinamide (NMND), creatine, and 2-aminoadipate increase with dosing whereas NMNA, hippurate, and 2-oxoglutarate decrease.

Inspection of Figure 4 reveals that *every* ion in the diagram that correlates with NMR signals for both NMND and NMNA does so with each in the opposite sense. Hence, the creatine parent ion at *m/z* 132 and its fragment at *m/z* 90 both correlate positively with NMND and negatively with NMNA, and the ions at *m/z* 98

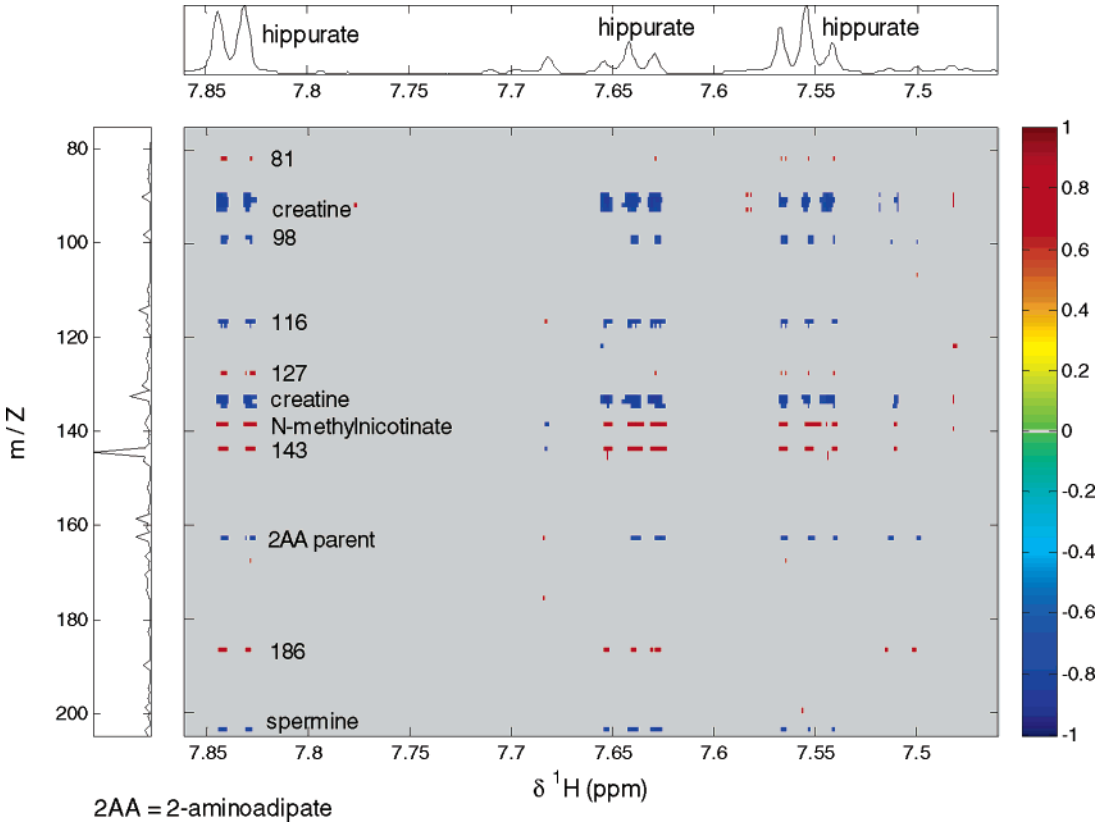

Figure 6. NMR−MS correlation for the high-dose hydrazine-treated sample group expanded to show a hippurate region of the ^{1}H NMR spectra against partial ion mass spectra for RT ≤ 1 (cutoff, 0.75). The filtering effect of the limited retention time range highlights several of the polar urinary components observed in the NMR spectrum.

Table 2. Accurate Masses and Retention Times for Some Biomarker Candidates

m/z	retention time/min
81.04	0.86
127.04	0.86
149.04	1.36
156.03	0.63
156.06	0.77
167.06	1.37
167.08	0.91
170.06	1.54
193.01	1.27
193.07	3.65

and 116 (consistent with 2-aminoadipate fragments) behave likewise. The NMNA parent ion at *m/z* 138 correlates negatively with NMND and positively with itself, while the hippurate parent at *m/z* 180 and its fragment at *m/z* 105 both correlate negatively with NMND and positively with NMNA. Interestingly, spermine, which is not commonly seen in NMR spectra of rat urine and has not been reported as perturbed by hydrazine toxicity, can also be observed (parent *m/z* 203), and this identification was confirmed by inspection of the MS spectra and by using fragment data from Table 1. The correlation diagram suggests that the spermine level increases with dosage, which is a novel toxicological biomarker finding. This is supported by evidence[33] relating the intracellular spermine level of isolated rat hepatocytes to hydrazine toxicity. Looking at the unidentified ions, at *m/z* 81, 127, 149, 156, 167, 170, and 193, these are consistent with being derived from compounds anticorrelated to hydrazine dosage. Similarly, Figures 5 and 6, which present correlations of ions with hippurate NMR signals, show consistent behavior throughout.

The accurate masses and retention times for the unidentified ions in Figure 4 were found by detailed inspection of the LC−MS data and are presented in Table 2. The matching of the retention times for *m/z* 81 and 127 suggests that these ions may derive from the same compound, and similarly for *m/z* 149 and 167 (consistent with a neutral loss of water). Because any ion that correlates with a biomarker is itself a biomarker these unidentified ions are all *biomarker candidates*; hence, the SHY approach offers outstanding opportunities for novel biomarker discovery using existing biomarkers as benchmarks.

Spectral Editing by Retention Time. While the mass histogram over all retention times (i.e., 0−6 min) yields an overview of all correlations, it is possible to focus attention on particular compounds by restricting the range of retention time windows summed. For example, in Figure 5, the NMR spectra are correlated with mass histograms obtained by summing over retention times between 1.0 and 6.0 min, and we see ions from hippurate together with unidentified ions at *m/z* 167, 170, and 193 common to Figure 4, but the ions for creatine and NMNA are absent. This is unsurprising given that the retention time window of hippurate in the LC−MS experiment was 4.0 min, whereas that for creatine and NMNA was 0.7 min. Figure 6 is

(33) Wua, C.; Miyagawaa, C.; Kennedya, D. O.; Yanob, Y.; Otanib, S.; Matsui-Yuasaa, I. *Chem.-Biol. Interact.* **1997**, *103* (3), 213−224.

complementary to Figure 5, being derived from retention times not exceeding 1.0 min, which restricts the coverage to the most polar compounds found in urine. The results agree well with Figure 4, showing creatine, 2-aminoadipate fragments, NMNA, and unidentified ions at m/z 81 and 127, in consistent relations of correlation or anticorrelation relative to hippurate. The correlations with the hippurate ions at m/z 105 and 180, however, are absent.

Correlations between Related Molecules in Similar Pathways. For the control samples in Figure 3, the NMR signal for 2-oxoglutarate does not correlate with any detectable ion from 2-oxoglutarate, but it does correlate positively with unidentified ions at m/z 93, 95, 107, and 134, and it would be reasonable to hypothesize that these arise from related compounds. The diagram also shows that the unidentified ions at m/z 93 and 134 correlate negatively with the NMR signal for creatinine. The fact that these two ions correlate with both creatinine and 2-oxoglutarate suggests that they may derive from the same compound, perhaps biochemically related to creatinine and 2-oxoglutarate. The negative correlation suggests that the as yet unidentified compound might be a product of a reaction in which creatinine is a reactant or a reactant where creatinine is a product.

It is noticeable that ions consistent with 2-aminoadipate fragments are present in both Figures 5 and 6: m/z 98 and 144 in Figure 5 and m/z 98 and 116 in Figure 6. However, the 2-aminoadipate parent ion (m/z 162) is only seen in Figure 6. As the retention time window for 2-aminoadipate in the LC−MS experiment was 0.7 min, we deduce that the ions at m/z 98 and 144 in Figure 5 are due to a compound (or compounds) eluting at a later time, giving rise to some of the same ions as 2-aminoadipate. It is known[28] that hydrazine treatment affects the levels of a number of metabolites (e.g., citrulline, *N*-acetylcitrulline, argininosuccinate) structurally and biochemically related to 2-aminoadipate. These data indicate the significant potential of the SHY approach to pathway activity exploration in drug or disease-induced disrupted metabolic states.

CONCLUSIONS

We have demonstrated that SHY can be used to reveal structural information inherent in NMR and LC−MS data sets, that biological relationships deriving from normal or toxin-disrupted metabolic pathways can be obtained, and that novel biomarker discovery can be achieved using known biomarkers as benchmarks. However, there are many other combinations of analytical probes that are amenable to SHY. With sufficient computer power, the approach could be extended into three dimensions by relating, for example, STOCSY to MS, NMR to LC to MS (by omitting summation over retention time windows), or NMR to MS to electrochemical measurements. Further work is ongoing to illustrate and extend the application of SHY to these new domains and to some important biomedical problems.

ACKNOWLEDGMENT

D.J.C. acknowledges support as part of a DTI Beacons Project. The authors are also grateful for access to urine samples and NMR spectra obtained by the Consortium for Metabonomic Toxicology (COMET). NMR spectra were preprocessed using NMRPROC, a program developed in MATLAB at Imperial College by T.M.D. Ebbels and H.C. Keun.

Received for review August 11, 2005. Accepted November 17, 2005.

AC051444M